

Especialista em Data Warehouse

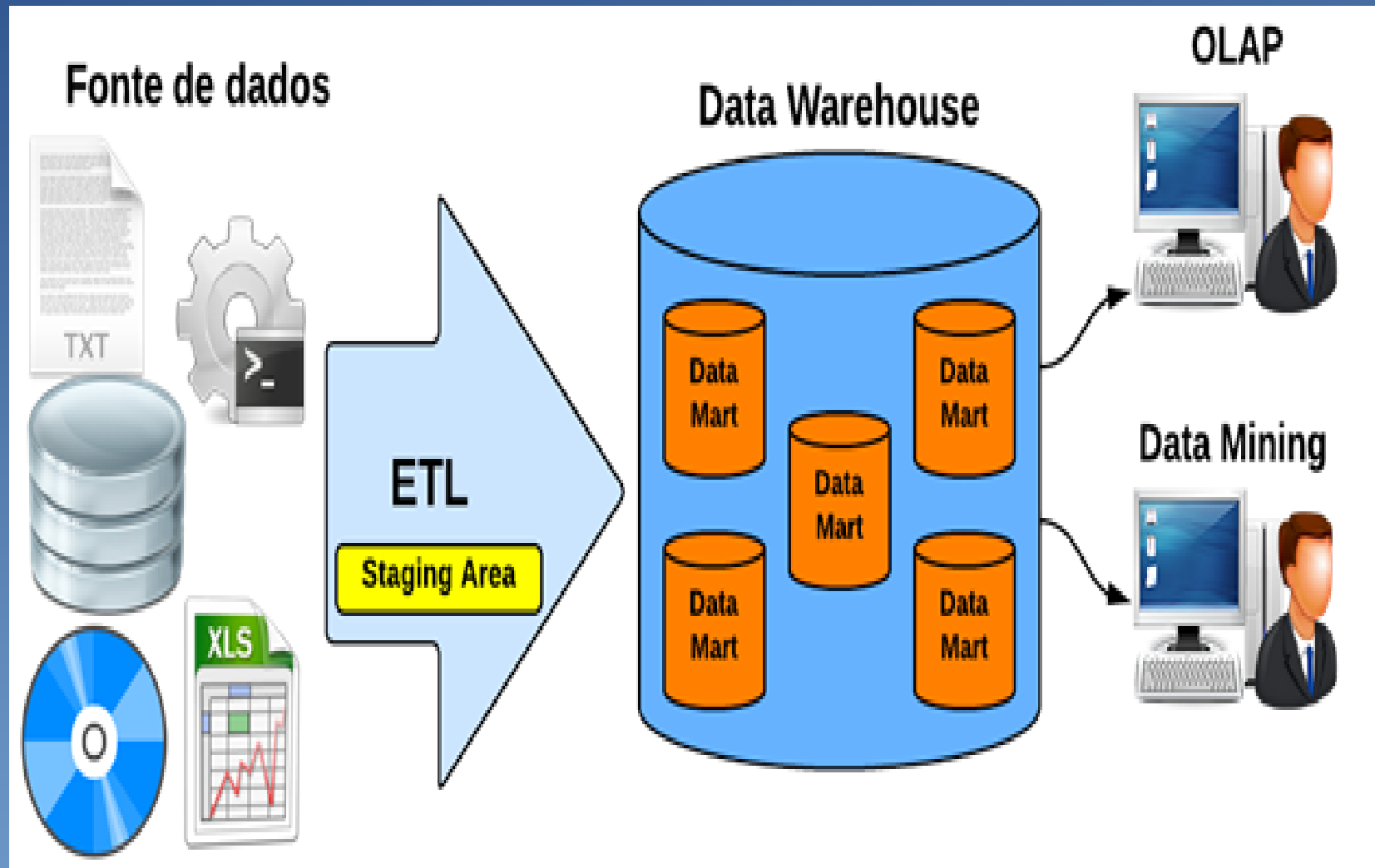
Autores

William Ricardo e Gabriel Luiz

O que é um Data Warehouse?

- Um Data Warehouse (armazém de dados) é utilizado para armazenar informações relativas às atividades de uma organização de forma consolidada. Possibilita a análise de grandes volumes de dados, que são coletados a partir de sistemas transacionais (OLTP – Online Transaction Processing) alguns exemplos de OLTP são processamento da folha de pagamento, faturamento, estoque e qualquer sistema de uma empresa que tenha como função a gravação ou leitura de processos e transações da mesma.
- O conceito de Data Warehouse surgiu da necessidade de integrar dados corporativos espalhados em diferentes máquinas e sistemas operacionais, para que fosse possível tornar os dados acessíveis a todos os usuários dos níveis decisórios.

Arquitetura



ETL

ETL como a própria tradução já diz é um processo de extração, transformação e carga, para a construção de um Data Warehouse ou Data Mart.

O processo de ETL (Extract, Transform and Load) é um processo que exige esforço e a maior parte do tempo de construção de um Data warehouse, este processo vai extrair dados de fontes de dados heterogêneas e tem que alimentar o Data Warehouse de forma homogênea e concisa, pois vai servir de base para gerar relatórios e gráficos de apoio à decisão para a gerencia da corporação e não pode trazer resultados errôneos.

Esse processo deve definir claramente quais informações são relevantes e que farão parte dos dados que vão alimentar o Data Warehouse, que deverá ter a estrutura desenhada com relação aos dados que serão extraídos dos vários sistemas e as informações que devem ser geradas a partir deste repositório.

Depois de selecionados os dados que serão carregados no Data Warehouse, vem a parte de tratamento ou transformação e limpeza dos dados, que consiste em padronizar os dados com relação ao tamanho e tipo, substituição de caracteres estranhos, correção de erros de digitação, comparação fonética para evitar duplicidade de informações, por exemplo o mesmo nome escrito de formas diferentes, com ou sem acentuação etc.

Data Mart

Data mart (repositório de dados) é sub-conjunto de dados de um Data warehouse (armazém de dados). Geralmente são dados referentes a um assunto em especial (ex: Vendas, Estoque, Controladoria) ou diferentes níveis de sumarização (ex: Vendas Anual, Vendas Mensal, Vendas 5 anos), que focalizam uma ou mais áreas específicas. Seus dados são obtidos do DW, desnormalizados e indexados para suportar intensa pesquisa. Data marts extraem e ajustam porções de DWs aos requisitos específicos de grupos/departamentos.

As questões de planejamento, projeto e implementação são as mesmas encontradas em qualquer fase da implementação de um DW, pois uma parte dos especialistas argumenta que o DW é uma evolução natural de um Data mart que começou localizado e cresceu para atender um escopo maior. Essa visão tem defesa menos técnica e mais intuitiva do que sua contra-parte: o Data mart é criado a posteriori do DW, pois ele seria montado com dados extraídos do DW para atender especificamente um setor da organização..

Numa visão comparativa dos dados, onde consideramos os requisitos escopo, integração, tempo, agregação, análise e dados voláteis, percebemos que a diferença está no escopo, pois enquanto o DW é feito para atender uma empresa como um todo, o data mart é criado para atender um sub-conjunto da empresa. Repare que atender um sub-conjunto da empresa pode significar reunir dados de outros setores, já que, na prática, raramente um único setor possui ou gera toda informação que precisa. Vem dessa observação a defesa da tese de que o Data mart é construído após o DW.

Ferramentas de um Data Warehouse

OLAP e Data Mining

O OLAP (Online Analytical Processing) é a capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas. As aplicações OLAP são usadas pelos gestores em qualquer nível da organização para lhes permitir análises comparativas que facilitem a sua tomada de decisões diárias.

O Data Mining (Mineração de dados) são sistemas que tem como função principal a procura de padrões e detectam os relacionamentos entre as informações dando origem a novos sub-grupos de dados organizados ou seja esse sistema é um organizador de dados.

Benefícios

Entre os vários benefícios decorrentes da implantação da tecnologia de Data Warehouse em uma organização, pode-se destacar os seguintes:

- Informações unificadas através de uma fonte centralizada
- Maior agilidade e independência na extração de informações para os usuários
- Ampliação da capacidade de análise do ambiente;
- Suporte à tomada de decisão
- Maior facilidade para apuração da qualidade dos dados dos sistemas transacionais
- Disponibilização de informações históricas para identificar tendências

Especialização e Certificação

- Formação superior em Marketing ou Administração com ênfase em análise de sistemas, pós-graduação em Banco de dados e um MBA em Business Intelligence fazem com que o profissional nessa área seja mais valorizado.
- Certificações em Business Intelligence (Oracle Business Intelligence e MCSE da Microsoft)

Salário de um Especialista em Data Warehouse

O Salário de um Especialista em Data Warehouse pode chegar a ganhar até R\$17 mil, dependendo do seu tempo de empresa pois como estamos falando de um banco de dados, contendo informações que acompanham um histórico mais importante esse profissional vai se tornando para a empresa e logo é mais valorizado pela mesma.

Porte da empresa	0 - 2 anos	3 - 5 anos	6 - 9 anos	>10 ano
Pequena ou Media	2.500 - 5.500	3.500 - 7.000	5.500 - 11.000	8.500 - 15.000
Grande	3.000 - 6.000	4.500 - 9.000	6.500 - 12.500	10.000 - 17.000

Referências

- <https://vivianeribeiro1.wordpress.com/2011/03/30/o-que-e-data-warehouse/>
- https://pt.wikipedia.org/wiki/Data_mart
- <http://exame.abril.com.br/carreira/ferramentas/tabela-de-salarios-rh/?empresa=ti>
- https://pt.wikipedia.org/wiki/Armaz%C3%A9m_de_dados#Ferramentas
- <http://corporate.canaltech.com.br/o-que-e/business-intelligence/o-que-significa-oltp-e-olap-na-pratica/>
- <https://pt.wikipedia.org/wiki/OLTP>
- <http://www.alvoconhecimento.com.br/saiba-o-que-e-data-mining/>
- <http://www.devmedia.com.br/extract-transformation-and-load-etl-ferramentas-bi/24408>
- <https://www.microsoft.com/learning/pt-br/mcse-sql-business-intelligence.aspx>
- <http://www.oracle.com/br/solutions/midsize/oracle-products/business-intelligence/index.html>