

FIAP

NBBA



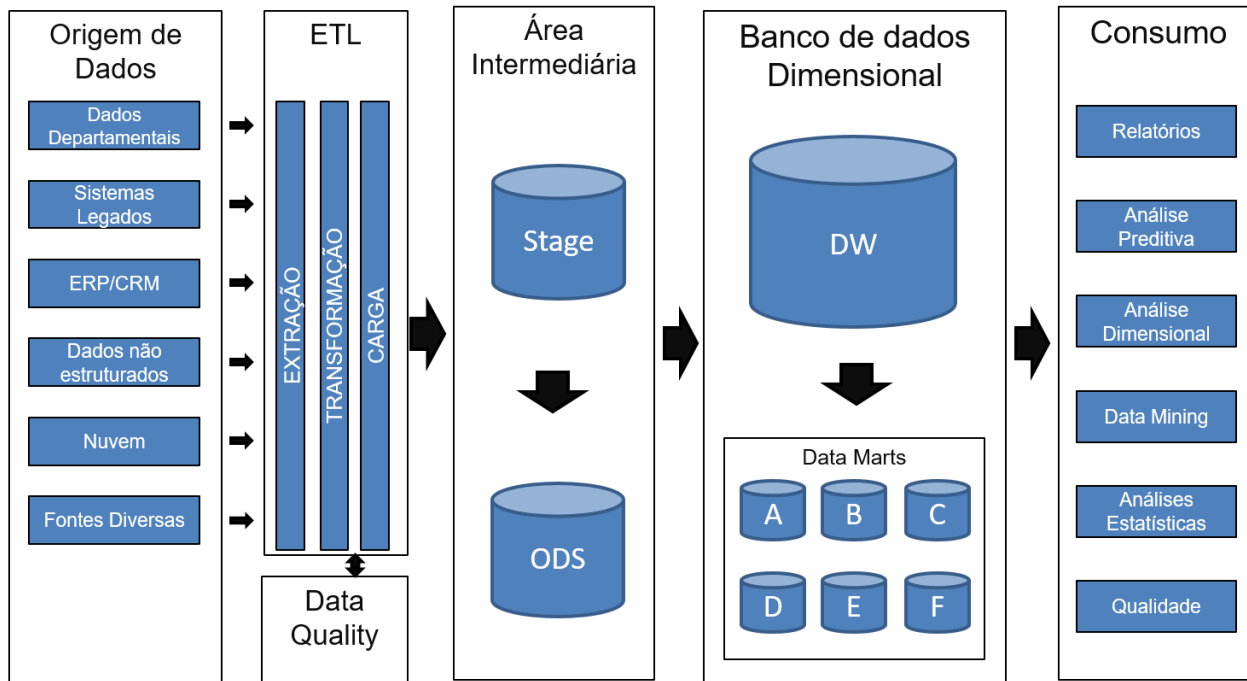
Arquitetura de DataLake Streaming de Dados



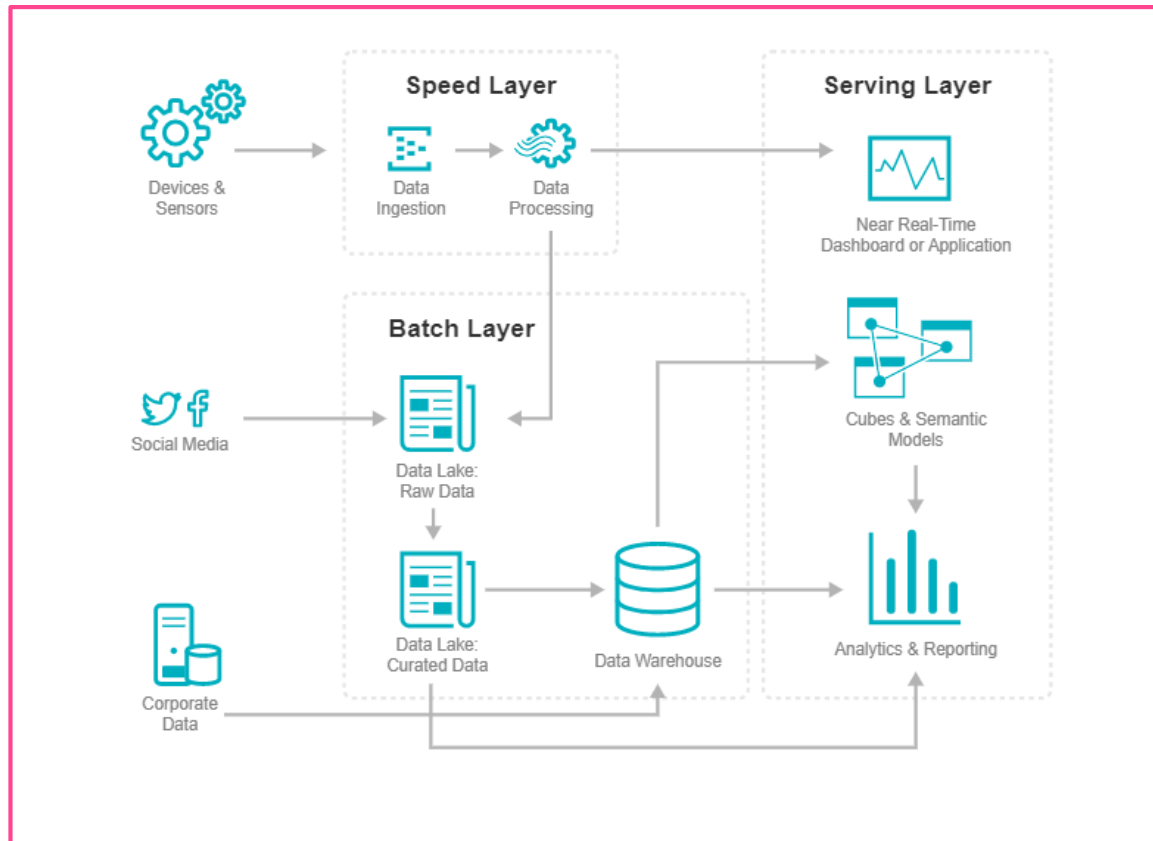
AULA 2 - OBJETIVOS

- Compreender a evolução arquitetural;
- Compreender a evolução da modelagem de dados;
- Comparar os cenários: OLTP, OLAP e OLAP com Big Data;
- Avaliar o processo de consulta a dados em HDFS;
- Desenhar os modelos de dados do Case de Trade.

ARQUITETURA TRADICIONAL – BI

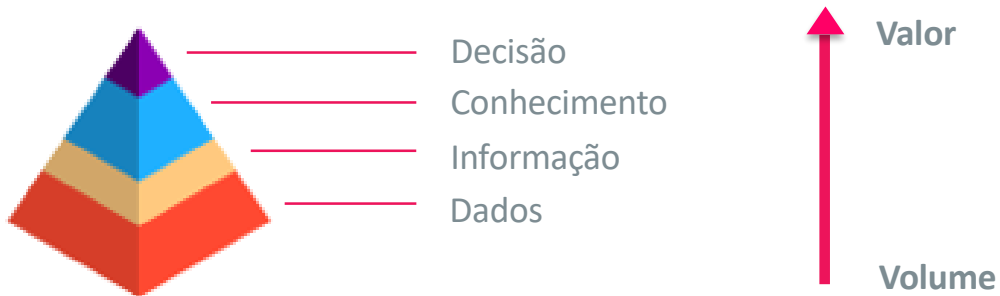


ARQUITETURA MODERNA (LAMBDA) – BIG DATA, BI & BA

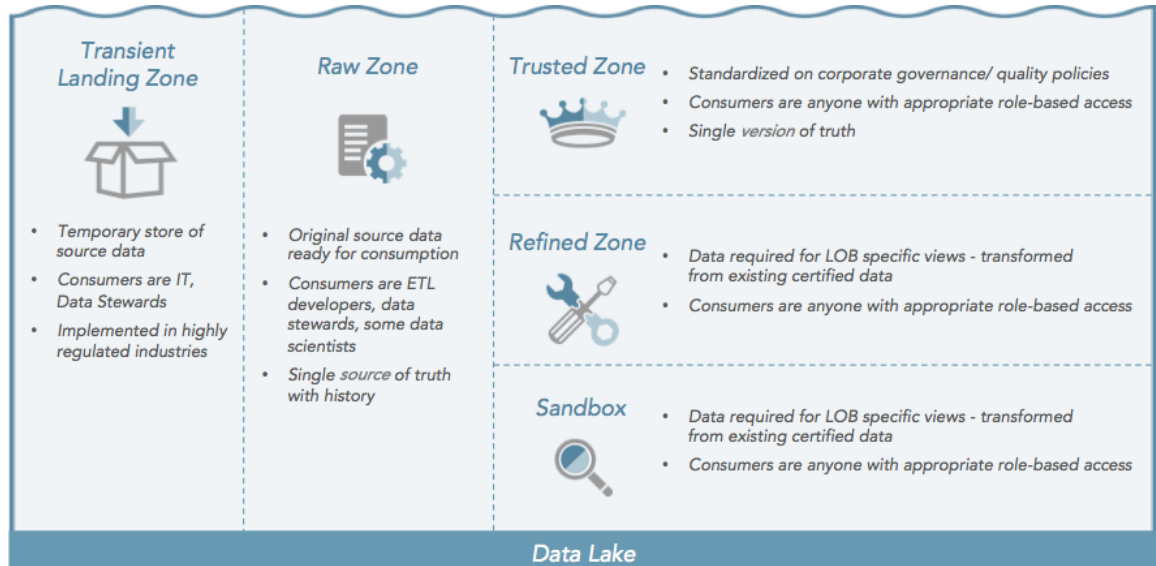
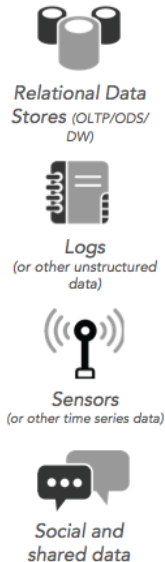


BI – PROPOSTA DO DATA WAREHOUSE

Seu objetivo é a conversão do **volume** de dados analisados em **valor** comercial por meio de **relatórios analíticos**.



BIG DATA – PROPOSTA DO DATA LAKE



DATA LAKE X DATA WAREHOUSE



Vantagens:

- ✓ Agilidade
- ✓ Flexibilidade
- ✓ Entrega rápida
- ✓ Exploração



Vantagens:

- ✓ Governança
- ✓ Padrões
- ✓ Segurança
- ✓ Confiabilidade



Esforço Menor

Aquisição dos dados



Esforço Maior



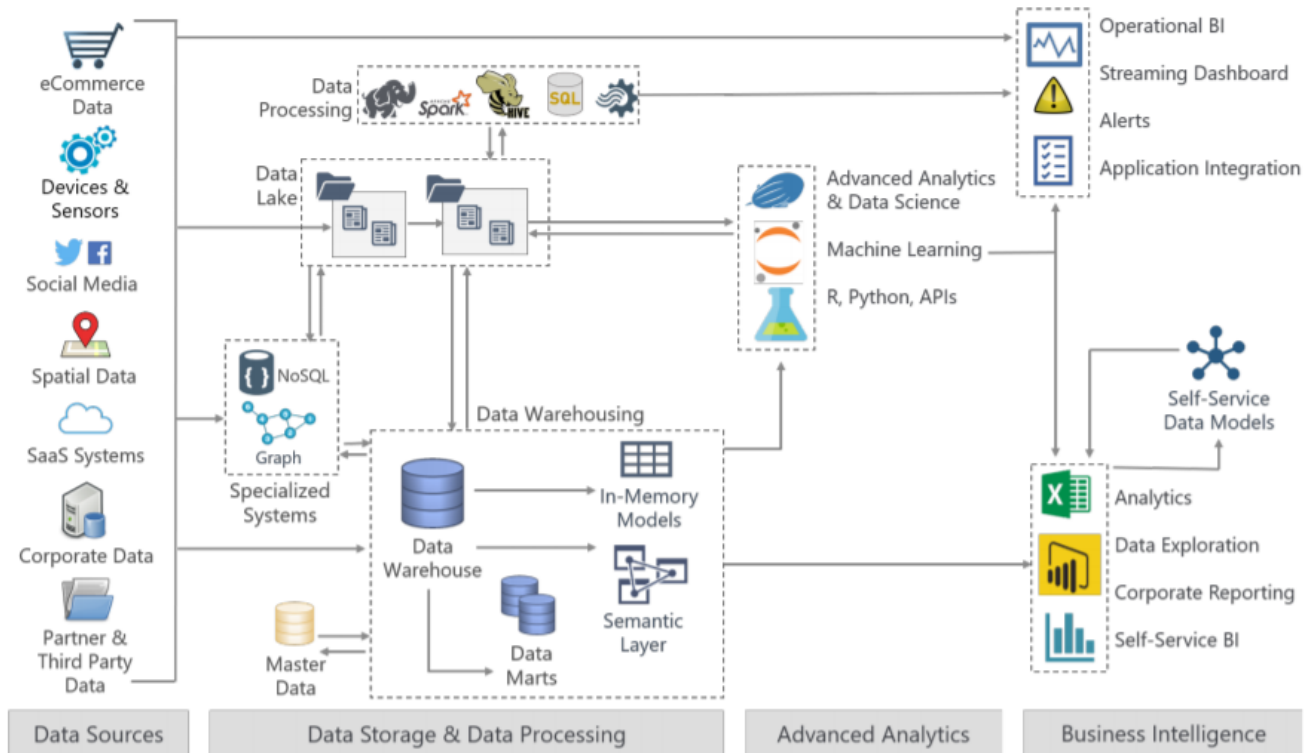
Esforço Maior

Recuperação dos dados

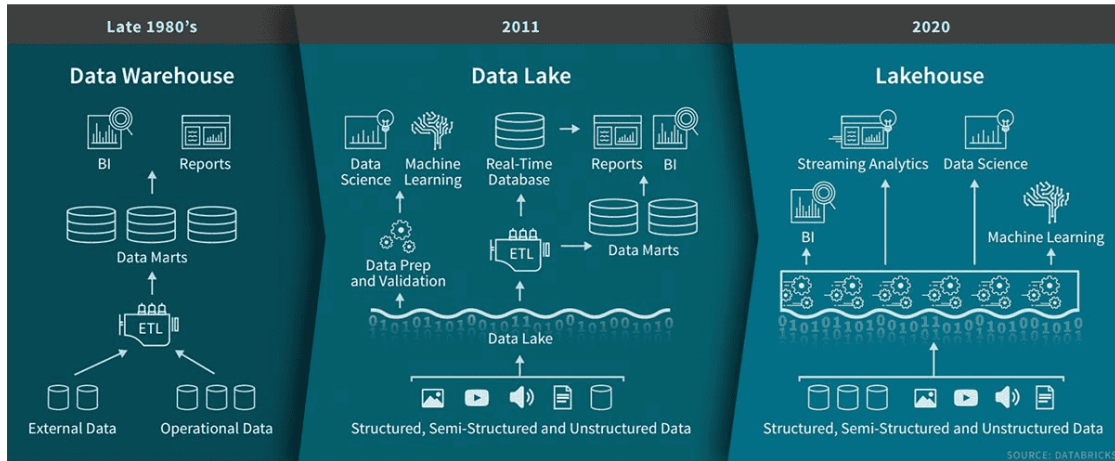


Esforço Menor

ARQUITETURA LAMBDA EXPANDIDA

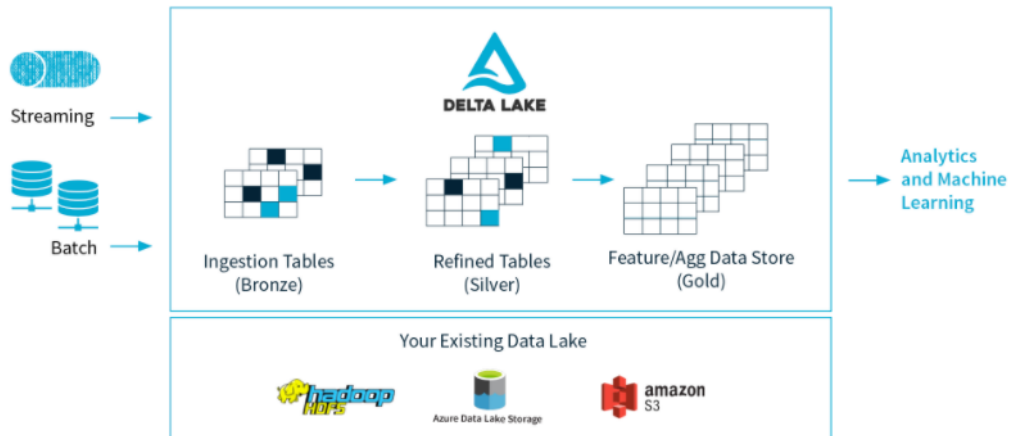


DATA WAREHOUSE X DATA LAKE X DATA LAKEHOUSE

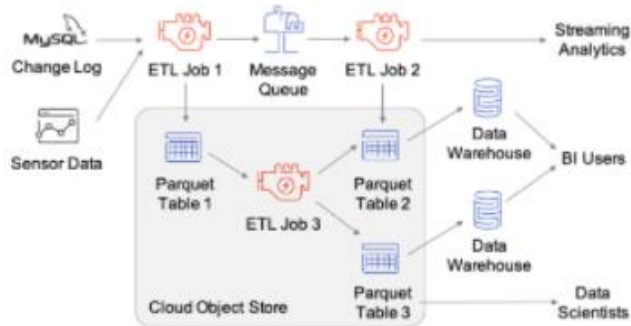


DELTA LAKE (DATA LAKEHOUSE)

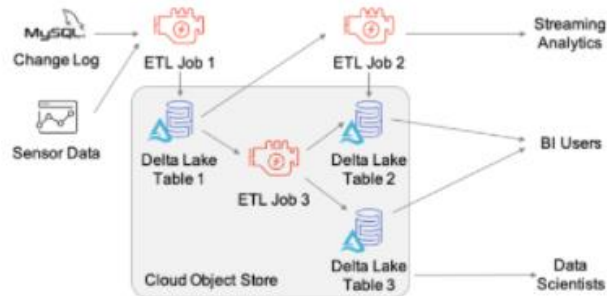
Delta Lake is an open-source storage layer that brings ACID transactions to Apache Spark™ and big data workloads.



ARQUITETURA LAMBDA COM DELTA LAKE



(a) Pipeline using separate storage systems.



(b) Using Delta Lake for both stream and table storage.

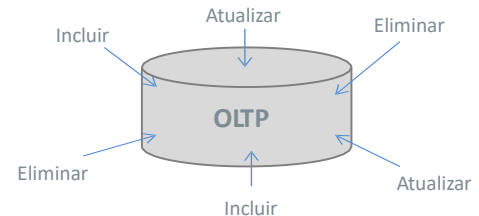
OLTP x OLAP x OLAP com Big Data



AMBIENTES TRANSACIONAIS

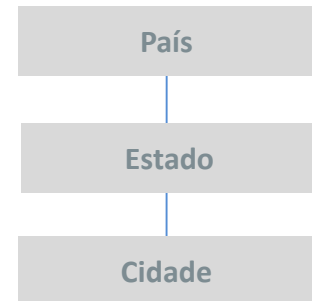
OLTP (*On-Line Transaction Processing*)

- Sistemas que tratam o negócio: CRMs, ERPs, entre outros.
- Ênfase nas operações diárias.
- Bases orientadas à escrita.



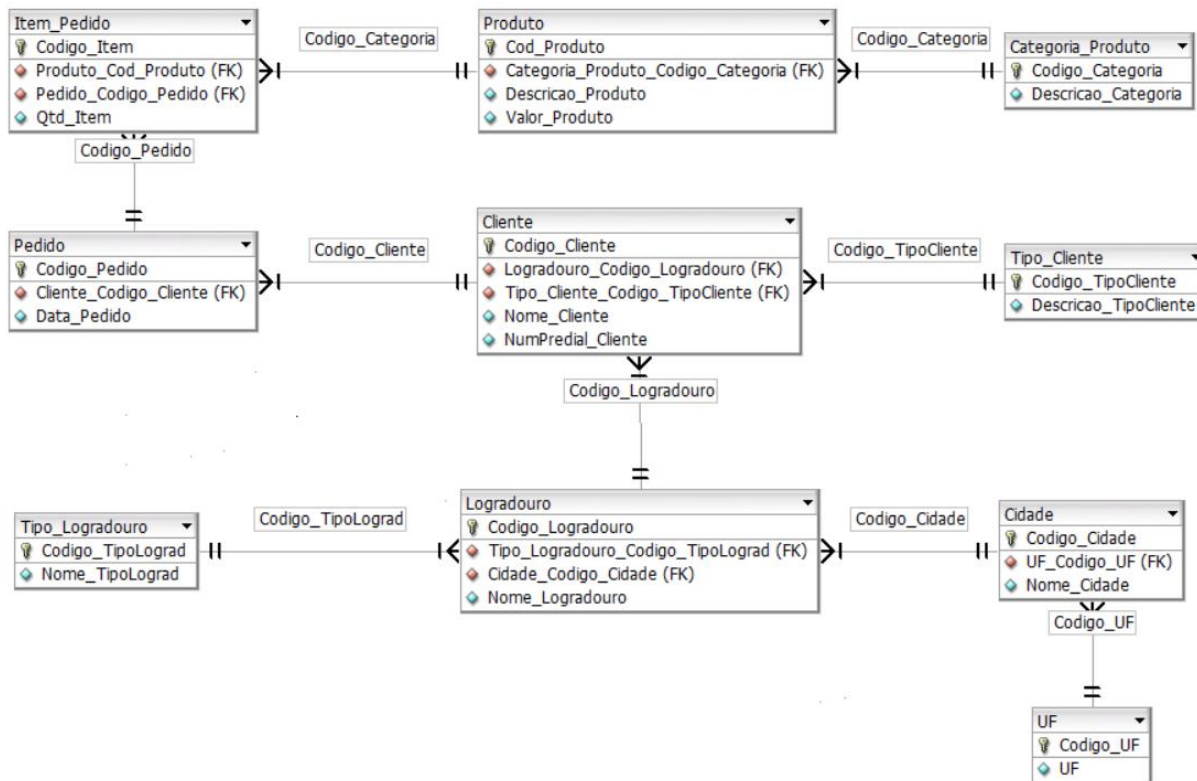
Modelagem de Dados – OLTP

- Sem repetições ou redundâncias.
- Terceira forma normal.
- Chaves de Negócio.
- Tabelas de Lookup.



AMBIENTES TRANSACIONAIS

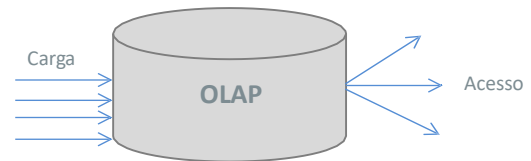
Modelagem de Dados OLTP



AMBIENTES ANALÍTICOS

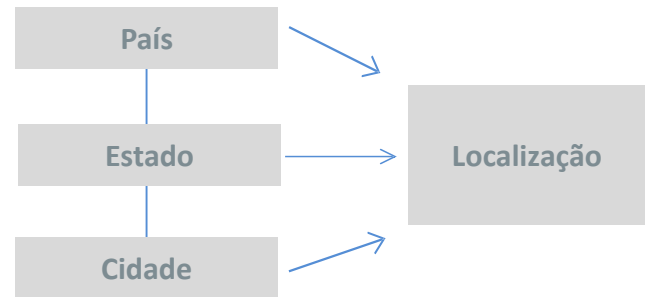
OLAP (On-Line Analytical Processing)

- Sistemas que analisam o negócio: relatórios, painéis, indicadores de desempenho.
- Dados Históricos.
- Bases orientadas à consulta.



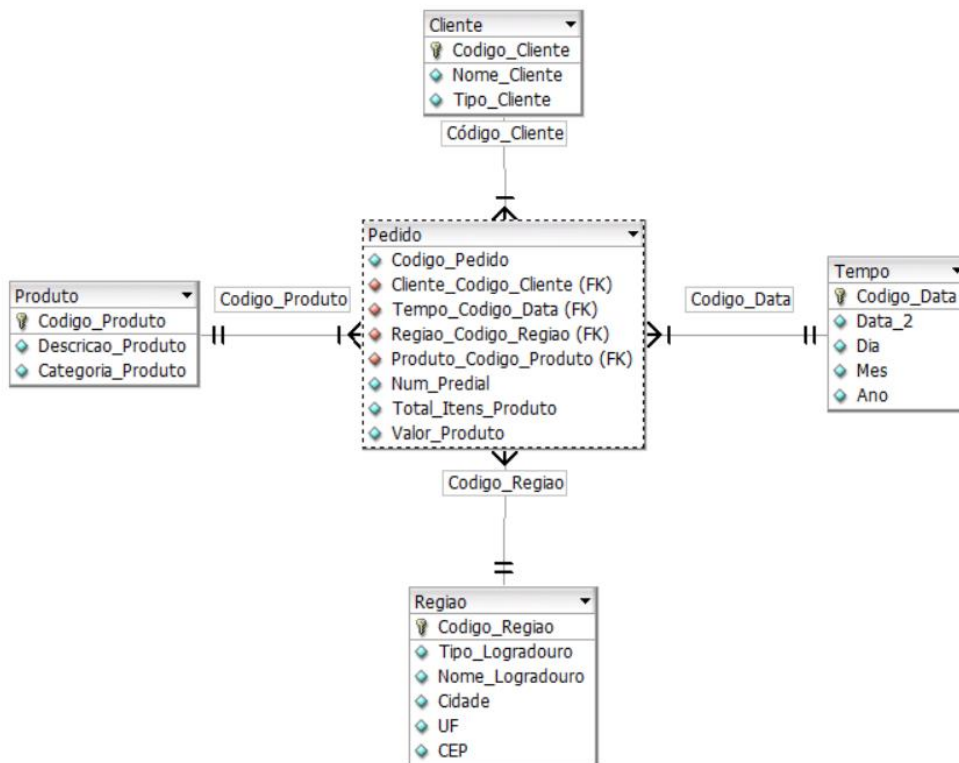
Modelagem Dimensional

- Permite redundâncias.
- Desnormalização.
- Chaves Substitutas.
- Modelo orientado aos processos de negócio.



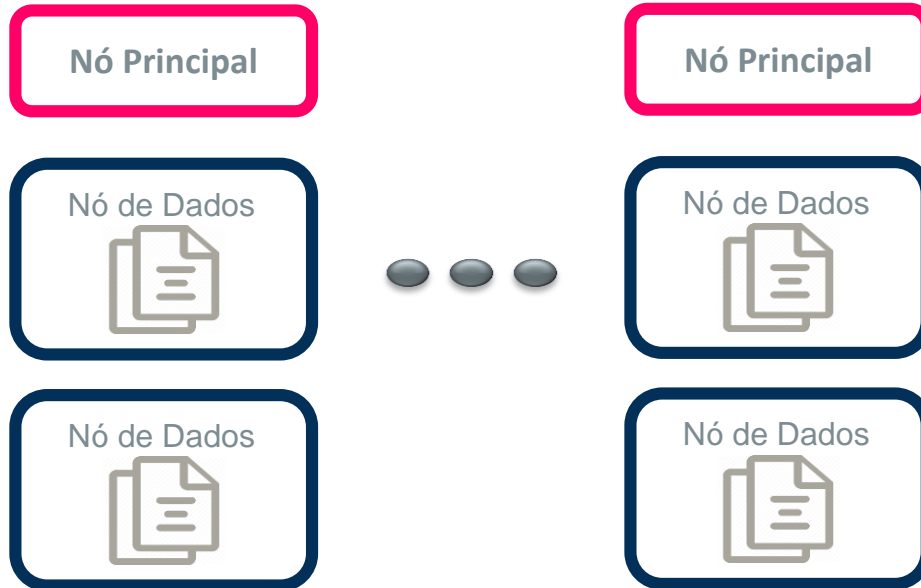
AMBIENTES ANALÍTICOS

Modelagem Multidimensional



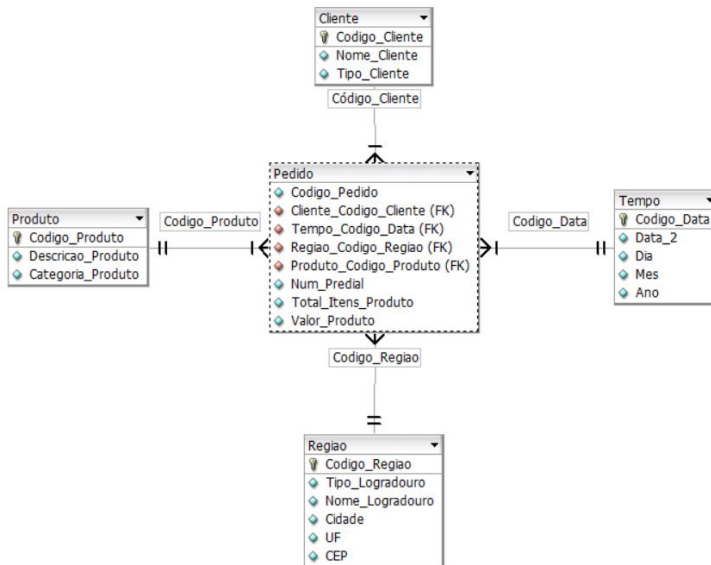
AMBIENTES ANALÍTICOS COM GRANDE VOLUME DE DADOS

HDFS - Hadoop Distributed File System



AMBIENTES ANALÍTICOS COM GRANDE VOLUME DE DADOS

Modelagem Multidimensional



Modelagem Flatten



AMBIENTES ANALÍTICOS COM GRANDE VOLUME DE DADOS

Modelo
Normalizado

Modelo
Dimensional
Star

Modelo
Flatten Star

Bom

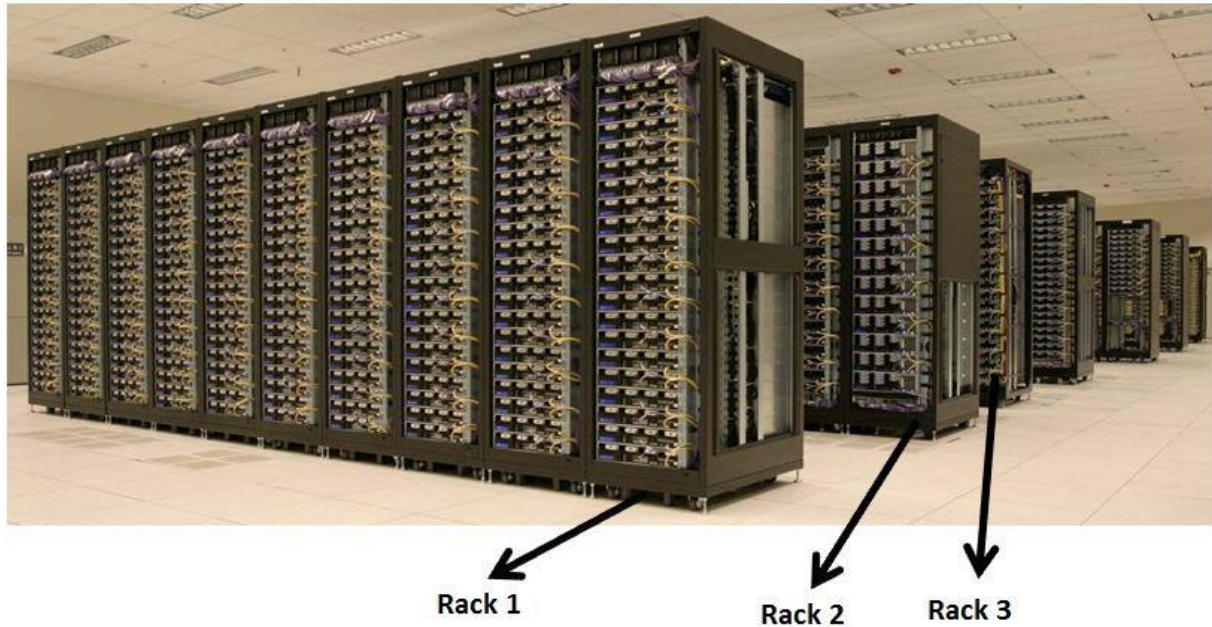
Mal

Feio

???????

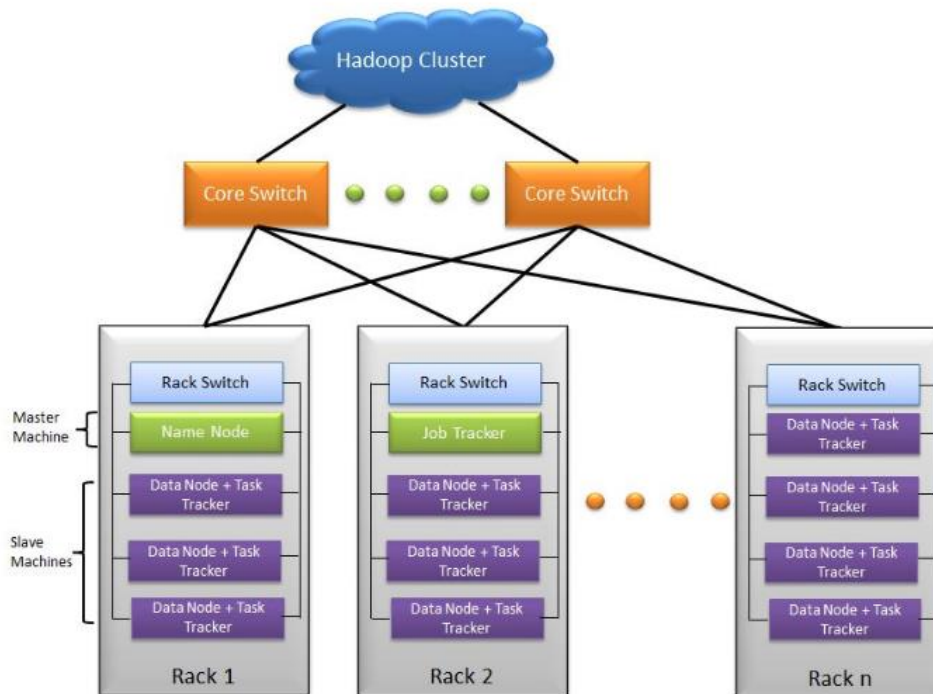
ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop



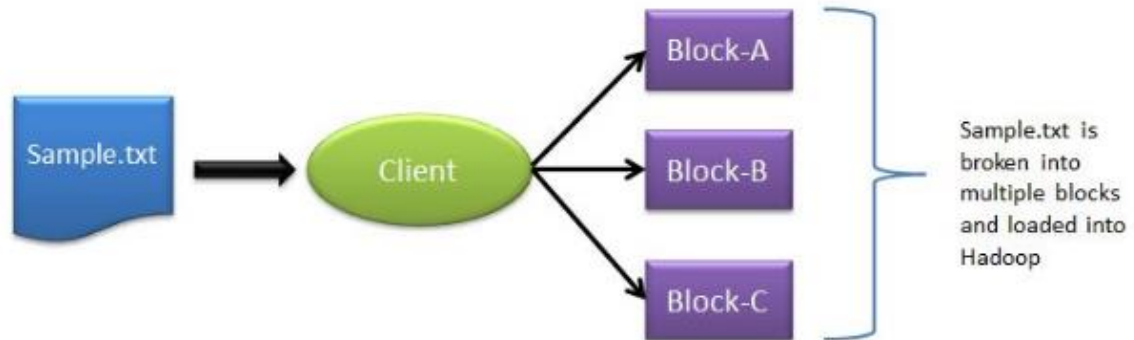
ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop - Nós



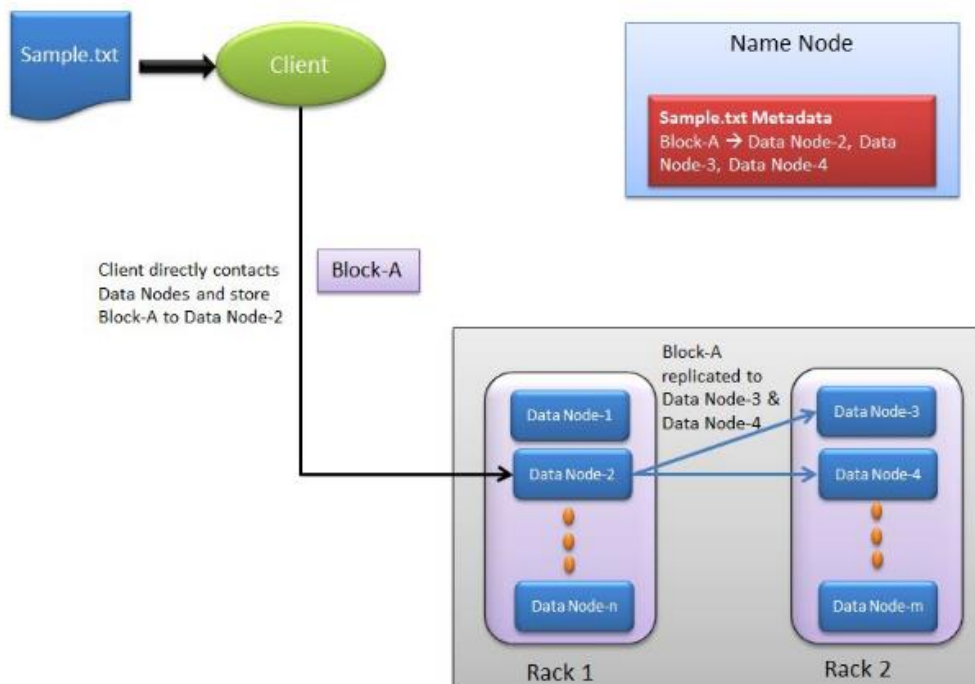
ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop – Armazenamento em Blocos



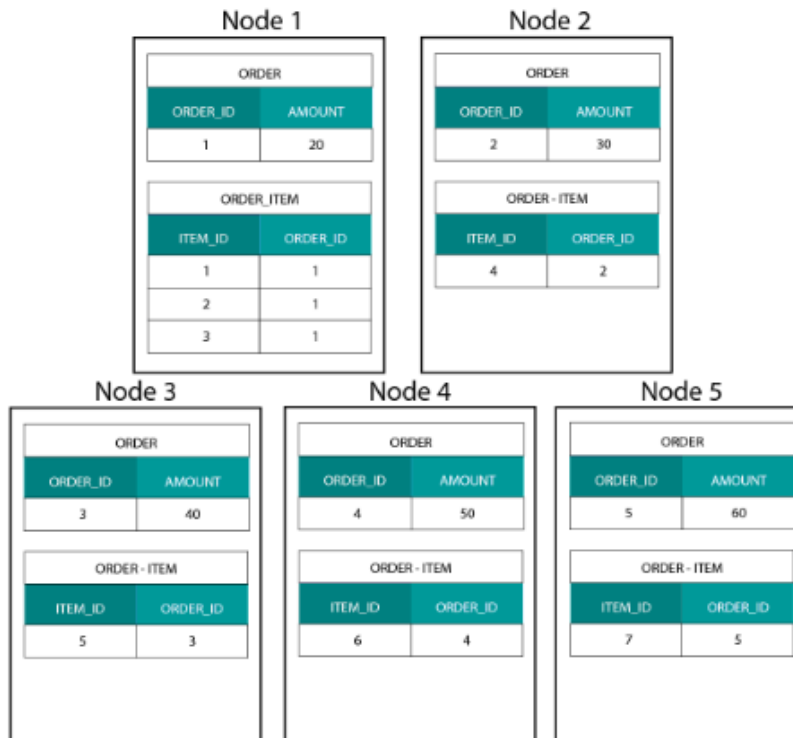
ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop – Redundância



ENTENDENDO A ARQUITETURA HADOOP

Bancos Relacionais MPP – Organização nos Blocos



ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop – Organização nos Blocos

Node 1

ORDER	
ORDER_ID	AMOUNT
1	20
2	30

Node 2

ORDER		ORDER_ITEM	
ORDER_ID	AMOUNT	ITEM_ID	ORDER_ID
5	60	4	2

Node 3

ORDER		ORDER_ITEM	
ORDER_ID	AMOUNT	ITEM_ID	ORDER_ID
3	40	5	3
4	50	6	4

Node 4

ORDER_ITEM	
ITEM_ID	ORDER_ID
7	5

Node 5

ORDER_ITEM	
ITEM_ID	ORDER_ID
1	1
2	1
3	1

ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop – Estrutura Flatten

```
SELECT
  order_id,
  date,
  flatten(kvgen(order_item))
FROM
  parquet.file;
```

ORDER_ID	DATE	ITEM_ID	AMOUNT
1	20-FEB-2017	1	50
1	20-FEB-2017	2	30

```
{
  "order":{
    "order_id":1,
    "total_amount":"80",
    "date":"20-Feb-2017",
    "country":"PT",
  },
  "order item":[
    {
      "item_id":1,
      "amount":"50"
    },
    {
      "item_id":2,
      "amount":"30"
    }
  ]
}
```

AMBIENTES ANALÍTICOS COM GRANDE VOLUME DE DADOS

Modelo
Normalizado

Modelo
Dimensional
Star

Modelo
Flatten Star

Complexo

Elegante

Simples



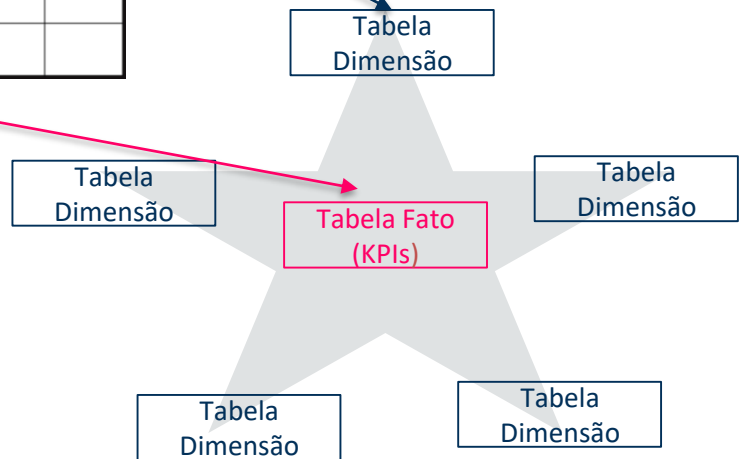
Uso do BPM na Construção do Modelo de Dados



IDENTIFICANDO INDICADORES E DIMENSÕES

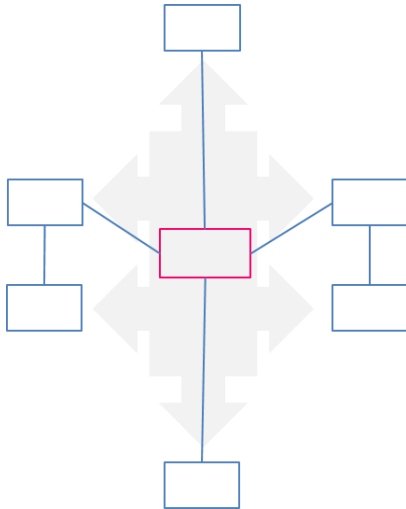
Business Process Matrix (BPM) – Star Schema

Processo	Indicador	Dim. Data	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6
Proc 1	KPI 1	x	x	x		x	x
Proc 2	KPI 2	x	x		x	x	
Proc 3	KPI 3	x					x
Proc 4	KPI 4	x		x			

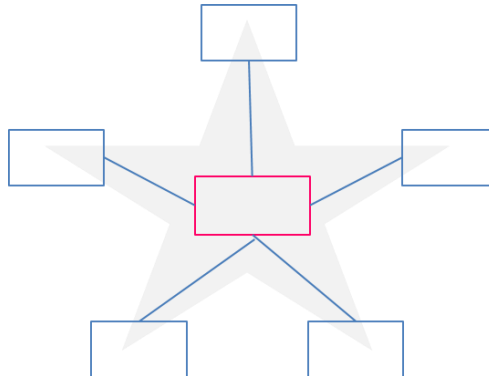


MODELOS DE DADOS

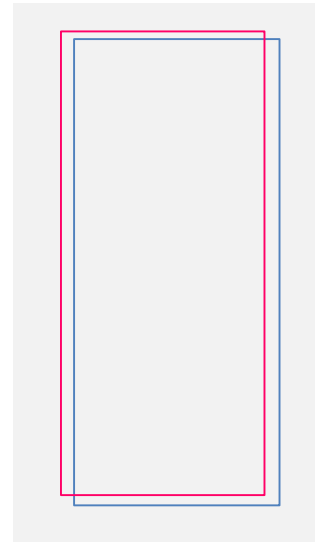
Snow Flake



Star Schema



Flatten



EXEMPLO – USO DO BPM

Fato 1: Qual é o total de Unidades Disponíveis e Unidades Vendidas dentro das torres de uma série de empreendimentos?

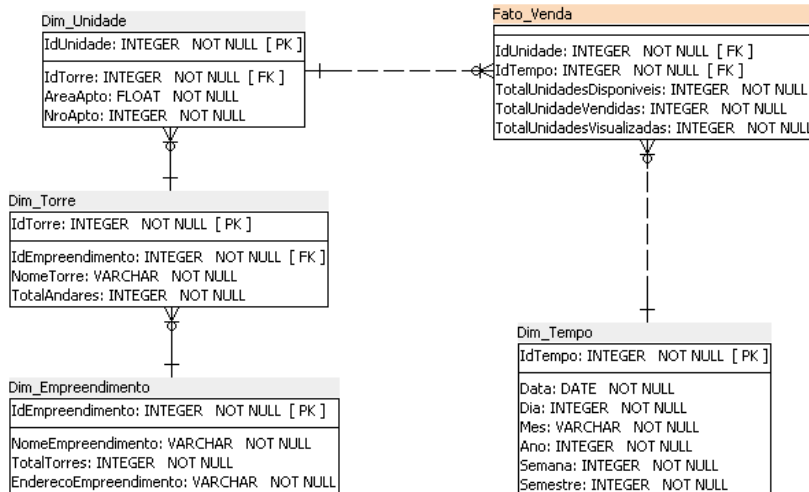
Fato 2: Fato 2 -- > Qual são as Unidades mais visualizadas no site?

Modelo:	Star Schema			
Grão:	Unidade / Data da Venda ou Lançamento			
Métrica	Tipo de Agregação	Tipo de Origem		
Total de Unidades Disponíveis	Soma/Média	Quente / Estruturada		
Total de Unidades Vendidas	Soma/Média	Quente / Estruturada		
Total de Unidades Visualizadas	Soma/Média	Quente / Semiestruturada		
Dimensão	Tipo de Dimensão			
Unidade	Hierárquica			
Torre				
Empreendimento				
Data	Tempo			
Matriz Métrica/Dimensão	Unidade	Torre	Empreendimento	Data
Total de Unidades Disponíveis	x	x	x	x
Total de Unidades Vendidas	x	x	x	x
Total de Unidades Visualizadas	x	x	x	x

MODELO SNOW FLAKE

Fato 1: Qual é o total de Unidades Disponíveis e Unidades Vendidas dentro das torres de uma série de empreendimentos?

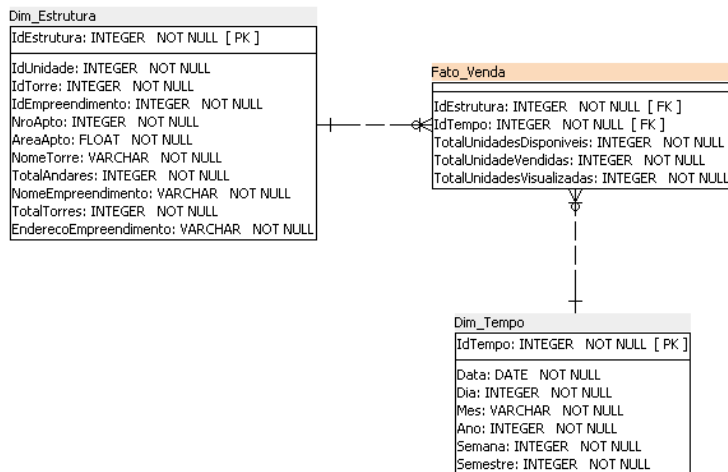
Fato 2: Fato 2 --> Qual são as Unidades mais visualizadas no site?



MODELO STAR SCHEMA

Fato 1: Qual é o total de Unidades Disponíveis e Unidades Vendidas dentro das torres de uma série de empreendimentos?

Fato 2: Fato 2 --> Qual são as Unidades mais visualizadas no site?



MODELO FLATTEN

Fato 1: Qual é o total de Unidades Disponíveis e Unidades Vendidas dentro das torres de uma série de empreendimentos?

Fato 2: Fato 2 -- > Qual são as Unidades mais visualizadas no site?

Flatten_Venda_Empreendimento	
IdVendaEmpreendimento:	BIGINT NOT NULL [PK]
IdUnidade:	INTEGER NOT NULL
IdTorre:	INTEGER NOT NULL
IdEmpreendimento:	INTEGER NOT NULL
NroApto:	INTEGER NOT NULL
AreaApto:	FLOAT NOT NULL
NomeTorre:	VARCHAR NOT NULL
TotalAndares:	INTEGER NOT NULL
NomeEmpreendimento:	VARCHAR NOT NULL
TotalTorres:	INTEGER NOT NULL
EnderecoEmpreendimento:	VARCHAR NOT NULL
IdTempo:	INTEGER NOT NULL
Data:	DATE NOT NULL
Dia:	INTEGER NOT NULL
Mes:	VARCHAR NOT NULL
Ano:	INTEGER NOT NULL
Semana:	INTEGER NOT NULL
Semestre:	INTEGER NOT NULL
TotalUnidadesDisponiveis:	INTEGER NOT NULL
TotalUnidadeVendidas:	INTEGER NOT NULL
TotalUnidadesVisualizadas:	INTEGER NOT NULL



HANDS ON



VAMOS APLICAR?

5. Considerando as matrizes métrica/dimensão criadas no exercício 4, desenhe dois modelos: star schema para os dados de origem estruturada, e flatten para os dados de origem semiestruturada.

**Importante:*

- *Insira o print dos Modelos no item de “3.1 Modelo Dimensional” do documento “Especificação da Solução de DW”.*
- *Como sugestão, utilize a ferramenta <https://www.dbdesigner.net/> para a construção dos modelos.*

VAMOS APLICAR?

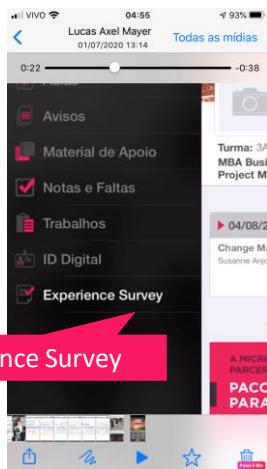
Teste seus conhecimentos!

[Clique aqui e responda.](#)



O que você achou da aula de hoje?

1. Aprendi satisfatoriamente com esta aula remota?
2. O professor conduziu bem a aula?
3. Me senti envolvido com as dinâmicas nesta aula remota?
4. O fluxo da aula foi bom?
5. Os materiais usados foram úteis na aula?



Pelo FIAPP

Pelo QR Code



Pelo Link

<https://fiap.me/Pesquisa-MBA>

Já estamos no meio da nossa jornada :o



<https://www.linkedin.com/in/tassianarugoni/>



profassiana.campos@fiap.com.br

FIAP

Copyright © 2021 | Professora Tassiana Rugoni de Campos, Regina Cantele, Jorge Surian, Marcelo Manzano.
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente
proibido sem consentimento formal, por escrito, do professor/autor.

