Tecnologias Oracle para DW
Visões Materializadas no Oracle

Prof. Dra. Cristina Dutra de Aguiar Ciferri

Prof. Dr. Ricardo Rodrigues Ciferri

Bruno Tomazela

Renata Miwa Tsuruda

Objetivo

- Criar e utilizar visões materializadas no Oracle
- Comparar o desempenho de consultas nas seguintes condições:
  - DW Desnormalizado x DW Normalizado
  - Tabela de Fatos x Visões Materializadas

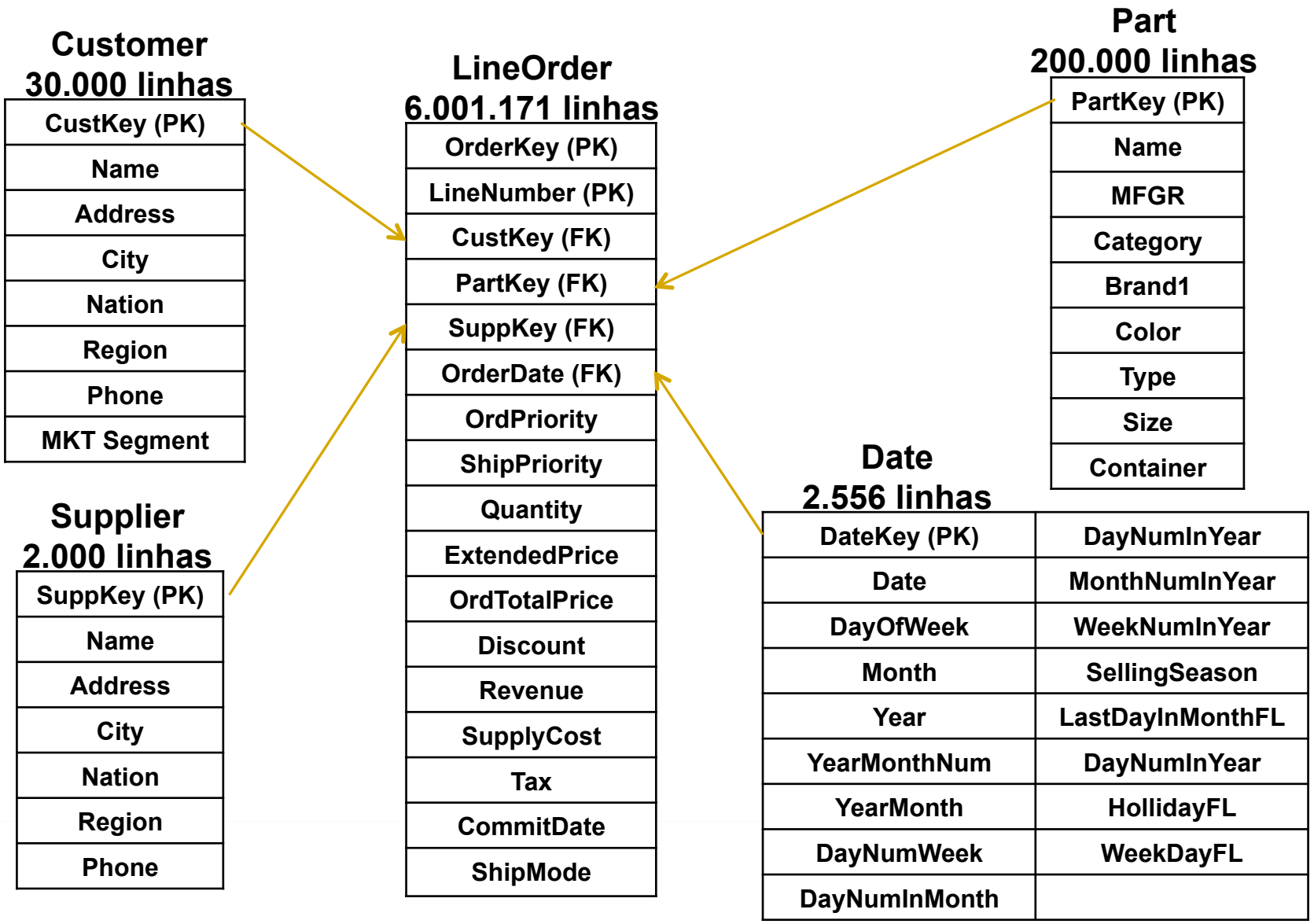
Base de Dados

- Data Warehouses construídos com base no Star Schema Benchmark (SSB) [O'NEIL, 2009]
  - Geração de tabelas de dimensão e de fatos
  - Fator de escala = 1 (aproximadamente 6 milhões de tuplas na tabela de fatos)

Data Warehouse Desnormalizado

- Redundância de dados (esquema estrela)
- Espera-se um bom desempenho na execução de consultas

Data Warehouse Desnormalizado



Data Warehouse Desnormalizado

- Hierarquias existentes:
  - Dimensões *Customer* e *Supplier*
    - *Address < City < Nation < Region*

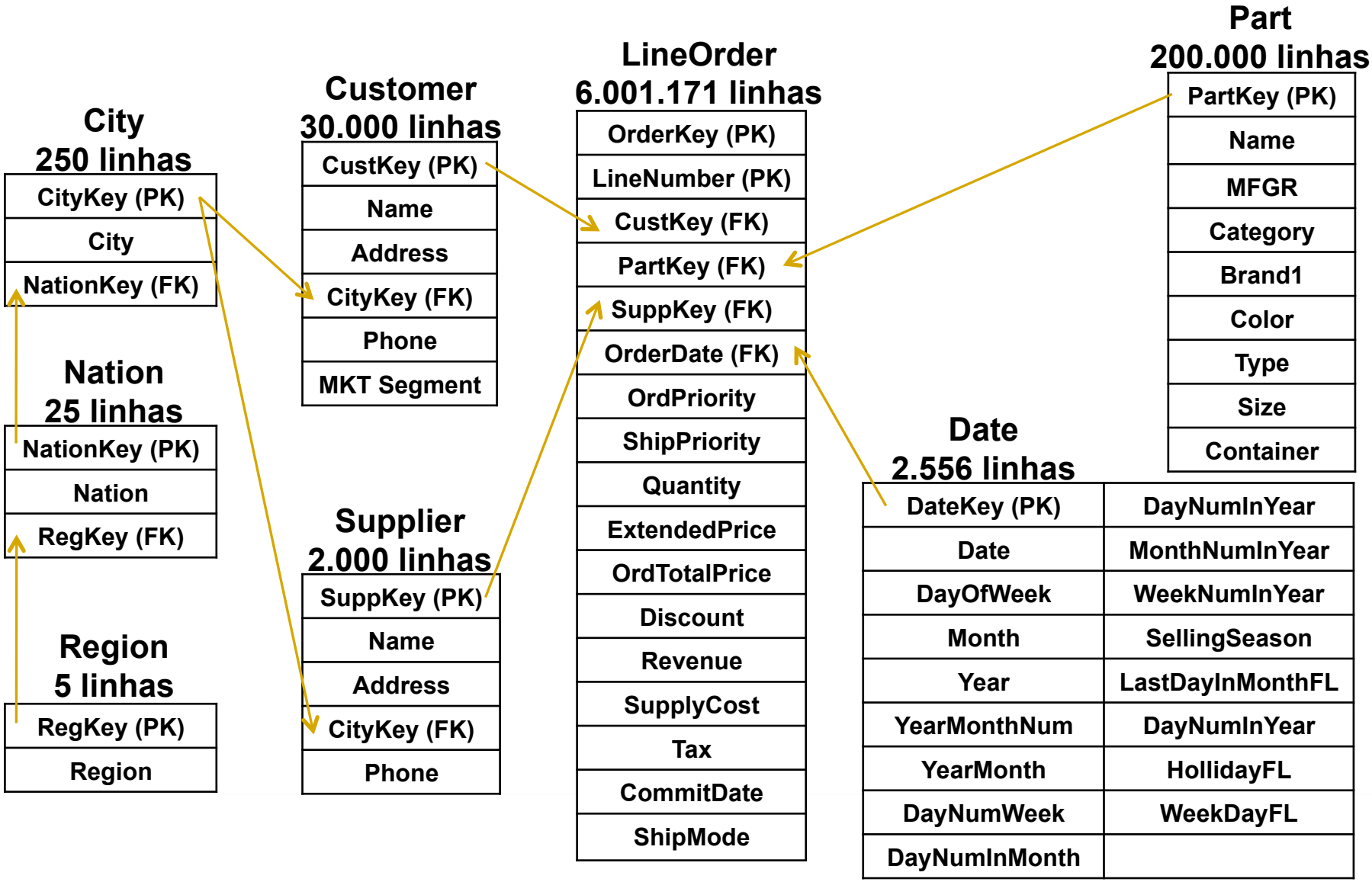
Customer 30.000 linhas
CustKey (PK)
Name
Address
City
Nation
Region
Phone
MKT Segment

Supplier 2.000 linhas
SuppKey (PK)
Name
Address
City
Nation
Region
Phone

Data Warehouse Normalizado

- Minimização do espaço de armazenamento
- Espera-se que consultas levem mais tempo para serem executadas
  - Operação de junção entre as tabelas
- Criação das tabelas *City*, *Nation* e *Region*

Data Warehouse Normalizado



Visões Materializadas

- Consultas pré-computadas e armazenadas no banco de dados
  - Aumento no desempenho do processamento de consultas
  - Espaço de armazenamento adicional
  - Custo para manter a consistência das visões
- Visões materializadas em DW
  - Agregação das medidas da tabela de fatos
  - GROUP BY nas dimensões

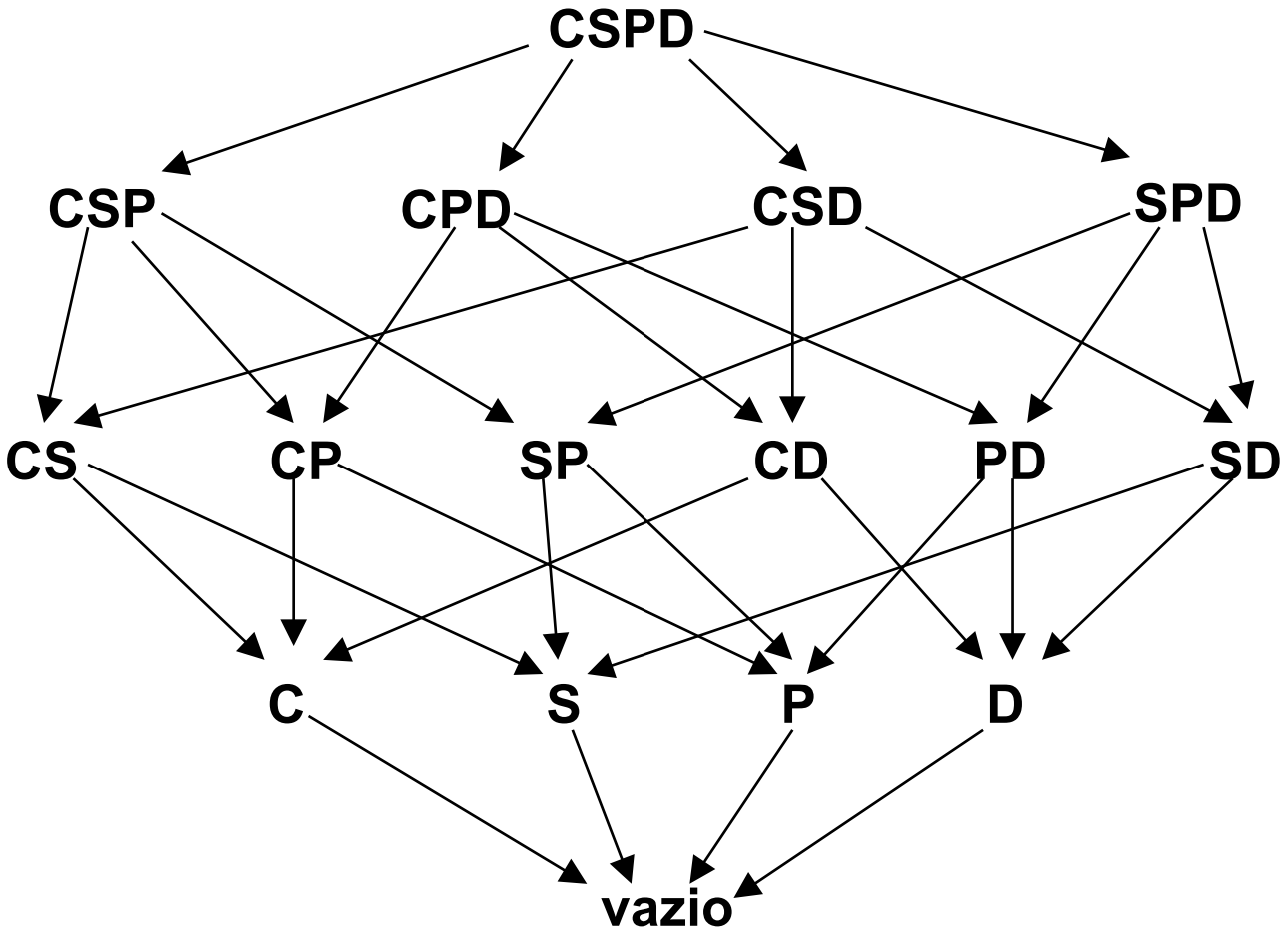
Visões Materializadas

- DW baseado no SSB (normalizado ou não)
  - GROUP BY nas dimensões:
    - *Customer*, *Supplier*, *Date* e *Part*
  - Agregação das medidas
    - SUM (*Quantity*)
    - SUM (*ExtendedPrice*)
    - SUM (*Revenue*)
    - SUM (*SupplyCost*)
    - Profit = SUM (*Revenue* – *SupplyCost*)

LineOrder

OrderKey (PK)
LineNumber (PK)
CustKey (FK)
PartKey (FK)
SuppKey (FK)
OrderDate (FK)
OrdPriority
ShipPriority
Quantity
ExtendedPrice
OrdTotalPrice
Discount
Revenue
SupplyCost
Tax
CommitDate
ShipMode

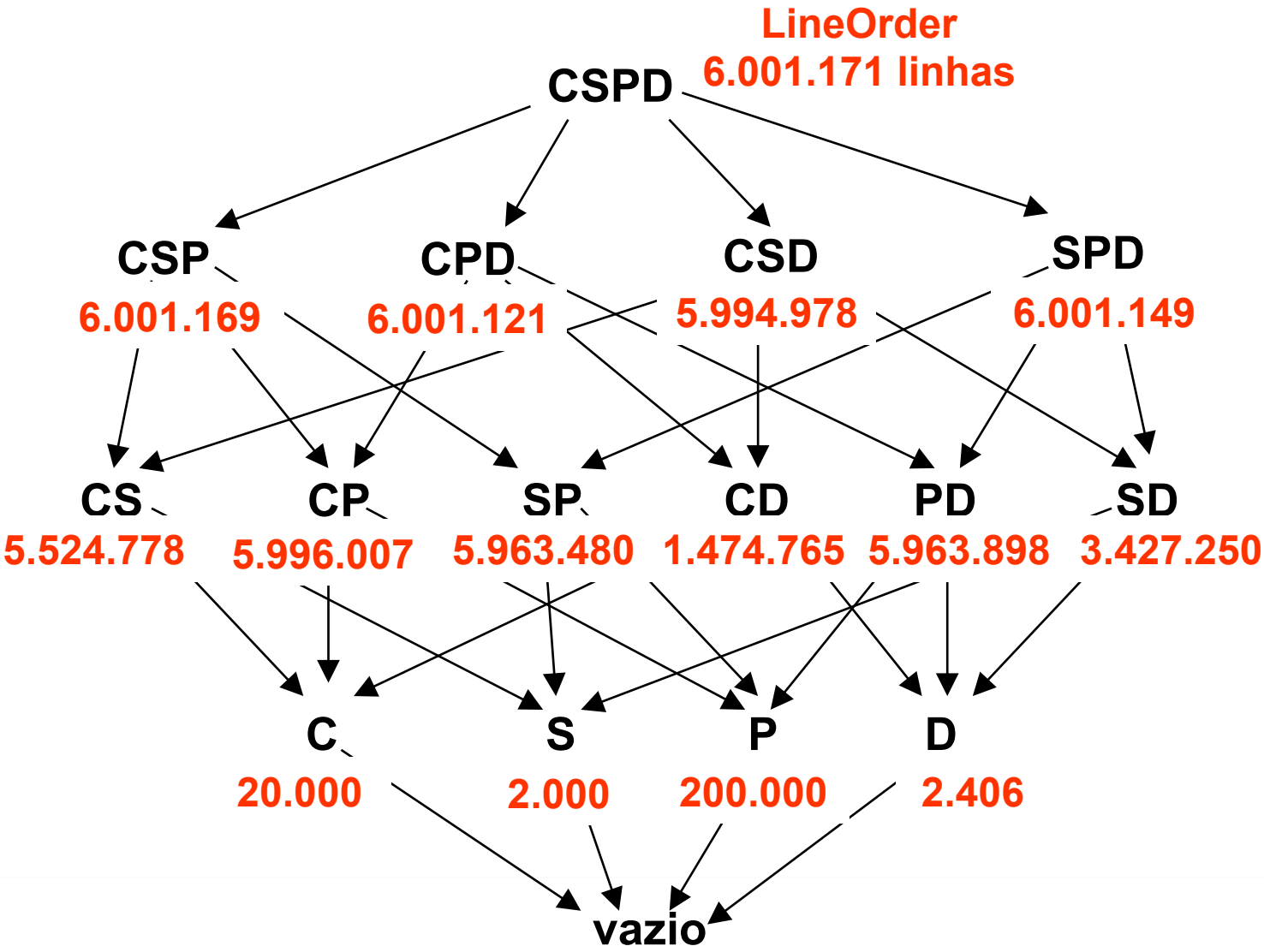
Grafo de Derivação



Dimensões

- *Customer* (C)
- *Supplier* (S)
- *Part* (P)
- *Date* (D)

Grafo de Derivação



Visões Materializadas - Exercício

- Criar uma visão materializada agrupando as medidas abaixo por *Customer*
  - *Quantity*
  - *ExtendedPrice*
  - *Revenue*
  - *SupplyCost*
  - *Profit = Revenue – SupplyCost*

Visões Materializadas - Exercício

- Criar uma visão materializada agrupando as medidas abaixo por *Customer* e *Supplier*
  - *Quantity*
  - *ExtendedPrice*
  - *Revenue*
  - *SupplyCost*
  - *Profit = Revenue – SupplyCost*

Visões Materializadas - Exercício

- Criar uma visão materializada agrupando as medidas abaixo por *Customer*, *Supplier* e *Part*
  - *Quantity*
  - *ExtendedPrice*
  - *Revenue*
  - *SupplyCost*
  - *Profit = Revenue – SupplyCost*

Consulta 1

- Objetivo:
  - Comparar o desempenho de consultas no DW normalizado, DW não normalizado e suas visões materializadas
- Estratégia:
  - Explorar a complexidade de consultas em bases normalizadas

Consulta 1

- Retornar
  - Nação de *Customer*
  - Nação de *Supplier*
  - Revenue
- Filtro: região de *Customer* = região de *Supplier*
- Agrupar por nação de *Customer* e nação de *Supplier*
- Ordenar por
  - Nação de *Customer* (ascendente)
  - Revenue (descendente)

Consulta 1

- Filtro aplicado
  - Região de *Customer* = Região de *Supplier*

Tabela de Fatos Utilizada	Não Normalizado	Normalizado
Customer_Supplier_V	**6,39 s**	7,74 s
Lineorder	18,58 s	18,99 s

Consulta 2

- Objetivo:
  - Comparar o desempenho de consultas no DW normalizado, DW não normalizado e suas visões materializadas
- Estratégia:
  - Explorar a complexidade de consultas em bases normalizadas e a agregação existente nas tabelas

Consulta 2

- Retornar
  - Região de *Customer*
  - Região de *Supplier*
  - Brand1 de *Part*
  - Revenue
  - Profit = Revenue – SupplyCost
- Filtro: sem filtro
- Agrupar por região de *Customer*, região de *Supplier* e brand1 de *Part*
- Ordenar por Brand1 (asc), Revenue (desc) e Profit (desc)

Consulta 2

- Filtro aplicado
  - Sem filtro

Tabela de Fatos Utilizada	Não Normalizado	Normalizado
Customer_Supplier_Part_V	**19,15 s**	25,04 s
Lineorder	25,37 s	35,40 s

Consulta 3

- Objetivo:
  - Comparar o desempenho no DW não normalizado com uma visão materializada
- Estratégia:
  - Explorar o baixo volume de dados da visão em relação à tabela de fatos

Consulta 3

- Retornar a chave de *Customer* e as medidas
  - Quantity
  - ExtendedPrice
  - Revenue
  - SupplyCost
  - Profit = Revenue – SupplyCost
- Filtro: região de *Customer* = 'AMERICA'
- Agrupar por chave de *Customer*
- Ordenar por chave de *Customer*

Consulta 3

- Filtro utilizado
  - Região de *Customer* = 'AMERICA'

Tabela de Fatos Utilizada	Tempo
Customer_V	**0,32 s**
Customer_Supplier_V	7,49 s
Customer_Supplier_Part_V	7,81 s
Lineorder	14,87 s

Testes realizados

	Dimensões Utilizadas			Tabela de Fatos Utilizada	Tempos Normalização		Filtros Aplicados
	C	S	D		Não	Sim	
1	X	X	X	Lineorder	7,13	10,56	Região de Customer = Região de Supplier; 1992 >= Ano <=
2	X	X	X	Customer_Supplier_Date_V	4,96	6,34	Região de Customer = Região de Supplier; 1992 >= Ano <=
3	X	X		Lineorder	8,63	9,88	Região de Customer = Região de Supplier
4	X	X		Customer_Supplier_V	4,67	6,41	Região de Customer = Região de Supplier
5	X		X	Lineorder	7,63	8,96	Região de Customer = 'AMERICA' e 1992 >= Ano <= 199
6	X		X	Customer_Date_V	1,20	1,00	Região de Customer = 'AMERICA' e 1992 >= Ano <= 199
7		X	X	Lineorder	6,00	6,01	Região de Supplier = 'AMERICA' e 1992 >= Ano <= 1997
8		X	X	Supplier_Date_V	2,88	3,03	Região de Supplier = 'AMERICA' e 1992 >= Ano <= 1997
9	X			Lineorder	5,80	6,40	Região de Customer = 'AMERICA'
10	X			Customer_V	0,24	0,04	Região de Customer = 'AMERICA'
11		X		Lineorder	6,91	7,38	Região de Supplier = 'AMERICA'
12		X		Supplier_V	0,04	0,05	Região de Supplier = 'AMERICA'

Testes realizados

	Dimensões Utilizadas			Tabela de Fatos Utilizada	Tempos		Filtros Aplicados
					Normalização		
	C	S	P		Não	Sim	
1	X	X	X	Lineorder	15,01	17,05	group by em Região e part.brand1
2	X	X	X	Customer_Supplier_Part_V	11,97	15,25	group by em Região e part.brand1
3	X	X		Lineorder	10,06	13,03	group by em Região
4	X	X		Customer_Supplier_V	6,74	9,25	group by em Região
5	X		X	Lineorder	10,00	14,33	group by em Região e part.brand1
6	X		X	Customer_Part_V	7,61	9,32	group by em Região e part.brand1
7		X	X	Lineorder	9,94	12,56	group by em Região e part.brand1
8		X	X	Supplier_Part_V	7,27	8,55	group by em Região e part.brand1
9	X			Lineorder	7,79	8,25	group by em Região
10	X			Customer_V	0,27	0,26	group by em Região
11		X		Lineorder	7,91	8,00	group by em Região
12		X		Supplier_V	0,25	0,36	group by em Região

Conclusão

- A normalização deteriora o desempenho de consultas, o que pode ser compensado por meio da utilização de Visões Materializadas
- A utilização de Visões Materializadas, quando escolhidas corretamente, aumenta o desempenho de consultas

Referências

- [HARINARAYAN, 1997] HARINARAYAN, V.; RAJARAMAN, A.; ULLMAN, J. D. Implementing Data Cubes Efficiently. **ACM SIGMOD Record**, v. 25, n. 2, p. 205-216, 1996.

- [O'NEIL, 2009] O'NEIL, P.; O'NEIL, E.; CHEN, X. **The Star Schema Benchmark**. Preprint, Revision 3, 2009. Disponível em <http://www.cs.umb.edu/~poneil/StarSchemaB.pdf>. Acesso em: 13 set. 2010.