



Formação Engenheiro de Dados

Estudo de Caso Parte I

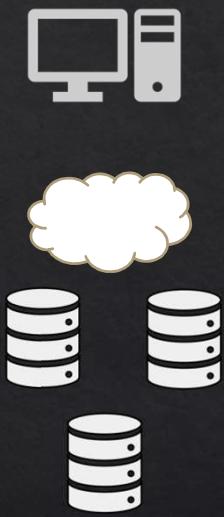
- ❖ Objetivo: construir um armazém de dados para que o administrador da JJBike possa compreender seu negócio

JJBike

FRAME

Road Bike
@rk~

Arquitetura Tradicional



Operação

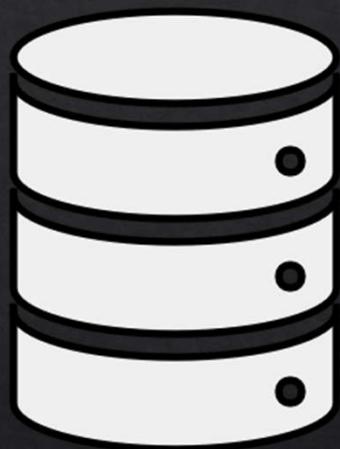


Staging



Armazém de
Dados

Nosso Caso Prático – Separação por Schemas



Relacional

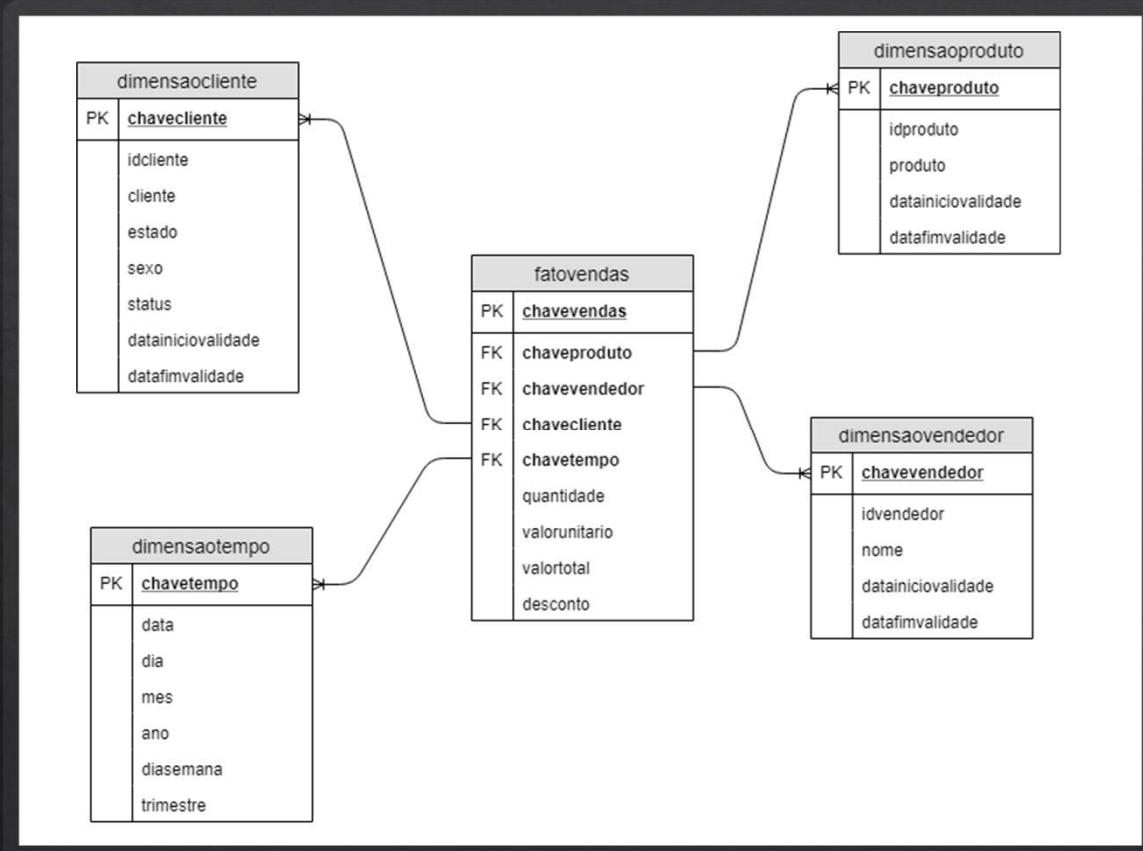
Dimensional

Ferramentas

- PostgreSQL 11 – Já instalado e com o banco transacional pronto!



Modelo Dimensional



Baixar do Ambiente do Curso

Dimensional.zip

Local padrão:
/home/cloudera/Downloads

Scripts

- ❖ 8.CreateTable.sql
- ❖ 9.InsertDimensaoTempo.sql
- ❖ 10.Interativo.sql
- ❖ 11.Desnormalização.sql
- ❖ 12.kpi.sql

8.CreateTable.sql

```
❖ CREATE SCHEMA Dimensional;  
  
❖ CREATE SEQUENCE Dimensional.ChaveVendedor;  
❖ CREATE TABLE Dimensional.DimensaoVendedor(  
❖     ChaveVendedor int default  
nextval('Dimensional.ChaveVendedor'::regclass) PRIMARY KEY,  
❖     IDVendedor int,  
❖     Nome Varchar(50),  
❖     DataInicioValidade date not null,  
❖     DataFimValidade date  
❖ );
```

9.InsertDimensaoTempo.sql

```
INSERT INTO Dimensional.DimensaoTempo(data, dia,
mes, ano, diasemana, trimestre)
SELECT datum AS Data,
       EXTRACT(DAY FROM datum) AS Dia,
       EXTRACT(MONTH FROM datum) AS Mes,
       EXTRACT(year FROM datum) AS Ano,
       EXTRACT(dow FROM datum) AS DiaSemana,
       EXTRACT(quarter FROM datum) AS Trimestre
FROM (SELECT '1970-01-01'::DATE+ SEQUENCE.DAY AS
      datum
      FROM GENERATE_SERIES (0,29219) AS SEQUENCE
(DAY)
      GROUP BY SEQUENCE.DAY) DQ
ORDER BY 1
```

chavetempo	data	dia	mes	ano	diasemana	trimestre
1	01-01-1970	1	1	1970	4	1
2	02-01-1970	2	1	1970	5	1
3	03-01-1970	3	1	1970	6	1
4	04-01-1970	4	1	1970	0	1
5	05-01-1970	5	1	1970	1	1
6	06-01-1970	6	1	1970	2	1
7	07-01-1970	7	1	1970	3	1
8	08-01-1970	8	1	1970	4	1
9	09-01-1970	9	1	1970	5	1
10	10-01-1970	10	1	1970	6	1
11	11-01-1970	11	1	1970	0	1
12	12-01-1970	12	1	1970	1	1
13	13-01-1970	13	1	1970	2	1
14	14-01-1970	14	1	1970	3	1
15	15-01-1970	15	1	1970	4	1
16	16-01-1970	16	1	1970	5	1
17	17-01-1970	17	1	1970	6	1
18	18-01-1970	18	1	1970	0	1
19	19-01-1970	19	1	1970	1	1
20	20-01-1970	20	1	1970	2	1
21	21-01-1970	21	1	1970	3	1
22	22-01-1970	22	1	1970	4	1
23	23-01-1970	23	1	1970	5	1

10.Interativo.sql

```
--CARREGAR OS DADOS DE CLIENTES (1ª CARGA)
WITH S AS (
    Select * From Relacional.Clientes
),
UPD AS (
    UPDATE Dimensional.DimensaoCliente T
    SET DataFimValidade = current_date
    FROM S
    WHERE (T.IDCliente = S.IDCliente AND T.DataFimValidade is null)
        AND (T.CLIENTE <> S.CLIENTE OR T.ESTADO <> S.ESTADO OR T.SEXO <> S.SEXO
    OR T.STATUS <> S.STATUS)
    RETURNING T.IDCLIENTE
)
INSERT INTO Dimensional.DimensaoCliente(IDCliente, Cliente, Estado, Sexo, Status,
DataInicioValidade, DataFimValidade)
SELECT IDCliente, Cliente, Estado, Sexo, Status, current_date, null FROM S
WHERE S.IDCliente IN (SELECT IDCliente FROM UPD) OR
S.IDCliente NOT IN (SELECT IDCliente FROM Dimensional.DimensaoCliente)
```

chavecliente	idcliente	cliente	estado	sexo	status	datainiciovalidade	datafimvalidade
1	1	Adelina Buenaventura	RJ	M	Silver	17-06-2019	
2	2	Adelino Gago	RJ	M	Silver	17-06-2019	
3	3	Adelio Lisboa	SE	M	Silver	17-06-2019	
4	4	Aderito Bahia	MA	M	Silver	17-06-2019	
5	5	Adelito Patrício	PE	M	Silver	17-06-2019	
6	6	Adelmo Medelha	RO	F	Platinum	17-06-2019	
7	7	Aida Dornelles	RN	F	Silver	17-06-2019	
8	8	Alarico Quinterno	AC	M	Silver	17-06-2019	
9	9	Alberto Cezimbra	AM	M	Silver	17-06-2019	
10	10	Alberto Monsanto	RN	M	Gold	17-06-2019	
11	11	Albino Canela	AC	M	Silver	17-06-2019	
12	12	Alcides Carvalho	PR	M	Silver	17-06-2019	
13	13	Alcides Carvalhais	RO	M	Silver	17-06-2019	
14	14	Aldo Martins	GO	M	Silver	17-06-2019	
15	15	Alexandra Tabares	MG	F	Silver	17-06-2019	
16	16	Alfredo Cotrim	SC	M	Silver	17-06-2019	
17	17	Almeno Figueira	SC	M	Silver	17-06-2019	
18	18	Alvito Peralta	AM	M	Silver	17-06-2019	
19	19	Amélia Martinho	RO	M	Silver	17-06-2019	
20	20	Amélia Estevez	PE	F	Silver	17-06-2019	
21	21	Ana Homem	RN	F	Silver	17-06-2019	

11.Desnormalização.sql

```
select
    dimensional.dimensaocliente.cliente,
    dimensional.dimensaocliente.estado,
    dimensional.dimensaocliente.sexo,
    dimensional.dimensaocliente.status,
    dimensional.fatovendas.quantidade,
    .....
    dimensional.dimensaoendereco.nome

    INTO dimensional.des_vendas

from (((dimensional.dimensaocliente
inner join dimensional.fatovendas
on (dimensional.fatovendas.chavecliente = dimensional.dimensaocliente.chavecliente))
.....
```

12.kpi.sql

```
select dimensional.dimensaotempo.mes as Mes,  
sum( dimensional.fatovendas.valortotal) as Realizado  
  
INTO dimensional.KPI  
from ( dimensional.fatovendas      inner join  
dimensional.dimensaotempo          on  
(dimensional.dimensaotempo.chavetempo =  
dimensional.fatovendas.chavetempo))  
group by dimensional.dimensaotempo.mes  
order by dimensional.dimensaotempo.mes
```