



**TREINAMENTO OFICIAL  
MICROSOFT**

# **CERTIFICAÇÃO DP-900**

**EP 1/4**



# Who I am?



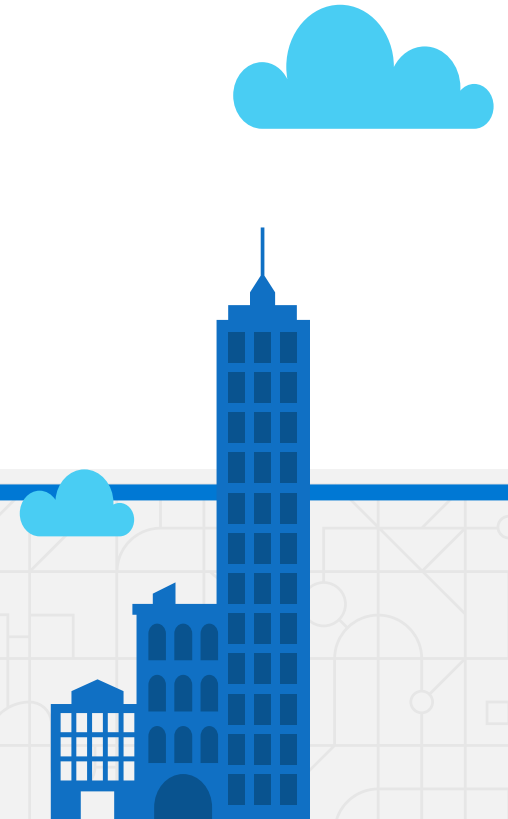
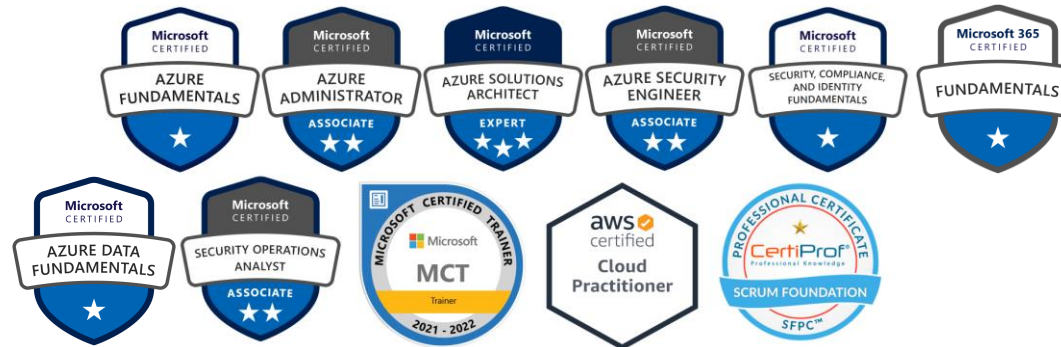
**Valéria Baptista**  
Azure Cloud Engineer

## Experiência

- +10 anos na TI

## Formação

- Ciências da Computação
- Pós em Cloud Computing
- Microsoft Certified Trainer



# Agenda



Explorar os principais conceitos de dados



Explorar funções e responsabilidades no mundo dos dados (opcional)



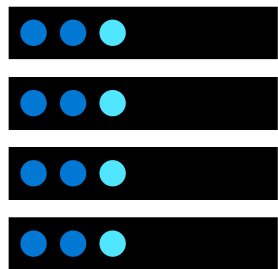
Descrever os conceitos de dados relacionais



Explorar os conceitos dos dados não relacionais



Explorar conceitos de análise de dados



# Módulo 1

Explorar os principais conceitos de dados





# O que são dados?

Coleção de fatos, números, descrições, objetos, armazenados de maneira estruturada, semiestruturada, não estruturada.





# O que são dados?

Coleção de fatos, números, descrições, objetos, armazenados de maneira estruturada, semiestruturada, não estruturada.

## Semiestruturado

```
## Document 1 ## {
  "customerID": "103248",
  "name": { "first": "AAA",
    "last": "BBB" }, "address": {
    "street": "Main Street",
    "number": "101", "city":
    "Acity", "state": "NY" },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003" }
## Document 2 ## {
  "customerID": "103249",
  "name": { "title": "Mr",
    "forename": "AAA",
    "lastname": "BBB" },
  "address": { "street":
    "Another Street", "number":
    "202", "city": "Bcity",
    "county": "Gloucestershire",
    "country-region": "UK" },
  "ccOnFile": "yes" }
```







# O que são dados?

Coleção de fatos, números, descrições, objetos, armazenados de maneira estruturada, semiestruturada, não estruturada.



# Armazenamento de dados transacional versus analítico

## Processamento transacional on-line (OLTP)

Cliente		
CustomerID	CustomerName	CustomerPhone

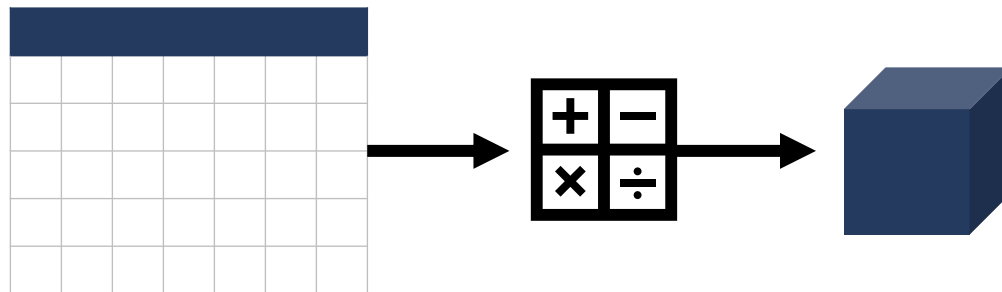
Pedidos		
OrderID	CustomerID	OrderDate

Os dados são armazenados em uma transação por vez



# Armazenamento de dados transacional versus analítico

## Processamento analítico on-line (OLAP)

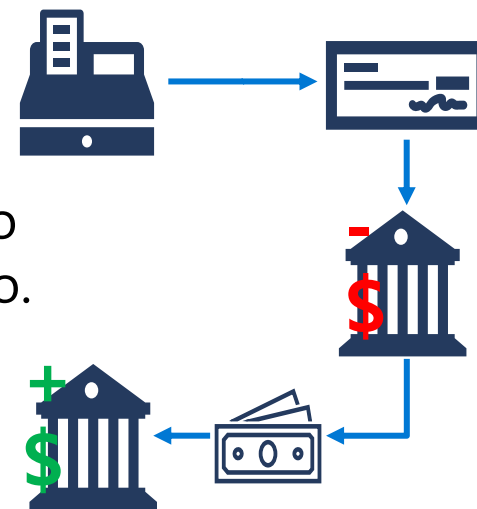


Os dados são periodicamente carregados, agregados e armazenados em um cubo

# Cargas de trabalho transacionais

Os dados transacionais são informações que rastreiam as interações relacionadas às atividades de uma organização.

- **Atomicidade** – cada transação é tratada como uma unidade independente que resulta em sucesso completo ou falha completa.
- **Consistência** – as transações só podem conduzir os dados do banco de dados de um estado válido para outro estado válido.
- **Isolamento** – a execução concorrente de transações deixa o banco de dados no mesmo estado.
- **Durabilidade** – assim que uma transação tiver sido confirmada, permanecerá assim.



# Cargas de trabalho analíticas

As cargas de trabalho analíticas são usadas para análise de dados e tomada de decisões.

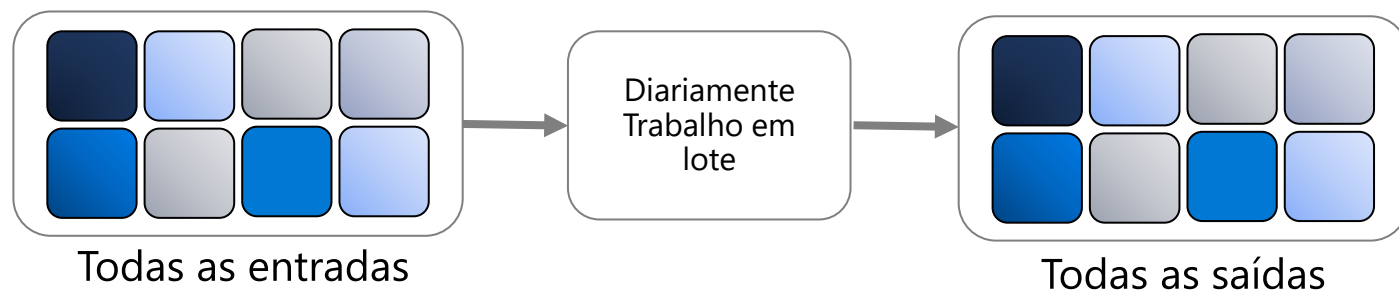
- Resumos
- Tendências
- Informações comerciais



# Processamento de dados

O processamento de dados é a conversão de dados brutos em informações relevantes por meio de um processo.

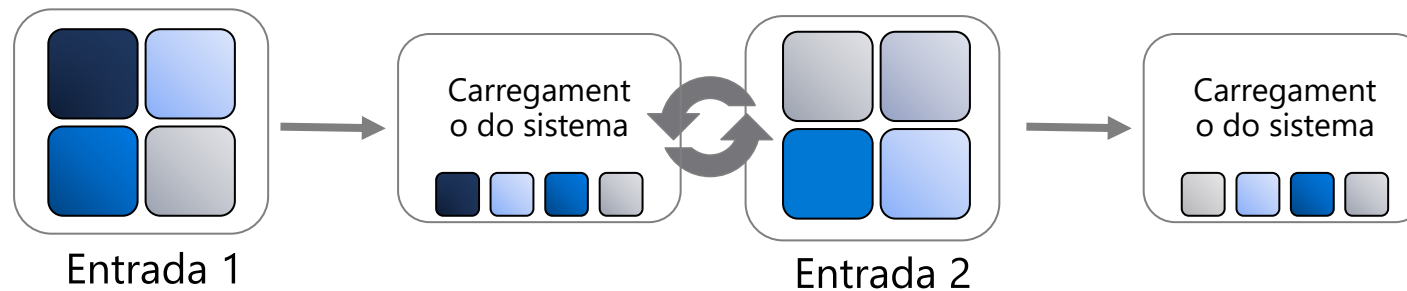
**Processamento em lotes:** os elementos de dados são reunidos em um grupo. Então, o grupo inteiro é processado em um momento futuro como um lote



# Processamento de dados

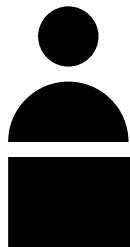
O processamento de dados é a conversão de dados brutos em informações relevantes por meio de um processo.

**Processamento de streaming:** cada novo dado é processado quando chega.





## Módulo 2



Explorar funções e responsabilidades no mundo dos dados





# Funções em dados

## Administrador de Banco de Dados

Gerenciamento de banco de dados

Segurança de dados de implementos

Backups

Acesso do usuário

Desempenho de monitores



## Engenheiro de dados

Pipelines e processos de dados

Armazenamento de ingestão de dados

Preparar dados para análise

Preparar dados para processamento de análise



## Analista de dados

Oferece insights sobre os dados

Geração de relatórios visuais

Modelagem de dados para análise

Combina dados para visualização e análise







# Ferramentas comuns

## Administrador de banco de dados

### Azure Data Studio

Interface gráfica para gerenciar serviços de dados no local e baseados na nuvem

Funciona no Windows, no macOS e no Linux.

### SQL Server Management Studio

Interface gráfica para gerenciar serviços de dados no local e baseados na nuvem

Funciona no Windows

Ferramenta de administração de banco de dados abrangente

### Portal/CLI do Azure

Ferramentas para gerenciamento e provisionamento de serviços de dados do Azure

Manual e automação de scripts usando o Azure Resource Manager ou criação de scripts em interface de linha de comando





# Ferramentas comuns

## Engenharia de dados

### Azure Synapse Studio

Portal do Azure integrado para gerenciar o Azure Synapse

Ingestão de dados (Azure Data Factory)

Gerenciamento de ativos do Azure Synapse (Pool de SQL/Pool do Spark)

### SQL Server Management Studio

Interface gráfica para gerenciar serviços de dados no local e baseados na nuvem

Funciona no Windows

Ferramenta de administração de banco de dados abrangente

### Portal/CLI do Azure

Ferramentas para gerenciamento e provisionamento de recursos do Azure

Manual e automação de scripts usando o Azure Resource Manager ou criação de scripts em interface de linha de comando



# Ferramentas comuns - Analista de dados

## Power BI Desktop

Ferramenta de visualização de dados

Modele e visualize dados

Gerenciamento de ativos do Azure Synapse (Pool de SQL/Pool do Spark)

## Power BI Portal/ serviço do Power BI

Criação e gerenciamento de relatórios do Power BI

Criação de painéis do Power BI

Compartilhe relatórios/conjuntos de dados

## Power BI Report Builder

Ferramenta de visualização de dados para relatórios paginados

Modele e visualize relatórios paginados

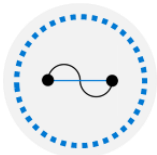


# • • • Módulo 3

## • Descrever os conceitos de dados relacionais



# Identificar casos de uso do banco de dados relacional



## **IoT:**

Embora sejam normalmente considerados como não relacionais, os dados de dispositivos IoT poderiam ser estruturados e consistentes

---



## **Processamento de transação on-line:**

Por exemplo, sistemas de pedidos que realizam muitas atualizações transacionais pequenas

---



## **Data warehousing:**

Grandes quantidades de dados podem ser importadas de várias fontes e estruturadas para permitir consultas de alto desempenho

# Tabelas



## Clientes

CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX
107	Francis Ribeiro	XXX-XXX-XXXX

Os dados são armazenados em uma tabela

A tabela consiste em linhas e colunas

Todas as linhas têm o mesmo número de colunas

Cada coluna é definida por um tipo de dados

# Entidades



Clientes		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Uma entidade é uma representação de um item que pode ser físico (como um cliente ou produto), ou virtual (como um pedido).

As entidades são conectadas por relações que permitem a interação.

Por exemplo: um cliente pode fazer um pedido de um produto





# Normalização



Clientes		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Pedidos		
OrderID	CustomerName	CustomerPhone
AD100	Noam Maoz	XXX-XXX-XXXX
AD101	Noam Maoz	XXX-XXX-XXXX
AD102	Noam Maoz	XXX-XXX-XXXX
AX103	Qamar Mounir	XXX-XXX-XXXX
AS104	Qamar Mounir	XXX-XXX-XXXX
AR105	Claude Paulet	XXX-XXX-XXXX
MK106	Muisto Linna	XXX-XXX-XXXX

Os dados são normalizados para:

Reduzir armazenamento

Evitar a duplicação de dados

Melhorar a qualidade dos dados



# Relações

Clientes		
CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

Pedidos		
OrderID	CustomerID	SalesPersonID
AD100	101	200
AD101	101	200
AD102	101	200
AX103	103	201
AS104	103	201
AR105	105	200
MK106	105	201



## Em um esquema de banco de dados normalizado:

As chaves primárias e estrangeiras são usadas para definir relacionamentos

Não há nenhuma duplicação de dados (a não ser os valores chave no 3º formulário normal (3NF))

Os dados são recuperados unindo as tabelas em uma consulta

# Índices



## Clientes

CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

## IDX-CustomerRegion

CustomerID	Região
100	França
101	Brasil
102	Croácia
103	Jordan
104	Espanha
105	França
106	EUA

## Um índice:

Otimiza consultas de pesquisa para recuperação mais rápida de dados

Reduz a quantidade de páginas de dados que precisam ser lidas para recuperar os dados em uma Declaração SQL

Os dados são recuperados unindo as tabelas em uma consulta





# Exibição

## Clientes

CustomerID	CustomerName	CustomerPhone
100	Muisto Linna	XXX-XXX-XXXX
101	Noam Maoz	XXX-XXX-XXXX
102	Vanja Matkovic	XXX-XXX-XXXX
103	Qamar Mounir	XXX-XXX-XXXX
104	Zhenis Omar	XXX-XXX-XXXX
105	Claude Paulet	XXX-XXX-XXXX
106	Alex Pettersen	XXX-XXX-XXXX

## Pedidos

OrderID	CustomerID	SalesPersonID
AD100	101	200
AD101	101	200
AD102	101	200
AX103	103	201
AS104	103	201
AR105	105	200
MK106	105	201
DB205	100	205

Criar a definição de uma exibição:

```
CREATE VIEW  
vw_customerorders AS  
SELECT  
Customers.CustomerID,  
Customers.CustomerName  
, Orders.OrderID FROM  
Customers JOIN Orders  
on  
Customers.CustomerID =  
Orders.CustomerID
```

Recuperar os pedidos feitos pelo cliente 102 usando a exibição:

```
SELECT CustomerName,  
OrderID from  
vw_customerorders  
WHERE CustomerID=102
```

Uma exibição é uma tabela virtual com base no conjunto de resultados de uma consulta:

As exibições são criadas para simplificar a consulta

Combine dados relacionados em uma exibição de painel simples





## Módulo 4

Explorar os conceitos dos dados não relacionais



# Explorar as características dos dados não relacionais



## Entidades

## Customer 1 ID: 1

Nome: Mark Hanson

Telefone: [ Residencial: 1-999-9999999, Comercial: 1-888-8888888, Celular: 1-777-7777777 ]

Endereço: [ Residencial: 121 Main Street, Some City, NY, 10110,  
Business: 87 Big Building, Some City, NY, 10111 ]

## Customer 2 ID: 2

Título: Sr

Nome: Jeff Hay

Telefone: [ Residencial: 0044-1999-333333, Celular: 0044-17545-444444 ]

Endereço: [ Reino Unido: 86 High Street, Some Town, A County, GL8888, UK,  
EUA: 777 7th Street, Another City, CA, 90111 ]

### Coleções não relacionais podem ter:

Várias entidades na mesma coleção ou contêiner com diferentes campos

Têm um esquema diferente, não tabular

São frequentemente definidos por identificar cada campo com o nome que ele representa



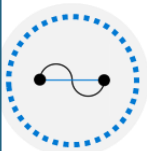
# Identificar casos de uso do banco de dados não relacional



## IoT e telemática:

Com frequência exigem ingerir grandes quantidades de dados em frequentes explosões de atividades, os dados são semiestruturados ou estruturados, com frequência exigem processamento em tempo real

---



## Varejo e marketing:

Cenários comuns para dados distribuídos globalmente, armazenamento de documentos

---



## Jogos:

Estatísticas de jogo, integração com mídias sociais, quadros de líderes, aplicativos de baixa latência

---



## Web e móvel:

Normalmente usado com análises de cliques na web, aplicativos modernos incluindo bots





# Tipos de dados não relacionais

## O que são dados semiestruturados?

A estrutura de dados é definida dentro dos dados reais por campos. O formato/tipos de arquivos incluem:

JSON

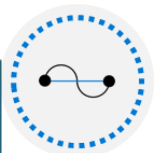
ORC

AVRO

Parquet



# O que são dados não estruturados?

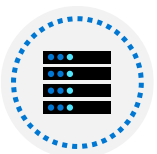


**Não contém campos naturalmente:**

*Exemplos: vídeo, áudio, streaming de mídia, documentos*



**Usado com frequência para extrair organização de dados e categorizar ou identificar “estruturas”**



**Frequentemente usado em combinação com recursos de Machine Learning ou Serviços Cognitivos para “extrair dados” usando:**

Análise de Texto

Análise de sentimento com APIs cognitivas

API de Visão

# O que é NoSQL?

Termo solto para descrever não relacional



Repositórios  
de chave-  
valor

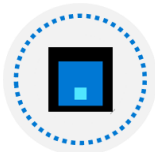
Bancos de  
dados de  
família de  
colunas

Baseado em  
documento

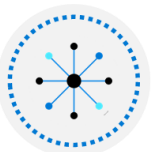
Bancos de  
dados de  
grafo



# O que é um repositório de chave valor?



Um armazenamento de valor-chave é o tipo mais simples (e geralmente mais rápido) de banco de dados NoSQL para inserção e consulta de dados. Cada item de dados em um armazenamento de valor-chave tem dois elementos, uma chave e um valor.



O foco de um armazenamento de valor-chave é a capacidade de ler e gravar dados muito rapidamente. Os recursos de pesquisa são secundários. Um armazenamento de valor-chave é uma excelente opção para ingestão de dados, quando um grande volume de dados chega como um fluxo contínuo e deve ser armazenado imediatamente.

O armazenamento de Tabelas do Azure é um exemplo de armazenamento de chave-valor. O Cosmos DB também implementa um armazenamento de chave-valor usando a Tabela de APIs.

Key	Value
AAAAA	1101001111010100110101111...
AABAB	1001100001011001101011110...
DFA766	0000000000101010110101010...
FABCC4	1110110110101010100101101...

Opaque to data store

# O que é um repositório baseado em documento?



Os bancos de dados de documentos normalmente armazenam dados no formato JSON, conforme descrito na unidade anterior, ou podem ser codificados usando outros formatos, como XML, YAML, JSON, BSON. Os documentos podem até ser armazenados como texto simples. Os campos nos documentos são expostos ao sistema de gerenciamento de armazenamento, permitindo que um aplicativo consulte e filtre dados usando os valores desses campos.

Um armazenamento de documentos não requer que todos os documentos tenham a mesma estrutura. Essa abordagem de forma livre oferece uma grande flexibilidade. Os aplicativos podem armazenar dados diferentes em documentos à medida que os requisitos de negócios mudam.

Key	Document
1001	{ "CustomerID": 99, "OrderItems": [ { "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 } ], "OrderDate": "04/01/2017" }
1002	{ "CustomerID": 220, "OrderItems": [ { "ProductID": 1285, "Quantity": 1, "Cost": 120 } ], "OrderDate": "05/08/2017" }



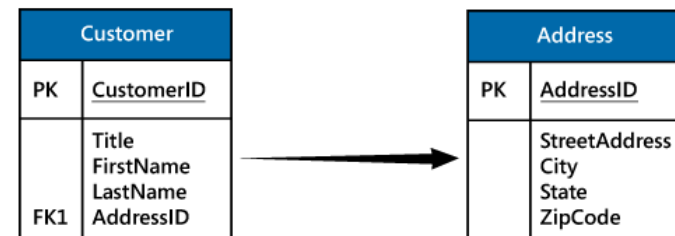
# O que é um banco de dados de famílias de colunas?



Um banco de dados de família de colunas organiza os dados em linhas e colunas. Exemplos dessa estrutura incluem arquivos ORC e Parquet.

Em sua forma mais simples, um banco de dados de família de colunas pode parecer muito semelhante a um banco de dados relacional, pelo menos conceitualmente. O verdadeiro poder de um banco de dados de família de colunas está em sua abordagem desnormalizada para estruturar dados esparsos.

O sistema de gerenciamento de banco de dados da família de colunas mais usado é o Apache Cassandra. O Azure Cosmos DB dá suporte à abordagem de família de colunas por meio da API do Cassandra.



Customer Table

CustomerID	Title	FirstName	LastName	AddressID
1	Mr	Mark	Hanson	500
2	Ms	Lisa	Andrews	501
3	Mr	Walter	Harp	500

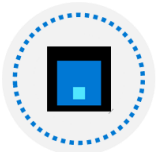
Address Table

AddressID	StreetAddress	City	State	ZipCode
500	999 500th Ave	Bellevue	WA	12345
501	888 W. Front St	Boise	ID	54321

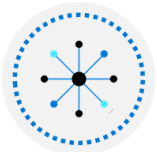




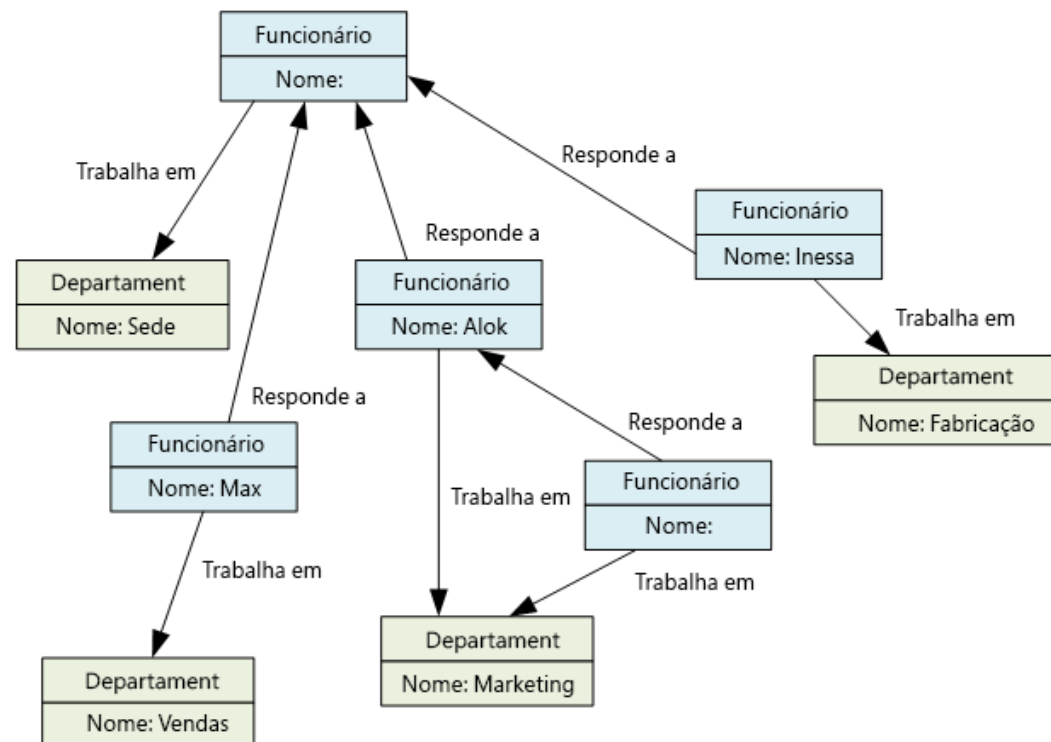
# O que é um banco de dados de grafo?



Armazena entidades centradas ao redor de relacionamentos



Permite que aplicativos realizem consultas percorrendo uma rede de nós e bordas







# Módulo 5

## Explorar conceitos de análise de dados

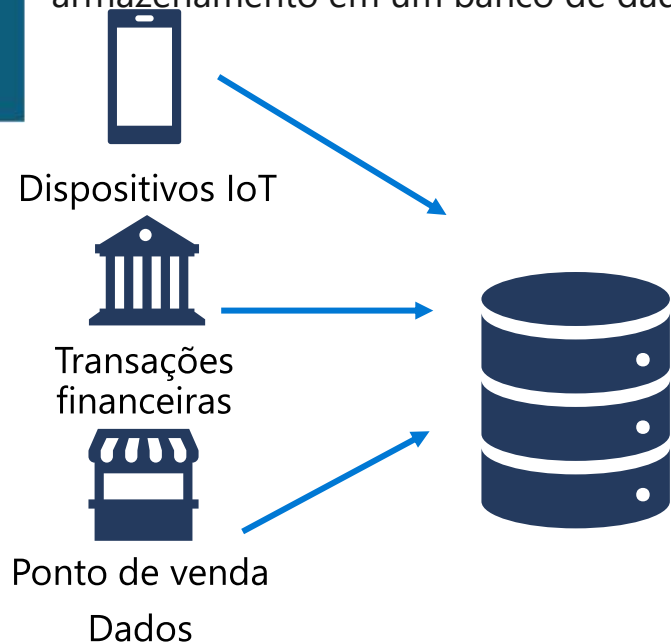




# A jornada dos dados

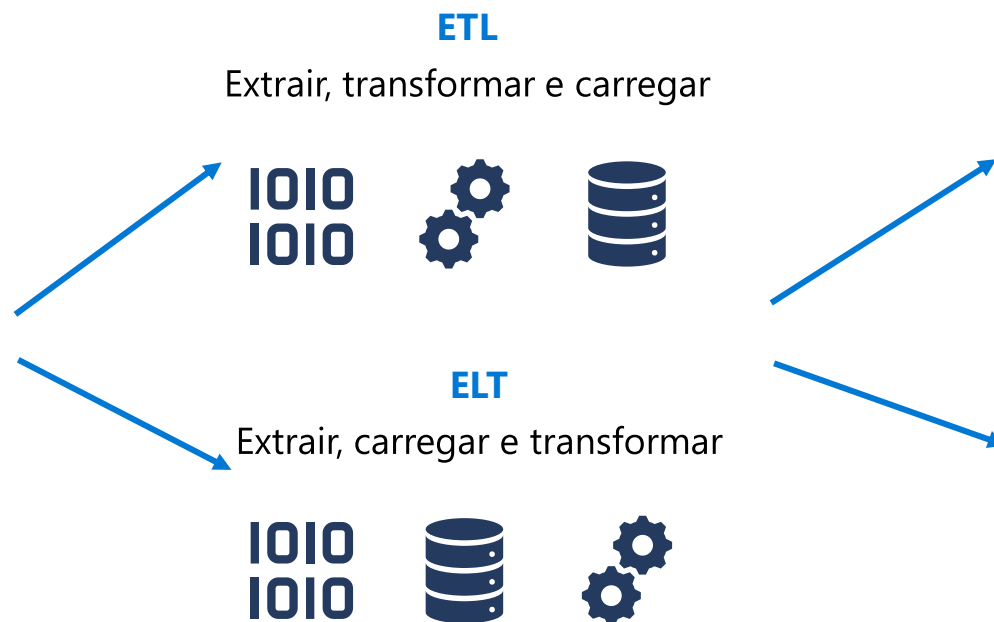
## Ingestão de dados

O processo de obtenção e importação de dados para uso imediato ou armazenamento em um banco de dados.



## Processamento de dados

Pega os dados em sua forma bruta, limpa-os e converte-os em um formato mais significativo



## Visualização de Dados

Consulte os dados e crie representações gráficas de informações e dados





# Visualização de dados

Um modelo de negócios pode conter uma enorme quantidade de informações – há técnicas para analisar e entender as informações em seus modelos



Relatórios



Business Intelligence (BI)



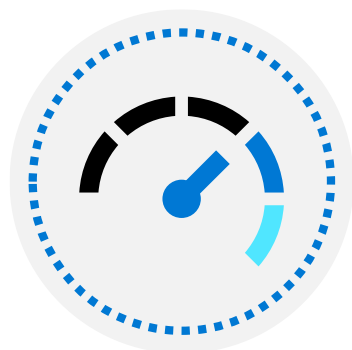
Visualização de dados



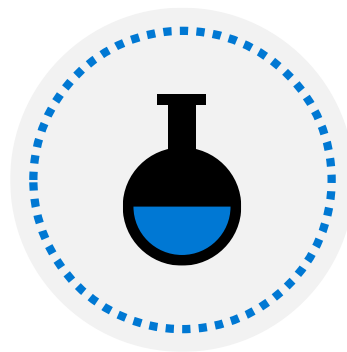
# Explorar a análise de dados



Descritiva



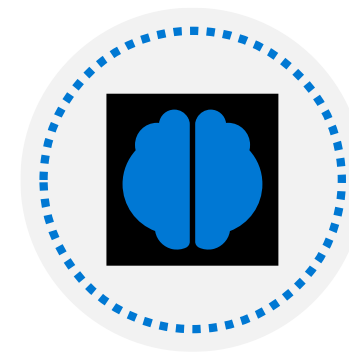
Diagnóstica



Preditiva



Prescritiva

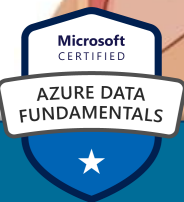


Cognitiva





# Teste de conhecimento



[www.youtube.com/canaldacloud](https://www.youtube.com/canaldacloud)



# Como os dados são organizados em uma tabela relacional?



☒ Linhas e colunas

☐ Cabeçalho e rodapé

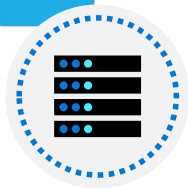
☐ Páginas e parágrafos



## Qual das alternativas a seguir é um exemplo de dados não estruturados?

- ☐ Uma tabela de Funcionários com as colunas ID do Funcionário, Nome do Funcionário e Designação do Funcionário
- ☒ Arquivos de áudio e vídeo
- ☐ Uma tabela dentro do banco de dados do SQL Server





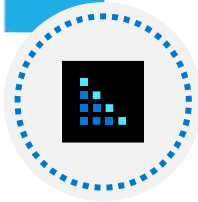
Qual das alternativas a seguir é um exemplo de conjunto de dados de streaming?

- ☒ Dados de feeds de sensores
- ☐ Dados de vendas do mês anterior
- ☐ Lista de funcionários que trabalham para uma empresa



Qual das ferramentas a seguir é usada para obter relatórios e visualização?

- ☐ SQL Server Management Studio
- ☒ Power BI
- ☐ Linguagem SQL



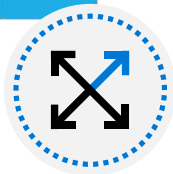
# Qual das afirmações a seguir é uma característica dos bancos de dados relacionais?

- ☐ Todos os dados precisam ser armazenados como cadeias de caracteres
- ☒ Uma linha em uma tabela representa apenas uma entidade
- ☐ Linhas diferentes na mesma tabela podem conter colunas diferentes



## O que é um índice?

- ☒ Uma estrutura que permite localizar linhas em uma tabela rapidamente, usando um valor indexado
- ☐ Uma tabela virtual com base no conjunto de resultados de uma consulta
- ☐ Uma estrutura que consiste em linhas e colunas que você usa para armazenar dados



Quais dos seguintes serviços você deve usar para implementar um banco de dados não relacional?



☒ Azure Cosmos DB

☐ Banco de dados SQL do Azure

☐ A API do Gremlin



## Qual das alternativas a seguir é uma característica dos bancos de dados não relacionais?

- ☐ Bancos de dados não relacionais contêm tabelas com registros de coluna fixa simples
- ☐ Bancos de dados não relacionais exigem que você use técnicas de normalização para reduzir a duplicação de dados
- ☒ Bancos de dados não relacionais são livres de esquema ou têm esquemas relaxados



## O que é ingestão de dados?

- ☐ O processo de transformar dados brutos em modelos que contêm informações significativas
- ☐ Analisar dados para encontrar anomalias
- ☒ Capturar fluxo de dados brutos de várias fontes e armazená-los





# Que tipo de análise ajuda a responder perguntas sobre o que aconteceu no passado?

- ☒ Análise descritiva
- ☐ Análise prescritiva
- ☐ Análise preditiva





# Obrigado!

