

Data Quality

Julio 2007

Los datos son un recurso crítico de una organización



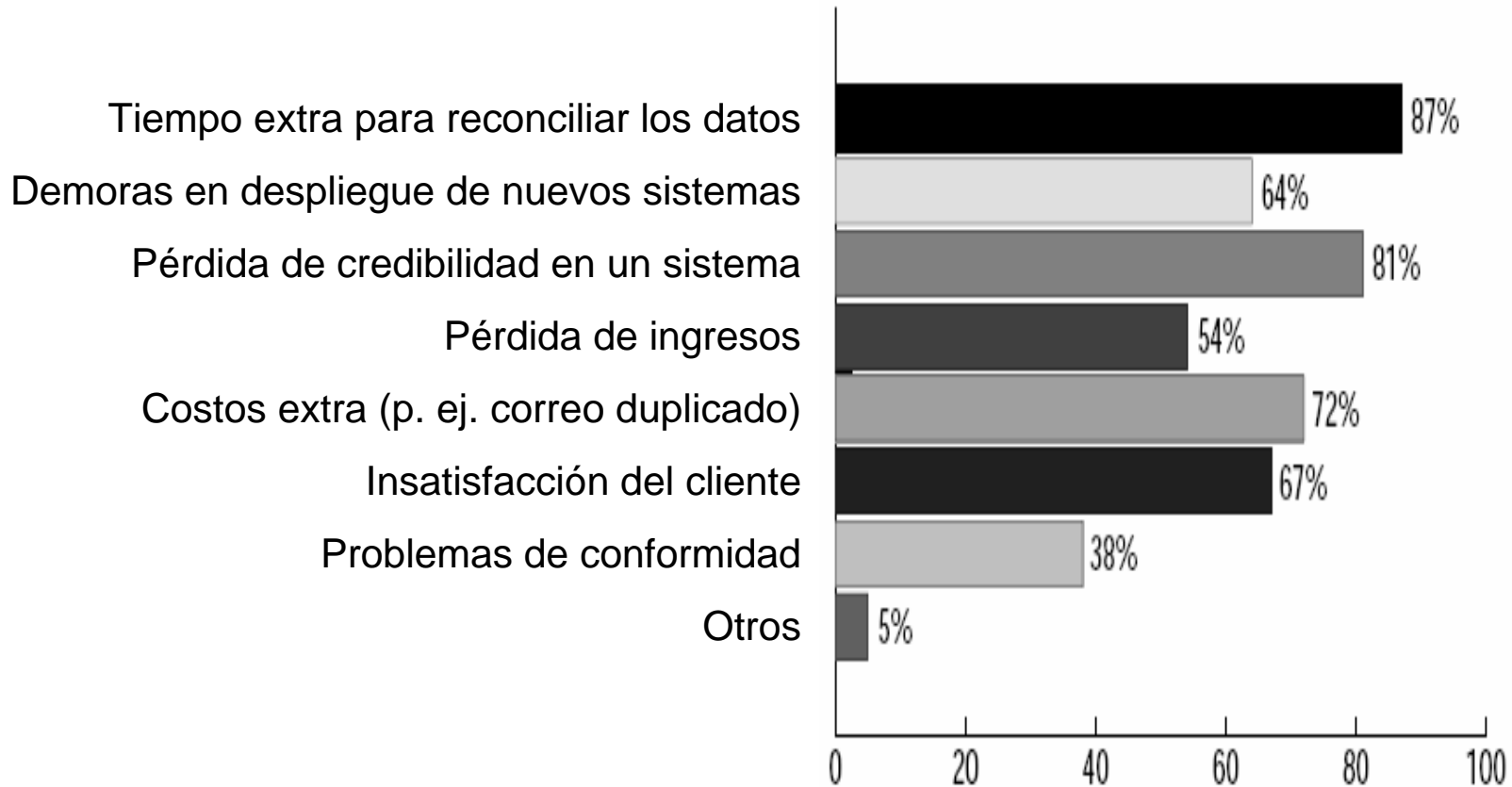
- Los datos y la información elaborada a partir de ellos son vitales para cualquier organización en el siglo XXI: son un factor fundamental para su supervivencia
- Iniciativas estratégicas como CRM, BI y supply chain management requieren grandes conjuntos de datos de buena calidad
- La mayoría de las organizaciones sobreestiman considerablemente la calidad de sus datos y subestiman el impacto de una baja calidad

Impacto de una baja calidad de datos - Problemas



- Una baja calidad de datos hace que las empresas incurran en costos innecesarios de imprenta, envíos postales y recursos humanos
- Erosiona la credibilidad de una organización desde el punto de vista de clientes y proveedores
- Impide o dificulta decisiones correctas basadas en información precisa
- El problema de una baja calidad de datos empeora con el tiempo: expertos estiman que un 2% de los registros de una base de clientes se vuelven obsoletos en un mes, debido a que estos se mueren, se divorcian, se casan, se mudan, etc.
- Los errores de data entry, las migraciones de sistemas, los cambios en los sistemas fuente, etc. generan muchísimos nuevos errores
- Según un estudio del DataWarehousing Institute, los dos principales desafíos que enfrentan las compañías que implementan soluciones de CRM son el manejo de la calidad de datos y consistencia de los mismos (46% de las empresas evaluadas), y reconciliar los registros de los clientes (40% de las empresas).
- En el mismo estudio se estima que un 40% de las empresas sufrieron pérdidas, problemas o costos debido a una baja calidad de datos y que un 43% de las empresas probablemente experimentaron problemas similares, pero no detectaron la cuestión

Impacto de una baja calidad de datos - Problemas



Fuente: *TDWI Data Quality Survey, December 2001*

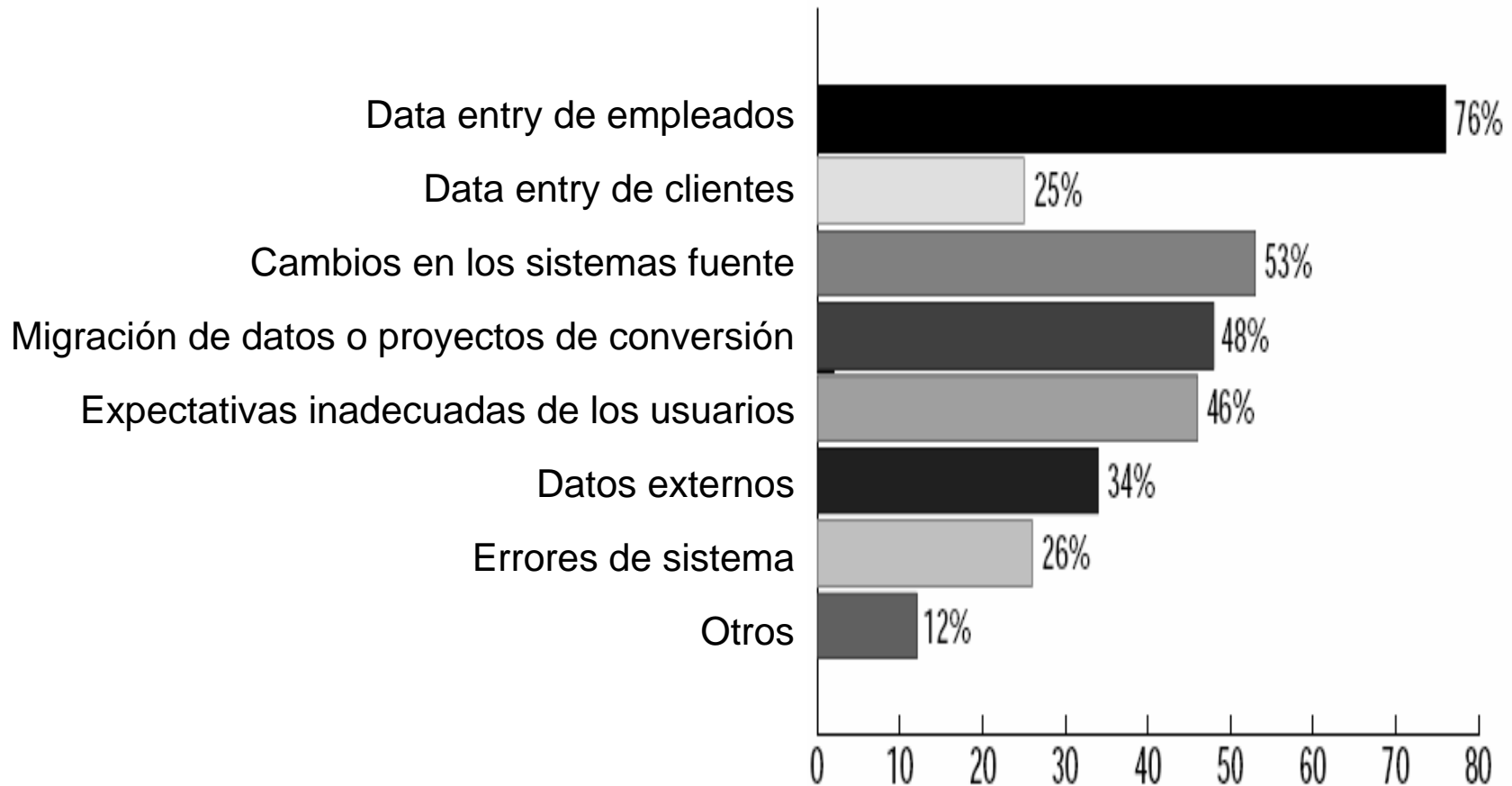
- El Data Warehousing Institute estimó en 2002 que los problemas de calidad de datos costaron a las empresas estadounidenses 611 mil millones de dólares anuales
- Larry English estima que de un 10 a un 25% de los ingresos operativos de una compañía se emplean en resolver los problemas ocasionados por una baja calidad de datos

Algunos ejemplos concretos



- Una compañía de seguros recibe 2 millones de reclamos mensuales, con 377 elementos de datos por reclamo. A una tasa de error de .001, los datos de los reclamos contienen 754.000 errores por mes y más de 9.04 millones por año. Una estimación muy básica de los costos incurridos debido a los errores es de 10 millones de dólares anuales
- Una empresa de telecomunicaciones perdió 8 millones de dólares en un mes solo en facturaciones erróneas a sus clientes debido a errores de data entry
- Un banco estimó el ROI de su proyecto de calidad de datos en 100.000 dólares anuales
- Una empresa internacional de educación en línea ahorra anualmente 500.000 dólares en correo, a partir de la implementación de un programa de calidad de datos. La empresa envía anualmente 20 millones de piezas de correo a clientes y prospectos

Fuentes de baja calidad en los datos



Fuente: *TDWI Data Quality Survey, December 2001*

Claves para la solución del problema de baja calidad de datos



- Considerar los datos como un recurso estratégico
- Desarrollar un programa de calidad de datos con un fuerte compromiso de los estamentos gerenciales superiores
- Utilizar profesionales experimentados en calidad de datos para monitorear y llevar a cabo el programa
- Utilizar herramientas computacionales apropiadas de calidad de datos
- Monitorear y mantener la calidad de datos una vez alcanzado un cierto nivel de calidad

¿Qué es calidad de datos?

- Calidad de datos es adecuar las características de los datos de un negocio a sus necesidades
- Las características de los datos que colectivamente conforman la calidad de datos son:
 - Precisión
 - Integridad
 - Consistencia
 - Completitud
 - Validez
 - Oportunidad
 - Accesibilidad
- Los conjuntos de datos totalmente libres de errores son sospechosos antes de un trabajo de calidad de datos y una utopía después de este

1. Lanzar un programa de calidad de datos que involucre a toda la empresa, especialmente la capa gerencial superior
2. Desarrollar un plan (o planes) del proyecto
3. Construir un equipo de calidad de datos
4. Revisar los procesos de negocios y la arquitectura de datos
5. Evaluar la calidad de datos
6. Limpieza de datos
7. Monitoreo de datos

- Oficial principal de calidad de datos
- Supervisor de calidad de datos
- Experto de dominio
- Líder de programa de calidad de datos
- Analista de calidad de datos
- Especialistas de herramientas
- Facilitador de mejoramiento de procesos
- Entrenador en calidad de datos

Etapas específicas de un proceso de calidad de datos



- Auditoría o profiling: evaluación de la calidad de datos en las áreas clave
- Construcción de reglas, esquemas, gramáticas y métricas de calidad de datos
- Aplicación de las herramientas construídas a la limpieza de datos
- Monitoreo y detección en tiempo real de nuevos errores

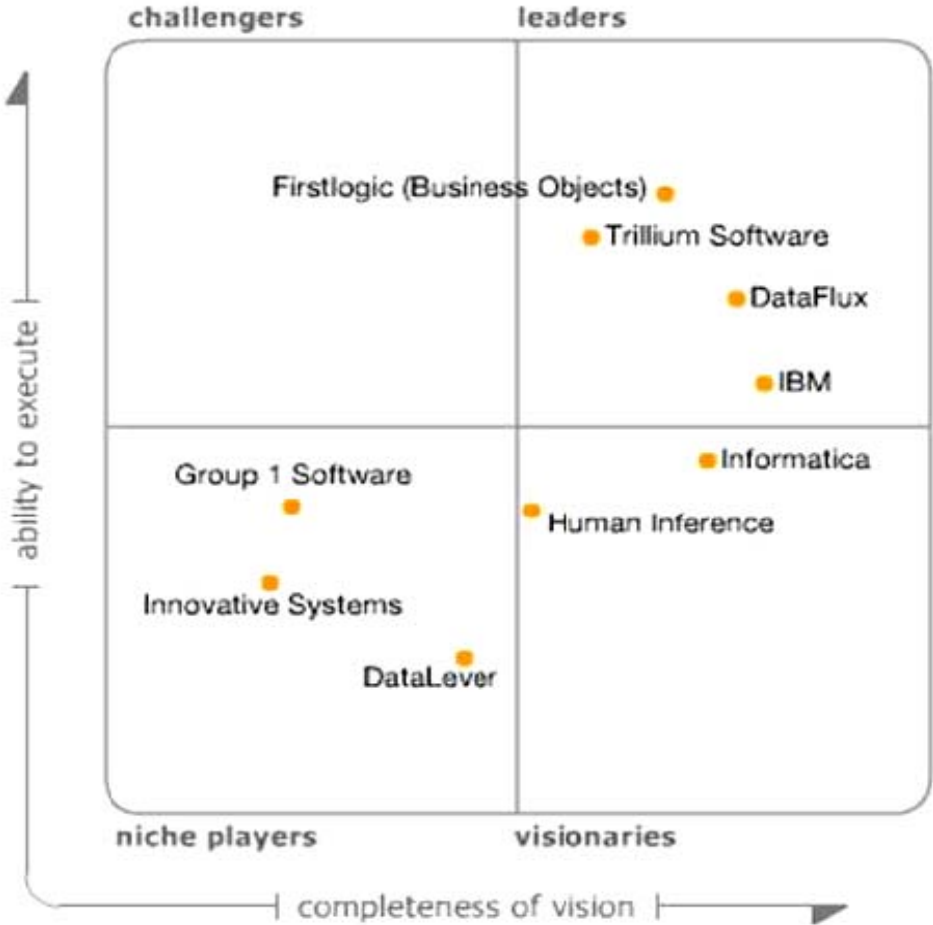
- Si bien un buen programa de calidad de datos es el resultado de una apropiada administración de personas y procesos, las herramientas tecnológicas tienen un papel importante
- Muchas empresas realizan tareas de limpieza de datos con herramientas caseras, programas en SQL o herramientas limitadas incluidas en productos de ETL
- El mercado de herramientas de calidad de datos es aun pequeño, pero se encuentra en expansión
- Aproximadamente un tercio de las empresas tienen actualmente herramientas específicas de calidad de datos

Funcionalidad de las herramientas de calidad de datos

- Profiling de datos
- Parsing de datos
- Estandarización o normalización
- Verificación
- Matching
- Consolidación

- Una herramienta de calidad de datos puede en forma “out of the box” resolver un 80% de los problemas de calidad de datos de una empresa, el resto debe resolverse complementando las herramientas con nuevas reglas, esquemas y gramáticas que surgen del análisis de los datos existentes
- Una herramienta debe incluir una base de conocimiento del lenguaje regional (español argentino, por ejemplo) y un “locale” apropiado (argentino). La aplicación de bases o locales más genéricos (p. ej. español) disminuyen notablemente la efectividad de las herramientas de calidad de datos
- Herramientas de ETL, text mining, data mining, estadísticas, etc. tienen funcionalidades sumamente limitadas de calidad de datos. No son apropiadas como herramientas centrales en un programa de calidad de datos
- Los criterios fundamentales para evaluar una herramienta de calidad de datos son: performance en relación con el precio, funcionalidades, facilidad de uso e integración con las herramientas existentes
- Algunas herramientas se concentraron históricamente en la calidad de datos de Nombres y Direcciones. Actualmente esto es insuficiente

Herramientas de calidad de datos



Próximamente,
evaluación de
herramientas

As of April 2006

Magic Quadrant for Data Quality Tools, 2006
Gartner: Abril 21, 2006

¿Cómo puede ayudar MAySA en calidad de datos?



- Evaluación del nivel de calidad de datos de una empresa y diagnóstico de problemas
- Asesoramiento y validación de proyectos de calidad de datos
- Mejoramiento de calidad de datos de una empresa inhouse
- Tercerización de mejoramiento de calidad de datos y mantenimiento
- Entrenamiento en técnicas y software de calidad de datos