

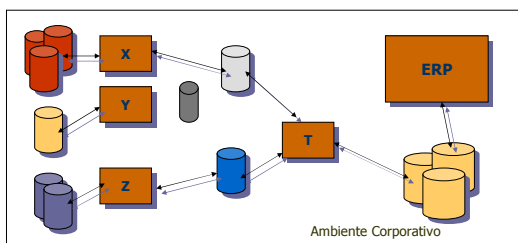
ETL

Extração, Tratamento e
Carga dos Dados em
DW

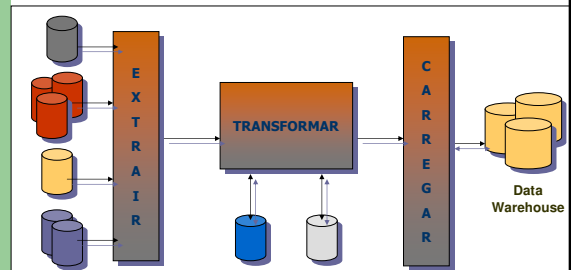
ETL

- DW caracteriza-se por ser um banco integrador de dados, histórico que sirva de suporte às decisões gerenciais.

ETL – Situação dos Dados no Ambiente Corporativo



ETL – Mais do que Copiar e Colar



ETL - Por quê tratar as informações

- Há várias origens para um mesmo dado: conflitos estruturais, de conteúdo e de formato;
- Permissividade dos sistemas OLTP;
- Não conformidade dos dados com as regras do negócio;
- Dados significativos em campos de entrada livre.

ETL – Barreiras/Dificuldades

Externas

- Exiguidade de prazos;
- Relacionamento com outras equipes:
 - » a falta de tempo dos analistas OLTP e
 - » a falta de paciência dos usuários;
- Falta de documentação:
 - » juras de fidelidade;
 - » esquecimentos e omissões;
- Necessidade de reduzir custos do projeto.

ETL – Barreiras/Dificuldades

Internas - Extração

- Erros humanos como troca de nomes e tipos de variáveis;
- Uso de regras de extração em desarmonia com as regras necessárias;
- Dados não esperados sendo extraídos (tipo, tamanho, formato e conteúdo).



Internas - transporte

- Incompatibilidade entre ambientes operacionais diferentes;
- Necessidade de quebrar o processo em etapas gerando dados estocados.



ETL – Barreiras/Dificuldades

Internas - Transformação

- Inserção de erros de transformação (escala de valor ou unidade);
- Uso de equações matemáticas.



Internas - Carga do Dados

- Aplicação de seleções não planejadas no momento da carga;
- Erros de codificação, tipo e formato de dados;
- Diferença entre modelagens de dados.



ETL – Conseqüências

Qual é o resultado de enganos sucessivos no ETL ?



- Perda da confiança dos usuários;
- Perda da motivação dos usuários para apoiar;
- Redução no moral da equipe de sistemas.

ETL - Conseqüências

Quanto custa a baixa qualidade nos dados ?

- Rever regras de negócio, distorções dos dados nas fontes e nos destinos;
- Rever o processo (etapa a etapa);
- Corrigir programas, rotinas e dados;
- Refazer a carga, analisar resultados e ajustar;
- Convencer os usuários, a gerência, etc a acreditar no projeto;
- Desenvolver aplicações fora do escopo original para ajudar no convencimento do usuário.



ETL - Responsabilidades

Todos!

- Nós;
- Nossos colegas dos sistemas OLTP;
- O usuários.

Todos estes agentes devem fazer parte do processo de uma forma estruturada.

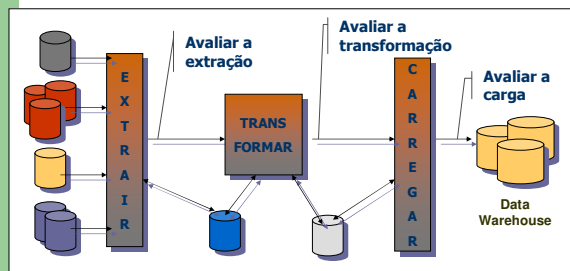


ETL – Metodologia para a qualidade dos dados

- Definir os requisitos e objetivos da qualidade de dados;
- Estabelecer um plano para qualidade de dados;
- Preparar os dados;
- Conduzir as análises de qualidade de dados;
- Desenvolver e comunicar recomendações e descobertas;
- Implementar o ETL;
- Analisar resultados das etapas do processo;

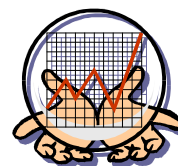


ETL – Metodologia para a qualidade dos dados



ETL – Preparação dos Dados

- Inspeccionar as fontes de dados;
- Levantar os problemas de baixa qualidade de dados;
- Identificar tabelas e arquivos a utilizar;
- Manter os dados sincronizados no tempo;
- Definir baterias de testes.



ETL – Métodos de Exploração e Análise dos Dados

- Análise de Domínio;
- Análise de integridade estrutural;
- Análise de obediência às regras de transformação.



ETL – Análise de Domínio

Método: Distribuição de frequência

Permite:

- Identificar formatos dos dados;
- Intervalos de valores;
- Uso de valores default;
- Presença de valores não legítimos;
- Dados de valor diferente e de mesmo significado;
- Presença de valores suspeitos;
- Revelar valores tecnicamente válidos mas fora das regras do negócio.

ETL – Análise de Integridade Estrutural

Integridade na chave primária	Esta situação pode aparecer na junção de dois ou mais arquivos.
Integridade referencial	Pode aparecer sempre que a integridade referencial não é obrigatória nos relacionamentos.
Cardinalidade dos relacionamentos	Verificar se ocorre na prática, a cardinalidade especificada para os relacionamentos.

ETL – Análise de Obediência às Regras de Transformação



Transformação Direta: Dados movidos da fonte para o destino sem transformação.	Transformação mapeada: Valores de dados são mapeados entre fonte e destino por tabelas de correlação.	Transformação Qualificada: Utilizam-se de regras de seleção para definir conteúdos das transformações.	Transformação Derivada: Valores resultam da aplicação de equações.
--	---	--	--

ETL – Comunicar Recomendações e Descobertas

- Processo de limpeza dos dados;
- Métodos para prevenir a proliferação de defeitos.
- Alterações nos sistemas-fonte;
- Construção de rotinas para detecção de defeitos;
- Relatórios com defeitos encontrados, causa e efeito

