
Universidad EAFIT

Proyecto de Grado

Ingeniería de Sistemas

Octubre 2008

Mejores Prácticas para Garantizar la Calidad de Datos en un Proyecto B.I.

Autores

Juliana Medina Sierra

Código: 200210005010

jmedinas@eafit.edu.co

Nancy Santacruz Arenas

Código: 200210132010

nsantacr@eafit.edu.co

Asesor

Sonia Cardona Ríos, MBA

scardona@eafit.edu.co

Nota de Aceptación

Firma del Presidente del Jurado

Firma del Jurado

Firma del Jurado

Medellín, Octubre de 2008

Dedicatoria

Para todos los que han tenido el deseo sincero de verme culminar esta etapa universitaria, a quienes consideraron éste como un paso muy importante para conseguir esa meta y para quienes anhelan que pueda obtener muchos más logros en mi vida.

Juliana

Todas las líneas acá contenidas van dirigidas a esas personas para quienes la gestión de datos y la seguridad de los mismos son indispensables para que los sistemas de información que diseñemos o desarrollemos cumplan a cabalidad con la finalidad para lo cual fueron creados.

Nancy

Agradecimientos

A quienes involucraron su tiempo de una u otra forma para que este Proyecto de Grado pudiera desarrollarse, a las personas que me motivaron a continuar trabajando cuando consideraba que no iba a alcanzar el final, a aquellos que entendieron y tomaron como propio el compromiso que implicó para mí este trabajo, a esas personas que como mi familia, comprendieron mi prioridad durante el desarrollo de este proyecto y que entendieron el tiempo que debía sacrificar para ello y a esos que aunque sólo fuera con cortas palabras de aliento hicieron que finalmente concretara este objetivo.

Juliana

A mi familia, a mis amigos, a mis profesores y mi asesora, quienes durante toda la carrera y especialmente durante la realización de este proyecto me apoyaron e incentivaron, siendo parte activa de mi formación profesional y personal.

Nancy

Contenido

| | |
|---|----|
| Introducción | 14 |
| 1. Conceptos Básicos | 15 |
| 1.1. Calidad de Datos | 16 |
| 1.2. Inteligencia de Negocios | 26 |
| 2. Calidad de Datos en un Entorno de Inteligencia de Negocios..... | 40 |
| 2.1. Antecedentes – Importancia del Tema | 41 |
| 2.2. Diferentes Propuestas Metodológicas que Existen | 42 |
| 2.3. Calidad de Datos en las Fuentes..... | 49 |
| 2.4. Calidad de Datos en el Proceso ETL..... | 57 |
| 2.5. Calidad de Datos en el Data Warehouse | 60 |
| 2.6. Calidad de Datos en las Vistas de Usuario..... | 62 |
| 2.7. Madurez de la Calidad de Datos en las Etapas de un Proyecto de Inteligencia de Negocios..... | 66 |
| 2.8. Factores Críticos de Éxito y Mejores Prácticas | 71 |
| 2.9. Lecciones Aprendidas | 76 |
| 3. Entrevistas | 78 |
| 4. Análisis Comparativo | 91 |
| 5. Conclusiones..... | 94 |
| 6. Bibliografía | 96 |

Tabla de Contenido

| | |
|--|----|
| Introducción | 14 |
| Calidad de Datos | 14 |
| Inteligencia de Negocios | 14 |
| Mejores Prácticas..... | 14 |
| 1. Conceptos Básicos | 15 |
| 1.1. Calidad de Datos..... | 16 |
| 1.1.1. Antecedentes..... | 16 |
| 1.1.1.1. Tendencias | 17 |
| 1.1.1.1.1. La Calidad de los Datos Ahora Incluye la Calidad de los Metadatos | 17 |
| 1.1.1.1.2. Perfilamiento de Datos como Primer Paso para Mejorar la Calidad de los Mismos | 17 |
| 1.1.1.1.3. Establecer Políticas Referentes a la Calidad de los Datos..... | 17 |
| 1.1.2. Atributos de Calidad en los Datos | 18 |
| 1.1.2.1. Facilidad de Acceso..... | 18 |
| 1.1.2.2. Cantidad Apropiaada de Datos..... | 18 |
| 1.1.2.3. Completitud..... | 18 |
| 1.1.2.4. Facilidad de Comprensión..... | 18 |
| 1.1.2.5. Credibilidad..... | 18 |
| 1.1.2.6. Disponibilidad Temporal | 19 |
| 1.1.2.7. Facilidad de Manipulación | 19 |
| 1.1.2.8. Facilidad de Interpretación | 19 |
| 1.1.2.9. Libre de Errores..... | 19 |
| 1.1.2.10. Objetividad | 19 |
| 1.1.2.11. Relevancia | 19 |
| 1.1.2.12. Representación Concisa..... | 19 |
| 1.1.2.13. Representación Consistente..... | 19 |
| 1.1.2.14. Reputación..... | 19 |
| 1.1.2.15. Seguridad | 19 |
| 1.1.2.16. Valor Agregado | 19 |

| | |
|---|----|
| 1.1.3. Ciclo de la Calidad de Datos | 20 |
| 1.1.3.1. Valoración | 20 |
| 1.1.3.2. Plan | 21 |
| 1.1.3.3. Implementar..... | 21 |
| 1.1.3.4. Evaluar | 21 |
| 1.1.3.5. Adaptar | 21 |
| 1.1.3.6. Educar | 21 |
| 1.1.4. Etapas de la Madurez en la Calidad de Datos | 22 |
| 1.1.4.1. Etapa 1: Incertidumbre | 22 |
| 1.1.4.2. Etapa 2: Despertar | 23 |
| 1.1.4.3. Etapa 3: Iluminación | 23 |
| 1.1.4.4. Etapa 4: Sabiduría..... | 23 |
| 1.1.4.5. Etapa 5: Certeza..... | 23 |
| 1.1.5. Gestión de la Madurez de la Calidad | 24 |
| 1.2. Inteligencia de Negocios | 26 |
| 1.2.1. Antecedentes..... | 27 |
| 1.2.2. Tendencias de la Inteligencia de Negocios..... | 28 |
| 1.2.2.1. Consolidación de las Compañías de BI | 28 |
| 1.2.2.2. Cambiando de BI Estratégico a BI Operacional o BI en el Momento Adecuado | 29 |
| 1.2.2.3. Análisis y Visualización de Datos más Sofisticados | 30 |
| 1.2.3. Proceso de Inteligencia de Negocios..... | 31 |
| 1.2.3.1. Necesidades | 32 |
| 1.2.3.2. Planeación y Dirección | 33 |
| 1.2.3.3. Colección de Fuentes..... | 33 |
| 1.2.3.4. Análisis y Producción | 33 |
| 1.2.3.5. Difusión | 34 |
| 1.2.4. Proyecto de Inteligencia de Negocios | 34 |
| 1.2.5. Metodologías Para La Implementación De Una Solución BI | 36 |
| 2. Calidad de Datos en un Entorno de Inteligencia de Negocios..... | 40 |
| 2.1. Antecedentes – Importancia del Tema | 41 |
| 2.2. Diferentes Propuestas Metodológicas que Existen | 42 |

| | |
|---|----|
| 2.2.1. Metodología para la Evaluación de la Calidad de Datos: Business Objects | 42 |
| 2.2.2. Metodología para el Análisis de la Calidad de Datos: IP-UML | 43 |
| 2.2.2.1. Análisis de Datos: | 43 |
| 2.2.2.2. Análisis de Calidad:..... | 44 |
| 2.2.2.3. Diseño del Mejoramiento de la Calidad: | 44 |
| 2.2.3. Metodología Total Information Quality Management (TIQM) | 44 |
| 2.2.3.1. P1. Valorar la Definición de Datos de Calidad de la Arquitectura de la Información | 45 |
| 2.2.3.2. P2. Valorar la Calidad de la Información | 45 |
| 2.2.3.3. P3. Medir Costos y Riesgos de Información sin Calidad..... | 45 |
| 2.2.3.4. P4. Re ingeniar y Corregir Datos | 46 |
| 2.2.3.5. P5. Mejorar la Calidad del Procesamiento de Información..... | 46 |
| 2.2.3.6. P6. Establecer un Ambiente de Calidad de la Información..... | 46 |
| 2.2.4. Proceso de Control de la Calidad de Datos | 46 |
| 2.2.4.1. Valoración de Datos | 47 |
| 2.2.4.2. Control de Calidad de Datos | 47 |
| 2.2.4.3. Seguimiento y Reporte | 47 |
| 2.3. Calidad de Datos en las Fuentes | 49 |
| 2.3.1. Introducción | 49 |
| 2.3.1.1. Los Usuarios | 50 |
| 2.3.1.2. La Aplicación | 51 |
| 2.3.1.3. La Base de Datos | 51 |
| 2.3.2. Técnicas para Evaluar la Calidad de Datos..... | 52 |
| 2.3.2.1. Técnica: Perfilamiento de Datos (Data Profiling)..... | 53 |
| 2.3.2.1.1. El Perfilamiento de Datos Ayuda a Planear Mejor, Así que los Proyectos se Ejecutan Mejor. | 53 |
| 2.3.2.1.2. El Perfilamiento de Datos es una Tarea que se Repite, no Algo que se Ejecuta una Vez..... | 53 |
| 2.3.2.1.3. El Perfilamiento de los Datos es en Realidad Dos Prácticas, Aplicadas en Diversas Fases del Proyecto. | 54 |
| 2.3.2.1.4. Los Datos que se Perfilan con una Herramienta Tienen Ventajas Sobre Métodos Manuales..... | 54 |

| | |
|--|----|
| 2.3.2.1.5. La Supervisión de los Datos es Casi Imposible Sin una Herramienta..... | 54 |
| 2.3.2.2. Técnica: Limpieza de Datos (Data Cleansing) | 54 |
| 2.4. Calidad de Datos en el Proceso ETL..... | 57 |
| Extracción, Transformación y Carga..... | 57 |
| Acumulación Simple | 58 |
| Rolling..... | 59 |
| 2.5. Calidad de Datos en el Data Warehouse | 60 |
| 2.6. Calidad de Datos en las Vistas de Usuario..... | 62 |
| 2.6.1. Técnica: Auditoría de Datos (Data Auditing)..... | 63 |
| 2.6.1.1. Cuantificar el Estado de los Datos | 63 |
| 2.6.1.2. Analizar el Estado de la Información | 63 |
| 2.6.2. Técnica: Prevención de Datos Defectuosos (Data Defect Prevention) . | 64 |
| 2.7. Madurez de la Calidad de Datos en las Etapas de un Proyecto de Inteligencia de Negocios..... | 66 |
| 2.7.1. Importancia del Estado de los Datos..... | 66 |
| 2.7.2. Madurez de la Calidad de Datos en un Proyecto de Inteligencia de Negocios que es el Primero en la Organización. | 67 |
| 2.7.2.1. Primera Etapa: Justificación..... | 67 |
| 2.7.2.2. Segunda Etapa: Planeación. | 68 |
| 2.7.2.3. Tercera Etapa: Ejecución..... | 68 |
| 2.7.2.4. Cuarta Etapa: Implementación..... | 69 |
| 2.7.3. Madurez de la Calidad de Datos en un Proyecto de Inteligencia de Negocios que es Expansivo/De Integración en la Organización..... | 69 |
| Quinta Etapa: Iteración/Expansión. | 69 |
| 2.8. Factores Críticos de Éxito y Mejores Prácticas | 71 |
| 2.8.1. Factores Críticos de Éxito | 71 |
| 2.8.2. Mejores Prácticas..... | 73 |
| 2.8.2.1. Mejores Prácticas - Entendimiento de la Tecnología | 73 |
| 2.8.2.2. Plan de Acción de Calidad de Datos..... | 73 |
| 2.8.2.3. Pasos para la Administración de la Calidad Total de Datos | 74 |
| 2.9. Lecciones Aprendidas | 76 |
| Problemas de Datos..... | 76 |

| | |
|--|----|
| 3. Entrevistas..... | 78 |
| 3.1. Breve Descripción de las Empresas Entrevistadas..... | 78 |
| 3.1.1. XM Compañía de Expertos en Mercados S.A ESP | 78 |
| 3.1.2. UNE Telecomunicaciones..... | 78 |
| 3.1.3. Valores Bancolombia | 79 |
| 3.1.4. Suramericana | 79 |
| 3.1.5. Intergrupo..... | 79 |
| 3.1.6. Serfinco | 79 |
| 3.1.7. Universidad EAFIT | 80 |
| 3.2. Formato de Entrevista | 81 |
| Preguntas..... | 81 |
| 3.3. Resultados Obtenidos | 82 |
| 3.3.1. Pregunta 1: ¿Cuáles son las etapas que se definen dentro de la compañía para llevar a cabo un proyecto de Inteligencia de Negocios? Y Pregunta 2: ¿Cuál es el objetivo o significado de cada una de las etapas? | 82 |
| 3.3.2 Pregunta 3: ¿Qué persona o área dentro de la organización es la encargada del tema de Calidad de Datos en el proyecto de Inteligencia de Negocios? | 83 |
| 3.3.3. Pregunta 4: ¿Qué actividades se realizan en cada una de las etapas respecto a la Calidad de Datos?..... | 84 |
| 3.3.3.1. En las Fuentes:..... | 84 |
| 3.3.3.2. En el Proceso de Extracción, Transformación y Carga:..... | 85 |
| 3.3.3.3. En el Data Warehouse:..... | 85 |
| 3.3.3.4. En la Visualización:..... | 85 |
| 3.3.4. Pregunta 5: De acuerdo a las etapas descritas ¿Qué técnicas utilizan para llevar a cabo las actividades que se refieren a Calidad de Datos en cada una de ellas?..... | 86 |
| 3.3.5. Pregunta 6: ¿Qué planea hacer para mejorar, garantizar o mantener la Calidad de Datos en su proceso de Inteligencia de Negocios? | 87 |
| 3.3.6. Pregunta 7: ¿Cuáles consideran ustedes que son lecciones aprendidas respecto a la Calidad de Datos en un proyecto de Inteligencia de Negocios? ¿En qué forma impacta el éxito o fracaso del proyecto la Calidad de Datos? | 88 |
| 3.3.6.1. XM | 88 |
| 3.3.6.2. Intergrupo..... | 89 |
| 3.3.6.3. UNE Telecomunicaciones..... | 89 |

| | |
|--|----|
| 3.3.6.4. Suramericana..... | 89 |
| 3.3.6.5. Serfinco..... | 89 |
| 3.3.6.6. Valores Bancolombia..... | 89 |
| 3.3.6.7. Universidad EAFIT..... | 90 |
| 3.3.7. Pregunta 8: ¿Cuáles considera como Factores Críticos de Éxito con respecto a la Calidad de Datos para un proyecto de Inteligencia de Negocios?..... | 90 |
| 4. Análisis Comparativo..... | 91 |
| 5. Conclusiones..... | 94 |
| 6. Bibliografía..... | 96 |

Tabla de Figuras

| | |
|---|----|
| Figura 1 - Ciclo del Mejoramiento de la Calidad de Datos | 20 |
| Figura 2 - Actividades para el Mejoramiento de la Calidad de Datos | 22 |
| Figura 3 - Modelo de la Madurez de la Capacidad de la Administración de Calidad de Información..... | 25 |
| Figura 4 - Ciclo de Inteligencia de Negocios..... | 32 |
| Figura 5 - Etapas de un proyecto de Inteligencia de Negocios | 36 |
| Figura 6 - Ciclo de la Metodología IP-UML..... | 43 |
| Figura 7 - Procesos de la Metodología TIQM | 45 |
| Figura 8 - Proceso de Control de la Calidad de Datos..... | 47 |
| Figura 9 - Monitoreo y Seguimiento de la Calidad de Datos | 48 |
| Figura 10 - Proceso de Adquisición de datos..... | 50 |
| Figura 11 - Técnicas aplicadas en las Fuentes | 52 |
| Figura 12 - Técnicas Aplicadas en el Proceso ETL..... | 57 |
| Figura 13 - Técnicas Aplicadas en el Data Warehouse | 60 |
| Figura 14 - Técnicas Aplicadas en las Vistas de Usuario..... | 62 |
| Figura 15 - Técnicas Aplicadas en Nuevos Proyectos | 64 |
| Figura 16 - Prevención de Datos Defectuosos..... | 65 |
| Figura 17 - Elementos del Entorno que Afectan la Calidad de Datos | 71 |
| Figura 18 - Etapas para Llevar a Cabo un Proyecto de Inteligencia de Negocios.. | 83 |
| Figura 19 - Actividades de Calidad de Datos que Realizan las empresas | 86 |

Tabla de Tablas

| | |
|--|----|
| Tabla 1 - Gestión de la Madurez de la Calidad de Datos | 25 |
| Tabla 2 - Comparación de Cuatro Metodologías en la Gestión de Proyectos BI.. | 38 |
| Tabla 3 - Metodología de Evaluación de Calidad de Datos de Business Objects.. | 43 |
| Tabla 4 - Elementos de Entorno que Afectan la Calidad de los Datos..... | 72 |
| Tabla 5 - Problemas de Datos..... | 77 |
| Tabla 6 - Encargado de la Calidad de Datos en las Empresas..... | 83 |
| Tabla 7 - Área de la que Depende el Encargado de Calidad de Datos en la Empresa | 84 |
| Tabla 8 - Personas o Áreas que Intervienen en la Calidad de Datos | 84 |
| Tabla 9 - Técnicas de Calidad de Datos que se Aplican a las Etapas del Proyecto BI | 87 |

Introducción

Para la toma de decisiones empresariales es muy importante contar con elementos de información confiables, que le brinden a los directivos el poder encaminar el negocio con plena conciencia de la situación real de los estados de la empresa y el entorno, en un momento determinado.

La Inteligencia de Negocios (*BI: Business Intelligence*) se centra en la información, pero la información a su vez, está compuesta por los datos. Para que estos datos, conformen dicha información deben cumplir con una serie de atributos que los definan como datos de calidad, así pues este documento presenta las mejores prácticas para lograr alcanzar, garantizar y mantener la calidad de los datos en un proyecto de Inteligencia de Negocios. Adicionalmente a lo anterior, se presentan también diversas metodologías para llevar a cabo el desarrollo de un proyecto de Inteligencia de Negocios y se muestran ciertas metodologías referentes a Calidad de Datos.

Calidad de Datos

La base de la información son los datos, pero éstos han de ser confiables para poder garantizar que las decisiones tomadas a partir de sus aportes son efectivas. Determinar que un dato posee calidad no es un elemento observable a simple vista en el dato, sino que es la verificación de una serie de atributos sobre un ambiente determinado por el uso que le de la empresa a éste. Uno de los objetivos de este documento es clarificar los atributos que hacen de un dato un elemento con calidad, que ayude a determinar la relevancia de estos atributos en cada organización.

Inteligencia de Negocios

Cada día el éxito en los negocios está marcado por un alto nivel de análisis y manejo de información, ya no es suficiente el instinto de negociante para salir adelante en un mundo globalizado. Con el uso de herramientas computacionales en todos los niveles de la organización se permite recolectar mayor cantidad de datos, los cuales crean una base de información la cual es analizada y sumariada por aplicaciones orientadas a la toma de decisiones empresariales, brindándole al negocio un nivel de inteligencia en la toma de dichas decisiones.

Mejores Prácticas

Para garantizar que a través del tiempo y el uso, los datos permanecen y se alcanzan con calidad dentro de una organización, es necesario identificar las mejores prácticas de la industria y definir los elementos que aplican para el interés particular de cada empresa. Aquí se presentarán una serie de lineamientos que ayudarán a determinar estos elementos.

1. Conceptos Básicos

1.1. Calidad de Datos

Día a día las empresas están generando constantemente datos, los cuales deben pasar por procesos definidos para poder garantizar que éstos sean correctos y precisos para que constituyan piezas de información y brinden valor a la empresa.

Si no se controla la calidad de los datos desde su origen las empresas requerirán de procesos de limpieza de datos adicionales al procesamiento mismo de los datos lo cual incrementa los costos y tiempo. En el mundo de los negocios acelerado de hoy muchas veces los costos pueden asumirse pero lo que sí es crítico es la pérdida de tiempo, y en un proceso de limpieza de datos esta pérdida es inevitable.

Los datos con calidad son los que apoyan y fortalecen los diferentes procesos empresariales permitiendo potenciar las oportunidades de negocio y la buena gestión en la organización.

La Calidad de Datos es la ciencia o arte de mantener la información en las organizaciones de forma completa, consistente, precisa, actualizada, única y válida para el fin por el cual se almacena.

1.1.1. Antecedentes

Las organizaciones invierten mucho dinero y esfuerzo en mejorar sus sistemas de información en pro de agilizar sus procesos y optimizar sus productos, cuando los datos no se encuentran en un estado óptimo estos proyectos no generan la rentabilidad esperada y el costo de devolverse en el proceso puede salir más elevado que el proyecto mismo.

Los datos son un activo muy importante de la empresa, siendo fundamental que éstos cumplan con todos los atributos de calidad manejados en la actualidad; desde que existen los sistemas de información ha existido la preocupación de que los datos sean correctos y se ha buscado la forma de prevenir y corregir los errores que se puedan presentar en ellos. Inicialmente la calidad se realizaba de forma manual o con programas que no fueron desarrollados en lenguajes para ese uso.

Los primeros sistemas de Calidad de Datos fueron desarrollados por el gobierno de Estados Unidos a principio de los 80, en un sistema denominado NCOA (National Change Of Address Registry), el cual se creó para regular los cambios de domicilio, fallecimientos, bodas, divorcios, etc.

A principio de los 90 se creó en España un sistema especializado en Calidad de Datos para corregir o informar el código postal en la dirección de los clientes de las principales entidades financieras del país, en ese entonces se utilizó una técnica muy primitiva llamada "Fuerza bruta" y no implementaba las metodologías que tienen ahora los sistemas modernos de Calidad de Datos.

Hoy en día aunque se está trabajando en nuevas metodologías y hay tecnología para el manejo de la Calidad de Datos falta crear un poco más de conciencia en

cuanto a su manejo, ya que aunque se tiene conocimiento de esto existen muchos problemas de calidad en las organizaciones.

1.1.1.1. Tendencias

Los datos son un elemento fundamental en la veracidad y calidad de la información que se suministra en las empresas, siendo esta la base para la toma de decisiones, definición de estrategias competitivas, mejoramiento de procesos, entre otros.

A menudo la detección de los problemas de calidad se da accidentalmente ya sea por los usuarios de los sistemas o por los clientes, también es posible que esto surja por nuevos proyectos o estrategias de negocio que se realizan en la organización y requieran un estudio del estado de los datos previo.

Existen numerosas técnicas y herramientas disponibles para ayudar a las empresas a trabajar la Calidad de Datos, es importante escoger herramientas que tengan un enfoque práctico y que estén acordes al nivel de madurez de la empresa en cuanto al manejo de los datos. A continuación se presentan algunas tendencias que se están presentando en cuanto al manejo de la Calidad de Datos en las organizaciones.(1)

1.1.1.1.1. La Calidad de los Datos Ahora Incluye la Calidad de los Metadatos

Las normas de Calidad de Datos y los métodos serán aplicados a los metadatos, la ausencia de los metadatos es el origen de muchos defectos, ya que éstos tienen información valiosa sobre los datos almacenados, estándares, reglas, validaciones. Profesionales reconocen la necesidad de aplicar normas estrictas a las reglas de negocio y datos relacionados con los metadatos y su calidad.

1.1.1.1.2. Perfilamiento de Datos como Primer Paso para Mejorar la Calidad de los Mismos

La identificación de las fuentes de los datos y de las necesidades de información requerida son un paso fundamental en el mejoramiento de la Calidad de Datos.

1.1.1.1.3. Establecer Políticas Referentes a la Calidad de los Datos

El diseño de los datos es la base de la calidad de los mismos y que ésta sea exitosa para el fin con el cual se use. Las políticas que definen las normas para la calidad de la información están relacionadas con la metodología de calidad de información que se lleve a cabo en la organización.

1.1.2. Atributos de Calidad en los Datos

Los datos deben cumplir con una serie de atributos que ayudan a determinar su nivel de calidad, esto es, que tan útiles son estos datos para la empresa.

La Licenciada **Fernanda Escalone** en su Tesis de Maestría “*Estudio Comparativo de los Modelos y Estándares de Calidad de Software*” (2), sugiere que los atributos de calidad en los datos más importantes son:

- Facilidad de acceso
- Cantidad apropiada de datos
- Completitud
- Facilidad de comprensión
- Credibilidad
- Disponibilidad temporal
- Facilidad de manipulación
- Facilidad de interpretación
- Libre de error
- Objetividad
- Relevancia
- Representación concisa
- Representación consistente
- Reputación
- Seguridad
- Valor agregado

1.1.2.1. Facilidad de Acceso

Una de las características observables en el entorno de negocios sobre la disponibilidad de datos es poder obtenerlos fácilmente, el promedio de ejecutivos y directivos de las empresas no cuentan con el tiempo y muchas veces con la capacidad de indagar en un sistema de información o en hojas de cálculo o incluso en una base de datos en busca de los datos que necesitan para análisis, por ello los datos deben ser recuperados de una manera ágil y fácil, ya que en momentos críticos pueden descartarse datos importantes porque simplemente no se pudo tener acceso a ellos en un tiempo prudente.

1.1.2.2. Cantidad Apropiaada de Datos

El volumen de datos es adecuado para la tarea que se está realizando.

1.1.2.3. Completitud

Los datos son completos y suficientes para la tarea que se está desarrollando.

1.1.2.4. Facilidad de Comprensión

Los datos son fácilmente comprensibles.

1.1.2.5. Credibilidad

Los datos pueden ser considerados como creíbles y verdaderos.

1.1.2.6. Disponibilidad Temporal

Los datos están lo suficientemente actualizados para la tarea que se está desarrollando.

1.1.2.7. Facilidad de Manipulación

Los datos son fácilmente aplicables y manipulables en diferentes tareas.

1.1.2.8. Facilidad de Interpretación

Los datos están representados en el idioma apropiado, con una simbología correcta y adecuada y con la definición apropiada.

1.1.2.9. Libre de Errores

Los datos son correctos y fiables.

1.1.2.10. Objetividad

Los datos son imparciales, sin prejuicios y connotaciones.

1.1.2.11. Relevancia

Los datos son útiles y aplicables en la tarea que se está desarrollando.

1.1.2.12. Representación Concisa

Los datos están representados de una forma compacta.

1.1.2.13. Representación Consistente

Todos los datos se representan en el mismo formato, que además es el más adecuado para la tarea que se está desarrollando.

1.1.2.14. Reputación

Los datos están altamente relacionados en términos de sus fuentes o contenidos.

1.1.2.15. Seguridad

El acceso a los datos está restringido apropiadamente para garantizar su seguridad.

1.1.2.16. Valor Agregado

Los datos son beneficiosos y ofrecen ventajas al usarlos.

Los anteriores atributos presentados, son constituyentes de un dato de calidad y éstos son los que deben conseguirse y conservarse para tener como base del proyecto de Inteligencia de Negocios fuentes sólidas de consulta donde se harán los requerimientos y cruces de información que generarán la inteligencia requerida por los tomadores de decisiones de la organización.

1.1.3. Ciclo de la Calidad de Datos

Aún cuando los datos cumplan con los atributos mencionados, se debe tener en cuenta que la Calidad de Datos no es algo que se conserve intacto a través del tiempo, ya que se pueden generar cambios sobre los datos ya existentes o se ingresan nuevos datos al sistema. Para conservar y mejorar la calidad de los datos se debe considerar que ésta es un ciclo de mejoramiento continuo. De acuerdo a la figura encontrada en el capítulo 3 del texto Data Strategy, el ciclo de Calidad de Datos es el siguiente:

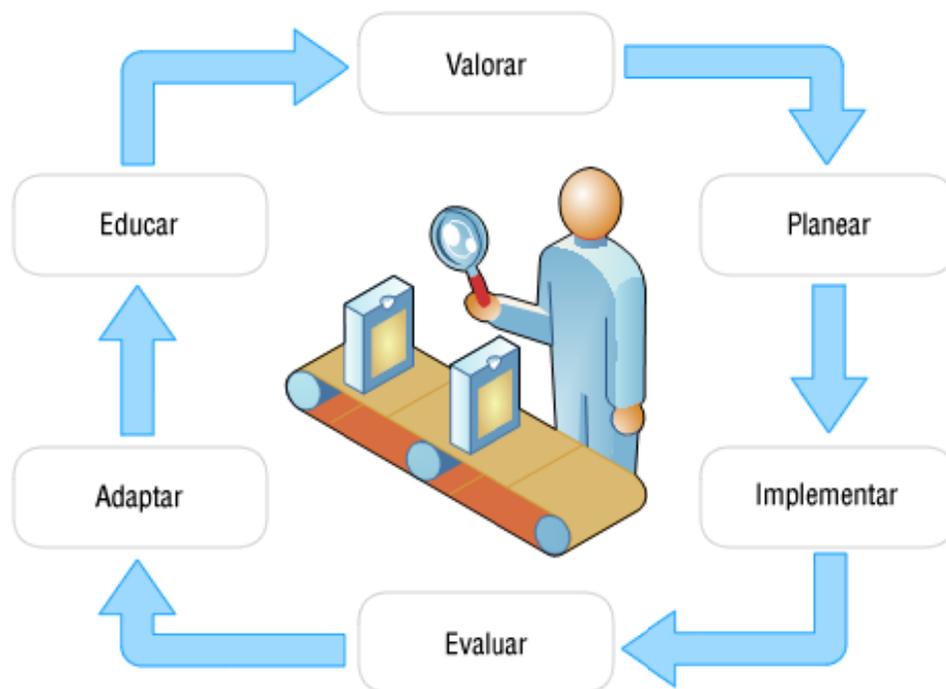


Figura 1 - Ciclo del Mejoramiento de la Calidad de Datos

1.1.3.1. Valoración

Cada ciclo de mejora comienza con una valoración. Esto puede ser una evaluación de calidad a la muestra de datos de toda la empresa, o evaluar la Calidad de Datos de uno de los departamentos. Al realizar la evaluación, no se deben limitar los esfuerzos a los perfiles de datos y la recogida de datos estadísticos sobre los defectos, se debe analizar toda la entrada de datos o proceso de manipulación de datos para encontrar las causas profundas de los errores y encontrar oportunidades de mejora de procesos.

Otro tipo de evaluación es una auditoria periódica de datos. Este tipo de evaluación se limita por lo general a un archivo o una base de datos a la vez. Se trata de datos de perfil, así como la validación manual de datos, contra los datos de dominio documentados (válidos los valores de los datos). Estos dominios ya deberían haber sido documentados como los metadatos, si no, se pueden encontrar en los programas, el código de traducción de libros, pantallas de ayuda en línea, hojas de cálculo y otros documentos.

1.1.3.2. Plan

Después de que las oportunidades de mejora se han definido, se deben analizar, priorizar, aprobar, financiar, asignar los recursos (personas) y programar. Dado que todas las inversiones no tienen el mismo retorno, son igual de prácticas o viables, se debe realizar un análisis de impacto que permita determinar que mejoras tienen mayores beneficios.

1.1.3.3. Implementar

En este momento es cuando se ponen en marcha las mejoras propuestas. En algunos casos, el grupo de calidad de los datos puede poner en práctica las mejoras aprobadas, pero en otros casos, otras personas ya sean del negocio o de TI pueden hacerlo.

Esto implicaría a las personas del negocio quienes están accediendo a los datos, el administrador de la base de datos quien hace el mantenimiento de ésta, y los desarrolladores que tienen acceso a ésta.

1.1.3.4. Evaluar

Las mejores ideas a veces son contraproducentes. Aunque algunos análisis de las repercusiones se han llevado a cabo durante la planificación, es posible que algo pase desapercibido. O peor, la implementación de la mejora puede estar creando un nuevo problema. Por lo tanto, es aconsejable monitorear la implementación de mejoras y evaluar su eficacia. Si se considera necesario, una mejora puede ser revertida.

1.1.3.5. Adaptar

Ojalá la mayoría de las mejoras no deban ser revertidas, sin embargo algunas pueden tener modificaciones antes de anunciar a toda la organización o ser convertidas en nuevas normas, directrices o procedimientos.

1.1.3.6. Educar

El último paso es difundir la información sobre el nuevo proceso de mejora que se ha implementado. Dependiendo del alcance del cambio, la educación se puede lograr a través de aulas de formación apoyando este trabajo en computadoras, un anuncio en la intranet de la organización, un boletín interno, o simple notificación por correo electrónico.

1.1.4. Etapas de la Madurez en la Calidad de Datos

Larry P. English presenta en su artículo *Information Quality Management Maturity: Toward the Intelligent Learning Organization* cinco etapas de la Administración de la Madurez de la Calidad de Información (IQMM). El IQMM se basa en la Gestión de madurez de la Calidad de Philip Crosby, la cual también fue utilizada por el Instituto de Ingeniería de Software como uno de los insumos para crear el Capability Maturity Model (CMM) para la calidad del software.

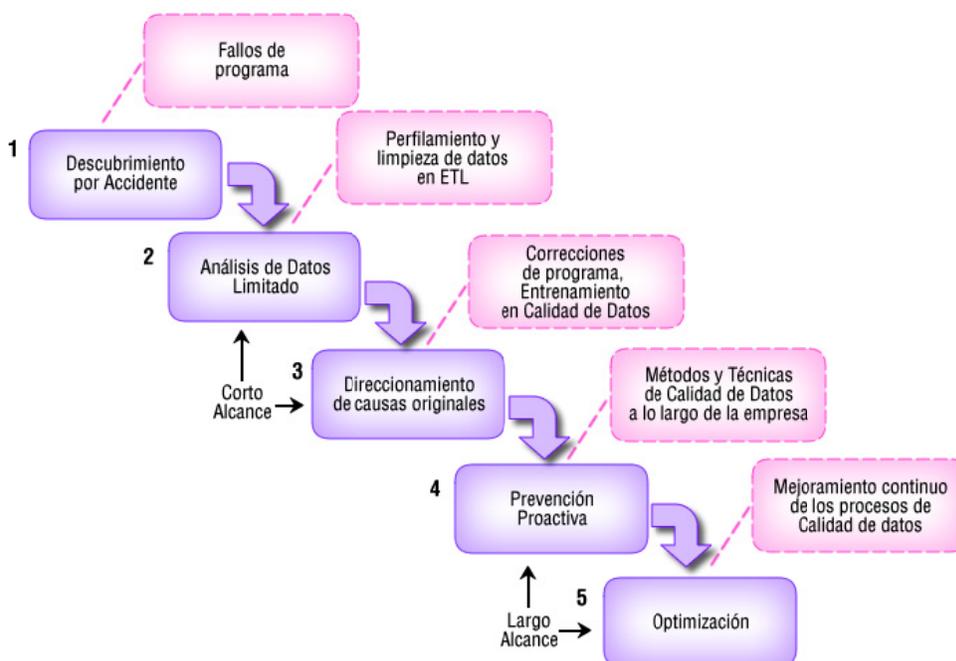


Figura 2 - Actividades para el Mejoramiento de la Calidad de Datos

(3)

Una manera fácil de determinar en una organización el nivel de la madurez respecto a la Calidad de Datos, es observar las actividades de mejora de la Calidad de Datos que se existen actualmente. La figura muestra las mejoras de la Calidad de Datos en cada una de las cinco etapas de los niveles de madurez de la Calidad de Datos basado en la adaptación de Larry English del modelo de capacidad de madurez (CMM) para la Calidad de Datos. Los cinco niveles son:

1.1.4.1. Etapa 1: Incertidumbre

Esta etapa se refiere a la falta de conocimiento sobre el porqué se presentan los problemas de calidad de información. No hay un proceso dinámico de mejora de la calidad, no hay un equipo de personas destinadas a trabajar en esto, al igual que tampoco hay financiación. La organización niega cualquier problema grave de Calidad de Datos y el análisis de datos se considera una pérdida de tiempo.

1.1.4.2. Etapa 2: Despertar

Es el momento cuando la organización se concientiza de los problemas generados por la mala calidad de los datos, se realizan algunos análisis de datos y corrección de datos de actividades, tales como los datos de perfil y limpieza de datos. Se recurre a solucionar el problema en el almacén de datos o en la zona de carga, en vez de hacer el esfuerzo por corregir el problema en la fuente.

1.1.4.3. Etapa 3: Iluminación

En esta etapa la organización lleva un proceso formal de gestión calidad, reconoce la causa de la mala calidad de información dividiendo sus procesos sistémicos y medidas de desempeño buscando el mejoramiento de la información, los ejecutivos de la empresa impulsan iniciativas de calidad de información y se trabaja para prevenir la repetición de los errores y causas fundamentales y la definición de las mejoras de las mismas. Un grupo de la calidad de los datos se crea y la financiación de proyectos de mejora para la Calidad de Datos está disponible. El grupo de Calidad de Datos realiza la evaluación de calidad de los archivos críticos y las bases de datos, y prioriza las actividades de mejora de la Calidad de Datos.

Esta etapa incluye capacitación formal para los directores y productores de información sobre los datos de sus clientes y la forma de mejorar sus procesos. Los resultados de esta etapa tienen como consecuencia reducciones en los costos de los procesos y de información defectuosa o innecesaria, dando un mayor control y mejorando la toma de decisiones.

1.1.4.4. Etapa 4: Sabiduría

Las organizaciones continúan mejorando sus sistemas de gestión de calidad y sus procesos. Cada gerente es responsable de la información producida y la actualización de la misma. La organización colabora y participa de forma activa en la prevención de futuros defectos de datos mediante una disciplina en el programa de mejora de Calidad de Datos. Hay un compromiso con la gestión de la calidad de la información en el sentido de mejorar al máximo su nivel de calidad y se empieza a ser conscientes de que es necesario prevenir los errores que se presentan hasta ahora.

1.1.4.5. Etapa 5: Certeza

La organización se encuentra en una optimización del ciclo mediante la supervisión continua y la mejora de los procesos de prevención de defectos de sus datos. La calidad de los datos es una parte integral de todos los procesos de negocio. Cada descripción del trabajo requiere la atención a la calidad de los datos, presentación de informes de defectos de datos, determinación de causas, seguimiento de los procesos de mejora. Básicamente, la cultura de la organización ha cambiado.

La única información que presenta problemas de calidad es aquella que es producida por procesos ajenos a la empresa.

1.1.5. Gestión de la Madurez de la Calidad

| | Etapa 1: Incertidumbre (Ad hoc) | Etapa 2: Despertar (Repetible) | Etapa 3: Iluminación (Definido) | Etapa 4: Sabiduría (Administrable) | Etapa 5: Certeza (Optimizado) |
|--|---|---|---|---|--|
| 1. Entendimiento y actitud administrativa | No se entiende la calidad de información como una herramienta de gestión. Se tiende a culpar la administración de los datos o los sistemas de información de la organización por los "problemas de calidad de información" o viceversa. | Reconocer que la gestión de la calidad de información puede tener valor pero sin disposición de proveer dinero a tiempo para hacer de esto posible. | Mientras se avanza en un programa de mejora de calidad de información se aprende más sobre la administración de calidad, siendo de apoyo y útil. | Participación. Entender el absolutismo de la gestión de calidad de información. Reconoce su papel personal en énfasis continuo. | Considera la gestión de calidad de información como una parte esencial del sistema de la compañía. |
| 2. Estado de organización de calidad de información | La calidad de "datos" está oculta en los departamentos de desarrollo de aplicaciones. La auditoria de datos probablemente no es parte de la organización. Énfasis en la corrección de datos erróneos. | Un fuerte papel de calidad de información es "designado" pero el énfasis principal es todavía corregir datos erróneos. | Existe organización de calidad de información, toda valoración es incorporada y el administrador tiene un papel en el desarrollo de aplicaciones. | El administrador de calidad de información rinde cuentas al CIO; reporte efectivo de estado y acciones preventivas. Involucrado con áreas de negocio. | El administrador de calidad de información es parte del equipo de gestión. Prevención es el objetivo principal. La calidad de información es un líder del pensamiento. |
| 3. Manejo del problema de calidad de información | Los problemas son atacados a medida que ocurren; sin resolución; definición inadecuada; muchos gritos y acusaciones. | Se configuran equipos para atacar los problemas mayores. Soluciones de largo alcance no se piden. | Comunicación de acción correctiva establecida. Los problemas son encarados abiertamente y resueltos de un modo ordenado. | Los problemas son identificados temprano en su desarrollo. Todas las funciones están abiertas a sugerencias y mejoramiento. | Sólo en los casos más inusuales, los problemas de calidad de información son prevenidos. |
| 4. Costo de información de calidad en términos de porcentaje de ingresos | Reportado: ? Actual: 20% | Reportado: 5% Actual: 18% | Reportado: 10% Actual: 15% | Reportado: 8% Actual: 10% | Reportado: 5% Actual: 5% |
| 5. Acciones de mejoramiento de información de calidad | Sin actividades organizadas. Sin entendimiento de dichas actividades. | Intento de esfuerzos obvios de "motivación" de corto alcance. | Implementación del programa de 14 puntos con entendimiento completo y establecimiento de cada paso. | Continuación del programa de 14 puntos e inicio de optimización. | Mejoramiento de calidad de información es una actividad normal y continua. |
| Resumen de la postura de información de calidad de la compañía | "No sabemos por qué tenemos problemas con calidad de información" | "¿Es absolutamente necesario siempre tener problemas con la calidad de la información?" | "A través de compromiso administrativo y mejoramiento de calidad de información estamos identificando y resolviendo nuestros problemas." | "La prevención de los problemas de calidad de información es una rutina que hace parte de nuestra operación" | "Sabemos por qué no tenemos problemas con la calidad de información" |

Tabla 1 - Gestión de la Madurez de la Calidad de Datos

(4)

A continuación, se hace una representación gráfica de lo que expresa la tabla anterior:

**Figura 3 - Modelo de la Madurez de la Capacidad de la Administración de Calidad de Información**

(5)

Las organizaciones deben ir avanzando en la escala de madurez para así conseguir llegar a un nivel adecuado de Calidad de Datos que les permita que la toma de sus decisiones los impulse y acerque a sus objetivos estratégicos y organizacionales.

1.2. Inteligencia de Negocios

Buscando obtener una definición concisa y adecuada del término Inteligencia de Negocios, se pretende hacer una agrupación de todos los elementos que se consideran importantes dentro de ésta. En la mayor parte de las definiciones que normalmente se encuentran para Inteligencia de Negocios, se tiende a tener en cuenta exclusivamente elementos tales como el software o el elemento tecnológico, sin embargo es de suma importancia reconocer que sin duda uno de los componentes que debe unirse al conjunto para la definición es el elemento humano.

Si vamos un poco más allá de esos componentes (software y tecnológico), vemos que el elemento humano es el que finalmente (y basado en la información que puede proveer una herramienta de inteligencia de Negocio), analiza y toma la decisión sobre el camino a seguir en una compañía.

Con el objetivo de tener un acercamiento a la definición del concepto de Inteligencia de Negocios e intentando identificar en diversas propuestas de definición cuales son las cosas que resaltan en ellas y consiguiendo finalmente una definición que contempla el conjunto de elementos requeridos, vamos a ver las siguientes opciones:

TechTarget define Inteligencia de Negocios como una "amplia categoría de aplicaciones y tecnologías para recolectar, almacenar, analizar y proveer acceso a los datos para ayudar a los usuarios de una empresa a tomar mejores decisiones de negocio."

(6)

Por su parte, para **Gartner Group**, "la clave para prosperar en un mercado competitivo es permanecer adelante de la competencia. Tomar las decisiones económicas adecuadas basadas en la información exacta y actual es algo más que intuición. El análisis de datos, reportes y herramientas de consulta pueden ayudar a los usuarios empresariales a avanzar a través de un mar de datos para sintetizar la información valiosa de ellos - estas herramientas entran hoy colectivamente dentro de la categoría llamada «inteligencia de Negocio»."

(7)

Como se ve, el elemento humano aunque es mencionado, no recibe el adecuado reconocimiento dentro de las definiciones anteriores, por lo cual se tiene aquí la definición que provee una persona que es de las más reconocidas en el tema de Inteligencia de Negocios actualmente:

Larry English propone lo siguiente:

“Inteligencia de Negocios: La capacidad de una empresa a actuar de forma eficaz a través de la explotación de sus recursos humanos y de información.”

A su vez, proporciona un elemento que es fundamental dentro de la Inteligencia de Negocios, su entorno, el cual lo define así:

“Entorno de Inteligencia de Negocios: Información de calidad en almacenes de datos bien diseñados, acoplada con herramientas de software amigables al negocio que le provee a los trabajadores del conocimiento acceso oportuno, análisis efectivo, y presentación intuitiva de la información correcta, permitiéndoles tomar las acciones adecuadas o tomar las decisiones correctas.”

(8)

Como se vio anteriormente en las definiciones planteadas, los datos son los que proveen la base para la toma de decisiones por parte de las personas que se involucran con la Inteligencia de Negocios de una organización. Más aún, como lo expresa *English* en su definición de *Entorno de Inteligencia de Negocios*, los datos y la información que estos proveen tiene que ser de calidad y que sin esta, las decisiones que se llevan a cabo no son beneficiosas e incluso pueden ser contraproducentes para la organización.

Dicho lo anterior, se considera entonces que el tema de la Calidad de Datos en un proyecto de Inteligencia de Negocios, es uno de los temas más importantes a tratar y por esta razón, se trabajará sobre él a lo largo de este documento.

1.2.1. Antecedentes

Para llegar a lo que hoy conocemos como Inteligencia de Negocios, se han atravesado diferentes instancias como las que White (9) identifica y presenta.

Estas instancias pueden agruparse en lo que son Consultas y Reportes basados en Host, Data Warehousing y Business Intelligence.

Las consultas y reportes basadas en Host eran usadas por un número reducido de personas proveedoras de información como analistas de negocios que eran las personas que tenían un gran conocimiento de los datos y una amplia experiencia en el manejo de computadores. El siguiente paso, el Data Warehousing mostraba una evidente separación en cuanto a volúmenes de datos comparado con la etapa anterior y Business Intelligence que se identificó como un sistema de información que debe incluir dentro de sus bondades el empaquetamiento de aplicaciones, las posibilidades de acceso de los usuarios, proveer apoyo para la toma de decisiones del negocio y la posibilidad de acceso a una amplia gama de información requerida que no se encuentra en el *Data Warehouse*.

Aparte de esto, no es sólo el cambio en la tecnología sino que también se han sucedido cambios en los nombres de los sistemas de información pasando así de ser Sistemas de Información Ejecutiva a Sistemas de Soporte a las Decisiones y posteriormente se han transformado en Business Intelligence.

Otro de los aspectos hace referencia a que anteriormente las consultas que se requerían se hacían directamente desde el puesto de trabajo de quien lo requería, ahora, se cuenta con accesos Web y gran disponibilidad de manera remota.

Adicionalmente, se tiene que para Kalakota y Robinson (2001)(10), los componentes anteriores a la Inteligencia de Negocios son el manejo que las compañías tenían acerca del conocimiento de sus empleados considerándolos a ambos como de gran importancia. En esta parte, el interés es el de compartir información al interior de la compañía. Seguidamente, se origina como segunda instancia el deseo de compartir información entre las diferentes áreas de la organización a través de Intranets. Continuando con los círculos de comunicación, ahora se pretende que las personas que se encuentran al exterior de la organización pero que hacen parte de ella como los proveedores, puedan conocer por medio de Extranet lo que ocurre en la compañía. Como penúltimo paso, tenemos el comercio electrónico y finalmente, abordamos lo que es Business Intelligence que pretende un acceso a la información sin importar el tiempo y el lugar.

1.2.2. Tendencias de la Inteligencia de Negocios

La evolución de las cosas es algo inevitable y la Inteligencia de Negocios no se queda atrás. A medida que se desarrollan más aspectos alrededor de un concepto, éste se diversifica y se transforma abarcando diferentes ámbitos de su interés para hacerlo cada vez más aplicable dentro de su entorno y ser mejor dentro de éste. Dado lo anterior, se van a exponer las siguientes tendencias de la Inteligencia de Negocios para la época que se está viviendo. (11)

1.2.2.1. Consolidación de las Compañías de BI

Para quienes trabajan en el mundo de los sistemas de información no es un tema ajeno la globalización de las empresas dedicadas a brindar soluciones de BI, en el mercado es cada vez más frecuente encontrar alianzas entre estas empresas o incluso adquisiciones de unas hacia otras. Esos movimientos están enfocados principalmente en tres razones:

- Los proveedores de soluciones de Calidad de Datos están siendo absorbidos por grandes creadores de soluciones integradas, en el punto que hoy en día son pocas las empresas especializadas en el mercado que ofrezcan soluciones exclusivamente de Calidad de Datos. Las empresas creadoras de soluciones completas o suites, ofrecen toda la gama de servicios integrados dentro de un solo paquete, esto trae beneficios en la simplificación de la cadena de procesamiento en el sentido de tener control sobre la calidad de los datos que están siendo integrados dentro de un mismo proceso.

- Los proveedores de servicios de integración de información empresarial también están siendo absorbidos por los grandes actores en el mundo de BI, pero en menos escala encontrándose mucho las alianzas estratégicas, las cuales ofrecen beneficios tanto para quienes son usuarios de soluciones completas como para quienes tienen soluciones específicas en términos de integración a medida que el proyecto crece o es migrado a una solución completa dentro de la cadena de la alianza.
- Las grandes corporaciones como IBM u Oracle, han llegado al punto de adquirir a los pequeños actores incluso antes que estos hagan historia o se formen una reputación en el mercado. Estas empresas están en la búsqueda constante de candidatos de innovación para incluir sus productos y servicios dentro de los paquetes de soluciones con los que ya cuentan estas compañías grandes.

Bien, ahora lo que resta preguntarse es si acaso ¿esto es conveniente? para lo cual la respuesta depende mucho del punto de vista desde el cual se desee darle importancia.

Por un lado, tener una gran herramienta integrada que ofrezca toda la gama de servicios y soluciones de inicio a fin dentro del proceso de BI es una ventaja, ya que esto permite que existan ahorros en términos de tiempo y dinero, ya que incluso la adquisición de una solución completa puede resultar en algunos casos mucho más rentable que cada pieza especializada por separado. Igualmente esta integración contiene una unificación de la manera de hacer las cosas, del lenguaje integrado en los procedimientos y una reducción en la curva de aprendizaje en el uso de la solución.

Por otra parte, las desventajas en este esquema, son la falta de "especialización" a las que se ven abocadas las soluciones con el pasar del tiempo, las empresas grandes no se pueden concentrar en especializar cada una de las ramas sino que su principal foco de atención es la integración entre las diferentes herramientas ofrecidas.

Esto entonces deja en el aire la respuesta si es mejor una solución completa que hace muchas cosas pero no necesariamente todo lo que se necesita para una solución particular o si lo mejor es pensar en adquirir cada una de las piezas del proceso de proveedores que ofrecen la mejor solución en su clase.

En el mundo tecnológico las grandes empresas no lo tienen todo, día a día las pequeñas empresas que aparecen pueden fácilmente poner en apuros muchas de las soluciones existentes, pero la tendencia del mercado es una alta consolidación ya sea por adquisición o por alianzas entre las diferentes empresas.

1.2.2.2. Cambiando de BI Estratégico a BI Operacional o BI en el Momento Adecuado

Actualmente, se produce un gap entre la ocurrencia de un evento alrededor de una organización y la decisión que se toma referente a dicho evento. Este gap, de acuerdo a lo planteado por Dick Hackathorn tiene tres componentes que lo definen y son el estado latente de la preparación de los datos que consiste en el tiempo que se requiere para coleccionar los datos que van a ser analizados, el

estado latente del análisis que es el tiempo que se requiere para obtener los resultados de una operación analítica y el estado latente de decisión que es el tiempo que se demora que la persona encargada de la toma de la decisión respecto al evento se apropie de la información obtenida del análisis y finalmente proponga la acción que debe ser llevada a cabo.

Para que la Inteligencia de Negocios operacional sea exitosa, los anteriores tiempos de latencia de los datos que hacen su aparición en el gap descrito anteriormente deben reducirse hasta prácticamente ser eliminados. Para lograr esto, día a día nuevos componentes virtuales de Inteligencia de Negocios son creados. Dentro de estos componentes, pueden enumerarse los Almacenes de datos Operacionales (ODS) y Data Marts. Si los datos operacionales de la compañía se encuentran en condiciones adecuadas (esto es que requieren mínima integración y/o mínima limpieza de datos), los componentes virtuales mencionados anteriormente pueden ser de gran utilidad para reducir el gap sin perder de vista los efectos en los sistemas.

Para el tiempo latente que se presenta en el análisis, están apareciendo tecnologías que ofrecen la supervisión de ciertas actividades empresariales tanto en términos operacionales como económicos y que ofrecen la posibilidad de emitir alarmas en el momento en que surge algún inconveniente con la información que monitorea. Otro de los factores que puede monitorearse ahora es el relacionado con los resultados operacionales consolidados de los Indicadores Clave de Desempeño (KPI's). La información que estos proveen se tiene en el momento actual y pueden ser reflejados en los portales empresariales para que las personas encargadas de dicha información tengan conocimiento de lo que está ocurriendo en ese momento.

Mientras agilizar la recolección, el análisis y el despliegue de datos operacionales es ciertamente útil para el negocio, debe recordarse que no todos los datos de Inteligencia de Negocios se deben incluir. Muchos implementadores de TI no son cuidadosos en las labores de determinar con precisión que datos deben ser incluidos en la aplicación de BI operacional. Ellos comenten el error catastrófico de incluir lo mas que puedan, forzando todos los datos a ser "en tiempo real", de este modo creando un mundo BI pesado y difícil de manejar. Aquí es donde se ve afectado el tiempo latente de decisión. El mensaje es evaluar cuidadosamente el empuje hacia un análisis en tiempo real. Entender las necesidades del negocio completamente y se podrá encontrar que es solo realmente un pequeño porcentaje de datos que deben apresurarse hacia las manos de los consumidores de negocio. En este punto es importante resaltar que la mayor parte de información analítica puede tener hasta meses de haber sido obtenida pero aún puede ser relevante para la toma de decisiones.

1.2.2.3. Análisis y Visualización de Datos más Sofisticados

Inicialmente la Inteligencia de Negocios consistía en proporcionar informes y responder a preguntas simples que se obtenían de consultas sin mucha elaboración; luego, vino el análisis multidimensional y aunque actualmente se continúe con el uso de cubos y esquemas de estrella, hacia lo que se dirige actualmente la Inteligencia de Negocios en términos de visualización es un

análisis más complejo de los datos que posee la organización. Técnicas como la explotación de información usando la minería de datos, los métodos estadísticos, etc. hacen parte de lo que actualmente se empieza a buscar para la Inteligencia de Negocios. Una de las facultades con las que se cuenta, es que las personas que no tienen amplio conocimiento en la forma de trabajar con las herramientas de visualización pueden hacerlo pues no son complejas de manejar y usar para suplir sus necesidades. Ahora, el análisis de datos en tiempo real permitiendo a quienes toman decisiones tener control sobre actividades diarias. Hay que tener presente que las actividades y logros que se obtienen con este análisis de datos debe ser incorporado completamente dentro de la organización para evitar caos y adicionalmente, se debe asumir que el conocimiento de los procesos de negocio y flujos de trabajo debe ser imperante ya que como consecuencia de no tener un conocimiento de lo anterior, las personas que son encargadas de la toma de decisiones, no sabrán dónde aplicar el conocimiento obtenido de los análisis y por ende, no se obtendrá provecho de dicho conocimiento.

Lo que se retiene de lo anterior, es que evidentemente hay que tener datos en un adecuado estado de calidad para tener la posibilidad de sacar provecho de las nuevas tecnologías que orientan a las organizaciones a tener cada vez más información (basada en sus datos) en tiempo real y sobre la que toman decisiones en el momento en que se obtiene. De no tener la total certeza sobre la calidad de los datos y requieran que se esté verificando si realmente la información refleja lo que son o deben ser los datos en términos de calidad, no se llegará a tener análisis y toma de decisiones sobre información en el momento en que ocurran los eventos.

1.2.3. Proceso de Inteligencia de Negocios

Un proceso hace referencia a una colección de actividades estructurales relacionadas que producen un valor para la organización, sus inversores o sus clientes.

El Proceso de Inteligencia de Negocios en las organizaciones, debe estar definido principalmente por las necesidades de información de la empresa y de las personas que son las encargadas de la toma de decisiones para la misma. Este proceso, está alimentado por los proyectos de Inteligencia de Negocios que se desarrollan al interior de la organización ya que de acuerdo a sus resultados, pueden modificar esas necesidades iniciales de la organización respecto a inteligencia. El ciclo de Inteligencia de Negocios que se adopta en las organizaciones más a menudo es el propuesto por Jan Herring (12), quien considerando el Modelo de inteligencia de la CIA propone el siguiente Ciclo de Inteligencia de Negocios:

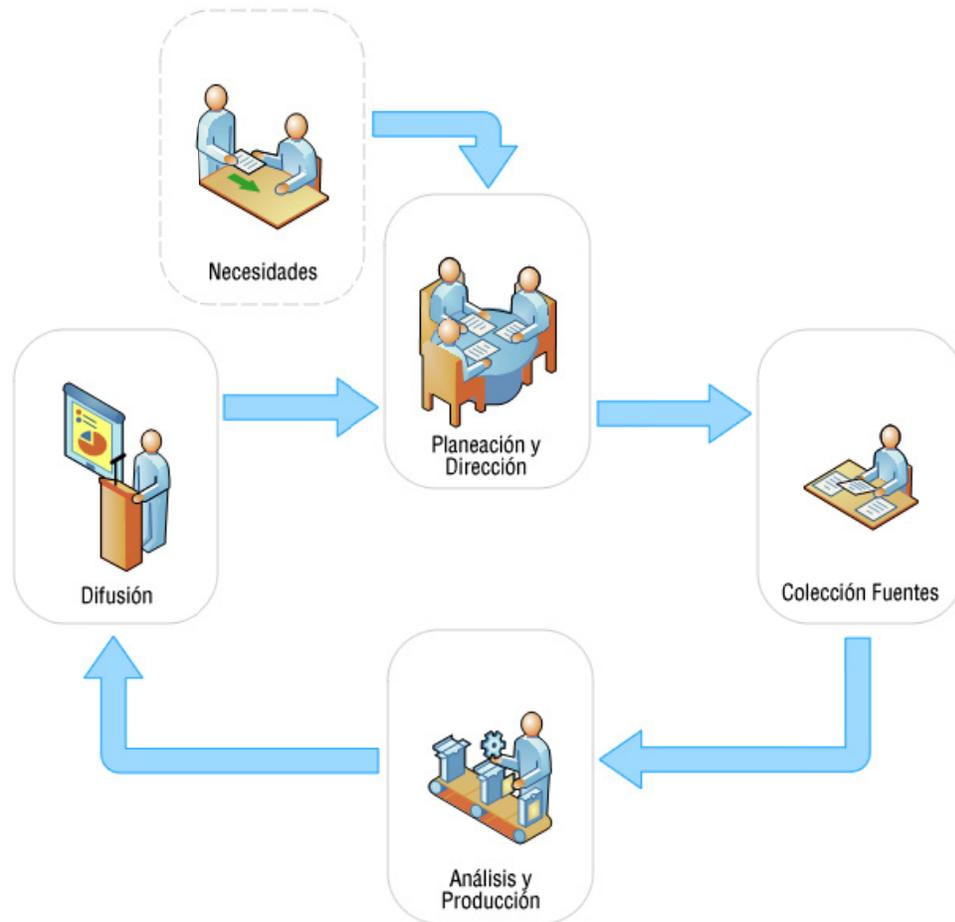


Figura 4 - Ciclo de Inteligencia de Negocios

Adaptado de la figura de Jan Herring (12)

1.2.3.1. Necesidades

El ciclo tiene su inicio en la definición de las necesidades de información por parte de los usuarios de la inteligencia en la organización y/o personas responsables de la toma de decisiones en la misma. En este punto los encargados de la planeación y dirección de las operaciones deben velar por que las necesidades de los usuarios sean satisfechas. En Para esta etapa es de vital importancia el acercamiento constante con los usuarios y la generación de discusiones que ayudan a identificar de forma exacta el foco del proyecto de inteligencia para así evitar que éste no logre los objetivos trazados.

Las necesidades mencionadas por los usuarios, son habitualmente definidas como Temas Clave de Inteligencia "KIT" (*Key Intelligence Topics*) las cuales ayudan a identificar de forma clara lo que se busca. De dichos KIT's se deriva lo que se conoce como Preguntas Clave de inteligencia "KIQ" (*Key Intelligence Questions*) las cuales son a las que debe apuntarse a resolver. Desde este punto, parte la identificación de lo que se requiere por parte de los usuarios y que debe ser obtenido de las fuentes de datos de la compañía para ser extractado y resolver los interrogantes propuestos.

1.2.3.2. Planeación y Dirección

Adicional a las necesidades propias de la inteligencia, expresadas por usuarios y la organización misma, es necesario que se tengan en cuenta aspectos como tiempos de entrega de la información requerida, medios de inteligencia a usar para llegar a obtener la información solicitada, etc. Luego de tener el compendio de lo anterior, puede recurrirse a la planeación de proyectos. Una vez en este punto, el director establece las fuentes a las cuales es necesario recurrir para extraer los datos indispensables para resolver las preguntas planteadas por las personas que toman las decisiones dentro de la organización, definir las reglas de negocio que deben aplicarse a dichos datos en la parte de la transformación para imprimirles calidad, etc.

1.2.3.3. Colección de Fuentes

Uno de los principales retos de un equipo de BI, es el de afrontar una gran cantidad de datos y a través de ellos llegar a lo que en realidad se desea. Muchas de las fuentes que se pueden encontrar pueden llegar a ser poco útiles o tener datos que no son completamente consistentes. Para dar solución a la situación anterior, es importante saber dónde más puede llegarse a obtener dichos datos y en caso de que no exista una fuente clara, comenzar a hacer una construcción partiendo de varios lugares.

Las fuentes a las cuales se recurrirá en busca de inteligencia, hacen referencia a dos formas de éstas: Las publicadas o literarias y las humanas. De las primeras, la información es obtenida a partir de documentación impresa o digital la cual constituye la mayor parte de los datos que van a ser extractados para ponerlos en el *Data Warehouse* para proveer a través de las consultas de los tomadores de decisiones la información requerida. De la segunda fuente, de la humana, que posee información por ser expertos del tema o porque poseen conocimiento que los capacita sobre el tema específico que se busca, pueden proveer dado el conocimiento que tienen de dichos datos, reglas de negocio aplicables a éstos para incrementar su calidad y dar significado a los datos que se llevarán al *Data Warehouse*.

1.2.3.4. Análisis y Producción

Habiendo extractado de las fuentes los datos requeridos y las reglas de negocio aplicables a dichos datos, se procede a hacer la transformación de los mismos para que evidentemente respondan a las necesidades de información de la organización.

Con esta transformación de los datos, se busca principalmente poder identificar hechos relevantes que sirvan como apoyo a la organización y de acuerdo a esto, analizar y determinar relaciones significantes entre los mismos que nos lleven a poder ampliar el panorama y poder tener bases más sólidas de información. Otro punto muy importante que se toca en esta etapa es la de poder comenzar a sacar conclusiones basándose en hallazgos que se vayan realizando a lo largo de dicho análisis.

Esta información es visualizada luego de ser cargada por medio de las consultas que se realizan en las herramientas que el proyecto de Inteligencia de Negocios provee para dicho fin.

1.2.3.5. Difusión

Una vez conseguida la inteligencia requerida por las personas que toman las decisiones, es enormemente relevante hacer la entrega y distribución de la misma. La inteligencia debe llegar a los interesados en ésta de forma oportuna, ser lo que realmente requieren, y debe ser una herramienta que verdaderamente pueda ser usada para encaminar a la organización en los aspectos que requiere.

Este es un paso crítico en todo el ciclo ya que este sintetiza todo lo realizado anteriormente. Esta difusión marca el camino de la conversión de información a inteligencia, en este momento es donde el usuario logra sacar todo el provecho que se busca y es cuando en realidad se transmite toda esta información de forma práctica y de fácil entendimiento para los usuarios.

Cada función del ciclo anteriormente expuesto y el conjunto como tal, debe ser adaptada a las necesidades específicas de la organización y ésta debe desarrollar de la mejor forma con respecto al cumplimiento de sus expectativas las funciones establecidas.

Como se verá luego cuando se trate el tema de las metodologías para la implementación de un proyecto de Inteligencia de Negocios, la metodología propuesta por *Gartner Research Group* tiene un doble bucle en este punto de Difusión que en su segunda parte, efectivamente propone una producción de cambios que lo que pretende es revisar problemas para ayudar a analizar y evaluar los cambios necesarios a realizar. Esta parte, efectivamente sugiere una reingeniería del proceso considerando esos cambios necesarios y así se retoma el Ciclo de Inteligencia de Negocios.

1.2.4. Proyecto de Inteligencia de Negocios

Para tener mayor claridad acerca de lo que es un proyecto, se va a definir éste como un esfuerzo único y temporal que culmina cuando se obtienen productos, servicios o resultados únicos; se elabora de forma gradual y es diferente a una actividad operativa.

Un proyecto de Inteligencia de Negocios es algo que se construye progresivamente y que no llega a una etapa donde se suspende, o por lo menos no debería ser así. Los resultados del o de los proyectos de Inteligencia de Negocios de la organización que los implementa, consciente o inconscientemente los incorporará a su diario vivir y se convertirá en componente de su proceso de Inteligencia de Negocios.

Desde que se capturan los datos de la compañía, hasta que éstos se convierten en información para apoyar la toma de decisiones de la organización, hay un camino que recorrer y que en cada uno de sus pasos, imprime calidad a esos datos que serán transformados hasta obtener la información relevante de la

compañía la cual será analizada por los interesados para tomar decisiones al respecto.

En términos de recursos tales como personas involucradas, dinero, recursos técnicos, etc. la empresa que se interesa en el desarrollo de un proyecto de Inteligencia de Negocios debe asegurarse de que efectivamente puede abarcar lo que implica emprender éste camino. Para tener una base de lo necesario, es posible recurrir a empresas del sector que tengan actualmente un proyecto de Inteligencia de Negocios en desarrollo y que estén en la capacidad de compartir la experiencia que se ha enfrentado pero eso sí, sin dejar de guardar las proporciones que se deban respetar entre organizaciones. En términos de la calidad de los datos, deberá hacerse un diagnóstico en el cual se define si la compañía que tiene planeada la implementación del proyecto puede avanzar en este o no. Haciendo un análisis de los datos que se encuentran en las bases de datos, hojas de cálculo y todas las fuentes donde la compañía pueda tener almacenados datos importantes para su proyecto de Inteligencia de Negocios, puede tomarse la determinación de continuar con el proyecto. Dentro de los aspectos básicos a evaluar, se encuentran la identidad única en las tablas de las bases de datos, las relaciones entre los atributos, la definición del dominio, entre otras. Teniendo esto, se requiere llevar a cabo la identificación de la información que requiere la organización para saber cuáles son los datos de las fuentes que deben ser extraídos, transformados y cargados en la bodega de datos que es de donde saldrán posteriormente dichos datos para ser visualizados por el usuario de la aplicación destinada para este fin.

Es muy importante tener presente que una vez se haga el diagnóstico de los datos de la compañía, las deficiencias encontradas pero que permiten que se avance con el proyecto de Inteligencia de Negocios deben mejorarse con el uso de mecanismos que no afecten en gran medida el tiempo de desarrollo del proyecto y que tampoco se conviertan en un obstáculo para el avance del mismo. Una vez se avanza en el proyecto, se piensa en el proceso de Extracción de los datos que el negocio requiere de acuerdo a los objetivos del proyecto. Un mecanismo usado, es almacenar estos datos requeridos en un contenedor temporal que es el que permitirá concentrarlos para posteriormente, transformarlos de acuerdo a las necesidades evidenciadas. Esta transformación, consiste en identificar las inconsistencias que los datos almacenados en el contenedor temporal puedan tener y crear ciertas "reglas" para resolverlas. Las reglas definidas para resolver las inconsistencias, son definidas por la compañía, áreas de negocio, o usuarios finales para los cuales se desarrolla el proyecto de Inteligencia de Negocios ya que ellos son los que conocen a qué hacen referencia sus datos.

Dentro de las inconsistencias que pueden encontrarse, se evidencian las abreviaciones de ciudades, las diferentes formas de ingresar las direcciones de los clientes, la falta de asociación entre ciertos datos que tienen un significado en otra tabla pero que cuando se hace la combinación de ambas, la tabla en la que se encuentran los datos a los que se hace referencia no están completos, problemas con las unidades manejadas en las diferentes áreas de la compañía, etc. En el caso en que durante la transformación, el dato que se captura no esté contenido dentro de las reglas que puedan corregir la inconsistencia que se presenta, ese dato debe ser revisado y debe posteriormente informarse a la persona autorizada para hacer el cambio requerido. Una vez se hacen estas

transformaciones requeridas, vuelven a almacenarse en el contenedor temporal y luego se procede a hacer la carga en la bodega de datos. Luego de esto, se continúa el proceso hacia la visualización.

Teniendo en cuenta lo anterior, se observa que durante la implementación de un proyecto de Inteligencia de Negocios, la Calidad de Datos está presente desde su inicio hasta el momento en que se incorpora completamente a la compañía. Si desde el comienzo del proyecto no se trabaja con datos con calidad, los resultados que se van a obtener cuando se estén haciendo consultas en la parte de la visualización, no van a corresponder con lo que realmente enfrenta la organización y las decisiones que se tomen basadas en la información errónea arrojada por los datos, serán igualmente erróneas.

1.2.5. Metodologías Para La Implementación De Una Solución BI

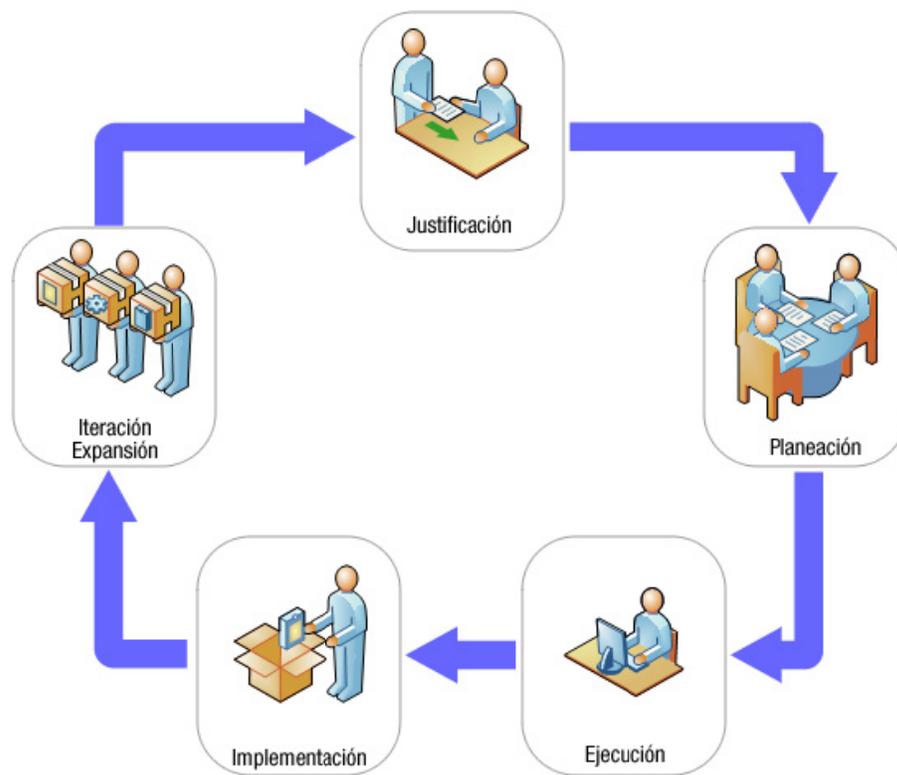


Figura 5 - Etapas de un proyecto de Inteligencia de Negocios

Para la implementación de un proyecto de Inteligencia de Negocios, existen metodologías que indican y facilitan la implementación de la solución requerida para la organización que así lo piensa hacer o en ocasiones, las organizaciones que se ponen en la tarea de implementar Inteligencia de Negocios en su interior, crean su propia metodología. Con el fin de tener una idea se presentan a continuación los puntos comunes de diferentes metodologías propuestas para la implementación de éste tipo de proyectos. Las metodologías analizadas fueron las siguientes: La propuesta por LARISSA T. MOSS, SHAKU ATRE., IBM, GARTNER RESEARCH GROUP y ATRE Group. (13)

El análisis de las cuatro metodologías realizado previamente por Carlos Esteban García y Nicolás Vélez Mesa en su proyecto de grado titulado: "Estudio sobre las metodologías para gestión de proyectos de BI, su aplicación y factores críticos de éxito" es el siguiente:

| Etapa | Metodologías | | | |
|----------------|--|---|--|--|
| | Larissa | IBM | Gartner | Atre |
| Justificación | Se encuentra en la etapa de justificación. | No tienen una etapa definida, pero resaltan la importancia de hacer análisis al ROI. | No cuentan con una etapa. Mencionan la necesidad de hacer análisis de costo y ROI. | No se define como una etapa separada, pero la trabajan dentro de la etapa de justificación. |
| Planeación | Se divide en 2 pasos: evaluación de la infraestructura (Técnica, no técnica), y planeación como tal. | Establecer/definir el proyecto y prepararlo. | Está dividida en 3 etapas: definición, identificación y preparación de datos, evaluación y selección de herramientas. | Toda la planeación se realiza en la etapa de justificación. |
| Ejecución | La dividen en tres etapas principalmente: análisis del negocio, diseño y construcción. | Iniciar la base de datos, lo que incluye toma de requisitos, análisis, diseño de implementación del DW. Además de explorar/probar la base de datos. | Se divide en tres etapas, definición de información del negocio, identificación y preparación de datos y evaluación y selección de herramientas. | La dividen en dos etapas, cada una de la cuales diseña y construye: infraestructura del proyecto y el almacenamiento de datos. |
| Implementación | La implementación se realiza en 2 procesos que denominan implementación y evaluación de la | Cuentan con una etapa para la implementación. | Cuenta con dos pasos: en el primero se desarrolla, implementa y capacita, en el siguiente los usuarios exploran y se descubren la | Se divide en 2 etapas: diseño y puesta en marcha de las herramientas para la generación de reportes, y manejo de |

| | efectividad. | | solución de BI. | capacitaciones y soporte. |
|---------------------|--|--|--|---|
| Iteración/Expansión | No existe una etapa clara, pero se menciona la necesidad de identificar deficiencias y nuevas necesidades. | Cuentan con una etapa para llevar un estricto control sobre el sistema luego de estar en producción. | Cuenta con una etapa llamada producir cambios. | Existe una etapa para la identificación de deficiencias o nuevas necesidades. |

Tabla 2 - Comparación de Cuatro Metodologías en la Gestión de Proyectos BI

De acuerdo a la tabla anterior donde se muestran las etapas con las cuales debe contar un proyecto de Inteligencia de Negocios para su implementación, se puede visualizar cuáles de dichas etapas requieren o presentan mayor criticidad con relación a la Calidad de Datos.

En la etapa de Justificación, lo que principalmente se busca es impulsar el proyecto dentro de la compañía. Se debe identificar lo que hace indispensable el llevar a cabo dicha inversión. Es posible que en esta etapa la Calidad de Datos no sea aún relevante pero sería adecuado considerar que de ésta dependerá gran parte del éxito del proyecto.

Se continúa con la Planeación, encontrando que se ponen en orden todos los factores para comenzar a desarrollar el proyecto y surge nuevamente la atención sobre la calidad de los datos que se requieren dentro del proyecto. Aparte de todas estas cosas, se consideraría importante comenzar a identificar si verdaderamente en el nivel de Calidad de Datos que se encuentra la organización, es viable el proyecto. De no ser adecuada para el proyecto, la Calidad de Datos es lo que primero debe mejorarse y llevarse a un nivel aceptable para posteriormente sí incorporar dichos datos en un proyecto de Inteligencia de Negocios. Indudablemente es de suma importancia considerar que la materia fundamental que es la que apalancará como resultado del proyecto la toma de decisiones de la compañía y la guiará en su objetivo de lograr grandes cosas, puede ser extremadamente contraproducente en caso de que esa materia fundamental no esté correcta.

En la etapa de Ejecución y teniendo en cuenta las necesidades expresadas por los usuarios respecto a los datos que se requieren visualizar en las aplicaciones resultantes del proyecto, deben identificarse las reglas de negocio que rigen dichos datos para llevar a cabo las transformaciones requeridas y así incrementar calidad a los datos que serán representados en los informes de prueba que se generarán durante el proceso de Extracción, Transformación y Carga. En este punto es indiscutible que se requiere de la Calidad de Datos para asegurar que la información que se obtiene a partir de los datos almacenados en el *Data Warehouse* luego del proceso ETL, refleja efectiva y fielmente la situación por la cual está pasando la compañía en un determinado momento y que esta

información de verdad contribuya con las decisiones que se tomen en la compañía.

En la parte de Implementación que es el momento en el cual se pone en marcha el lanzamiento de las herramientas para generación de reportes, se capacita a las personas que harán uso de ellas, etc., si el proceso ETL se ha llevado a cabo de una forma adecuada, podríamos confiar en que los datos que se cargan en el almacén son datos con calidad.

Cuando se llega a la etapa de Iteración/Expansión, en caso de presentarse fallas cuando se visualiza la información que fue solicitada por los usuarios, debe recurrirse a la verificación de la calidad de los datos que están siendo los proveedores de dicha información. Es posible que las reglas del negocio no hayan sido adecuadamente establecidas, que los datos solicitados requieren otros nuevos para mostrar lo que se quiere o que las reglas de negocio se hayan modificado, entre otras. Aunque la verificación de la Calidad de Datos se incrementa de manera permanente en el proceso ETL, es muy importante hacer comprobaciones posteriores como en la visualización.

2. Calidad de Datos en un Entorno de Inteligencia de Negocios

2.1. Antecedentes – Importancia del Tema

Las empresas hoy en día requieren confiar en los resultados que arrojan sus sistemas de información para la toma de decisiones correctas y oportunas, todos estos resultados son provenientes de una serie de datos que se encuentran almacenados la mayoría de las veces en diferentes fuentes, lo que hace de la tarea de recolectar, limpiar y transformar los datos todo un reto en las organizaciones. Al tomar la decisión de abordar una estrategia de Inteligencia de Negocios y siendo el objetivo principal de ésta el aprovechamiento máximo de los datos almacenados en la empresa se debe pensar en el mejor estado de los mismos.

A medida que aumenta la capacidad de almacenamiento de los ordenadores y servidores, las empresas almacenan más y más datos perdiendo muchas veces el control sobre estos y generando la necesidad de mayor almacenamiento, igualmente los datos almacenados son la base para generar la información sobre la cual se toman las decisiones tanto de la operación diaria como de proyección y estrategia del negocio, todo este proceso de recolección, transformación y análisis de datos se debe realizar con precisión y rapidez con el fin de reaccionar a tiempo a los cambios del mercado, como respuesta a esta necesidad surge el *Data Warehouse* como un componente de Inteligencia de Negocios.

Es un factor común en las organizaciones tener los datos almacenados en diferentes repositorios y bajo distintas condiciones, lo que conlleva a que se presenten numerosos problemas en el momento de unificar conceptos y datos. Para tener una respuesta asertiva del proyecto de Inteligencia de Negocios se debe asegurar la calidad de los datos almacenados en el *Data Warehouse*, ya que este es un factor trascendental para la toma de decisiones correctas y la orientación estratégica del negocio. Según el **1º Estudio sobre Calidad de Datos en la Empresa Española**(14), las amenazas generadas por la baja calidad de los datos que más afectarían a la gestión de sus negocios serían, “*Dificultad en la toma de decisiones*” (84%), “*Pérdida de oportunidades de negocio*” (77%) y “*Aumento de costes*” (76,6%).

2.2. Diferentes Propuestas Metodológicas que Existen

Para garantizar la Calidad de Datos en los proyectos de Inteligencia de Negocios, no existe una metodología propuesta que se pueda usar ya que la forma como se afrontan estos proyectos depende del tipo de empresa, el estado de los datos, el equipo de trabajo y los recursos que se tengan destinados para esto. Sin embargo, existen empresas proveedoras que han definido metodologías para hacer evaluaciones sobre la calidad de los datos de una forma genérica (no para un proyecto de Inteligencia de Negocios), a continuación se presentan varias metodologías para realizar esta evaluación y en ocasiones hasta el mejoramiento de dicha calidad.

2.2.1. Metodología para la Evaluación de la Calidad de Datos: Business Objects

La metodología de evaluación de calidad definida por **Business Objects** está compuesta por cuatro etapas:

- 1- Planificación de la pre-evaluación.
- 2- Realización de la evaluación de datos.
- 3- Presentación de los resultados de la evaluación.
- 4- Repetir el proceso.

Cada etapa del proceso de la evaluación tiene una serie de pasos, como en cualquier proyecto de TI una evaluación de este tipo requiere de una planeación para garantizar el éxito, por lo cual durante la primera fase se establecen los requisitos del proyecto, su duración y las personas que van a participar en él.

| Metodología de Evaluación de Calidad de Datos de Business Objects | |
|--|---|
| 1- Planificación de la pre-evaluación | 2- Realización de la evaluación de datos |
| <ul style="list-style-type: none"> • Crear el equipo de trabajo • Seleccionar la muestra de los datos • Revisión de la documentación • Identificar las metas u objetivos de la evaluación • Definir las necesidades del proyecto • Establecer el alcance de la evaluación de datos | <ul style="list-style-type: none"> • Perfil de los datos • Analizar y poner a prueba los datos • Realizar pruebas de integridad referencial • Examen de Calidad de Datos • Revisar que los datos estén de acuerdo con las reglas del negocio |
| 3- Presentación de los resultados de la evaluación | 4- Repetir el proceso |
| <ul style="list-style-type: none"> • Compilar los resultados • Revisión con los expertos en el tema • Preparar informes escritos y exposiciones orales. | <ul style="list-style-type: none"> • Repetir periódicamente evaluaciones • Mantener a los expertos de los temas relacionados implicados en la identificación y evolución de las necesidades. |

Tabla 3 - Metodología de Evaluación de Calidad de Datos de Business Objects

(15)

2.2.2. Metodología para el Análisis de la Calidad de Datos: IP-UML

Esta metodología, usa un lenguaje de modelado para hacer la evaluación de la Calidad de Datos. El proceso y la relación de los diagramas que se usan para hacer dicha evaluación en cada etapa del proceso se encuentran a continuación.(16)

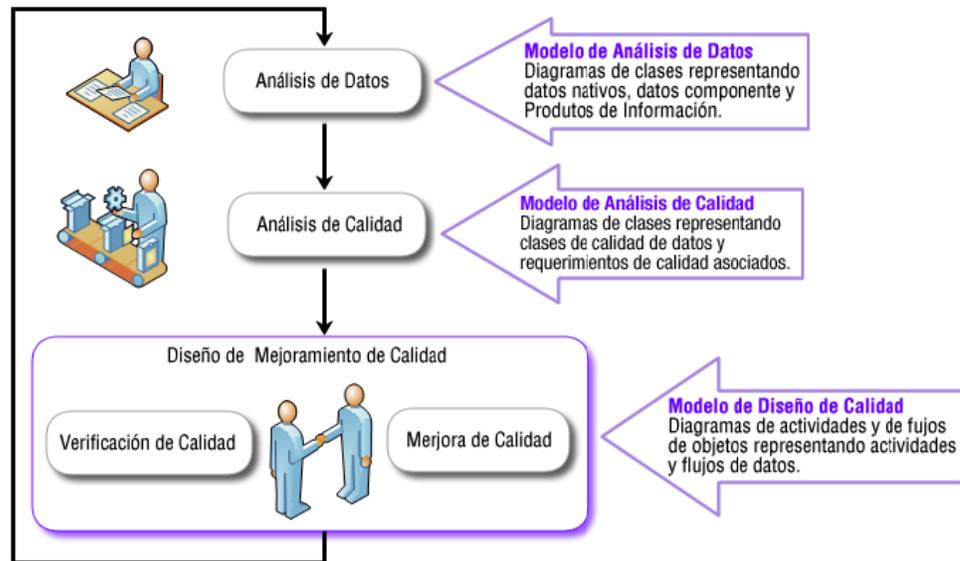


Figura 6 - Ciclo de la Metodología IP-UML

2.2.2.1. Análisis de Datos:

Implica el análisis de los datos que son intercambiados, almacenados, actualizados y modificados por procesos de una sola organización (o sistemas de información) o de múltiples organizaciones (o sistemas de información). El foco principal está en los datos que son críticos para el éxito de las organizaciones ya que la mala calidad puede causar costos en términos, por ejemplo, de pérdidas de ventas o generar multas cuando la calidad se debe garantizar según requisitos contractuales. El resultado de la fase de análisis de datos es el modelo del análisis de datos. Tal modelo consiste en un conjunto de diagramas de clase que describen productos de información identificados y sus componentes en términos de datos nativos (sin transformar) y componentes de datos, modelados por el perfil de la Calidad de Datos.

2.2.2.2. Análisis de Calidad:

La fase de análisis de la calidad especifica la calidad requerida en términos de dimensiones de la calidad, para los datos identificados en la fase de análisis de datos. Para cada dimensión, el perfil de la calidad de los datos introduce las clases estereotipadas asociadas a una clase de datos de calidad. El resultado de esta fase es un modelo del análisis de la calidad, que consiste en un conjunto de diagramas de clase, describiendo clases de Calidad de Datos con los requisitos específicos de calidad.

2.2.2.3. Diseño del Mejoramiento de la Calidad:

Esta fase se puede dividir en dos:

- Verificación de Calidad
- Mejoramiento de Calidad

En la fase de la verificación de la calidad, se describe el proceso por el cual el producto de información es elaborado. Por otra parte, si se detectan embotellamientos de calidad en tal proceso, los chequeos de calidad se pueden introducir para enfrentar directamente el problema controlando los datos en los puntos más débiles del proceso. El resultado de esta sub-fase es un conjunto de diagramas de actividades, en los cuales los flujos del objeto se describen para modelar intercambios de datos y manipulaciones. Los diagramas de actividades son dibujados usando las actividades estereotipo, los agentes y las relaciones establecidas previamente. Aunque la fase de la verificación de la calidad soluciona problemas de calidad en muchos casos, acciones más complejas de mejoramiento de calidad necesitan a menudo ser comprometidas, implicando no simplemente la introducción de puntos de control, sino un proceso completo de reingeniería conducido por requisitos de calidad. En la fase de mejoramiento de calidad, las acciones de mejoramiento se definen a tal alcance. Específicamente, esta fase consiste en un diseño a nivel de procesos que realizan la mejora de la calidad. Como en la fase de verificación de la calidad, los diagramas de actividades y los flujos del objeto son dibujados usando lenguajes como UML. Los diagramas de verificación de la calidad y de las fases del mejoramiento de calidad constituyen el modelo del diseño de la calidad.

2.2.3. Metodología Total Information Quality Management (TIQM)

Esta metodología fue desarrollada basándose en principios de Calidad de Datos y en técnicas que han obtenido resultados óptimos en las industrias manufactureras y otras industrias.(17)

Esta metodología busca el mejoramiento continuo de la Calidad de Datos y la creación de hábitos de análisis en los mismos mediante procesos certificados y buscando un cambio cultural en el manejo de la información.

TIQM define procesos sobre los mismos procesos de mejoramiento continuo divididos en 6 pasos.

A continuación se presentan los pasos que constituyen esta metodología.

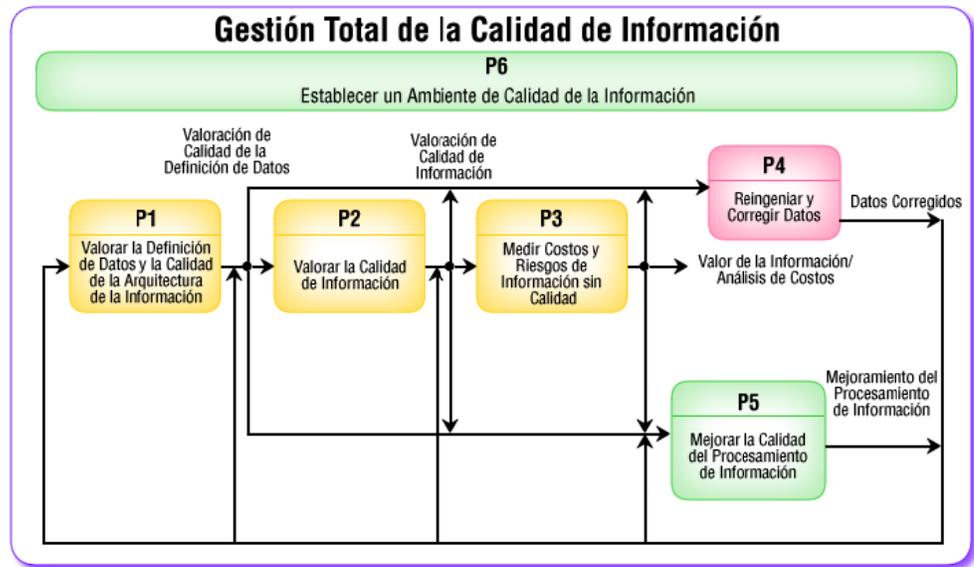


Figura 7 - Procesos de la Metodología TIQM

2.2.3.1. P1. Valorar la Definición de Datos de Calidad de la Arquitectura de la Información

Este proceso precisa la forma de medir la calidad de los datos definidos así como la información que los empleados requieren y la forma como ellos la entienden. Este proceso también permite medir la calidad de los modelos de datos y diseños de bases de datos.

2.2.3.2. P2. Valorar la Calidad de la Información

Este proceso define como medir la calidad de la información, permite identificar varias características de calidad, como la exactitud, la entereza, la no duplicación y la consistencia a través de múltiples bases de datos. Este proceso mide el estado de la calidad de los datos dentro de una base de datos o colección de datos. Igualmente permite valorar la calidad de los datos producidos por un proceso corriente.

2.2.3.3. P3. Medir Costos y Riesgos de Información sin Calidad

Este proceso define cómo establecer un caso de negocio para la gestión de la calidad de información. Midiendo los costos de las fallas en los procesos, fragmentación y reproceso de información, pérdida de clientes y negocios, y oportunidades perdidas. Usted puede "hablar con datos" (como dice Kaoru Ishikawa). Esto es, puede presentar esta información en términos de negocio hacia la administración de negocios.

2.2.3.4. P4. Re ingeniar y Corregir Datos

Este proceso define como conducir proyectos de corrección de datos, transformar la información y controlar los procesos de movimiento de datos para depósitos de datos o proyectos de conversión de datos. El proceso de corrección de datos no es un proceso independiente.

2.2.3.5. P5. Mejorar la Calidad del Procesamiento de Información

En este proceso se hace un llamado a una técnica muy conocida de prevención de datos defectuosos. Este proceso implementa el PDCA (Plan-Do-Check-Act). Es por esto que en este proceso se identifican y mejoran los procesos que entregan las empresas datos defectuosos.

2.2.3.6. P6. Establecer un Ambiente de Calidad de la Información

Este paso no es un proceso que tiene definido un principio y un final. Aquí se describen los 14 Puntos propuestos por Deming que deben ser inculcados en la cultura de la empresa para poder tener un buen manejo de los datos y para crear y sostener un ambiente adecuado para el funcionamiento del negocio y el hábito de la mejora continua.

La técnica que usa esta metodología, se encuentra en los títulos siguientes como Data Defect Prevention.

2.2.4. Proceso de Control de la Calidad de Datos

Cualquier iniciativa de Calidad de Datos necesita procesos bien definidos para ser seguida y permitir maximizar el control de calidad. La siguiente figura proporciona los detalles del proceso del control de Calidad de Datos implicado en la valoración, limpieza e incremento de datos, y seguimiento y reporte de las anomalías de los datos.(18)

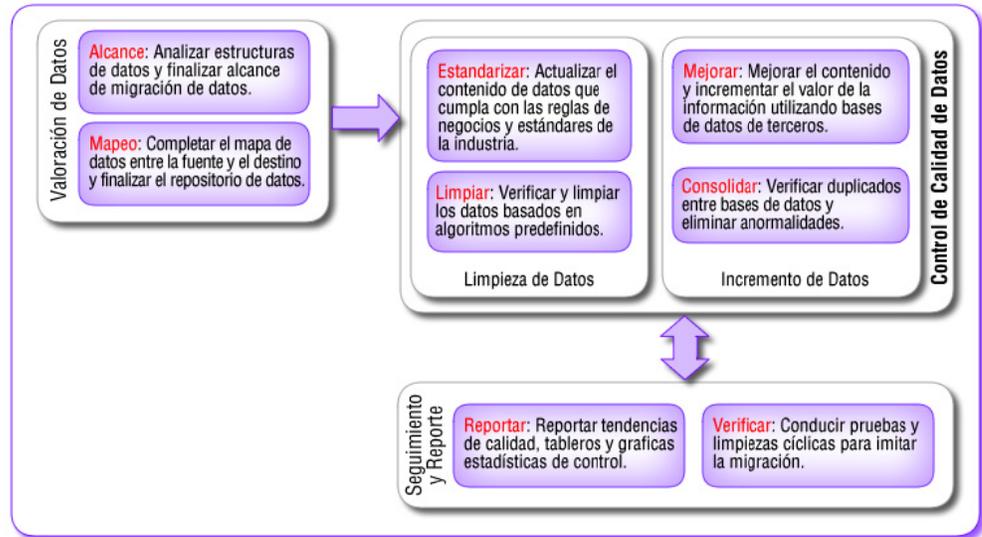


Figura 8 - Proceso de Control de la Calidad de Datos

2.2.4.1. Valoración de Datos

La fase de valoración de los datos consiste en el analizar las estructuras de datos y el concluir del alcance de las intervenciones de Calidad de Datos. Esto es una fase importante para la gestión de datos pues las prioridades de esfuerzos de la corrección se determinan en esta fase. En el mapeo de datos o la fase de perfilamiento, el mapeo de punta a punta es llevado a cabo entre los sistemas origen y destino. El perfilamiento proporciona más visibilidad a los requisitos de limpieza. Los informes de limpieza pueden ser creados y priorizados en las entidades que son críticas para la migración (por ejemplo, la entidad del cliente se debe limpiar a fondo para evitar que afecte entidades dependientes tales como contratos o cuentas pendientes).

2.2.4.2. Control de Calidad de Datos

La fase del control de Calidad de Datos se centra en la corrección y estandarización de los datos e implica la limpieza e incremento de los datos. Los datos se evalúan para determinar los errores, las anomalías, las duplicaciones y las insuficiencias. La fase implica la detección de errores y la corrección de datos para controlar la integridad de datos en un cierto plazo.

2.2.4.3. Seguimiento y Reporte

La siguiente figura habla sobre el monitoreo y seguimiento del ejercicio de la limpieza de datos, lo cual es importante entender la disponibilidad de datos y también el progreso. Permite a los usuarios saber si la corrección ha sido exitosa o no. Provee la habilidad de priorizar la Calidad de Datos basados en la severidad de evento y la línea de tendencia. El desempeño de los equipos de Calidad de Datos puede ser valorado en la acción tomada en la calidad de los eventos reportados. El seguimiento también ayuda a identificar las causas más comunes

de error y ayuda a explorar la posibilidad de procedimientos automatizados de corrección.

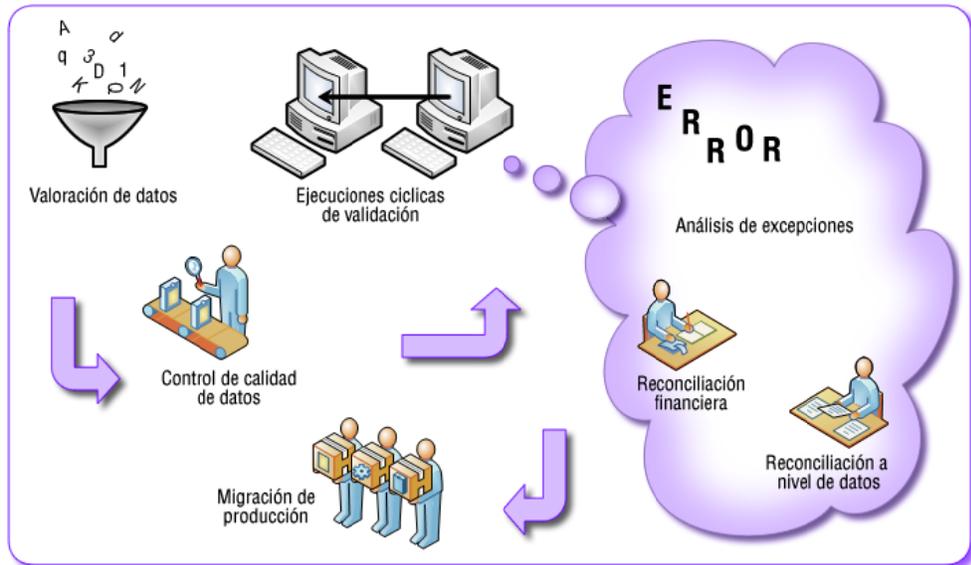


Figura 9 - Monitoreo y Seguimiento de la Calidad de Datos

2.3. Calidad de Datos en las Fuentes

2.3.1. Introducción

“El activo más importante de una empresa es la información. La capacidad de las empresas para competir, adaptarse y crecer en un clima de negocios con rápidos cambios es dependiente en gran medida de lo bien que la empresa los utiliza para la toma de decisiones. El intercambio de información que no está limpia y consolidada en su máxima medida, puede reducir sustancialmente la efectividad de un sistema de inversión significativo y el considerable potencial de la rentabilidad.”

(Stoker, 1999)

Para el desarrollo de un proyecto de Inteligencia de Negocios, los datos que proporcionarán la información requerida por los usuarios son indudablemente de gran importancia y debido a esto, hay que tener mucho cuidado con las características que poseen los datos a usar en el proyecto.(19)

Un sinnúmero de factores tales como abreviaturas, diferencia de significado de acuerdo a la forma en la que se denotan los datos, la no completitud de información de registros y falta de definición de dominio, entre otras, son algunas de las causas por las cuales los datos pueden considerarse faltos de calidad. Así entonces, es necesario hacer una limpieza de los datos que serán usados para proporcionar la solución a la consulta de los usuarios del producto software resultante del proyecto de Inteligencia de Negocios.

Más concretamente los problemas más comunes con los datos son los que se enumeran a continuación:

- Valores simulados.
- Ausencia de datos.
- Campos multipropósito.
- datos cifrados.
- datos contradictorios.
- Uso inadecuado de las líneas de dirección.
- Violación de las reglas de negocio.
- Reutilización de claves primarias.
- Falta de identificadores exclusivos.
- Problemas de integración de datos.

Como dijimos anteriormente, los datos son un producto que puede ser caracterizado como con calidad o sin calidad. La habilidad para tomar decisiones de calidad depende de los datos que sean utilizados para obtener la información requerida y que ésta sea de calidad.

Sin embargo, con mucha frecuencia, los datos que se manejan en las organizaciones no son de calidad y tienen problemas como los mencionados anteriormente de completitud, significado, etc. Dichos errores, provienen tanto de fallas en la definición inicial de la arquitectura de las estructuras encargadas del almacenamiento de los registros con los cuales opera la empresa, como de los errores que introducen las personas que son las encargadas de digitar la información a almacenar para conservar datos importantes como los de un cliente por ejemplo.

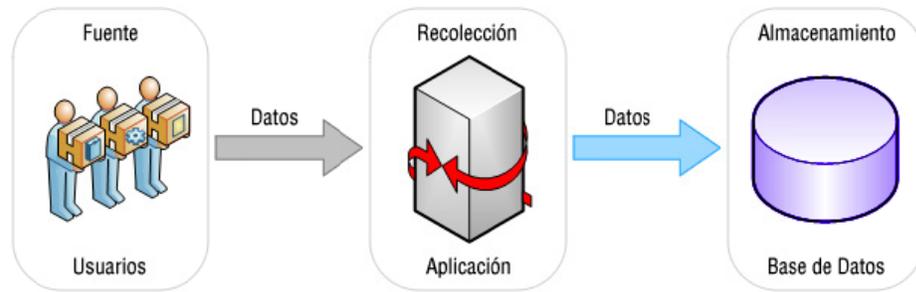


Figura 10 - Proceso de Adquisición de datos

Para asegurarnos que los datos que estamos usando como base de la información para la toma de decisiones dentro de una organización sean efectivamente datos de calidad, debemos recurrir a la limpieza de datos. La limpieza de datos es el proceso que asegura que la misma pieza de información es referenciada sólo de una forma dentro de un almacén de datos. Cuando los datos están limpios, sus usuarios se pueden concentrar en su uso y no en la credibilidad que pueda brindársele a éstos.

Se identifican 3 niveles en los cuales se introducen errores en los datos, errores que requerirán posteriormente procesos de limpieza para poder usarlos en la toma de decisiones.

2.3.1.1. Los Usuarios

Los usuarios quienes pasan los datos al nivel de aplicación son los primeros que deben tener cuidado en introducir datos correctos, ellos toman los datos directamente del entorno de producción, ya sea en un punto de venta en el momento de atender a un cliente, o en el almacén en el momento de realizar inventarios o registro de entrada de nuevos insumos, en la toma de información en líneas de soporte, o incluso (aunque se creería que pocas veces suceden errores) en la taquilla del banco el cajero introduciendo los datos de una transacción.

Dentro de este proceso para controlar el ingreso de errores es necesario aparte de ayudar a los usuarios haciendo uso en el desarrollo de aplicaciones con validaciones que les impidan en cierto modo ingresar datos defectuoso, educarlos para que respeten los datos, ya que estos constituyen la

información de la empresa, información de alto valor, por el respeto de los datos nos referimos a valorarlos, es decir, evitar abreviaturas no definidas por el sistema, enseñarlos a seguir estándares que se hayan definido para el ingreso de esos datos, seguir las recomendaciones definidas en el uso de los aplicativos, o en casos donde no exista un aplicativo sino que es una hoja electrónica lo que se utiliza para depositar los datos, diligenciar la información de modo correcto, escribir números telefónicos en el formato adecuado, direcciones de correspondencia como deben ser, números de identificación, entre otros. Es muy importante reconocer que comúnmente la labor de las personas encargadas de la digitación de los datos para ingresarlos a un lugar de almacenamiento no es incentivada de forma correcta. Dichas personas no son conscientes de la importancia de la labor que ejecutan simplemente porque nadie se los ha dicho y tampoco les han expresado las consecuencias posteriores a una mala práctica de digitación de datos.

2.3.1.2. La Aplicación

La aplicación, cuando son casos diferentes a hojas electrónicas, y demás formatos libres de almacenamiento de datos, tiene una responsabilidad de verificar que los datos que ingresa el usuario cumplen con una definición o un formato adecuado, esta responsabilidad cae en manos del desarrollador de la aplicación, ya sea una aplicación desarrollada a la medida para la empresa o una aplicación comercial adquirida en el mercado, aunque los usuarios sean conscientes del valor de los datos, bien dice el adagio popular *Errare humanum est* (es propio del hombre el engañarse), esto pueden cometer errores durante la digitación de los datos en el sistema, por tal motivo y buscando minimizar los errores en el ingreso de la información la tecnología ha buscado elementos que automaticen muchas labores que eran asignadas a usuarios, por ejemplo la recepción de suministros puede realizarse de manera automática insertando sensores en los productos para que estos se “autoregistren” ante el sistema de inventario, tanto al momento de ingresar como al momento de salir, mantenido el inventario actualizado en tiempo real.

Una forma más en la cual las aplicaciones protegen la limpieza de los datos es ofrecer valores predeterminados y validaciones para campos como nombres de ciudades, genero, tipos de producto, entre otras opciones, las cuales han de estar almacenadas en la base de datos o directamente codificadas en la aplicación que restringen el rango de posibles valores que el usuario pueda darle a los datos en un conjunto de información reduciendo así la fragmentación de datos en el sistema.

2.3.1.3. La Base de Datos

El tercer elemento pero no menos importante es la base de datos, igualmente bajo responsabilidad del desarrollador, o en muchos casos de un administrador de base de datos, el cual debe velar por que la base de datos refleje las reglas de los requisitos de información que requiere la empresa en sus aplicaciones, esto es manteniendo la integridad referencial de los datos,

verificando que los tipos de datos son adecuados a lo que se espera almacenar en los diferentes campos, evitar al máximo campos de propósito general, ya que son difíciles de procesar de forma automática, verificando que el diseño de la base de datos este normalizado al nivel adecuado, entre otras funciones con el objetivo de evitar que las aplicaciones almacenen datos que no correspondan al nivel de limpieza esperado.

2.3.2. Técnicas para Evaluar la Calidad de Datos

Existen numerosas técnicas para abordar el tema de la Calidad de Datos en las organizaciones. La selección de la técnica a utilizar depende del nivel de madurez que la empresa tenga respecto a la gestión de los datos, funcionamiento del negocio y los procesos que lo componen, al igual que del presupuesto del que se disponga en la empresa para llevar a cabo dichas técnicas y del estado actual de los datos. Estas técnicas pueden ser usadas en diferente orden y combinaciones según el tipo de empresa, estado de los datos, empresa proveedora (en caso que exista), entre otros. Dichas técnicas se verán reflejadas en las diferentes etapas definidas para llevar a cabo un proyecto de Inteligencia de Negocios y que tienen actividades sobre la Calidad de Datos.

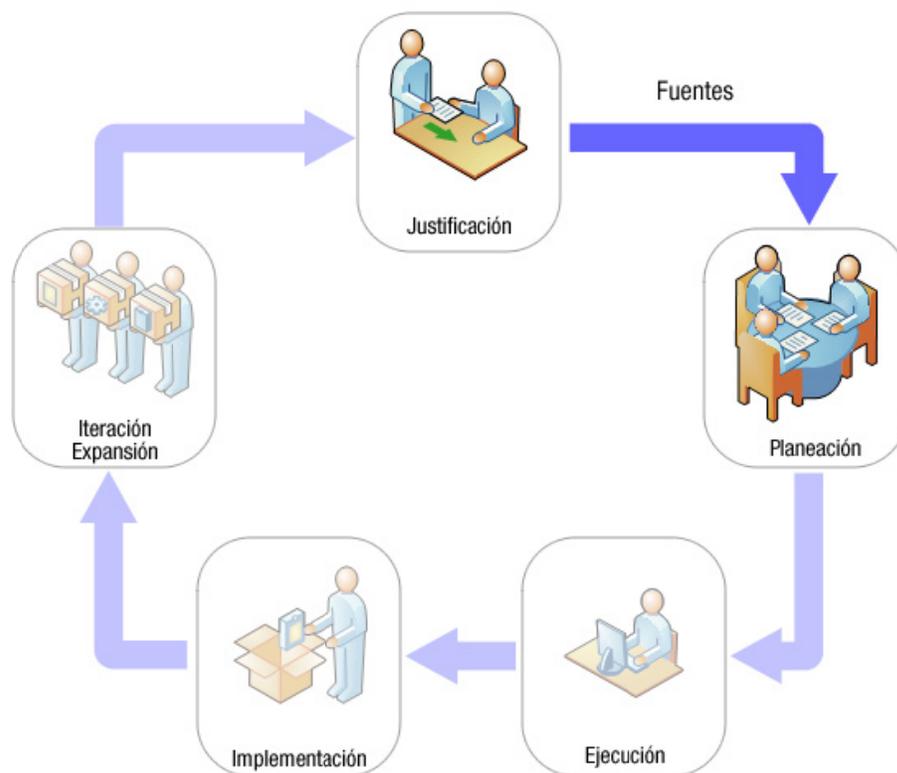


Figura 11 - Técnicas aplicadas en las Fuentes

2.3.2.1. Técnica: Perfilamiento de Datos (Data Profiling)

“Data profiling es el proceso de examinar los datos que existen en las fuentes de origen de una organización y recopilar estadísticas e información sobre los mismos”. (20)

El propósito de dichas estadísticas es:

- Conseguir métricas de Calidad de Datos que incluyen si los datos cumplen los estándares de la organización.
- Reduce el riesgo de integrar información a nuevas aplicaciones dado que conocemos su estado.
- Permite hacer un seguimiento de la Calidad de Datos.
- Capacidad de entender problemas derivados de los datos en proyectos que hagan uso intensivo de los mismos.
- Tener una visión global de los datos de la organización para desplegar políticas de Data Governance.
- Encontrar patrones en campos, porcentaje de nulos, distribución de valores, comparación con expresiones regulares, al igual que patrones con tablas relacionadas, columnas candidatas a clave, dependencias funcionales entre columnas.

(21)

Para llevar a cabo dicha técnica, es recomendable tener en cuenta lo siguiente:

2.3.2.1.1. El Perfilamiento de Datos Ayuda a Planear Mejor, Así que los Proyectos se Ejecutan Mejor.

Los productos a entregar esperados de la integración de datos y de las soluciones de Calidad de Datos son datos integrados y datos limpios – No perfiles de datos. Puesto que el perfilamiento de datos no produce el último entregable, es duro racionalizar el tiempo y los recursos destinados a él. Sin embargo, no se debe escatimar en el perfilamiento de los datos, porque éste da la integración de datos y soluciones de Calidad de Datos de mayor calidad, más el alcance exacto del proyecto, y una reducción de "atrapadas" que pueden surgir en las pruebas y despliegue.

2.3.2.1.2. El Perfilamiento de Datos es una Tarea que se Repite, no Algo que se Ejecuta una Vez.

Cuando se definen nuevas fuentes de datos se requiere crear perfiles nuevos con base en estas fuentes. Es por esto que cuando se trabaja de forma proactiva y se busca darle valor a los datos, estos resultados se expanden dentro de la empresa.

2.3.2.1.3. El Perfilamiento de los Datos es en Realidad Dos Prácticas, Aplicadas en Diversas Fases del Proyecto.

La mayoría del perfilamiento de datos se realiza con el fin de realizar el descubrimiento de datos y documentación en fases tempranas de un proyecto, como por ejemplo en las fases de planeación, diseño y desarrollo. Luego cuando se avanza en las fases el perfilamiento de datos cambia de rol y cumple una función de monitoreo donde diariamente se realizan integración de datos nuevos o actualizaciones con el fin de asegurar la calidad de los datos existentes. De ahí que esta sea conocida como una de las etapas más extensas que hay.

2.3.2.1.4. Los Datos que se Perfilan con una Herramienta Tienen Ventajas Sobre Métodos Manuales.

La mayoría de los usuarios de perfilamiento de datos normalmente ejecutan queries para ejecutar búsquedas de información y la van registrando en archivos para luego ser procesada. Mucha de esta información se amontona y luego es difícil de realizar una integración de datos o asegurar la calidad de los mismos. Por esto los métodos manuales son menos eficientes y a su vez gastan más tiempo en datos que luego pueden ser no fiables. Lo ideal o en lo posible los usuarios deberían usar una herramienta para conseguir mayor precisión y productividad a la hora de perfilar datos.

2.3.2.1.5. La Supervisión de los Datos es Casi Imposible Sin una Herramienta.

Diariamente se realizan operaciones de monitoreo de datos para luego registrar su estado. Estos procesos toman mucho tiempo y son costosos. Para mayor eficacia es recomendable la utilización de herramientas de automatización de dichas tareas, es el caso de herramientas de software, estas herramientas fuera de agilizar las tareas ayudan a apoyar las reglas del negocio y a identificar errores que de forma manual no son verificados.

2.3.2.2. Técnica: Limpieza de Datos (Data Cleansing)

“Data cleansing es el proceso de detectar o descubrir y corregir datos corruptos, incoherentes o erróneos de un conjunto de datos. Después del proceso la información será consistente con otros conjuntos similares de datos. La validación de datos puede ser estricta o mediante el uso de fuzzy logic.

Este proceso permite detectar entradas duplicadas, incompletas, etc. y establecer reglas para corregirlas. El objetivo no es borrar información pero sí mejorar la calidad de los datos construyendo un proceso de mejora continua.”
(20)

Las herramientas de Data Cleansing, permiten limpiar, validar y corregir datos de las fuentes de origen durante el proceso de incorporación y creación a un *Data Warehouse*. Por lo que es importante analizar las fuentes de los datos y clasificar los que son críticos, importantes o insignificantes para la empresa, para de este modo concentrarse en primera instancia en la limpieza de los datos críticos.

Es importante tener en cuenta que la limpieza no es algo que en ocasiones sea aplicada a todos los repositorios de datos de la organización. Es claro identificar

el hecho de que no puede tomarse todo el tiempo para hacer una limpieza exhaustiva de los registros que una compañía posee, por lo anterior es vital comprender que algunas veces la respuesta a la pregunta acerca de la posibilidad de limpiar todas las fuentes es negativa, al igual que la de la pregunta sobre si debe ser limpiada. (22)

La limpieza de datos no se refiere únicamente a la actualización de datos incorrectos por unos correctos en un registro. Para llevarla a cabo de forma seria, es necesario desensamblar y re ensamblar los datos lo que posibilita una corrección más óptima de éstos.

Los pasos en los que podemos descomponer la limpieza de datos son: Separación de registros en elementos, Estandarización o Normalización, Verificación, Coincidencia, Consolidación y Documentación.

La primera etapa, la de Separación de registros, hace referencia a la división en las mínimas unidades en las que pueden ser desagregados los datos que conforman un registro. Localiza e identifica los distintos elementos de datos en los archivos fuente y, a continuación, estos datos son puestos en el archivo objetivo.

Observando esto de una forma ejemplificada, tomaremos como registro inicial el siguiente:

Santiago H. y Verónica Cárdenas acudientes de Cárdenas, Diego A.
Apt 116
Kr 34 No. 32B-28
Patio Bonito, Poblado
Medellín, Amz

Lo que se hace en el primer paso para llevar a cabo la limpieza de datos es desagregar el anterior registro de la siguiente forma:

Primer Nombre Destinatario (1): *Santiago*
Inicial Segundo Nombre Destinatario (1): *H*
Apellido Destinatario (1): *Cárdenas*
Primer Nombre Destinatario (2): *Verónica*
Apellido Destinatario (2): *Cárdenas*
Relación Destinatarios: *Acudientes de*
Primer Nombre Persona Relacionada: *Diego*
Inicial Segundo Nombre Persona Relacionada: *A*
Apellido Persona Relacionada: *Cárdenas*
Tipo de la Calle: *Carrera*
Numero de la Calle: *34*
Intersección: *32B*
Distancia desde la esquina: *28*
Apartamento: *116*
Barrio: *Patio Bonito*
Comuna: *El Poblado*
Ciudad: *Medellín*
Departamento: *Amazonas*

Los elementos resultantes de la desagregación, corresponden a los campos que se encuentran en el registro que se está analizando. Continuando con el segundo paso de la limpieza, se procede a normalizar o estandarizar los elementos resultantes. Tres de los elementos que deben ponerse en una forma más estándar. Se ha descartado "Apt" porque lo hemos reconocido como "Apartamento", se puede sospechar que "Kr 34" en realidad debería decir "Carrera 34", se realiza este cambio provisional, y en el ciclo de verificación asegúrese que el valor actual de la calle sea "Carrera 34", igualmente cambiar "Amz" por "Amazonas".

El tercer paso consiste en verificar la coherencia de los datos que fueron normalizados o simplemente revisar si hay algún error en el contenido de lo que obtuvimos. Aunque puede que no sea obvio, hay un error evidente contenido en la dirección. Medellín, El Poblado y Patio Bonito son Ciudad, Comuna y Barrio de Antioquia no de Amazonas. Dado que tres partes de la información apuntan a Antioquia, haremos este cambio de Departamento. Es de importancia tener presente este registro y verificar si se tienen otras direcciones de los destinatarios para corroborar si el Departamento es correcto.

Posteriormente de la realización de los tres pasos anteriores, es perfectamente posible dirigirse hacia los siguientes pasos: Coincidencia y Familiaridad. La coincidencia consiste en encontrar otros registros de clientes en los cuales aparezcan Santiago Cárdenas y Verónica Cárdenas y asegurarse de que todos los elementos de la dirección son idénticos. Cabe anotar que se manejó el problema de cambiar la dirección de manera legítima asignando tipos de elementos separados a dirección "anterior" y dirección "actual".

Por su parte, la Familiaridad consiste en reconocer que Santiago y Verónica constituyen un hogar porque comparten la misma dirección, aunque hay que tener cuidado de excluir a las personas que viven en diferentes apartamentos de un edificio. Es posible tener información interna o externa de otra fuente de datos que indique que Santiago y Verónica están casados.

El sexto paso de la limpieza de datos hace referencia a la documentación de los datos que resultaron de los pasos anteriores en metadatos para conseguir así que cuando se hagan limpiezas de datos posteriores éstas estén mejor adaptadas a la elaboración del proceso que vimos en este caso que fue el de la corrección de direcciones y para que el usuario final de las aplicaciones esté en mejores condiciones para entender la base de datos en la que realiza su consulta.

Aunque la Transformación que se lleva a cabo en el proceso ETL permite que se imprima calidad a los datos mediante su limpieza y aplicación de reglas de negocio, es muy importante reconocer que de donde definitivamente tiene que partir la Calidad de los datos es de las fuentes ya que si desde allí hay "basura" eso mismo será lo que se obtiene en las salidas. El tratamiento de las fuentes en términos de Calidad de Datos es obviamente más costoso en términos de tiempo, dinero, recursos, etc. pero al no hacerse puede acarrear problemas cuando se hacen transformaciones grandes en el ETL ya que como es evidente, sin modificar las fuentes desde donde se origina la información que se obtiene posteriormente pueden presentarse desfases y eventualmente pérdidas de dinero.

2.4. Calidad de Datos en el Proceso ETL

Extracción, Transformación y Carga

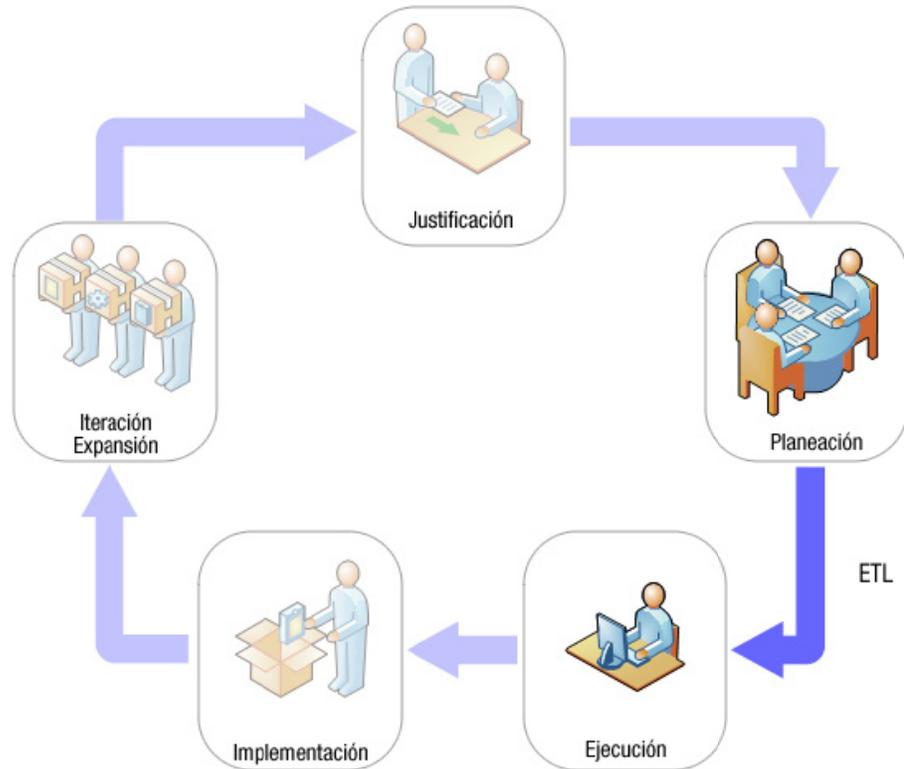


Figura 12 - Técnicas Aplicadas en el Proceso ETL

El proceso *ETL* (*Extract, Transform and Load* por sus siglas en Inglés) hace referencia a la Extracción, Transformación y Carga de los datos que se definen como necesarios por parte de los usuarios de la inteligencia de una organización y que es proporcionada por las herramientas de visualización B.I.

La primera parte, la parte de Extracción se lleva a cabo recolectando los datos requeridos de las diferentes fuentes en las cuales se encuentran. Frecuentemente, los datos se encuentran en diferentes sistemas de la compañía y por tal motivo, hay necesidad de hacer una fusión de éstos posteriormente en el *Data Warehouse*. Las diferentes estructuras de las cuales pueden provenir los datos son por ejemplo bases de datos relacionales, hojas de cálculo, etc. Es importante resaltar que lo que hace compleja esta fusión en sí es el formato que tengan los datos de las diferentes fuentes, ya que para que se obtenga información útil de éstos, es necesario que estén unificados.

En la etapa de Transformación, se involucran ciertas reglas del negocio referentes a los datos o simplemente se aplican ciertas conversiones a los datos que se extrajeron para posteriormente cargarlos. Estas reglas de negocio lo que definen

es la forma como serán tratados los registros incompletos, la información que no tiene significado dentro de la organización o el tratamiento de valores nulos. Dependiendo del estado en el que se encuentren los datos que se extrajeron de las fuentes es necesaria una transformación en mayor o menor grado y dentro de estas posibles transformaciones encontramos las siguientes:

- Seleccionar sólo ciertas columnas de una tabla para su carga (Ej. que las columnas con valores nulos no se carguen).
- Traducir códigos (Ej. Si la fuente almacena una "H" para Hombre y "M" para Mujer pero el destino tiene que guardar "1" para Hombre y "2" para Mujer).
- Codificar valores libres (Ej. convertir "Hombre" en "H" o "Sr" en "1").
- Obtener nuevos valores calculados (Ej. $total_venta = cantidad * precio$).
- Unir datos de múltiples fuentes (Ej. búsquedas, combinaciones, etc.).
- Calcular totales de múltiples filas de datos (Ej. ventas totales de cada región).
- Generación de campos clave en el destino.
- Transponer o pivotar (girando múltiples columnas en filas o viceversa).
- Dividir una columna en varias (Ej. columna "Nombre: García, Miguel"; pasar a dos columnas "Nombre: Miguel" y "Apellido: García").
- La aplicación de cualquier forma, simple o compleja, de validación de datos, y la consiguiente aplicación de la acción que en cada caso se requiera:
 - ❖ Datos OK: Entregar datos a la siguiente etapa (Carga).
 - ❖ Datos erróneos: Ejecutar políticas de tratamiento de excepciones (Ej. Rechazar el registro completo, dar al campo erróneo un valor nulo o un valor provisional).

Podría decirse, que dentro del proceso ETL, la actividad más importante que se realiza es la Transformación en términos de Calidad de Datos. En la Transformación es donde se lleva a cabo, por decirlo así, una segunda limpieza de datos que van a ser usados para producir la inteligencia que requieren las personas que toman las decisiones en la organización. En esta etapa, se logra incrementar la calidad a los datos que puedan tener problemas provenientes de su origen o los adquiridos por degeneración.

La fase de Carga hace referencia a la etapa donde los datos que han sido transformados en la etapa previa, se almacenan en el *Data Warehouse*. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes. En algunas bases de datos se sobrescribe la información antigua con nuevos datos. Los *Data Warehouse* mantienen un historial de los registros de manera que se pueda hacer una auditoría de los mismos y disponer de un rastro de toda la historia de un valor a lo largo del tiempo.

Existen dos formas básicas de desarrollar el proceso de carga:

Acumulación Simple

La acumulación simple es la más sencilla y común, y consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo

seleccionado y transportar el resultado como una única transacción hacia el *Data Warehouse*, almacenando un valor calculado que consistirá típicamente en una sumatoria o un promedio de la magnitud considerada.

Rolling

El proceso de Rolling por su parte, se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (Ej. Totales diarios, totales semanales, totales mensuales, etc.).

La fase de carga interactúa directamente con la base de datos de destino. Al realizar esta operación se aplicarán todas las restricciones y *triggers* (disparadores) que se hayan definido en ésta (Ej. valores únicos, integridad referencial, campos obligatorios, rangos de valores). Estas restricciones y *trigger* (si están bien definidos) contribuyen a que se garantice la calidad de los datos en el proceso *ETL*, y deben ser tenidos en cuenta.

Debe tenerse presente que de acuerdo a la forma en que cambian las necesidades del negocio en términos de inteligencia, también habrán cambios dentro del proceso *ETL* que se use dentro de la organización. Igualmente, en el momento en que se considere que deben hacerse ajustes al proceso, se generarán cambios. A esto puede llamarse el mantenimiento del proceso *ETL*.

2.5. Calidad de Datos en el Data Warehouse

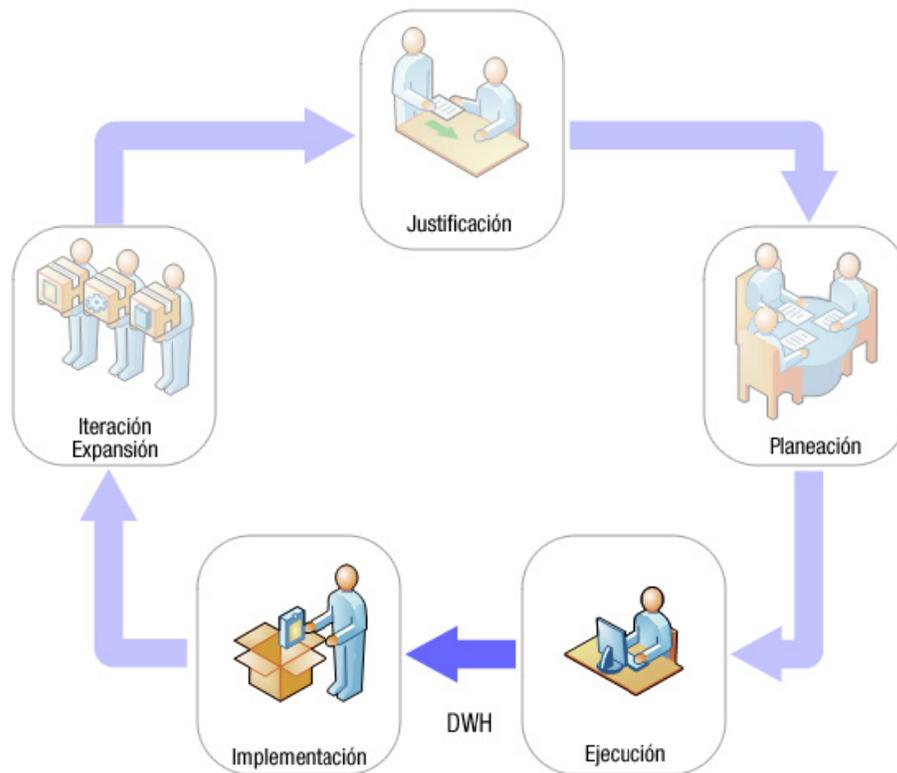


Figura 13 - Técnicas Aplicadas en el Data Warehouse

Luego de haber pasado por reconocer las acciones referentes a la Calidad de Datos que pueden ejecutarse sobre los datos en las fuentes y las transformaciones que se realizan luego de extraerlos de las mismas para que se encuentren de acuerdo a las necesidades que la organización plantea en términos de información, hay que considerar que también se requiere tener un tratamiento adecuado de esos datos que han sido mejorados por los pasos anteriores para conservar la calidad que se les ha definido en el momento en que se almacenan. La calidad en el *Data Warehouse* parte desde el diseño que se defina para éste, la forma en cómo será mantenido a través del tiempo y la estrategia que se tenga para desarrollar su evolución. Como se sabe, en el *Data Warehouse* es donde van a situarse los datos que han venido siendo procesados y que son necesarios para la Inteligencia de Negocios y por tal razón y como se mencionó anteriormente, la calidad que éstos posean también hasta llegar a almacenarse va a afectar el repositorio. (23)

Adicionalmente al impacto que se genera en el *Data Warehouse* producto de la calidad resultante de la limpieza en las fuentes y las transformaciones en el proceso ETL, debe hacerse mención de otro factor que se define como que efectivamente está relacionado con dicha calidad en el almacenamiento y que es llamado los componentes propios del *Data Warehouse*. Como es evidente, si los

datos que entran al *Data Warehouse* son datos con calidad, éstos seguirán conservando su calidad una vez ingresen, pero del tema que no se conoce o no se ha hablado específicamente es acerca de esos componentes propios del *Data Warehouse* y la forma como efectivamente pueden afectar dichos datos. Estos componentes propios del *Data Warehouse*, a su vez son impactados por Modelos y Métricas de Calidad que se relacionan con los requerimientos de calidad relevantes de acuerdo al rol de cada usuario del *Data Warehouse*, la forma de medir esos requerimientos y la forma en que se comportarían dichos requerimientos frente a la evolución del *Data Warehouse*. Cada uno de los anteriores, deben estar efectivamente definidos para que el *Data Warehouse* pueda considerarse como contenedor de datos con calidad.

Pero, ¿cuáles son esas características que deben encontrarse en los datos almacenados en el *Data Warehouse* para decir que efectivamente el repositorio es de calidad? Llegando al punto donde se definen esas características con las que requiere contar un *Data Warehouse* respecto a su información almacenada, podemos recordar atributos como los siguientes: Completitud, Relevancia, Fiabilidad, Cantidad de datos, Consistencia, Precisión, Exactitud, entre otros muchos que se describieron en capítulos anteriores y que están definidos como atributos que efectivamente evidencian la Calidad de Datos. Ahora, para poder medir atributos como los anteriores en un *Data Warehouse*, se hace uso de dos métodos que son nombrados como de experimentación o estudio en organizaciones reales o puede ejecutarse también haciendo uso de modelos de métricas de calidad. Las métricas dan información de gran utilidad que permite determinar la complejidad y tiempo de desarrollo de los procesos de carga. Para definir la métrica es recomendable agrupar los problemas de calidad en categorías y definir las medidas relevantes sobre cada una de ellas, igualmente se deben definir unos niveles mínimos y máximos de aceptación respecto a lo que se está evaluando. La elaboración o definición de las métricas incluyen aspectos como identificar las principales características de los datos relacionados con las políticas de negocio, determinar como esas características se relacionan cuantitativamente con el negocio, evaluar como las fallas identificadas se categorizar dentro del conjunto de dimensiones de calidad definidos, cuantificar la contribución de cada falla de acuerdo con las políticas de cada empresa, y definir y aplicar las normas para datos dentro de un marco de presentación de informes.

La forma más usada para medir la Calidad de Datos en los *Data Warehouse* es la experimentación o estudio en organizaciones reales. El anterior método, hace referencia a llevar a cabo, por ejemplo, un análisis de los factores que afectan la Calidad de Datos hasta llegar al *Data Warehouse*, o incluso puede llevarse a cabo la evaluación por parte de los usuarios acerca de la calidad que perciben de los datos que se evidencian en los informes.

2.6. Calidad de Datos en las Vistas de Usuario

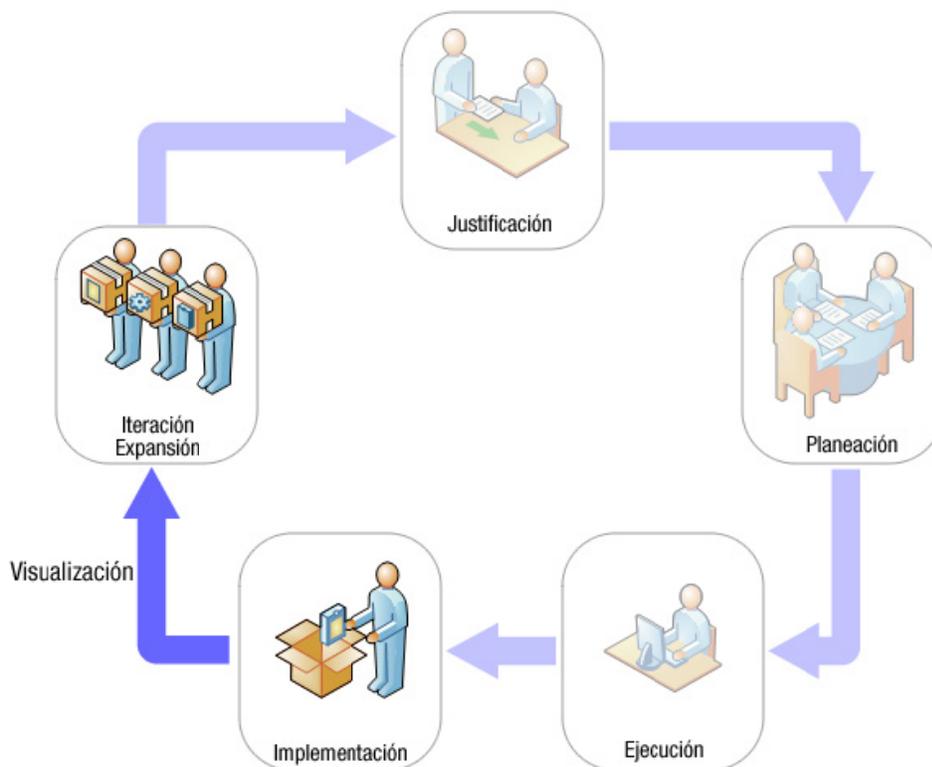


Figura 14 - Técnicas Aplicadas en las Vistas de Usuario

Para poder verificar que la Calidad de Datos que se ha alcanzado en las diferentes instancias por las que atraviesan los datos desde las fuentes hasta el momento en que finalmente se presentan los informes a las personas interesadas en ellos se mantiene, es necesario definir con las personas que conocen los datos y los resultados que deben arrojar luego de todo el proceso, si lo que se está obteniendo al final de la ejecución del proceso que despliega los informes de Inteligencia de Negocios son los correctos.

Las personas que son usuarias de los reportes, conocen esos posibles valores a entregar, conocen el significado y la forma en cómo tiene que ser presentado un dato, conocen las operaciones que tuvieron que haber sido realizadas para obtener determinado dato y así mismo los registros que se debieron involucrar, etc. y por tanto es con ellos con quienes se valida esa salida. Para tener un mayor control acerca de las transformaciones que se llevan a cabo durante el proceso de obtención de los informes y para corroborar que efectivamente la información proviene de donde se requiere, es necesario devolverse en operaciones sobre la información que se obtiene y así poder ejecutar un tracking hasta llegar prácticamente a las fuentes de datos para validar lo que se está alcanzando. En este momento, la Calidad de Datos no depende propiamente de los datos en sí, sino de las variaciones que se ejecutan durante todo el proceso para obtener

respuesta a las preguntas estratégicas de negocio a las que se apunta a responder en el proyecto. En este momento, las discrepancias que puedan presentarse ya tienen que ver directamente con las personas involucradas en el proyecto de Inteligencia de Negocios y en ese orden de ideas, éstos son los responsables de corregir por ejemplo, los cálculos de información.

2.6.1. Técnica: Auditoría de Datos (Data Auditing)

Data Auditing es el proceso de gestionar cómo los datos se ajustan a los propósitos definidos por la organización. Se establecen políticas para gestionar los criterios de datos para la organización. No es suficiente con actuar sino que se debe vigilar. (20)

Para llevar a cabo la técnica antes mencionada, se debe tener en cuenta lo siguiente:

2.6.1.1. Cuantificar el Estado de los Datos

Se debe comenzar desarrollando métricas para medir la calidad de los datos. Estas pueden ser jerárquicas, en esta medida se puede ver el estado individual de cada columna (un buena forma para perfilar y supervisar) que luego se puede aplicar a cada tabla. Como el monitoreo de datos re cuantifican estas métricas, es aconsejable guardar un historial en una Base de datos de tiempo.

2.6.1.2. Analizar el Estado de la Información

Haciendo uso su base de datos de métricas, se puede divulgar y analizar tendencias en la calidad de los datos, lo que proporciona ayudas para determinar si los datos están continuamente mejorando (una de las principales metas de la Calidad de Datos). También puede hacerse uso de estos reportes para validar si se está cumpliendo con las normas reguladoras de calidad como los son ISO o Six-Sigma.

Normalmente el monitoreo de datos es aplicado a la Calidad de Datos pero también sirve para verificar que los trabajos de integración de datos se ejecuten de forma correcta.

2.6.2. Técnica: Prevención de Datos Defectuosos (Data Defect Prevention)

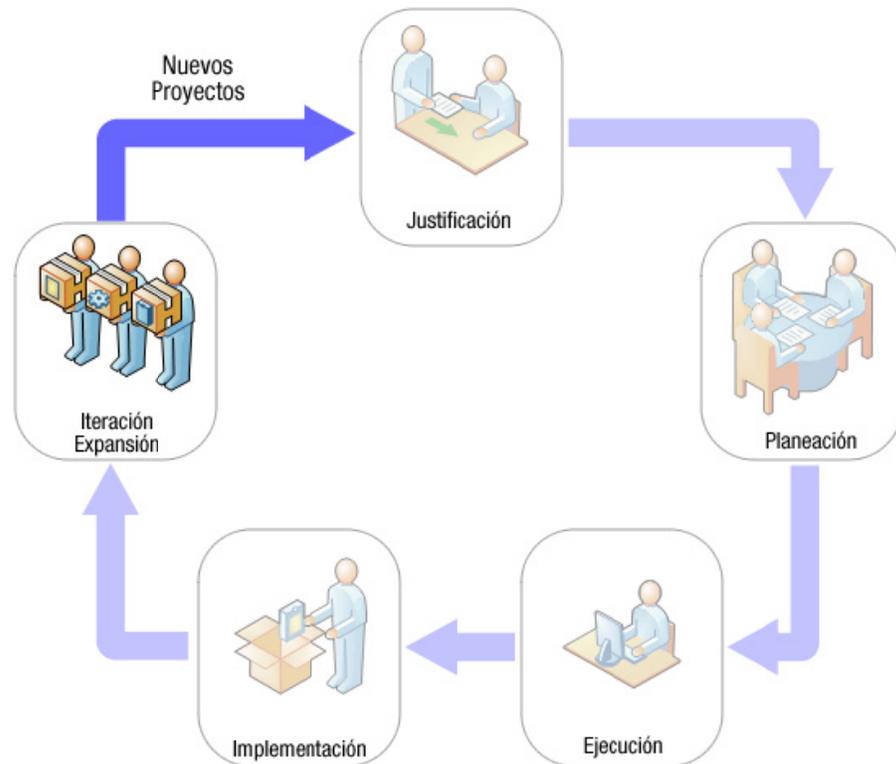


Figura 15 - Técnicas Aplicadas en Nuevos Proyectos

Es fundamental asegurar la integridad, consistencia, completitud y demás atributos que deben tener los datos para ser considerados de calidad, evitando que en ocasiones futuras se filtren “datos sucios”; esto comienza con la identificación de las causas de los posibles errores o defectos, los cuales pueden ser combinaciones de:

- Defectos de programación.
- No tener claridad sobre el significado de los datos.
- No tener metadatos comunes.
- No tener definiciones de dominio.
- No hacer verificación de datos.
- No dedicar el tiempo necesario para entrar los datos.
- No tener un incentivo para la calidad de la entrada de datos.

La identificación de datos faltantes, la corrección de información inexacta o la eliminación de información duplicada son solo algunos pasos para lograr la Calidad de Datos. Para minimizar costos y mejorar resultados se debe buscar una sola actividad que corte el problema desde la raíz, es decir identificar los orígenes de los datos de mala calidad. Para lograr esto es necesario el uso de una técnica o metodología.(17)

Se conocen dos causas que pueden ocasionar datos defectuosos: La primera de ellas es donde los datos se crean incorrectamente, la segunda se conoce como "van mal" en donde los datos cambian en algún momento y no se actualizan correctamente.(24)

Para combatir estos problemas Walter Stewart diseñó un proceso, el cual es de los más ejecutados en estos momentos, este se conoce como el PDCA (Plan-Do-Check-Act).

En la primera etapa (PLAN) se analizan las causas del origen de los datos defectuosos, las cuales pueden ser duplicación de datos, mala escritura entre otros, luego se ejecuta el segundo paso (DO) en el cual se implementa de forma controlada la forma como se analizan los datos, garantizando así sus resultados, el tercer paso (CHECK) se chequea o verifica que se hayan logrado las metas propuestas con el análisis hecho previamente y por último se ejecuta el paso 4 (ACT) en donde se actúa en el control del proceso para así retener las ganancias obtenidas y poder transmitir las a toda la organización.

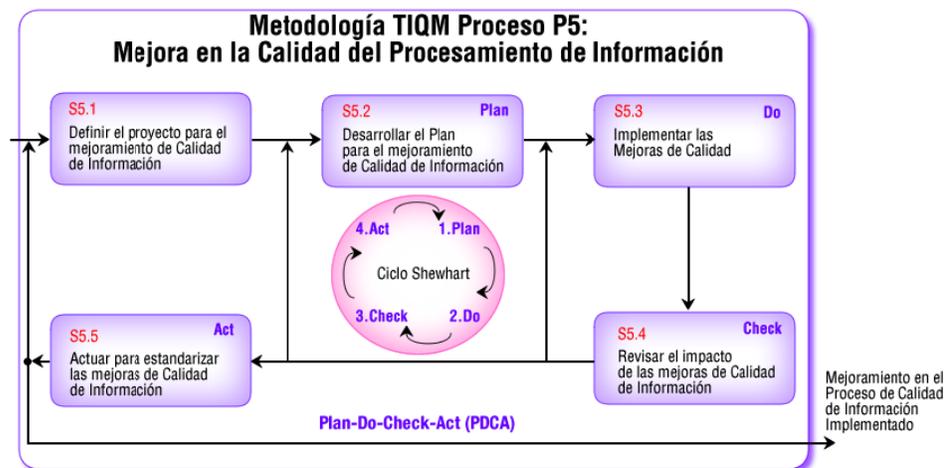


Figura 16 - Prevención de Datos Defectuosos

La técnica previamente mencionada, tiene su aparición en la Metodología TIQM que también se explicó antes.

2.7. Madurez de la Calidad de Datos en las Etapas de un Proyecto de Inteligencia de Negocios

De acuerdo a las búsquedas realizadas para el tema de Madurez de la Calidad de Datos en las Etapas de un Proyecto de Inteligencia de Negocios se tiene que decir que no se obtuvo bibliografía al respecto.

Con motivo de lo anterior, a continuación se hace la presentación de una propuesta propia acerca de cómo unir la teoría vista en los dos capítulos conceptuales previos.

Se verá, que es clara la diferencia cuando desde cero una organización está emprendiendo un proyecto de Inteligencia de Negocios versus cuando el proyecto de Inteligencia de Negocios está siendo ampliado dentro de la organización.

En lo que se describirá a continuación, se verá cómo las Etapas de la Madurez en la Calidad de Datos hacen presencia o se relacionan con las etapas con las que cuenta un proyecto de Inteligencia de Negocios de acuerdo a las Metodologías para la Implementación de un Proyecto BI que se tienen previamente expresadas en este documento.

2.7.1. Importancia del Estado de los Datos

Dentro de lo que se define en un proyecto de Inteligencia de Negocios, la información y los datos son la piedra angular sobre la cual se apoyan los resultados, que son las decisiones tomadas o que se puedan tomar basadas en los análisis del sistema. Identificar cual es la información que se necesita para poder tener elementos con protagonismo en la toma de decisiones es fundamental, pero es allí donde se define cuál es el nivel mínimo de calidad que se espera de los datos para la necesidad que se tiene.

Lo fundamental y más importante, es que desde las fuentes se tenga una adecuada Calidad de Datos la cual permita dar inicio al proyecto y que lo que se requiera hacer posteriormente en términos de transformación de dichos datos sea muy poco.

Se deben definir reglas claras en el proceso de *ETL*, y estandarizarlas a lo largo de la empresa. Se debe hacer un recorrido enumerando las aplicaciones que recolectan los datos que se necesitan para determinar la forma en que éstas lo hacen, ya que en la calidad de los datos la generación es el primer punto a controlar, si por ejemplo se tiene una aplicación en la cual se recolectan datos geográficos, y en ésta se mezclan indistintamente zonas de una gran ciudad con ciudades pequeñas esto podrá acarrear problemas a futuro, ya que aunque para la aplicación en particular esto no representa un elemento que incida en su funcionamiento para el análisis posterior de la información puede ser un elemento influyente, para esto entonces es necesario realizar un estudio de todas las aplicaciones existentes que estén en la cadena de datos del proyecto, y definir si los datos que están aplicaciones generan requerirán conversión o si es posible

realizar cambios en la aplicación para evitar el procesamiento adicional que involucra la transformación.

Del anterior proceso, se debe documentar y definir para futuras aplicaciones que generen datos del mismo tipo, cuál es el estado óptimo o de limpieza en el que se espera que estos datos se almacenen, es decir, continuando con el ejemplo anterior, si se planea desarrollar una nueva aplicación que igualmente haga uso de datos geográficos, definir muy bien que no se pueden mezclar nombres de diferentes elementos como si fueran iguales dentro de lo que se considera como un nivel de organización territorial, esto es, almacenar ciudades con ciudades, y zonas o barrios con elementos de su mismo tipo. Si pensáramos en el caso particular de nuestra región, es muy habitual tener a la ciudad Medellín fragmentada en zonas o barrios, como Belén, El Poblado, La América, Estadio, y cuando dentro de los datos se involucran los municipios aledaños se utiliza Itagüí, Envigado o Bello, en un mismo nivel de los barrios o zonas de Medellín, sabiendo de por sí que en la organización territorial estos deberían estar al mismo nivel que Medellín, esto se hace basado en el hecho que para el usuario la fragmentación en barrios de los municipios aledaños no es un elemento de uso habitual, pero para efectos futuros del análisis de datos esto constituye en una imprecisión de la información.

Una vez se hayan definido estándares en los datos en su formato y nivel, para nuevas aplicaciones, y se hayan identificado estos datos en las aplicaciones existentes se continúa con la planeación de los procesos de transformación, es necesario que estos procesos estén bien documentados y en lo posible incrustados en sistemas de automatización porque si actualmente existe una aplicación que por así decirlo corrompe los datos por que los recolecta en una forma inadecuada, el proceso de transformación será una tarea de uso constante para poder mantener los datos actualizados en el sistema de Inteligencia de Negocios.

2.7.2. Madurez de la Calidad de Datos en un Proyecto de Inteligencia de Negocios que es el Primero en la Organización.

2.7.2.1. Primera Etapa: Justificación.

Cuando recién una organización define dentro de sus necesidades el emprender un proyecto de Inteligencia de Negocios que apalanque la toma de decisiones para darle rumbo a sus objetivos estratégicos, se comienza a gestar la definición de los componentes que se requieren para la puesta en marcha de dicho proyecto. Como se ha evidenciado a lo largo de este documento, la Calidad de Datos es uno de los aspectos fundamentales dentro de un proyecto de Inteligencia de Negocios y tarde o temprano hará parte de las preocupaciones referentes al proyecto si la organización no ha sido cuidadosa con el manejo de sus datos a lo largo del tiempo. En este primer momento, cuando se está evaluando la viabilidad del desarrollo y se está haciendo un análisis sobre su ROI, puede decirse que no hay preocupación aún por los datos.

2.7.2.2. Segunda Etapa: Planeación.

Generalmente y como es de esperarse, la primera vez que se tiene un acercamiento con un proyecto de Inteligencia de Negocios dentro de una compañía, la concepción que se alcanza acerca de la Calidad de Datos es bastante incipiente. Cuando el proyecto se encuentra en su etapa de Planeación que es donde se hace una definición de las herramientas con las que se va a contar, se preparan los datos, se define el alcance, presupuesto, recursos, etc., puede decirse que la madurez de la Calidad de Datos se encuentra en su etapa de Incertidumbre. En este momento no es claro lo que se tiene frente al tema de Calidad de Datos y tampoco se asume de la manera como sería beneficioso. Aquí, la preocupación se encamina a buscar culpables y posteriormente es cuando se empieza a hacer someramente una corrección de los datos erróneos. A medida que se presentan los problemas es cuando se hace la manera de solucionarlos pero no hay un plan de acción establecido sino que se trabaja ad hoc. Aquí se toma la decisión de avanzar o replantear el proyecto.

2.7.2.3. Tercera Etapa: Ejecución.

Cuando se avanza en el desarrollo del proyecto y éste se ubica en la etapa de Ejecución todavía hay vestigios de lo que se presentaba referente a la madurez en su etapa previa. Aunque ya en este punto se puede decir que se encuentra en la etapa del Despertar, el movimiento alrededor de la Calidad de Datos no es muy decidido y aún no hay una disposición importante y definitiva frente al tema. Ahora, ya se está estableciendo la configuración para enfrentar los problemas grandes que se presenten referentes a la Calidad de Datos y esto se enfrenta con el uso del proceso ETL donde en su actividad de Transformación se ejecuta una corrección de problemas que dichos datos tengan. Hay que tener en cuenta, que dicha Transformación ayuda a mejorar la Calidad de Datos que viene desde las fuentes o simplemente da el formato requerido a éstos. Este proceso ETL ayuda con la calidad en los datos de la organización, pero es importante recordar que sólo se realiza sobre los datos que son requeridos para ser visualizados para la toma de decisiones de la organización. En este momento, se decide hacer un tratamiento alrededor de los datos que se involucran para la toma de decisiones del proyecto. La definición de reglas de negocio que se aplican a los datos para la Transformación de los mismos están adecuadamente definidas y hay acciones que efectivamente corrigen los problemas que se presentan. De acuerdo a las situaciones que se presentan referentes a los datos, se evalúa la posibilidad de llevar la transformación de los datos del proceso ETL a la corrección y limpieza de las fuentes. Aunque se reconoce que es costoso en términos de tiempo, dinero, recursos, etc., hay que hacer una valoración referente a la decisión a tomar y definir dependiendo de los pros y contras si se hace el trabajo desde las fuentes o definitivamente se decide continuar con el control de la Calidad de Datos en el proceso ETL. Sin embargo, no se descarta que posteriormente sí se haga control de Calidad estrictamente desde la entrada de los datos por parte de los responsables. Es de hacer notar, que la Calidad de Datos en las fuentes puede obtenerse haciendo uso de restricciones desde las bases de datos, no dejando campos abiertos en las aplicaciones de no ser necesario, usar validaciones respecto a longitudes, tipos de datos y caracteres de separación de los mismos, usando listas de selección desplegadas para los posibles valores que pueda

tener un campo, etc. Estas formas de control, pueden usarse a las aplicaciones existentes por medio de modificaciones en éstas o en las aplicaciones futuras en las cuales si se tiene conciencia se evita gran número de inconvenientes respecto a calidad.

Como anotación hay que recordar que los datos que se quedan en los sistemas de la compañía y que no se ven involucrados dentro del proyecto de Inteligencia de Negocios no van a sufrir ninguna transformación y por tal motivo a éstos no se les mejorará su calidad.

2.7.2.4. Cuarta Etapa: Implementación.

Cuando se avanza, todavía en la etapa de Implementación del proyecto la madurez respecto a la Calidad de Datos se encuentra en la etapa del Despertar. Se siguen encontrando errores que son corregidos a través de las transformaciones ejecutadas en el proceso ETL y aunque empieza a verse un cambio respecto a la apreciación de la temática de la Calidad de Datos, ésta se sigue viendo como un tema “de otros” por parte de las personas de la organización.

Uno de los aspectos importantes para hacer notar a los empleados de una organización, es que los datos que digitan diariamente en sus sistemas son de gran importancia y por tal motivo deben ser ingresados de la mejor manera posible. Para el momento de encontrarse el proyecto en la parte de Implementación que es donde se da un vistazo de los resultados alcanzados y más aún de los resultados a alcanzar, de acuerdo a las etapas de madurez se vería reflejada la Iluminación. En este punto es donde debería haber un involucramiento en el tema de Calidad de Datos por parte de las personas que son las que toman las decisiones estratégicas en las organizaciones, generalmente los gerentes y éstos deberían estar al tanto de la forma en cómo se manejan los datos dentro de la organización para obtener esa información que ellos requieren para tomar decisiones.

2.7.3. Madurez de la Calidad de Datos en un Proyecto de Inteligencia de Negocios que es Expansivo/De Integración en la Organización.

Quinta Etapa: Iteración/Expansión.

Infortunadamente, el llegar a que la Calidad de Datos se trate con estrategias de prevención y que efectivamente una organización pueda asegurar que no existen problemas con la Calidad de Datos en su interior implica un arduo trabajo que muchos no están preparados para hacer o no disponen de los recursos para realizarlo. En el punto de llegar a reconocer la Calidad de Datos como un asunto de peso dentro de Inteligencia de Negocios es cuando la compañía se encamina a desarrollar ya sea un proyecto expansivo en términos de áreas dentro de la organización o similares. Ya existe una experiencia previa proveniente del desarrollo de un proyecto de Inteligencia de Negocios para un área dentro de la organización y ahora teniendo en cuenta las lecciones aprendidas del mismo, se disponen a la implementación de nuevos proyectos del mismo tipo para otras dependencias que lo requieran. Como es evidente, ya se ha adquirido un

conocimiento respecto a lo que se enfrenta al embarcarse en el desafío de un proyecto de esta magnitud, se tienen mejor definidos los aspectos que lo componen y definitivamente se sabe que la Calidad de Datos es una pieza importante para que finalmente el proyecto funcione. Hay que anotar, que no puede esperarse que todo salga perfecto pero es bueno reconocer que hay un camino previo que ya se ha recorrido y que aporta para el nuevo proyecto.

El llegar a un nivel de madurez que se considera como Sabiduría, implica que la compañía se encamina a la Integración de sus proyectos de Inteligencia de Negocios con el resto de sistemas que se encuentran en su interior como lo son ERP's, CRM's, etc. En este momento y considerando que se ha hecho un verdadero proyecto expansivo previo de la Inteligencia de Negocios dentro de la compañía hay que hacer uso de las experiencias que se recolectaron en cada uno de los proyectos. Aquí, ya no es sólo conocimiento lo que se tiene respecto a la Calidad de Datos, en este momento debe considerarse una de las prioridades dentro de la organización si se tiene en cuenta y se dimensiona realmente a lo que se refiere y en lo que se refleja para el logro de objetivos estratégicos dentro de la compañía. En este momento, las aplicaciones que se usan son ampliamente adaptadas al control de la Calidad de Datos, las personas encargadas de hacer la digitación de los mismos tienen plena conciencia de su trabajo y hay un área dentro de la compañía que efectivamente impulsa el tema de la Calidad de Datos. Éste tema se convierte en un fin por sí mismo y ahora hay quien lo controle y se haga cargo de él realmente, es un proceso importante al interior de la compañía.

El alcanzar el nivel de Certeza en la escala de madurez de la Calidad de Datos es sumamente costoso en términos de tiempo y dinero para una organización. Como se dijo antes, durante el proceso ETL solamente se trabaja sobre los datos que requieren ser involucrados para la toma de decisiones y los demás, conservan su estado a través del tiempo. El asegurar que dentro de una organización no existen problemas con la Calidad de Datos, tomar conciencia y llegar a prevenir los problemas referentes a la Calidad de Datos o incluso hacer que dentro de la organización se comprometan con ayudar a identificar los problemas de los datos y corregirlos requiere un gran cambio en la forma de hacer las cosas al interior de una organización y un compromiso de los empleados de la misma. Prácticamente un cambio cultural al interior de la compañía.

2.8. Factores Críticos de Éxito y Mejores Prácticas

A medida que en el medio se van desarrollando proyectos de Inteligencia de Negocios van surgiendo aspectos críticos referentes a la calidad de los datos, así como se van estableciendo una serie de mejores prácticas alrededor del tema. A continuación se hace referencia a estos aspectos.

2.8.1. Factores Críticos de Éxito

- En la Edición N° 94 de la ACIS, en el **artículo Lecciones aprendidas en Inteligencia de Negocios** la señora Olga Lucía Prieto comenta, “Es factor crítico de éxito el compromiso de la alta Gerencia. La información que maneja el operario será diferente del área del vicepresidente o el presidente. Pero cada uno tendrá información acorde al rol que deba ejecutar en la compañía. Respecto al nuevo gerente, las soluciones de tecnología hoy le ofrecen herramientas útiles para que pueda adelantar una gestión basada en valor, enfocada al logro de resultados para anticiparse a las necesidades del mercado y llevar a su organización a ser más competitiva, más rentable. Cuenta con herramientas como el Balanced Scorecard, tableros de control, entre otras, que lo van a apoyar en la definición de estrategias, metas y objetivos, su planificación, ejecución y monitoreo.” (25)
- De acuerdo al marco funcional propuesto por Mark Mosley en el DAMA-DMBOK, Versión 3.02 se definen siete elementos del entorno que afectan la Calidad de Datos: Tecnología, Organización y Cultura, Actividades, Entregables, Roles y Responsabilidades, y Practicas y Técnicas. En la siguiente tabla de muestra el ámbito de aplicación de cada uno de los elementos, el orden de estos comienza en el centro y su forma de desplazamiento es alrededor del círculo desde la posición 1:00 hasta la posición 12:00. (26)



Figura 17 - Elementos del Entorno que Afectan la Calidad de Datos

- En el curso **Creando el próximo Data Warehouse: Integración y Calidad de Datos** dictado por Alejandro Mínguez Porras, se definen unos factores de éxito agrupados en los elementos de Tecnología, Metodología y apoyo, a continuación se hace explícito los factores atribuidos a cada elemento. (27)

| | |
|--------------------|---|
| Tecnología | <p>La tecnología es fundamental para la resolución de problemas de Calidad de Datos.</p> <p>En la elección de la tecnología deben exigirse los siguientes requisitos:</p> <ul style="list-style-type: none"> ▪ Combinación con una plataforma de integración de datos: acceso universal a los datos ▪ Tratamiento de cualquier tipo de datos. No sólo nombres y direcciones. ▪ Solución de reporting, scorecarding y monitorización ▪ Orientado a negocio |
| Metodología | <p>Decidir la metodología adecuada en función del nivel de apoyo logrado.</p> <ul style="list-style-type: none"> ▪ Proyectos estratégicos con apoyo de la dirección: metodologías de Data Governance. ▪ Proyectos tácticos con apoyo de una o varias unidades de negocio. Proyectos relacionados de Calidad de Datos, con establecimiento de mejores prácticas para cada uno de ellos. ▪ Proyectos departamentales, orientados a la resolución de un problema determinado. <p><i>Establecer objetivos, definir métricas, monitorizar, corregir, repetir y evolucionar.</i></p> |
| Apoyo | <ul style="list-style-type: none"> - Divulgación interna, principalmente a las unidades de negocio, tecnología y a dirección - Auditoría de Calidad de Datos: <ul style="list-style-type: none"> ▪ Análisis y medición de los problemas de Calidad de Datos de la Compañía ▪ Conclusiones de los problemas de Calidad de Datos ▪ Propuesta de soluciones ▪ Análisis del ROI |

Tabla 4 - Elementos de Entorno que Afectan la Calidad de los Datos

2.8.2. Mejores Prácticas

A continuación se presentan las mejores prácticas descritas por los expertos para asegurar la Calidad de Datos en las organizaciones.

2.8.2.1. Mejores Prácticas - Entendimiento de la Tecnología

Este conjunto de mejores prácticas son una adaptación de B. Stackpole , “Dirty Data Is the Dirty Little Secret That Can Jeopardize Your CRM Effort,” CIO, February 15, 2001, pp. 101–114; y J. Moad, “Mopping Up Dirty Data,” Baseline, Issue 25, December 1, 2003, pp. 74–75. (28)

- *La limpieza de los datos no es suficiente.* Sea consciente que el software de limpieza de datos maneja solamente algunas ediciones: números inexactos, errores de ortografía, y campos incompletos. Los programas de calidad de datos globales se acercan a la estandarización de los datos para mantener integridad de la información.
- *Comience en lo alto.* Sea consciente del asunto de calidad de los datos y cómo afectan a la organización. Los altos mandos de la organización deben asumir cualquier esfuerzo por reparar los datos ya que estos recursos serán necesarios por un largo tiempo.
- *Conozca sus datos.* Entienda qué datos tiene y para qué se utilizan. Determine el nivel apropiado de precisión necesario para cada tipo de datos.
- *Hacer de esto un proceso continuo.* Desarrolle una cultura de la calidad de los datos. Institucionalice una metodología y las mejores prácticas para incorporar y comprobar la información.
- *Resultados de la medida.* Revise regularmente los resultados para asegurarse que se están cumpliendo los estándares de calidad y para estimar impactos de fondo.

2.8.2.2. Plan de Acción de Calidad de Datos

Un plan de acción de calidad de los datos es recomendado para la mejora de la calidad de los datos. Este plan de acción implica los siguientes pasos: (28)

1. Determine las funciones que se consideran críticas del negocio.
2. Identifique los criterios para seleccionar datos críticos.
3. Señale los datos críticos.
4. Identifique las preocupaciones conocidas de calidad de los datos, datos críticos y sus causas.
5. Determine los estándares de calidad que se aplicarán a los datos críticos.
6. Diseñe un método a la medida para cada estándar.

7. Identifique y ejecute rápido las iniciativas de mejora de Calidad de Datos.
8. Ejecute los métodos definidos para mejorar los problemas de Calidad de Datos desde la base.
9. Evaluar las medidas y también las preocupaciones de la Calidad de Datos y sus causas.
10. Planear las iniciativas adicionales de mejora.
11. Continúe midiendo los niveles de calidad e iniciativas de mejora.
12. Amplíe el proceso para incluir los datos adicionales.

2.8.2.3. Pasos para la Administración de la Calidad Total de Datos

Estos 14 pasos para la mejora y sostenibilidad de la Calidad de Datos son una adaptación de Deming's 14 Points, descritos por Larry English en *Improving Data Warehouse & Business Information Quality*. (29)

1. Crear la conciencia del propósito para el mejoramiento de la información de los productos y servicios: Plan a largo plazo.
2. Adoptar la nueva filosofía de calidad información compartida como una herramienta para el mejoramiento del negocio: "la información compartida reduce costos"
3. Evitar la dependencia entre la información y la aplicación para así obtener información de calidad: diseñar calidad en los datos de diseño y los del proceso de producción, y establecer una asignación de responsabilidades.
4. Terminar la práctica de desarrollo de aplicaciones "a tiempo" y "en el presupuesto".
5. Mejora constante de procesos de aplicación y desarrollo de información, al igual que los procesos de servicio, producción y mantenimiento: implementar la metodología planear-hacer-evaluar-actuar con el fin de generar un mejoramiento continuo de la Calidad de Datos.
6. Capacitar a los empleados en el tema de Calidad de Datos, especialmente a los dueños de la información.
7. Tener un líder de Calidad de Datos de tiempo completo en la organización.
8. Exteriorizar el miedo por la incertidumbre de los datos o la corrección de estos: implementar programas de incentivos para localización y corrección de los problemas fuentes que producen defectos en los datos.
9. Romper las barreras entre áreas de negocios: administración de información y el desarrollo de la aplicación; Tecnología de Información y negocios; áreas de negocio y unidades de área de negocio.

10. Eliminar eslogan y advertencias, remplazar esto con acciones para el mejoramiento de la Calidad de Datos: desarrollar un hábito de datos “prevención de defectos”.
11. Cambiar las cuotas de productividad con métricas de calidad: satisfacer al cliente de información de productos.
12. Remover las barreras entre los compañeros de trabajo, permitiendo a los productores de la información solucionar los problemas en los procesos.
13. Estimular la educación y mejora personal para todas las personas en la cadena de valor de información.
14. Tomar acción para lograr una transformación por Calidad de Información: los directivos de la organización deben sentir el dolor de los problemas de calidad, organizarse y comunicar esto a las personas a su cargo. Cada proceso es candidato al mejoramiento.

2.9. Lecciones Aprendidas

Cada proyecto trae consigo una serie de aprendizajes que aunque responden a las necesidades particulares de una empresa, puede replicarse en otras organizaciones, esos detalles o experiencias que quedan de estos desarrollos son puntos valiosos para el desarrollo de nuevos proyectos. A continuación se presentan algunos casos.

Problemas de Datos

En esta tabla se presentan los problemas habituales que se dan en las organizaciones, junto con el tipo de causa y la posible solución. *Source: Based on S.L. Alter, Decision Support Systems: Current Practices and Continuing Challenges. Addison-Wesley, Reading, MA, 1980, p. 130. (28)*

| Problema | Tipo de causa | Posible Solución |
|---|--|---|
| Los datos no son correctos | <ul style="list-style-type: none"> ▪ Los datos fueron generados por descuido. ▪ Los datos primarios se han introducido incorrectamente. ▪ Los datos fueron manipulados. | <ul style="list-style-type: none"> ▪ Desarrollar de forma sistemática la introducción de datos. ▪ Automatizar la entrada de datos. ▪ Introducir controles de calidad en la generación de datos. ▪ Establecer programas de seguridad. |
| Los datos no son oportunos | <ul style="list-style-type: none"> ▪ El método para la generación de datos no es suficientemente rápido para satisfacer la necesidad de datos. | <ul style="list-style-type: none"> ▪ Modificar el sistema para la generación de datos. ▪ Usar la web para obtener nuevos datos. |
| Los datos no son medidos o indexados correctamente | <ul style="list-style-type: none"> ▪ Los datos primarios son incompatibles con los propósitos del análisis. ▪ Se usan modelos complejos. | <ul style="list-style-type: none"> ▪ Desarrollar un sistema de reajuste o recombinación de los datos indexados. ▪ Utilizar un almacén de datos. ▪ Utilizar los motores de búsqueda adecuados. ▪ Desarrollar modelos más sencillos y globales. |
| Los datos necesarios no | <ul style="list-style-type: none"> ▪ Los datos que se necesitan en el momento no fueron | <ul style="list-style-type: none"> ▪ Predecir lo que los datos puedan ser necesarios en el futuro. |

| | | |
|----------------|--|---|
| existen | almacenados. ▪ Los datos requeridos no existen. | ▪ Utilice un almacén de datos. ▪ Generar nuevos datos o estimación de ellos. |
|----------------|--|---|

Tabla 5 - Problemas de Datos

3. Entrevistas

3.1. Breve Descripción de las Empresas Entrevistadas

En este capítulo se presentan los resultados de las entrevistas realizadas en las visitas a un grupo de siete empresas seleccionadas pertenecientes a diferentes sectores como lo son Financiero, Energía, Telecomunicaciones, Seguros, Tecnología y Educación, quienes trabajan proyectos de Inteligencia de Negocios y cuyo manejo de los datos es un factor crítico en sus procesos operacionales y administrativos. A continuación se presenta una breve descripción de estas empresas con el fin de contextualizar y dimensionar el manejo de los datos de que éstas requieren y la dimensión de los proyectos que se realizan en estas.

Las empresas en las cuales se llevaron a cabo las entrevistas son las que se describen a continuación:

3.1.1. XM Compañía de Expertos en Mercados S.A ESP

“Es una filial del grupo empresarial ISA, encargada de prestar los servicios de planeación y coordinación de la operación de los recursos del Sistema Interconectado Nacional y la administración del sistema de intercambios comerciales de energía eléctrica en el Mercado Mayorista, así como la liquidación y administración de los cargos por uso de las redes del sistema interconectado nacional. Igualmente, presta servicios especializados afines para diferentes sectores productivos del país y la región.” (30)

3.1.2. UNE Telecomunicaciones

“En el año 2005, Empresas Públicas de Medellín, después de observar el entorno, analizar las señales de los clientes y determinar las posibilidades del negocio, replanteó el mercado de servicios públicos domiciliarios y configuró tres grupos estratégicos de negocio: Energía, Aguas y Telecomunicaciones. Cada uno muy diferente del otro, con dinámicas competitivas distintas y retos específicos.”(31)

“Para cumplir entonces con la estrategia definida por EPM, el primero de julio de 2006 se escinde la UEN Telecomunicaciones, dando origen a EPM Telecomunicaciones S.A. E.S.P. con su marca UNE. Así, queda constituida como una empresa 100% pública, propiedad de Empresas Públicas de Medellín, conformada como una sociedad anónima por acciones, que presta servicios de tecnologías de información y comunicaciones a sus clientes en todo el país. Al igual que para los otros negocios de EPM (Aguas y Energía), para UNE se definió una Meta Estratégica Grande y Ambiciosa (MEGA) a largo plazo que debe cumplir en el 2015.”

3.1.3. Valores Bancolombia

“Valores Bancolombia, Comisionista de Bolsa pertenece al Grupo Bancolombia S.A. y es una compañía del sector bursátil que canaliza todos sus esfuerzos profesionales y experiencia en la solución y propuesta de alternativas financieras. (32)

Valores Bancolombia, antes Suvalor, al ser miembro de la Bolsa de Valores de Colombia, es vigilada por la Superintendencia Financiera, antes Superintendencia de Valores, que inició operaciones en mayo de 1991. En octubre de 2005 se fusionó con la compañía Comisionista de Colombia, también del Grupo Bancolombia S.A. y que operó desde 1992.

Valores Bancolombia está posicionada como una de las principales firmas comisionistas de Bolsa de Valores de Colombia y presta sus servicios en ocho ciudades: Medellín, Bogotá, Cali, Barranquilla, Cartagena, Manizales, Pereira y Bucaramanga.”

3.1.4. Suramericana

“Está conformada por las Compañías Suramericana de Seguros S.A, Suramericana de Seguros de Vida S.A, y Suramericana de Capitalización S.A las cuales se dedican a satisfacer las necesidades de protección, ahorro y servicios afines de sus clientes en el país.” (33)

3.1.5. Intergrupo

“InterGrupo tiene como promesa hacia sus clientes el ser una empresa que provee soluciones que les ayudan a transformar sus operaciones y que les permitan construir nuevos modelos de negocios usando tecnologías de punta, permitiéndoles ahorrar costos y mejorar la eficiencia de su negocio. (34)

Su misión es la de ser líderes en la prestación de servicios de consultoría, desarrollo de software e integración de productos y servicios en tecnología de punta, con un excelente servicio al cliente, promoviendo el crecimiento profesional y humano de nuestros colaboradores.”

3.1.6. Serfinco

“Firma Comisionista de Bolsa, miembro de la Bolsa de Colombia, vigilada por la Superintendencia Financiera, que presta servicios de asesoría financiera integral e intermediación de títulos valores en el mercado financiero colombiano a personas naturales y jurídicas.” (35)

3.1.7. Universidad EAFIT

“La Universidad **EAFIT** es un Centro de Educación Superior aprobado por el Ministerio de Educación Nacional. Actualmente cuenta con una población estudiantil de más de diez mil alumnos entre los cuales se incluyen estudiantes de pregrado y posgrado.”(36)

3.2. Formato de Entrevista

Preguntas

1. ¿Cuáles son las etapas que se definen dentro de la compañía para llevar a cabo un proyecto de Inteligencia de Negocios?
2. ¿Cuál es el objetivo o significado de cada una de las etapas?
3. ¿Qué persona o área dentro de la organización es la encargada del tema de Calidad de Datos en el proyecto de Inteligencia de Negocios?
4. ¿Qué actividades se realizan en cada una de las etapas respecto a la Calidad de Datos?
5. De acuerdo a las etapas descritas ¿Qué técnicas utilizan para llevar a cabo las actividades que se refieren a Calidad de Datos en cada una de ellas?
6. ¿Qué planea hacer para mejorar, garantizar o mantener la Calidad de Datos en su proceso de Inteligencia de Negocios?
7. ¿Cuáles consideran ustedes que son lecciones aprendidas respecto a la Calidad de Datos en un proyecto de Inteligencia de Negocios? ¿En qué forma impacta el éxito o fracaso del proyecto la Calidad de Datos?
8. ¿Cuáles considera como Factores Críticos de Éxito con respecto a la Calidad de Datos para un proyecto de Inteligencia de Negocios?

3.3. Resultados Obtenidos

3.3.1. Pregunta 1: ¿Cuáles son las etapas que se definen dentro de la compañía para llevar a cabo un proyecto de Inteligencia de Negocios? Y Pregunta 2: ¿Cuál es el objetivo o significado de cada una de las etapas?

Para estas preguntas que se refieren a las Etapas que define cada una de las empresas para llevar a cabo un Proyecto de Inteligencia de Negocios y el objetivo que tiene cada una de ellas se tiene lo siguiente:

- La totalidad de las empresas coinciden en que dentro de las etapas que involucra un Proyecto de Inteligencia de Negocios están la Identificación de Necesidades que es dónde se define que se quiere con el proyecto, Definir las fuentes necesarias para responder a las necesidades donde se identifica cuáles datos son los que se requieren para responder a las necesidades planteadas, el Diseñar los módulos a construir donde se vislumbra a lo que se quiere llegar y cómo se va a hacer en términos de arquitectura de la solución, Implementar la solución refiriéndose a la puesta en marcha del desarrollo del proyecto como tal y Realizar pruebas donde se verifica con los usuarios finales de los reportes que efectivamente la información que se está obteniendo es la correcta.
- Cinco de las empresas consideran que adicionalmente a las anteriores etapas se debe realizar la Selección de la plataforma de trabajo que se refiere al análisis y posterior toma de la decisión acerca de la herramienta que se usará para llevar a cabo el proyecto, la Definición del Proceso ETL donde se definen adecuadamente las reglas de negocio y las transformaciones que se requieren para obtener el resultado requerido y se debe realizar la Medición de la Calidad de Datos en las fuentes para efectivamente ejecutar acciones al respecto.
- Tres empresas, consideran importante definir las Herramientas de visualización para Usuarios.
- Sólo dos de las empresas hacen un análisis de Factibilidad que hace referencia a la evaluación efectivamente de definir si se ejecuta el proyecto o no y el mismo número de empresas consideran la Capacitación dentro de las etapas del proyecto, esto es, enseñarle a los usuarios finales el manejo de la herramienta de visualización cuando el proyecto se encuentra en producción.
- Sólo una empresa expresa la consideración del Diseño e Implementación que se contratan con proveedores por separado haciendo referencia a reunirse son las personas que los ayudarán a diseñar el proyecto y posteriormente realizar una reunión con quienes lo desarrollarán.
- Sólo una empresa expresa específicamente la Definición del Recurso Humano como etapa dentro del proyecto donde se hace la selección de los integrantes del proyecto.

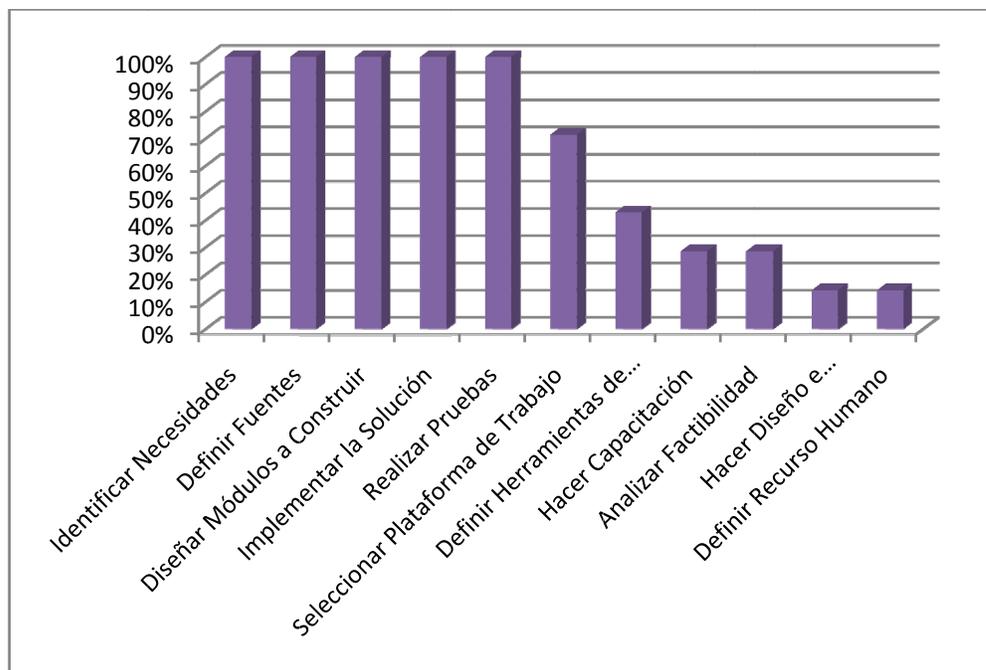


Figura 18 - Etapas para Llevar a Cabo un Proyecto de Inteligencia de Negocios

3.3.2 Pregunta 3: ¿Qué persona o área dentro de la organización es la encargada del tema de Calidad de Datos en el proyecto de Inteligencia de Negocios?

Para todas las empresas involucradas en el estudio la Calidad de Datos es responsabilidad de quien genera los datos, sin embargo es cada una de ellas hay un responsable de realizar las tareas de aseguramiento de la calidad y unas áreas que interfieren en el proceso.

Áreas como Gestión de Información, Inteligencia de Negocios, Equipo de Calidad de Datos, Mercadeo, Tecnología son las encargadas en las diferentes empresas de identificar, solucionar, validar, supervisar y documentar los defectos que se generan respecto a los datos, ya sea como área encargada o apoyo del proceso de calidad, a continuación se presenta porcentualmente la relación entre las empresas y los roles, áreas responsables y diferentes áreas que intervienen.

¿Quién se encarga?

29% Gestión de Información

57% Equipo de Calidad

14% Tecnología

Tabla 6 - Encargado de la Calidad de Datos en las Empresas

¿De quién depende el área encargada?

| | |
|-------------------------------|------------------------------|
| Gestión de Información | 50% Gestión de Información |
| | 50% Tecnología |
| Equipo de Calidad | 50% Mercadeo |
| | 25% Tecnología |
| | 25% Inteligencia de Negocios |
| Tecnología | 100% Tecnología |

Tabla 7 - Área de la que Depende el Encargado de Calidad de Datos en la Empresa

¿Quiénes intervienen en el proceso?

| |
|---|
| 100% Usuarios generadores de información |
| 100% Tecnología |

Tabla 8 - Personas o Áreas que Intervienen en la Calidad de Datos

- El 57% de las empresas tienen definido un equipo de Calidad de Datos, lo cual refleja que la calidad de los datos son un factor fundamental para el buen desempeño de las actividades de la empresa y por eso requiere de un equipo específico para la administración de los datos.
- 50% de los equipos de calidad dependen del área de mercadeo, esto indica la responsabilidad de los datos no son responsabilidad de tecnología propiamente si no de los dueños de la información, en estos casos el equipo de calidad se encuentra en la dependencia para la cual es más crítico la calidad de los datos.
- En el 100% de las empresas la dirección de tecnología interfiere en los procesos de Calidad de Datos, al igual que los usuarios generadores de la información.

3.3.3. Pregunta 4: ¿Qué actividades se realizan en cada una de las etapas respecto a la Calidad de Datos?

De acuerdo a lo evidenciado en las entrevistas, y dado que evidentemente aunque en las empresas tengan etapas similares es difícil unificar, se va a presentar el resultado discriminado por actividades en las Fuentes, Data Warehouse, Proceso ETL y Visualización y adicionalmente, se hará la discriminación por empresa.

3.3.3.1. En las Fuentes:

- Cuatro de las empresas manifiestan que se hace una corrección de errores en las fuentes.

- Una de las empresas dice hacer un análisis previo a iniciar el proyecto que pretende identificar el impacto que pueden generar los errores que poseen las fuentes, en los sistemas de la empresa.
- Una de las empresas dice realizar un diccionario de datos para dar sentido a lo que se está obteniendo de las fuentes.
- Dos empresas verifican que la misma cantidad de datos tomados de las fuentes sea la que efectivamente se carga en el Data Warehouse.

3.3.3.2. En el Proceso de Extracción, Transformación y Carga:

- Tres empresas manifiestan que hay una definición de reglas de validación que se aplican en la parte del Proceso ETL.
- Una de las empresas dice hacer correcciones de acuerdo a informes o alertas de error generados en las cargas.

3.3.3.3. En el Data Warehouse:

- Una de las empresas dice realizar un seguimiento manual cuando se generan resultados atípicos que provienen del Data Warehouse y dicen ejecutar reglas de validación en los metadatos.
- Dos empresas dicen tener control de inconsistencias como tal en el Data Warehouse.

3.3.3.4. En la Visualización:

- Cuatro de las empresas que consideran como una actividad de aseguramiento de la calidad el comparar lo que entregan los informes contra lo que conocen los usuarios dicen que se ejecuta dicha validación.

Ahora bien, teniendo los anteriores resultados y buscando identificar cuántas empresas hacen control de Calidad de Datos en los diferentes puntos discriminados anteriormente, se tiene lo siguiente:

- Tres de las siete empresas entrevistadas hacen control de Calidad de Datos tanto en las Fuentes como en el Proceso ETL, Data Warehouse y en la Visualización con usuarios.
- Una empresa de acuerdo a lo expresado hace control en el Proceso ETL y en el Data Warehouse.
- Una empresa, hace control en las Fuentes, el Proceso ETL y el Data Warehouse.
- Una empresa, hace control en las Fuentes y Visualización con usuarios.
- Una empresa hace según lo dicho únicamente control en la Visualización con usuarios.

De lo anterior, se obtiene el siguiente gráfico:

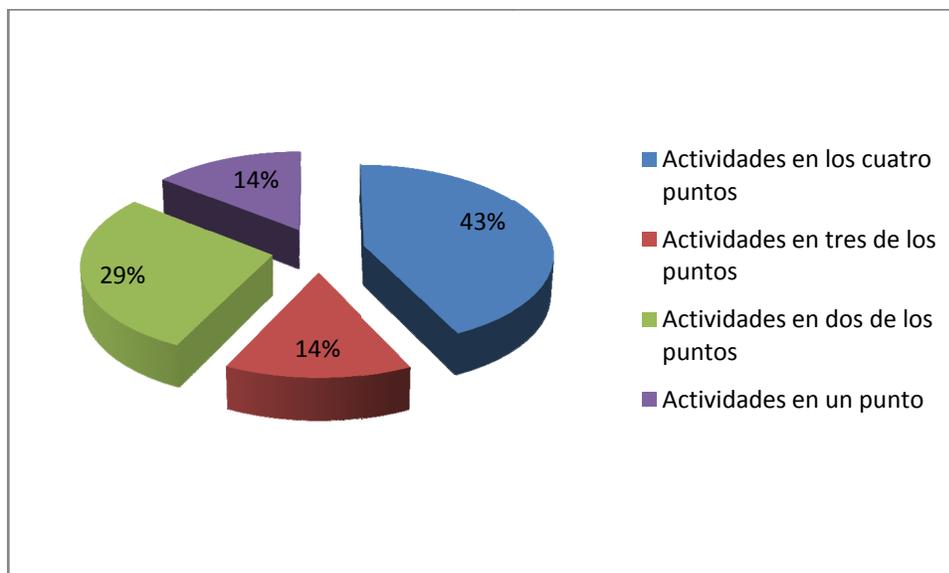


Figura 19 - Actividades de Calidad de Datos que Realizan las empresas

3.3.4. Pregunta 5: De acuerdo a las etapas descritas ¿Qué técnicas utilizan para llevar a cabo las actividades que se refieren a Calidad de Datos en cada una de ellas?

Aunque es posible definir unas técnicas estándar para las diferentes etapas de BI cada organización selecciona la que más se ajuste a sus necesidades, las técnicas utilizadas para la administración de la calidad dependen del estado de los datos fuentes, la cantidad de los mismos, la complejidad del negocio y los recursos que se tengan destinados para este fin (presupuesto, tiempo, equipo de trabajo).

A continuación se presentan las técnicas utilizadas por las diferentes empresas clasificadas en las etapas de un proyecto de Inteligencia de Negocios definidas en la teoría.

| Etapa | Técnica |
|-----------------------|--|
| Justificación | 14% Perfilamiento de datos y Limpieza de datos (En las Fuentes) |
| Planeación | 57% Perfilamiento de datos 43% Limpieza de datos (En las Fuentes) |
| Ejecución | 14% Limpieza de datos (En ETL) 57% Limpieza de datos (En las Fuentes) |
| Implementación | 57% Limpieza de datos (En las Fuentes) |

| | |
|----------------------------|---|
| | 29% Limpieza de datos (En ETL) |
| | 71% Auditoria de datos |
| Iteración/Expansión | 29% Limpieza de datos (En ETL) |
| | 71% Limpieza de datos (En las Fuentes o en ETL) |
| | 100% Auditoria de datos |
| | 57% Prevención de datos Defectuosos |

Tabla 9 - Técnicas de Calidad de Datos que se Aplican a las Etapas del Proyecto BI

- El hecho de que en la etapa de justificación se refleje un porcentaje de empresas que realizan técnicas de Calidad de Datos responde a la conciencia que tenía la empresa respecto al estado de sus datos, así que tomaron la decisión de trabajar en la calidad antes de formalizar el proyecto de Inteligencia de Negocios.
- Un alto porcentaje de las empresas vinculadas al estudio de campo realizan Perfilamiento de datos (57%) y Limpieza de datos desde la fuente (43%) en la etapa de Planeación, si se quiere sacar un mayor provecho del proyecto en esta etapa el porcentaje de utilización de estas técnicas debería ser el 100%.
- En las etapas de Ejecución e Implementación un 29% de las empresas solo realizan Limpieza de datos en el ETL, eso corresponde a la complejidad del negocio (el tiempo de reprocesar los datos alteraría el comportamiento de la operación diaria) o la falta de presupuesto para realizar un proyecto de este tipo.

3.3.5. Pregunta 6: ¿Qué planea hacer para mejorar, garantizar o mantener la Calidad de Datos en su proceso de Inteligencia de Negocios?

Según el nivel de madurez que la empresa tenga respecto a la Calidad de Datos se plantean diferentes actividades o estrategias para el mejoramiento o mantenibilidad de la Calidad de Datos. Presentándose lo siguiente:

No tiene un esquema definido: Dos de las empresas entrevistadas no tienen definido un plan para el manejo de los datos, estas planean como actividades de mejora lo siguiente:

- Realizar plantillas para hacer cuadros de resultados vs sistema fuente.
- Utilizar una herramienta para evidenciar los defectos en los datos y corregir los que ya han sido detectados.

Están comenzando a definir el esquema: Una de las empresas está desarrollando actualmente su proyecto de Inteligencia de Negocios y lo que definen como actividades de mejora de la Calidad de Datos es:

- Asignar un área para control de calidad.

- Crear una mayor conciencia y sensibilización en cuanto al manejo de los datos en los dueños de la información.
- Crear un diccionario de datos.

Tienen un esquema de calidad definido: Las empresas cuyo proceso de Calidad de Datos ya está definido presentan las siguientes actividades:

- Ampliación del diccionario de datos.
- Mejoramiento de informes.
- Mejorar el apoyo de los usuarios en el manejo de los datos.
- Implementar un módulo para reportar informes de integridad que sea enviado al responsable del dato automáticamente.
- Tener claros los orígenes y conservar su calidad, esto se logra realizando las modificaciones necesarias y realizar un control sobre los datos que se almacenan en los sistemas transaccionales, con el fin de garantizar integridad, consistencia y limpieza en las fuentes.
- Crear una conciencia sobre el manejo de los datos al interior de la compañía.

3.3.6. Pregunta 7: ¿Cuáles consideran ustedes que son lecciones aprendidas respecto a la Calidad de Datos en un proyecto de Inteligencia de Negocios? ¿En qué forma impacta el éxito o fracaso del proyecto la Calidad de Datos?

Para todas las empresas entrevistadas la Calidad de Datos es vital para el éxito del proyecto de Inteligencia de Negocios en sus organizaciones, ya que los reportes que son generados desde el proyecto son el material para la toma de decisiones tanto operativas como estratégicas de la organización.

A continuación se presenta a modo de tips las lecciones aprendidas de las empresas involucradas en el estudio de campo.

3.3.6.1. XM

- La Calidad de Datos corresponde a toda la organización desde los diferentes generadores de datos.
- Se debe buscar tener una terminología estándar entre los diferentes generadores de información.
- La Calidad de Datos se debe asegurar desde los sistemas fuente, ya que la falta de calidad en esta puede presentar inconsistencias con la información generada en los reportes.
- Cuando se va a elegir un proveedor al hacer las licitaciones se debe proporcionar un mayor conocimiento sobre las necesidades ya que esto afecta la propuesta en cuanto a tiempos de implementación y es preferible que el proveedor se instale en el lugar donde se va a realizar el proyecto.

3.3.6.2. Intergrupo

- La corrección de datos en la fuente, el evitar la duplicidad de datos y controlar la cantidad de nulos definidos en el metadato son factores fundamentales y de gran impacto en el proyecto de Calidad de Datos, en caso de que esto no se realice la información que este genera no será de utilidad.
- Se debe aprender a gestionar y mostrarle al usuario el impacto que tiene en sus datos y reportes y por ende en las decisiones que se tomen sobre estos.

**3.3.6.3. UNE
Telecomunicaciones**

- Se debe definir claramente el alcance de lo que se quiere para no desgastarse en cosas que no se usan.
- El diseño del proyecto sea llevado a cabo por alguien que tenga conocimiento del negocio

3.3.6.4. Suramericana

- Ejecutar control de Calidad de Datos desde las fuentes y no al final es tal vez la mejor práctica en este tipo de proyectos al igual que realizar la corrección de los mismos en este punto.

3.3.6.5. Serfinco

- Almacenar aparte los datos inconsistentes que se detectan en las diferentes etapas de proceso es una buena práctica, aunque la utilización de esta técnica depende de la cantidad de datos que de manejen y de los recursos que se tengan para esta actividad, ya que de lo contrario puede no ser útil para el control de calidad de los datos.
- Establecer un buen plan de inconsistencias con el fin de presentar los datos de la forma más completa y consistente posible.

**3.3.6.6. Valores
Bancolombia**

- Se debe definir una estrategia para adquirir los datos de las fuentes.
- La participación de las personas que tienen conocimiento de los datos es fundamental al momento de hacer la selección de los datos a utilizar ya que se pueden generar errores o retardar los procesos si no se obtienen los queries directamente de los que conocen los datos.

3.3.6.7.
Universidad
EAFIT

- Se ha adquirido una mayor sensibilización respecto al estado de los datos en los sistemas de información de la universidad, al igual que se ha aprendido sobre la información almacenada en los diferentes repositorios y de los procesos mismos.
- Es importante contar con herramientas para la recolección, validación y seguimiento de los datos ya que los procesos manuales pueden traer consigo errores.

3.3.7. Pregunta 8: ¿Cuáles considera como Factores Críticos de Éxito con respecto a la Calidad de Datos para un proyecto de Inteligencia de Negocios?

Para el desarrollo de un Proyecto de Inteligencia de Negocios las diferentes empresas entrevistadas consideran como Factores Críticos de Éxito los siguientes ítems:

- El conocimiento de la empresa y del proyecto por parte de las personas que hacen el diseño y luego desarrollan el proyecto.
- El tener continuidad por parte de las personas que lideran el proyecto dentro de la compañía.
- Desarrollar un plan de trabajo que incluya las demás áreas de la compañía que se involucran en el proyecto.
- El proyecto de Calidad de Datos debe ser liderado por un área diferente a tecnología, debe ser una labor de los dueños de la información apoyado por el área de tecnología y sustentado en el área de calidad.
- Que el proyecto de Calidad de Datos sea algo establecido como un proyecto real de la empresa.
- La diligencia de los usuarios, que quieran estar involucrados en el proyecto y que sean conscientes de la importancia de los datos.
- Realizar correcciones en los aplicativos cuando se ven las falencias.
- Adaptar los cambios de las reglas de negocio.
- Realizar la definición adecuada del diccionario de datos para que haya un entendimiento por parte de los interesados de esos datos que hay en las fuentes.
- Considerar la adecuada definición de los datos en las aplicaciones origen.
- Que haya intervención de pocos usuarios finales en la definición de las necesidades.
- Asegurar la Calidad de Datos desde las fuentes.
- Asegurar una adecuada administración de las fuentes.
- Revisar las transformaciones a ejecutar.
- Definir la forma de hacer corrección de problemas de forma automática.
- Realizar pruebas adecuadas encaminadas a los resultados que se esperan obtener.

4. Análisis Comparativo

En el presente capítulo, se hará una comparación entre lo que se encontró en el marco teórico respecto al tema de Calidad de Datos y lo referente a Inteligencia de Negocios versus lo que se obtuvo en las entrevistas a las diferentes empresas anteriormente mencionadas.

- Haciendo referencia a las etapas que se definen para llevar a cabo un proyecto de Inteligencia de Negocios las cuales fueron definidas previamente cuando se habló de las metodologías que se usan para este fin, se obtuvo que las etapas son: Justificación, Planeación, Ejecución, Implementación e Iteración/Expansión.
Según lo que se recogió en las entrevistas, puede decirse que todas las empresas realizan aunque con nombres diferentes y con diferencia en las actividades que define la teoría, casi todas las etapas con las que cuenta un proyecto de Inteligencia de Negocios con excepción de la Justificación. Esta etapa de Justificación, sólo se realiza en dos empresas.
- Hay que recordar que la etapa de Implementación que se define en las metodologías, abarca también la parte de capacitación y de acuerdo a lo que se tiene de las entrevistas, sólo dos empresas son las que realizan capacitación para las herramientas de Inteligencia de Negocios.
- La etapa que en las empresas se denomina de Pruebas, hace referencia a la que se llama en las metodologías de Iteración/Expansión. A lo que se refiere esta etapa aparte de identificar deficiencias y producir cambios es a que se definen nuevas necesidades que abarcan diferentes áreas del negocio en nuevos proyectos de Inteligencia de Negocios. En varias empresas, realmente no se busca involucrar esas nuevas áreas en proyectos de inteligencia.
- Es importante resaltar que aunque todas las empresas definen casi todas las etapas que se consideran en las diferentes metodologías para el desarrollo de un proyecto de Inteligencia de Negocios, las empresas no necesariamente llevan a cabo todas las actividades que se definen en las metodologías como se evidenció previamente.
- A medida que las empresas definen e implementan sus planes de Calidad de Datos realizan una serie de actividades que aunque ellos no las enmarquen con un nombre determinado van muy de la mano con las técnicas que se definieron en la parte teórica de Calidad de Datos como lo son el Perfilamiento de datos, la Limpieza de datos, Auditoría de datos y la Prevención de datos defectuosos.
- De las técnicas mencionadas anteriormente las más utilizadas por las empresas mencionadas son el Perfilamiento de datos (Data Profiling) y la Limpieza de datos (Data Cleansing) ya que cuando se decide

implementar un proyecto de Inteligencia de Negocios ya se parte de unas fuentes de datos que posee la organización y es necesario evaluar el estado actual de los datos almacenados y según los resultados de esto se procede a hacer una limpieza, depuración y validación de los datos.

- Muchas de las actividades planteadas para el mejoramiento de la Calidad de Datos corresponden a actividades de técnicas de calidad definidas en la teoría, por ejemplo, garantizar la integridad y consistencia de los datos forma parte de las actividades de Auditoría de datos (Data Auditing) así como el tener claros los orígenes de datos corresponde a la técnica de Perfilamiento de datos (Data Profiling).
- Al igual que como lo reconoce la teoría, las empresas entrevistadas aseguran que la Calidad de Datos es un elemento esencial para alcanzar el éxito con un proyecto de Inteligencia de Negocios. La forma como lo definió la gran mayoría, fue diciendo enfáticamente que si en las fuentes hay basura, entonces saldrá basura una vez se tenga en producción el proyecto. Por lo anterior, es importante resaltar que cinco de las empresas que respondieron a la entrevista hacen medición de Calidad de Datos en las fuentes y que de hecho, hacen correcciones en las mismas.

Haciendo un compendio de los anteriores resultados y teniendo en cuenta lo que se evidenció en las entrevistas, se tienen las siguientes conclusiones:

- Las empresas no usan para poner en marcha sus proyectos de Inteligencia de Negocios, las etapas propiamente definidas por las metodologías que se vieron en los capítulos previos. Las empresas toman actividades de esas etapas y las desarrollan y en ocasiones, hacen uso es de las mismas etapas que se llevan a cabo para un proyecto software.
- Todas las empresas coinciden en que se debe tener un área de Calidad de Datos aunque en muchas no esté adecuadamente definida. No es desarrollada propiamente en el área de Tecnología pero todas las empresas dicen que debe ser apoyada por dicha área.
- No todas las empresas llevan a cabo controles de Calidad de Datos en todas las partes como lo son Fuentes, Proceso ETL, Data Warehouse y Visualización con usuarios. De hecho, sólo tres realizan esta labor en todos los puntos.
- En general, las empresas usan técnicas más no metodologías para llevar a cabo la parte del aseguramiento de la Calidad de Datos dentro de sus proyectos de Inteligencia de Negocios.
- Se pudo observar, que las empresas hacen proyecto de Calidad de Datos como algo general dentro de las mismas.
- El tener una persona encarga de tiempo completo del proyecto de calidad, al igual que el hecho de realizar limpieza de datos en la fuente,

son aspectos importantes entre los factores críticos de éxito y las mejores prácticas. Teniendo mayor peso la limpieza de datos en la fuente.

- Automatizar los procesos de mejoramiento de los datos, por lo menos en la auditoría de estos es otro factor común entre la práctica y la teoría de las mejores prácticas.

5. Conclusiones

Como se buscó desde el comienzo, el objetivo de este proyecto consistió en identificar dentro del proceso de implementación de un proyecto BI las mejores prácticas para garantizar y/o mejorar la calidad de los datos con el fin de potenciar el éxito de un proyecto de BI. A través del marco teórico levantado en los capítulos iniciales y por medio de la investigación de experiencias en algunas empresas de la ciudad de Medellín que se resumen en el capítulo anterior, se logró conseguir lo que sigue.

- Para saber si un dato puede clasificarse como con calidad o no, es indispensable hacer una discriminación de los atributos que debe poseer. En los capítulos iniciales, evidentemente se hizo una descripción de las características requeridas para que dentro de un proyecto de Inteligencia de Negocios se obtengan resultados óptimos por el lado de la Calidad de Datos.
- Luego de hacer un establecimiento en lo referente a la Inteligencia de Negocios e igualmente teniendo lo definido con el tema de Calidad de Datos, se realiza la construcción de una propuesta propia que evidencia la correlación que se encuentra entre las etapas que se requieren para llevar a cabo un proyecto de Inteligencia de Negocios y la Madurez de la Calidad de Datos. Dicha asociación, hace evidente que a medida que se desarrollan proyectos de Inteligencia de Negocios dentro de una compañía, la madurez de Calidad de Datos tiende a escalar posiciones. A medida que las empresas se hacen responsables del tema de Calidad de Datos, éste adquiere mejor trato y por lo tanto se obtienen mejores resultados en el proyecto.
- Llevar a cabo la implementación de un proyecto de Inteligencia de Negocios es algo para lo que puede decirse que las empresas de nuestro entorno están haciendo grandes esfuerzos. Aunque las empresas de la ciudad de Medellín que fueron entrevistadas acerca del tema en ocasiones muestran falencias sobre todo en temas como Calidad de Datos, el cual se evidenció y se reconoció por todas como un factor indudablemente indispensable para el éxito de un proyecto de este tipo tienen dificultades con él. Aunque lo consideran importante, hay oportunidades en las que expresan que definitivamente por cultura organizacional no es posible lograr que se tenga en cuenta realmente el tema y por el contrario, en otras empresas realmente responden al compromiso y se involucran en el proyecto aportando desde el mismo cuidado al ingresar los datos.
- Con base en el estudio de campo que consistió en la realización de entrevistas a diferentes empresas de la ciudad de Medellín, se tiene que la forma en que dichas empresas llevan a cabo los proyectos de Inteligencia de Negocios, se acerca bastante a lo que proponen las

diversas metodologías para dicho fin que se encuentran en capítulos previos. Adicional a esto, hay que reconocer que dichas empresas ejecutan diversas prácticas que también tienen coincidencia con las técnicas para el mejoramiento de la Calidad de Datos que se describen antes lo que puede llevar a decir que la Calidad de Datos es un tema sobre el que se está trabajando realmente.

- Existe una significativa correlación entre la teoría y lo que se hace en las empresas que hicieron parte del estudio de campo. El aspecto más relevante para todas las organizaciones es la evaluación, limpieza y seguimiento de las fuentes de datos, ya que el éxito de un proyecto de BI depende en gran medida del estado de los datos que se procesan y las fuentes son el insumo sobre el cual se trabaja. Es importante resaltar que aunque existe conciencia sobre la importancia del control de los datos en la fuente hay empresas que no lo practican debido a la complejidad del negocio o restricciones de presupuesto, es decir, existen problemas de calidad en la fuente pero las actividades necesarias para solucionarlos implicarían un impacto mayor en la operación del negocio y el costo sería muy alto, razón por la cual optan por “solucionar” los problemas de calidad por medio de mapeos, transformaciones y validaciones en el proceso de ETL.
- Sería interesante que posterior a este trabajo, pueda continuarse investigando sobre nuevas metodologías que sean específicamente para ser aplicadas a la Calidad de Datos en un proyecto de Inteligencia de Negocios ya que lo que se encontró no fue prácticamente lo que se buscaba del tema.
- Adicionalmente y teniendo en cuenta que uno de los factores críticos de éxito para los proyectos de Inteligencia de Negocios respecto a la Calidad de Datos es realizar limpieza en las fuentes y considerando empresas que debido a su operación diaria no pueden realizar este proceso, sería de gran utilidad proponer métodos o soluciones que contribuyan al mejoramiento de la calidad de los datos en la fuente sin interferir con la operación diaria de la empresa.

6. Bibliografía

1. **Agosta, Lou.** Trends in Data Quality. The Data Strategy Advisor. *DMReview.com*. [Online] Febrero 2005. [Cited: Octubre 17, 2008.] <http://www.dmreview.com/issues/20050201/1018111-1.html>.
2. **Lic. Escalone, Fernanda.** Laboratorio de Sistemas Inteligentes. *Universidad de Buenos Aires*. [En línea] 15 de Diciembre de 2006. [Citado el: 6 de Septiembre de 2008.] <http://www.fi.uba.ar/laboratorios/lsi/scalone-tesis-maestria-ingenieria-en-calidad.pdf>.
3. **Larissa Terpeluk Moss, Majid Abai, Sid Adelman.** How to Improve Data Quality. *informIT*. [En línea] 22 de Julio de 2005. [Citado el: 17 de Octubre de 2008.] <http://www.informit.com/articles/article.aspx?p=399325&seqNum=5>.
4. **English, Larry P.** Information Quality Management Maturity: Toward the Intelligent Learning Organization. *TDAN - THE DATA ADMINISTRATION NEWSLETTER*. [En línea] 1 de Junio de 2004. [Citado el: 17 de Octubre de 2008.] <http://www.tdan.com/view-special-features/5409>.
5. *Towards a Comprehensive Approach for Evaluating and Improving Information (Quality) Management Capability Maturity: A Case Study of Integrated Logistic Support in a Large Australian Engineering Asset Management Organisation.* **Baškarada, S, Koronios, A & Gao, J.** Harrogate, United Kingdom : s.n., 2007. Second World Congress on Engineering Asset Management and the Fourth International Conference on Condition Monitoring.
6. **TechTarget.** What is business intelligence? - a definition from Whatis.com. *SearchDataManagement.com*. [En línea] 21 de Noviembre de 2006. [Citado el: 10 de Septiembre de 2008.] http://searchdatamanagement.techtarget.com/sDefinition/0,,sid91_gci213571,00.html.
7. **Gartner, Inc.** *Reporte de Gartner Group Septiembre*. Stamford : Gartner Group, 1996.
8. **English, Larry P.** Business Intelligence Defined. *Business Intelligence Network*. [En línea] 6 de Julio de 2005. [Citado el: 16 de Septiembre de 2008.] <http://www.b-eye-network.com/view/1119>.
9. **White, Colin J.** The IBM Business Intelligence Software Solution. *ACM SIGMOD*. [En línea] Marzo de 1999. [Citado el: 19 de Septiembre de 2008.] <http://www.sigmod.org/disc/disc99/disc/ibm/bisolution.pdf>.
10. **Eduardo, Arrezola Moráz.** Factores Críticos de Éxito/Fracaso en la Implementación de Herramientas de Business Intelligence. *Phronesis Biblioteca Digital*. [En línea] Enero de 2004. [Citado el: 17 de Octubre de 2008.] <http://copernico.mty.itesm.mx/phronesis/mty/tmp/ITESMMTY2004459.pdf>.

11. **Imhoff, Claudia.** Three Trends in Business Intelligence Technology. *Business Intelligence Network*. [En línea] 4 de Abril de 2006. [Citado el: 17 de Octubre de 2008.] <http://www.b-eye-network.com/view/2608>.
12. **Correia, Cynthia Cheng.** Competitive Intelligence: A Practical Primer for Location Intelligence. *Directions Magazine*. [En línea] 21 de Febrero de 2006. [Citado el: 19 de Septiembre de 2008.] http://www.directionsmag.com/printer.php?article_id=2102.
13. **García, Carlos Esteban and Vélez Mesa, Nicolás.** *ESTUDIO SOBRE LAS METODOLOGIAS PARA GESTION DE PROYECTOS DE INTELIGENCIA DE NEGOCIOS, SU APLICACIÓN Y FACTORES CRITICOS DE EXITO*. Universidad EAFIT, Medellín : 2007.
14. **Iberia, Schober PDM.** Las empresas que invierten en Calidad de Datos aseguran haber obtenido beneficios cuantificables. *Financial Tech Magazine*. [En línea] 9 de Enero de 2007. [Citado el: 17 de Octubre de 2008.] http://www.financialtech-mag.com/000_estructura/index.php?id=24&idb=104&ntt=6872&sec=17&vn=1.
15. **Dravis, F.,C. Colbert, D. McBurnie, M. Meckler, D. Townley.** *Data Quality Assessment: A Methodology for Success*. Septiembre de 2006.
16. **Scannapieco, Monica.** DaQuinCIS : Exchanging and Improving Data Quality in Cooperative Information Systems. *School of Computing, Dublin City University*. [En línea] 2004. [Citado el: 17 de Octubre de 2008.] <http://www.computing.dcu.ie/research/dataquality/thesis/thesis/Monica%20Scannapieco.thesis.pdf>.
17. **English, Larry.** Total Information Quality Management → A Complete Methodology for IQ Management. *Data Management Review*. [En línea] Septiembre de 2003. [Citado el: 17 de Octubre de 2008.] <http://www.dmreview.com/issues/20030901/7320-1.html>.
18. *DATA QUALITY CONTROL STRATEGIES*. **Ninan, Sandeep Savla y Mathew.** Adelaide, Australia : University of South Australia, 2006. 1st Australasian Workshop on Information Quality.
19. **Watson, H.** Data Cleansing. *Terry College of Business*. [Online] 2001. www.terry.uga.edu/~hwatson/Data%20Cleansing%20Jan.%202001.ppt - .
20. **Curto, Josep.** Data quality: definiciones. *Information Management*. [En línea] 30 de Julio de 2008. [Citado el: 17 de Octubre de 2008.] <http://informationmanagement.wordpress.com/2008/07/30/data-quality-definiciones/>.
21. **Russom, Philip.** Unifying the Practices of Data Profiling, Integration, and Quality (dPIQ). *DataFlux and SAS*. [En línea] Octubre de 2007. [Citado el: 17 de Octubre de 2008.] http://download.101com.com/pub/tdwi/Files/TDWI_Monograph_DataFlux_Oct2007.pdf.
22. **Kimball, Ralph.** Dealing with Dirty Data. *Intelligent Enterprise*. [En línea] 20 de Septiembre de 2006. <http://www.dbmsmag.com/9609d14.html>.

23. **Daza Oyuela, Andrés Enrique.** *Calidad en Data Warehouses*. Universidad Nacional de Colombia, Bogotá : s.n.
24. **English, Larry.** Data Quality: Standardize, Validate and Improve Your Information Assets. *ZDNet: Tech News, Blogs and White Papers for IT Professionals*. [En línea] Febrero de 2004.
<http://whitepapers.zdnet.com/abstract.aspx?docid=84018>.
25. **M., Sara Gallardo.** Lecciones aprendidas en inteligencia de negocios. *ACIS*. [En línea] 29 de Noviembre de 2005. [Citado el: 17 de Octubre de 2008.]
<http://www.acis.org.co/index.php?id=610>.
26. **Mosley, Mark.** DAMA-DMBOK Functional Framework. *DAMA International*. [En línea] 10 de Septiembre de 2008. [Citado el: 17 de Octubre de 2008.]
http://www.dama.org/files/public/DAMA-DMBOK_Functional_Framework_v3_02_20080910.pdf.
27. **Porras, Alejandro Mínguez.** Fundamentos de la Calidad de Datos. *Aules d'Empresa*. [En línea] 2008. [Citado el: 17 de Octubre de 2008.]
http://www.aulesempresa.upc.edu/programes/PWD_DWH_S2_FundamentosCalidadDatos.pdf.
28. Online File W3.1: Databases. *Pearson Higher Education*. [En línea] [Citado el: 17 de Octubre de 2008.]
http://wps.prenhall.com/wps/media/objects/3778/3869053/additional_online/Turban_OnlineW3.pdf.
29. **English, Larry P.** Deming's 14 points of quality applied to information quality management. *Information Quality at MIT*. [En línea] Julio de 2007. [Citado el: 17 de Octubre de 2008.]
http://mitiq.mit.edu/IQIS/2007/iq_sym_07/Sessions/Session%201D/Session%201D%20-%20Tutorial%20-%20Deming's%2014%20Points%20-%20Larry%20English.pdf.
30. **XM Compañía de Expertos en Mercados S.A. E.S.P.** Quienes Somos. [En línea] [Citado el: 17 de Octubre de 2008.]
<http://www.xm.com.co/Pages/QuienesSomos.aspx>.
31. **EPM Telecomunicaciones S.A E.S.P.** HISTORIA UNE EPM TELECOMUNICACIONES S.A. E.S.P. [En línea] [Citado el: 17 de Octubre de 2008.]
<http://www.une.com.co/nuestracompania/Informacion-Corporativa/Historia.html>.
32. **VALORES BANCOLOMBIA.** Conozca Valores Bancolombia | Historia. [En línea] [Citado el: 17 de Octubre de 2008.]
<http://www.valoresbancolombia.com/valoresBancolombia/conValBancolombia/historia/historia.asp>.
33. **SURAMERICANA.** Sucursal Virtual. [En línea] [Citado el: 17 de Octubre de 2008.] <http://www.suramericana.com/Default.aspx>.
34. **Intergupo.** Acerca de Nosotros. [En línea] [Citado el: 17 de Octubre de 2008.]
http://www.intergrupo.com/Cor_IG_AcercaNosotros.aspx.

35. **Serfinco S.A.** Nuestra Compañía . [Online] 2001. [Cited: Octubre 17, 2008.] <http://www.serfinco.com/ourCompany/default.asp>.

36. **Universidad EAFIT.** Información General. [Online] 2008. [Cited: Octubre 17, 2008.] <http://www.eafit.edu.co/eafitcn/institucional/index.shtm>.