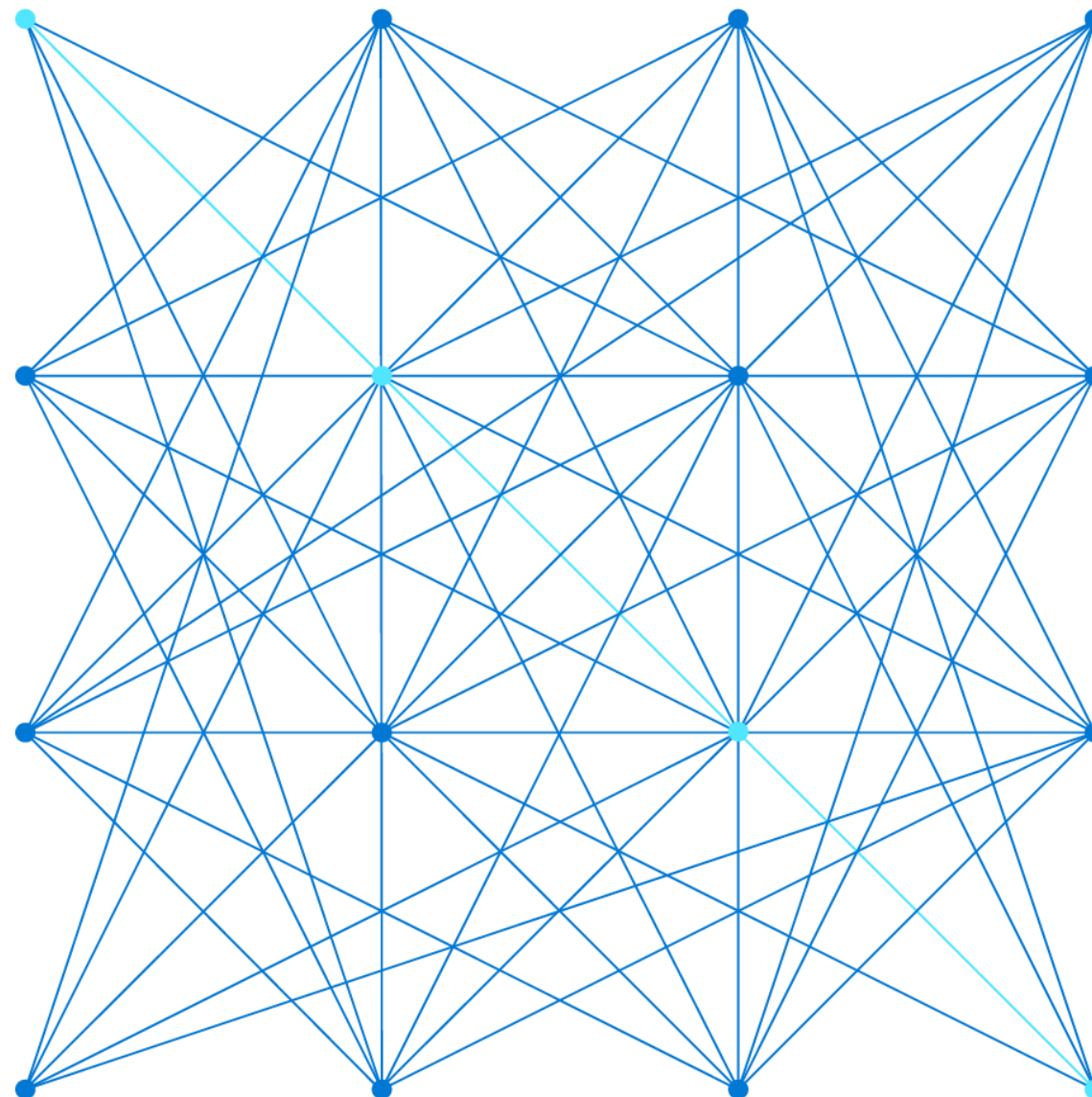


Módulo 1: Explorar conceitos básicos sobre dados



Agenda



Principais conceitos de dados



Funções e serviços de dados

Lição 1: Principais conceitos de dados



O que são dados?

Valores usados para registrar informações, geralmente representando *entidades* que têm um ou mais *atributos*

Estruturados

Cliente				
ID	FirstName	LastName	Email	Endereço
1	Joe	Jones	joe@litware.com	1 Main St.
2	Samir	Nadoy	samir@northwind.com	123 Elm Pl.

Produto		
ID	Nome	Preço
123	Martelo	2.99
162	Screwdriver	3.49
201	Chave inglesa	4.25

Semiestruturados

```
{
  "firstName": "Joe",
  "lastName": "Jones",
  "address":
  {
    "streetAddress": "1 Main
St.",
    "city": "New York",
    "state": "NY",
    "postalCode": "10099"
  },
  "contact":
  [
    {
      "type": "home",
      "number": "555 123-1234"
    },
    {
      "type": "email",
      "address":
        "joe@litware.com"
    }
  ]
}
```

```
{
  "firstName": "Samir",
  "lastName": "Nadoy",
  "address":
  {
    "streetAddress": "123 Elm
Pl.",
    "unit": "500",
    "city": "Seattle",
    "state": "WA",
    "postalCode": "98999"
  },
  "contact":
  [
    {
      "type": "email",
      "address":
        "samir@northwind.com"
    }
  ]
}
```

Não estruturados

Prezado Joe,

Obrigado por pedir suas ferramentas em nossa loja online (pedido número 1000) em 01/01/2022.

Seu pedido foi enviado e deve chegar em 3 a 5 dias úteis.

Contoso Hardware

Nossos produtos são da mais alta qualidade e usados por profissionais. Temos chaves de fenda incríveis, que são muito úteis para apertar e soltar parafusos.



Também temos chaves inglesas (ou, se preferir, chaves de porca)...



Como os dados são armazenados?

Arquivos

Texto delimitado

```
FirstName,LastName,Email
Joe,Jones,joe@litware.com
Samir,Nadoy,samir@northwind.com
```

JSON (JavaScript Object Notation)

```
{
  "customers":
  [
    { "firstName": "Joe", "lastName": "Jones"},
    { "firstName": "Samir", "lastName": "Nadoy"}
  ]
}
```

Linguagem XML

```
<Customer firstName="Joe" lastName="Jones"/>
```

BLOB (objeto binário grande)

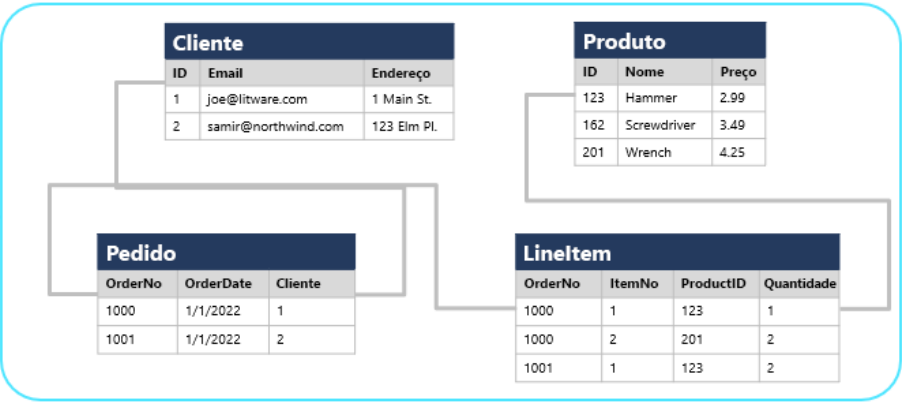
```
10110101101010110010...
```

Formatos otimizados:

- Avro, ORC, Parquet

Bancos de dados

Relacional



Não relacional

Produtos	
Chave	Valor
123	"Hammer (\$2.99)"
162	"Screwdriver (\$3.49)"
201	"Wrench (\$4.25)"

Chave-valor

Clientes	
Chave	Documento
1	{ "name": "Joe Jones" }
2	{ "name": "Samir Nadoy" }

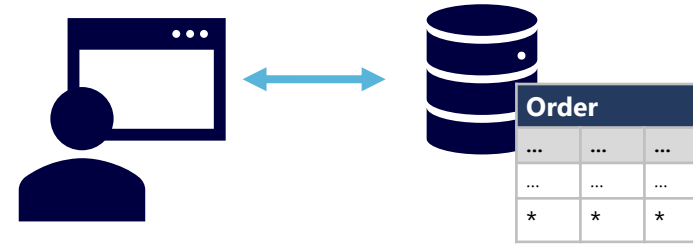
Documento

Pedidos				
Chave	Cliente		Produto	
	Nome	Endereço	Nome	Preço
1000	Joe Jones	1 Main St.	Hammer	2.99
1001	Samir Nadoy	123 Elm Pl.	Wrench	4.25

Família de colunas



Cargas de trabalho de dados transacionais



Os dados são armazenados em um banco de dados otimizado para *OLTP* (processamento de transações online) com suporte para aplicativos

Uma combinação de atividade de *leitura* e *gravação*

Por exemplo:

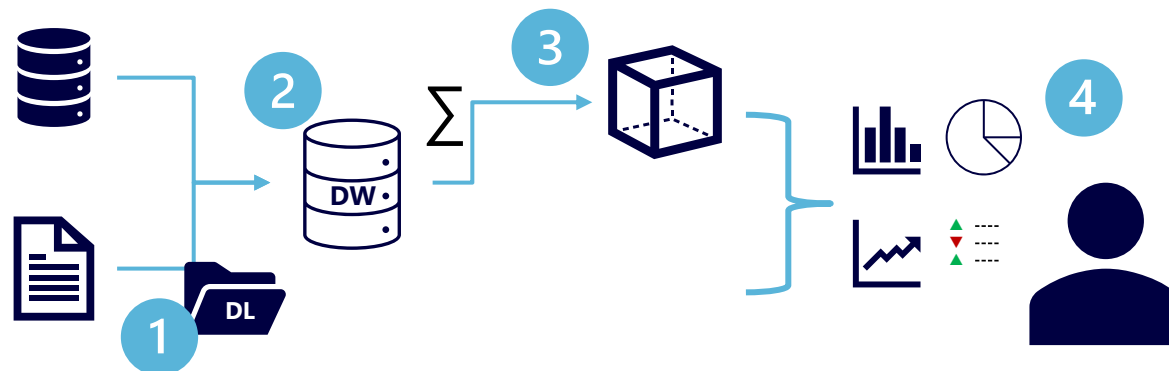
- Ler a tabela *Produto* para exibir um catálogo
- Gravar na tabela *Pedido* para registrar uma compra

Os dados são armazenados usando *transações*

As transações são baseadas em "ACID":

- **Atomicidade** – cada transação é tratada como uma unidade de trabalho, que é totalmente bem-sucedida ou que falha completamente
- **Consistência** – as transações só podem conduzir os dados do banco de dados de um estado válido para outro estado válido
- **Isolamento** – transações simultâneas não podem interferir umas com as outras
- **Durabilidade** – quando uma transação tiver sido bem-sucedida, as alterações de dados serão mantidas no banco de dados

Cargas de trabalho de dados analíticos



1. Os arquivos de dados podem ser armazenados em um *data lake* central para análise
2. Um processo de ETL (extração, transformação e carregamento) copia dados de arquivos e bancos de dados OLTP para um *data warehouse* otimizado para atividade de *leitura*
3. Os dados no data warehouse podem ser agregados e carregados em um modelo OLAP (processamento analítico online) ou *cubo*
4. Os dados no data lake, no data warehouse e no modelo analítico podem ser consultados para produzir relatórios e painéis.

Lição 2: Funções e serviços de dados



Funções dos profissionais de dados



Administrador de banco de dados

Provisionamento, configuração e gerenciamento de banco de dados

Segurança do banco de dados e acesso do usuário

Backups e resiliência de banco de dados

Monitoramento e otimização de desempenho do banco de dados



Engenheiro de dados

Pipelines de integração de dados e processos de ETL

Limpeza de dados e transformação

Esquemas de armazenamento de dados analíticos e cargas de dados



Analista de Dados

Modelagem analítica

Relatório e resumo de dados

Visualização de dados

Serviços de nuvem da Microsoft para dados

Armazenamentos de dados



SQL do Azure

- Família de serviços de banco de dados relacionais baseados no SQL Server



Banco de Dados do Azure para código aberto

- Maria DB, MySQL, PostgreSQL



Azure Cosmos DB

- Sistema de banco de dados não relacional altamente escalonável



Armazenamento do Azure

- Armazenamento de arquivos, blobs e tabelas
- Namespace hierárquico para armazenamento de data lake

Engenharia e análise de dados



Fábrica de dados do Azure

- Pipelines de dados



Azure Synapse Analytics

- Análise integrada de ponta a ponta
- Pipelines, SQL, Apache Spark, Data Explorer...



Azure Databricks

- Análise e processamento de dados do Apache Spark



Azure HDInsight

- Plataforma de código aberto Apache



Stream Analytics do Azure

- Processamento de dados em tempo real para soluções de IoT



Azure Data Explorer

- Análise de dados em tempo real para logs e telemetria



Microsoft Purview

- Governança de dados corporativos
- Mapeamento e descoberta de dados



Microsoft Power BI

- Modelagem de dados analíticos
- Visualização de dados interativa

outros...

