

Benchmark SSB

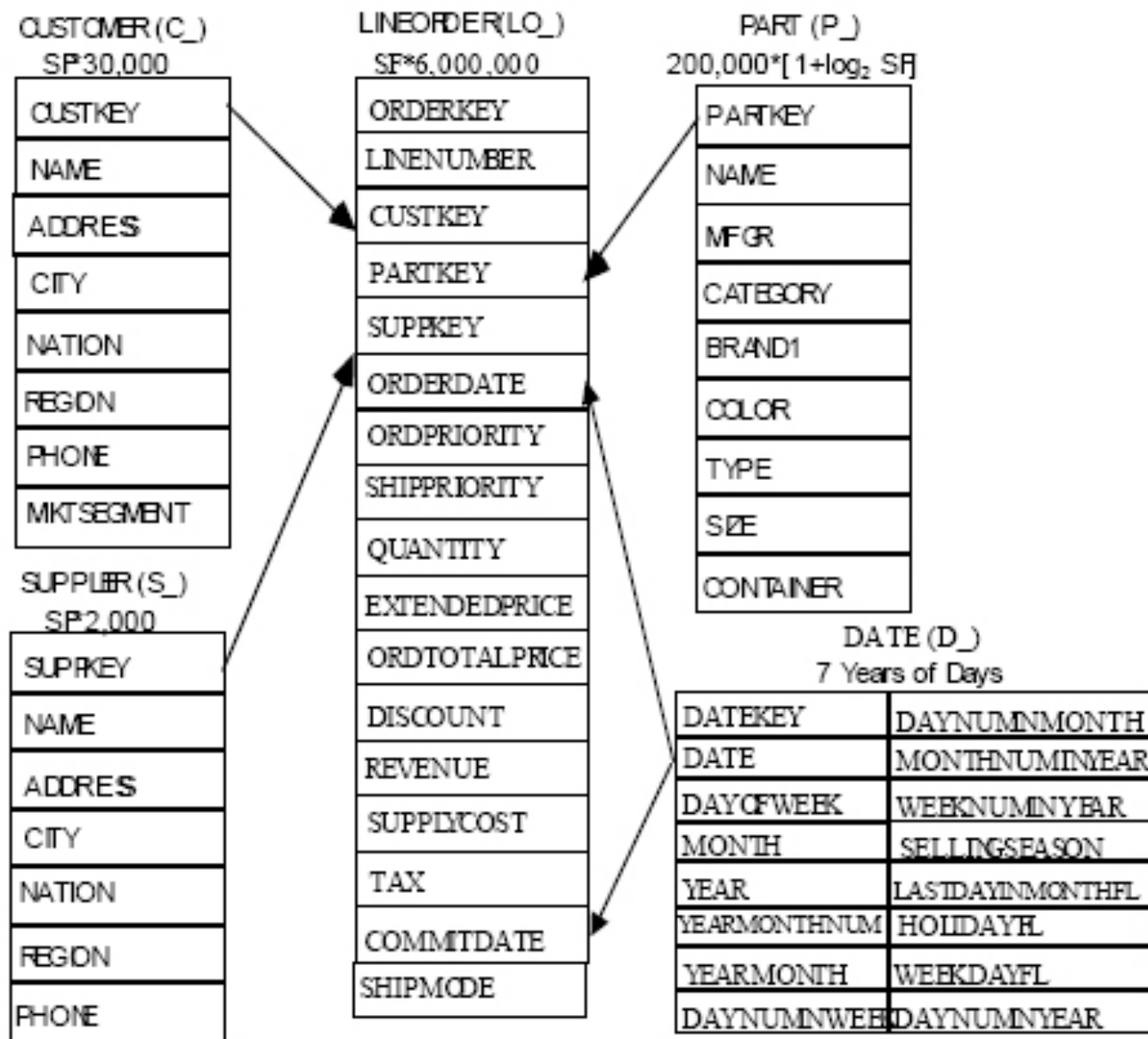
Processamento Analítico de Dados
Profa. Dra. Cristina Dutra de Aguiar Ciferri
Prof. Dr. Ricardo Rodrigues Ciferri
Samara Martins

Características

- Usado para avaliar o desempenho de um DW
- Baseado no *benchmark* TPC-H
 - Foram juntadas as tabelas de fatos Lineitem e Order, gerando a tabela de fatos Lineorder
 - Foi excluída a tabela de fatos Partsupp
 - Foi adicionada a tabela de dimensão Date

<http://www.cs.umb.edu/~poneil/StarSchemaB.pdf>

TPC-S



LINEORDER Table Layout SF*6,000,000

LO_ORDERKEY numeric (int up to SF 300) first 8 of each 32 keys populated

LO_LINENUMBER numeric 1-7

LO_CUSTKEY numeric identifier FK to C_CUSTKEY

LO_PARTKEY identifier FK to P_PARTKEY

LO_SUPPKEY numeric identifier FK to S_SUPPKEY

LO_ORDERDATE identifier FK to D_DATEKEY

LO_ORDERPRIORITY fixed text, size 15 (See pg 91: 5 Priorities: 1-URGENT, etc.)

LO_SHIPPRIORITY fixed text, size 1

LO_QUANTITY numeric 1-50 (for PART)

LO_EXTENDEDPRICE numeric $\leq 55,450$ (for PART)

LO_ORDTOTALPRICE numeric $\leq 388,000$ (ORDER)

LO_DISCOUNT numeric 0-10 (for PART, percent)

LO_REVENUE numeric (for PART:
(lo_extendedprice*(100-lo_discnt))/100)

LO_SUPPLYCOST numeric (for PART)

LO_TAX numeric 0-8 (for PART)

LO_COMMITDATE FK to D_DATEKEY

LO_SHIPMODE fixed text, size 10 (See pg. 91: 7 Modes: REG AIR, AIR, etc.)

Compound Primary Key: LO_ORDERKEY,
LO_LINENUMBER

Consultas

- Define diferentes tipos de consulta
- Exemplo: Q1
 - quantifica a renda a partir da eliminação de certos descontos da empresa, dada uma porcentagem de produtos enviados em um determinado ano

Consulta Q1.1

Q1.1 YEAR = 1993, DISCOUNT = 2, QUANTITY = 25, so predicates are d_year = 1993, lo_quantity < 25, lo_discount between 1 and 3.

```
select sum(lo_extendedprice*lo_discount) as revenue
from lineorder, date
where lo_orderdate = d_datekey
and d_year = 1993
and lo_discount between 1 and 3
and lo_quantity < 25;
```

$FF = (1/7)*0.5*(3/11) = 0.0194805$. Number of lineorder rows selected, for SF = 1, is $0.0194805*6,000,000 \approx 116,883$.

Consulta Q1.2

Q1.2 d_yearmonthnum = 199401, lo_quantity between 26 and 35, lo_discount between 4 and 6.

```
select sum(lo_extendedprice*lo_discount) as revenue
from lineorder, date
where lo_orderdate = d_datekey
and d_yearmonthnum = 199401
and lo_discount between 4 and 6
and lo_quantity between 26 and 35;
```

$FF = (1/84)*(3/11)*0.2 = 0.00064935$. Number of lineorder rows selected, for SF = 1:
 $0.00064935*6,000,000 \approx 3896$.

Consulta Q1.3

Q1.3 d_weeknuminyear = 6 and d_year = 1994,
lo_quantity between 36 and 40, lo_discount between 5
and 7.

```
select sum(lo_extendedprice*lo_discount) as revenue
from lineorder, date
where lo_orderdate = d_datekey
and d_weeknuminyear = 6
and d_year = 1994
and lo_discount between 5 and 7
and lo_quantity between 26 and 35;
```

$FF = (1/364)*(3/11)*0.1 = .000075$. Number of li-
neorder rows selected, for $SF = 1$, is
 $.000075*6,000,000 \approx 450$.

Medidas de Desempenho

- Os relatórios devem conter
 - planos de consulta
 - número de linhas acessadas
 - tempo execução da consulta
 - tempo da CPU utilizada
 - uso de entrada e saída