



Facisa

Faculdade de Ciências Sociais Aplicadas

FCM

Faculdade de Ciências Médicas de Campina Grande



ESAC - Escola Superior de Aviação Civil

Av. Senador Argemiro de Figueiredo, 1901 - Itararé - CEP: 58104-590 - Campina Grande, PB - Fone: 83 2101 8800

www.cesed.br - facisa@cesed.br - fcm@cesed.br - esac@cesed.br

CNPJ: 02.108.023/0001-40

Banco de dados 2

Data warehouse

Agenda

- Por que?
- O que é
- Histórico
- Elementos básicos
 - Source system, data staging area, etc.
- Servidor de apresentação
- Modelo dimensional
- Processos de manipulação dos dados

Banco de Brasil S/A

http://www.dadosenegocios.com.br/tecnologia_11.html

Data Warehouse

- Concebido para providenciar um modelo de arquitetura visando obter o fluxo de dados de sistemas operacionais e gerar informação para dar suporte à decisões.

http://en.wikipedia.org/wiki/Data_warehouse

Histórico

- Conceito na década de 80;
- Barry Devlin e Paul Murphy desenvolvem:
The business Data warehouse for IBM;
- 1990 – Red brick System;
- 1991 – Prism Warehouse manager;
- 1997 – ORACLE com suporte a *Star queries* é lançado;

Elementos básicos

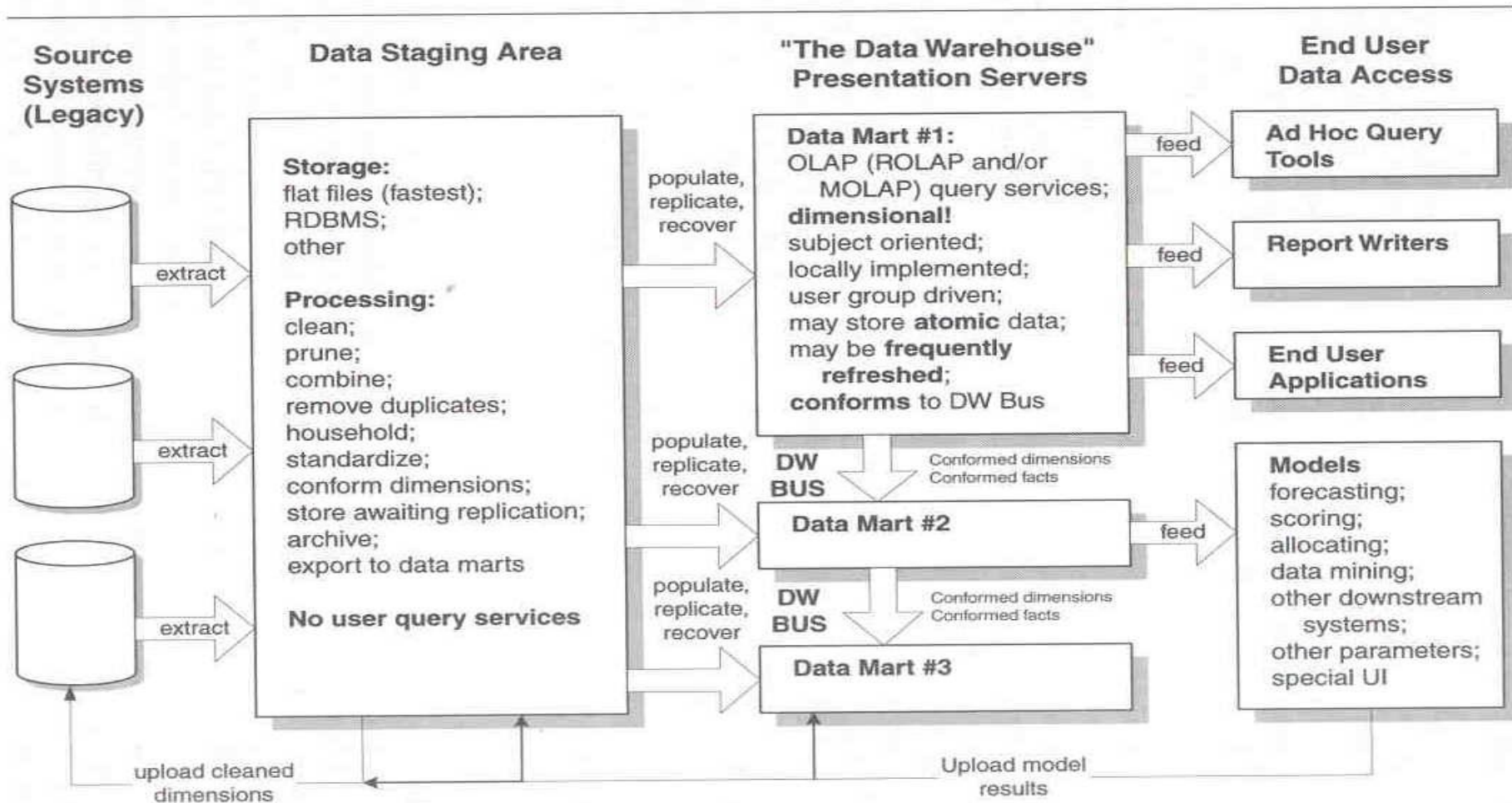


FIGURE 1.1 The basic elements of the data warehouse.

Source System

- Principal função é capturar as transações do negócio;
- Prioridades são disponibilidade e performance;

Data Staging Area

- Área de armazenamento: Repository;
- Conjunto de processos que limpam, transformam, e preparam os dados da fonte para uso;
- Provê serviços de consulta e apresentação;
- Não necessariamente utiliza dados relacionais.

Servidor de Apresentação

- A máquina física alvo no qual o data warehouse é organizado e armazenado para consulta direta pelos usuários finais, geradores de relatórios, e outras aplicações
- Pode estar organizada no formato de esquemas estrelas ou de cubos dimensionais

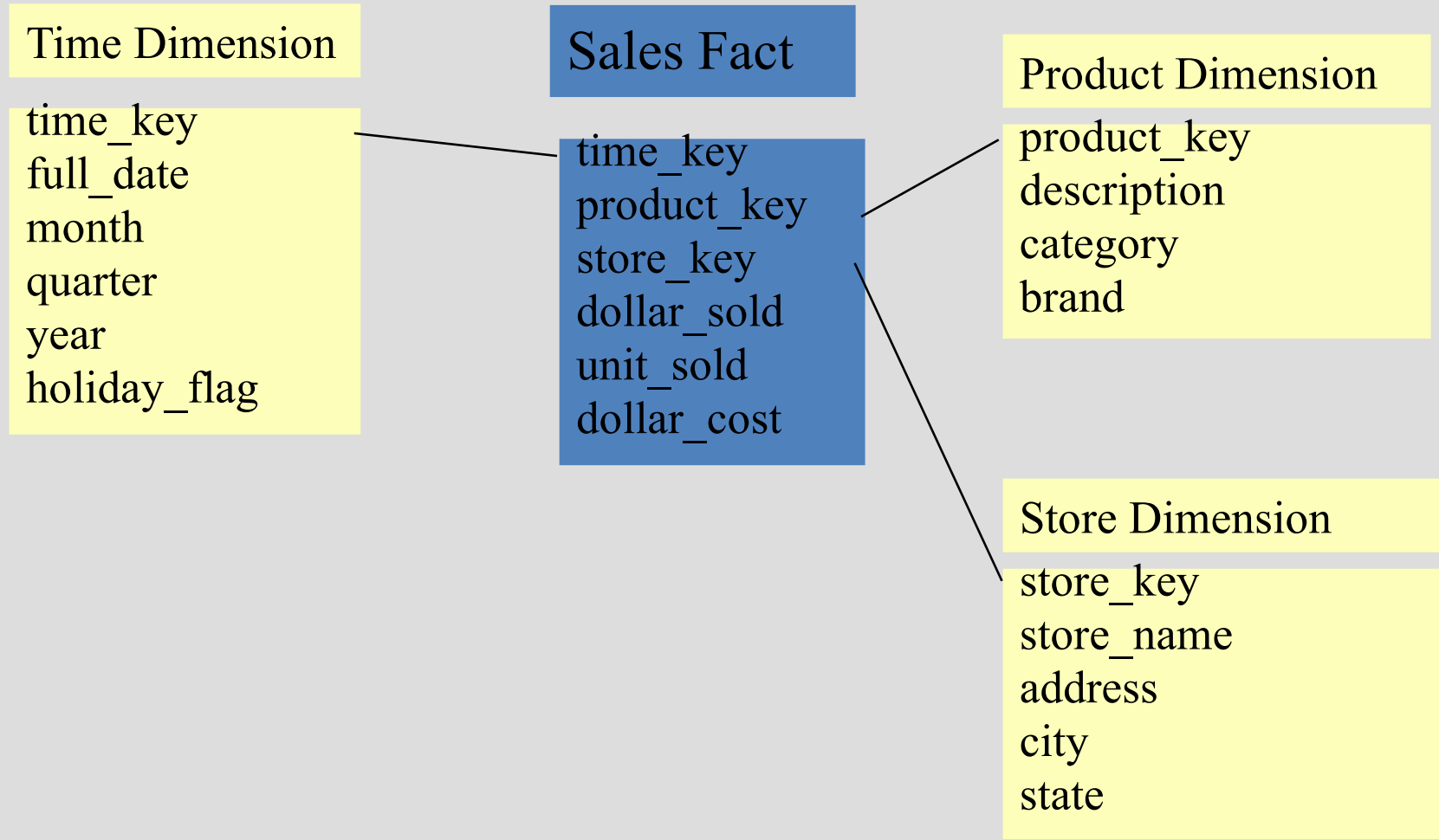
Modelo Dimensional

- Uma disciplina específica para modelagem de dados que é um alternativa à modelagem entidade-relacionamento
- Objetivos: compreensão dos dados e flexibilidade para mudanças
- Principais componentes: tabelas de fatos e tabela de dimensões

Modelo Dimensional

- Tabela de fatos
 - Tabela principal em todo modelo dimensional que contém medidas sobre o negócio (fatos)
 - Os fatos mais úteis são numéricos e aditivos
 - Representa relacionamentos muitos pra muitos
 - Contém um conjunto de duas ou mais chaves estrangeiras para a junção das respectivas tabelas de dimensões
- Tabela de Dimensão
 - Complementam as tabelas de fatos
 - A maioria contém muitos atributos textuais que são a base para restringir e agrupar dentro de consultas ao data warehouse
 - Cada dimensão é definida por sua chave primária que serve de base para restrições de integridade referenciais nas tabelas de fatos.

Modelo Dimensional

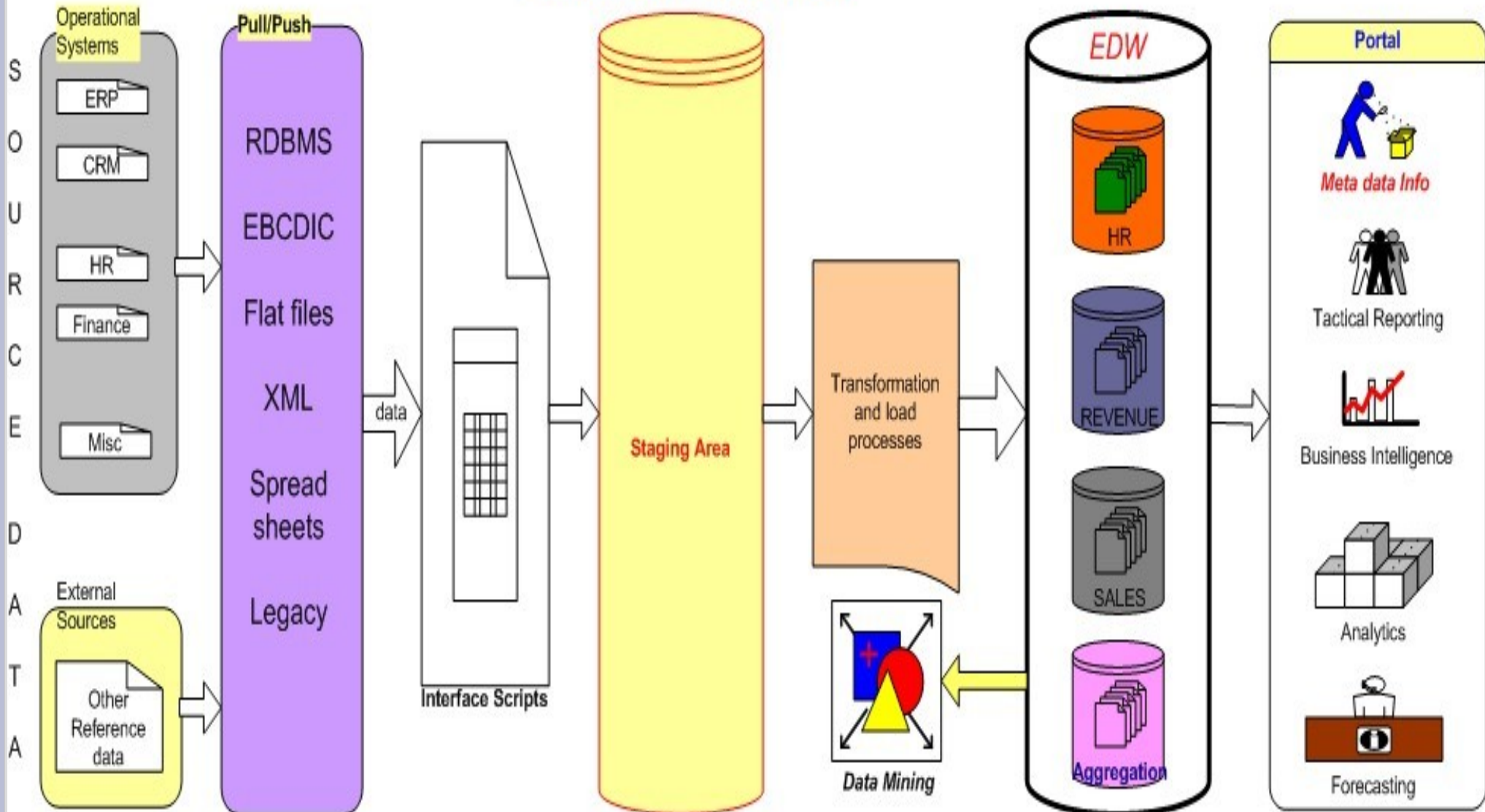


Data Mart

- Um subconjunto lógico do data warehouse completo
- O data warehouse é constituído da união de todos os seus data marts
- Pode ser visto como a restrição do data warehouse a um único processo de negócios

Data Mart

Data Warehouse Architect



Data Mart

- Recomendações:
 - Todo data mart deve ser representado por um modelo dimensional e todo data mart deve ser construído de dimensões e fatos “em conformidade” possibilitando seu re-uso
 - Projetar com um mix de top-down e botton-up design
 - Data marts são baseados em dados granulares e podem ou não conter “sumarios” ou “agregados” por razões de performance

QED DataMart

Business Intelligence

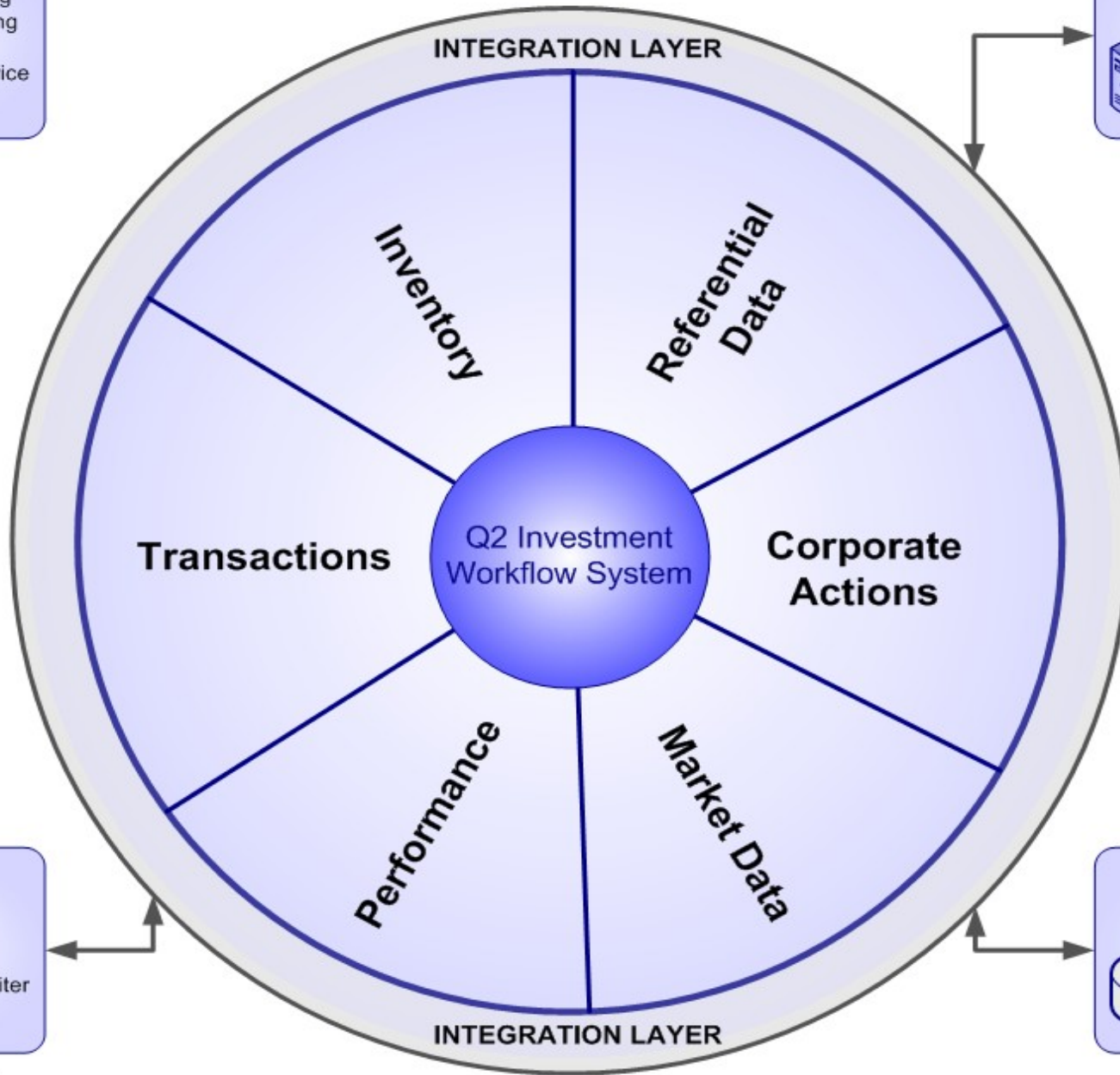


- Executive
- Marketing
- Accounting
- Audit
- Cust. Service
- ...

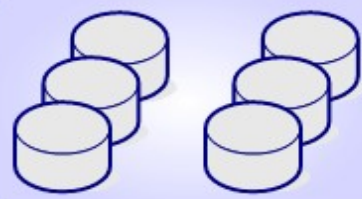
Online Transaction Processing Systems (OLTP)



- Banking
- Order Proc.
- Ecommerce
- ETrading
- ...



- CRM
- G/L
- TOM
- Report Writer
- ...



External Decision Support Systems

Data Warehouses & DataMarts

On-Line Analytic Processing - OLAP

- A atividade de consultar e apresentar dados textuais e numéricos dos data warehouses em um estilo dimensional específico;
- Quase sempre baseado em um cubo de dados multidimensional explícito.



OLAP

- ROLAP
 - Um conjunto de interfaces e aplicações que dão a um banco de dados relacional um “sabor” dimensional
- MOLAP
 - Um conjunto de tecnologias de interfaces, aplicações, e bancos de dados proprietários que têm um forte viés multidimensional

- Databases
 - Oracle11g (sh) - OLAP11g
 - Schemas
 - SH
 - Analytic Workspaces
 - SALES (attached RW)
 - Dimensions
 - PRODUCT
 - CUSTOMER
 - Cubes
 - SALESCUBE
 - Measures
 - Calculated Measu
 - Mappings
 - Views
 - SALESCUBE_VI
 - Cube Scripts
 - Load_and_Agg
 - Data Security
 - Measure Folders
 - Data Security Roles
 - Reports

General Aggregation Partitioning Storage Materialized Views

Specify General Cube Information

Measure Data Viewer

File

Dialog 12 B i U

Page Items

| | Sales |
|---|------------|
| | P |
| F | 349,540.00 |
| M | 569,303.00 |

Aharoni 8 B i U

| Dimension | Sales P |
|-----------|------------|
| F | 349,540.00 |
| M | 569,303.00 |

Input fields for configuration or data entry.

to the Selected Dimensions list

```

SELECT CUST_GENDER, CUST_FIRST_NAME||CUST_LAST_NAME, PROD_PACK_SIZE, PROD_NAME, FISCAL_YEAR, FISCAL_MONTH_DESC, DAY_NAME, SUM
(QUANTITY_SOLD) AS QUANTITY, SUM(AMOUNT_SOLD) AS SALES
FROM SALES a, CUSTOMERS b, PRODUCTS c, TIMES d
where a.CUST_ID = b.CUST_ID and ...

```

Last Execution Details

[SQL Repair Advisor](#)
[SQL Details](#)
[Schedule SQL Tuning Advisor](#)

[Results](#)
[Statistics](#)
Plan

Data Source **Cursor Cache**

Capture Time **Aug 30, 2007 8:49:01 PM (UTC-04:00)**


Parsing Schema **SH**

Optimizer Mode **ALL_ROWS**

Additional Information

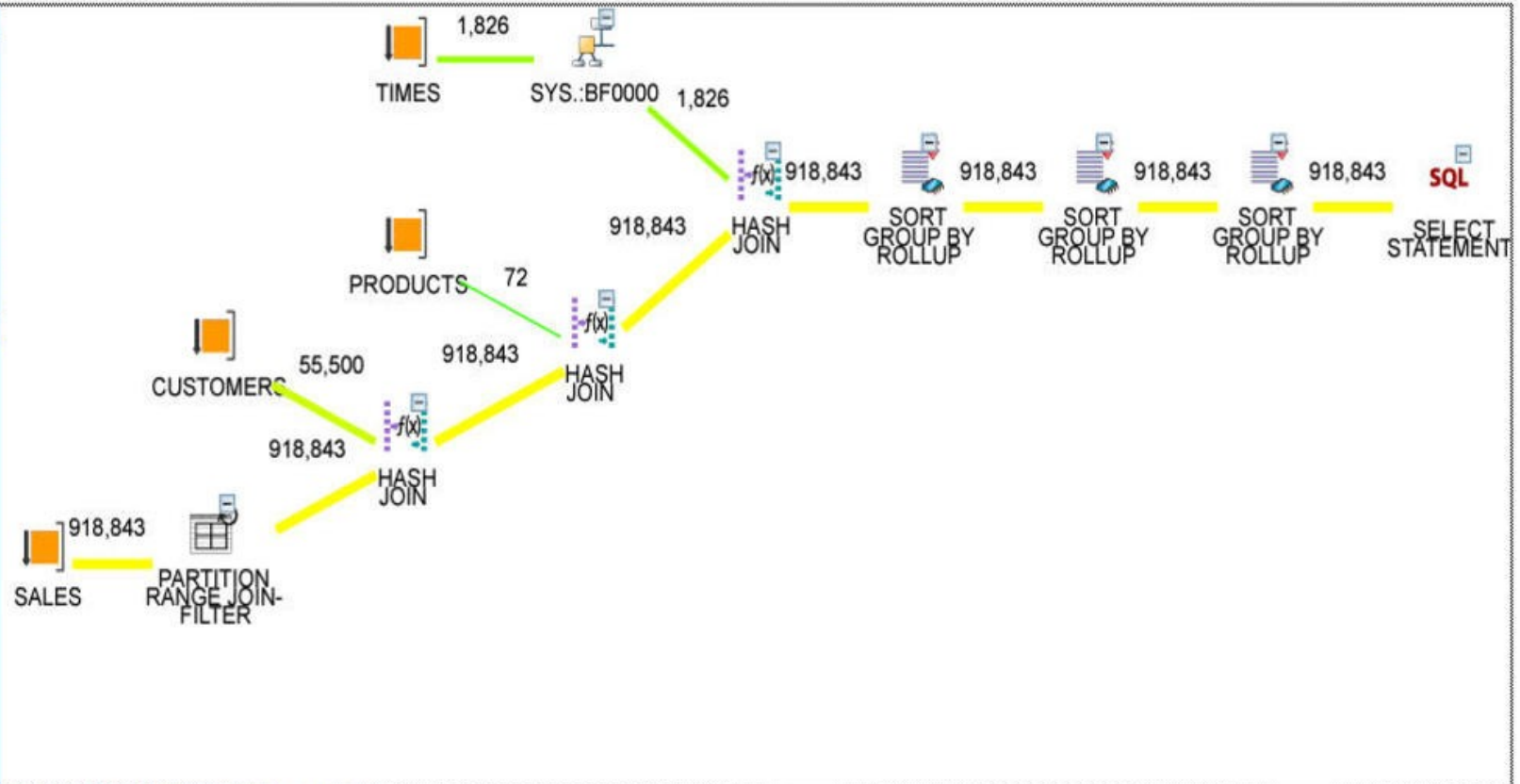
View Graph Table

Overview



Selection Details

Nothing Selected



TIP For an explanation of the icons and symbols used in this page, see the [Icon Key](#)

Hide Explain Rewrite

Explain Rewrite

Query rewrite transforms a SQL statement expressed in terms of tables or views into a statement accessing one or more materialized views that are defined on the detail tables. Oracle compares the cost for the query with and without rewrite and selects the least costly alternative.

Query Rewritten **No**

Materialized Views Used In Query Rewrite

Processos

- Extração
 - Ler e entender o dado fonte e copiar partes que são necessárias para a área de data staging no futuro
- Transformação
 - Limpar os dados através de:
 - Correção de erros de digitação
 - Resolução de conflitos
 - Tratamento de dados perdidos
 - Conversão para formatos padrão

Processos

- Carregamento e Indexação
 - Envolve a replicação das tabelas de dimensões e fatos e a apresentação dessas tabelas para as facilidades de ingestão de cada data mart
 - Normalmente efetuado em um único passo de carga
 - O data mart alvo tem que indexar os novos dados para garantir performance

Processos

- Checagem de Qualidade
 - Efetuada rodando ferramentas de relatório de exceções sobre o conjunto de dados carregados
 - Checar por inconsistências em contagens, somas, séries de tempos
 - Ferramentas de relatório de exceções vêm normalmente com as facilidade de geração de relatórios do usuário final dos data marts

Processos

- Publicação
 - Notificação de que os novos dados estão prontos
 - Comunicação de quaisquer mudanças que tenham ocorrido nas dimensões e novas “assumptions” que tenham sido incluídas nos fatos medidos e calculados

Processos

- Atualização
 - Envolve:
 - Correção de dados
 - Adequação a mudanças em status, hierarquias, corporate ownership
- Consulta
 - Envolve
 - Consultas ad hoc, escrita de relatórios, aplicações de tomadas de decisão, mineração de dados, etc.

Processos

- Feedback/Realimentação
 - Pode envolver a ingestão de dados limpos para os sistemas legados e o resultado de processos/consultas complexas para o data mart
- Auditoria:
 - Processo de conhecer de onde os dados vieram e como foram obtidos/calculados, etc (metadados de linhagem)

Processos

- Segurança
 - Dilema: a necessidade de publicar os dados da forma mais ampla possível e ao mesmo tempo proteger dados sensíveis contra hackers, espiões industriais, etc.
- Backup e Recuperação de Falhas
 - Envolve aspectos tais como onde e quando tirar os “snapshots” de dados legados e dos metadados para propósitos de arquivamento e recuperação de desastres

Transformação

- Remoção de campos selecionados de dados legados que não são úteis para o data warehouse
- Combinação de fontes de dados através de casamento exato em valores chave ou casamentos “fuzzy” em atributos não-chave
- Criação de chaves *surrogate* para cada registro de dimensão de forma a evitar a dependência em chaves legadas,
- Construção de agregados para melhoria de performance em consultas comuns

Estudo de caso

SAP Business Information Warehouse 3.0
for
Petrobrás S/A

Oracle DataWarehouse

Oracle Exadata V2
On
Oracle 11g for Linux

<http://www.oracle.com/us/solutions/datawarehousing/index.htm>

Obrigado!

<http://gmedeiros.net/site/index.php/tutoriais/71-data-warehouse>