



Analytics - Visualização, Relatórios e Tomada de Decisões com Big Data



AZURE SQL WAREHOUSE

PARTE 1

Analytics, Visualização, Relatórios e Tomada de Decisões com Big Data



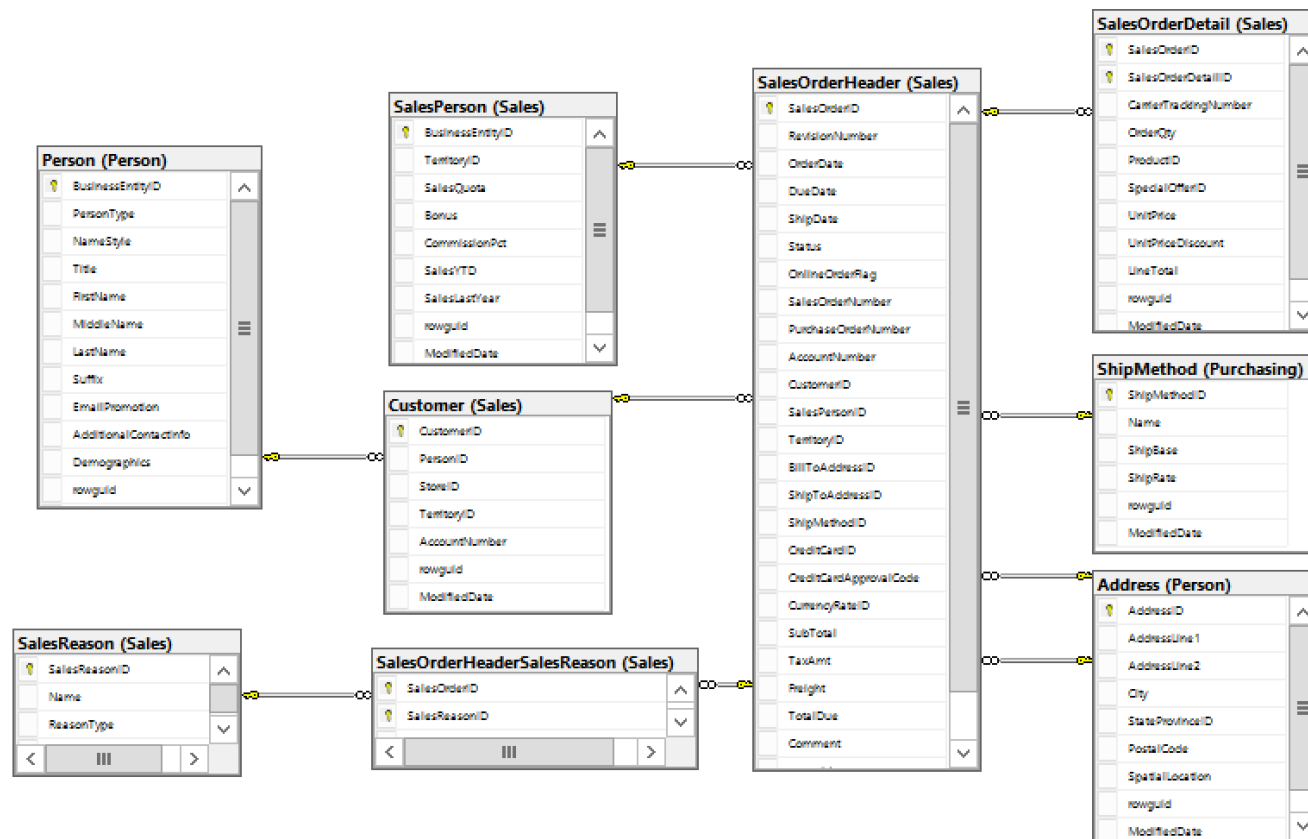


Azure SQL Data Warehouse – Parte 1

- Data WareHouses
- Azure SQL Data Warehouse
 - Arquitetura de Processamento Paralelo Massivo
 - Tabelas distribuídas e colunas de distribuição
 - Níveis de desempenho
- Criação de um Data Warehouse no Microsoft Azure
 - Conexão com C# e ADO.NET
 - Conexão com Tableau



Esquema de um banco de dados transacional



<https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>



Data WareHouse

- Banco de Dados relacional projeto para consultas, relatórios, análise de dados e outras tarefas de BI
- Usualmente contém dados históricos de banco de dados transacionais, podendo inclusive consolidar dados de diferentes fontes
- Usualmente possui dados não-voláteis
- Grandes cargas de dados
- Obs.: Não é voltado para processamento de transações



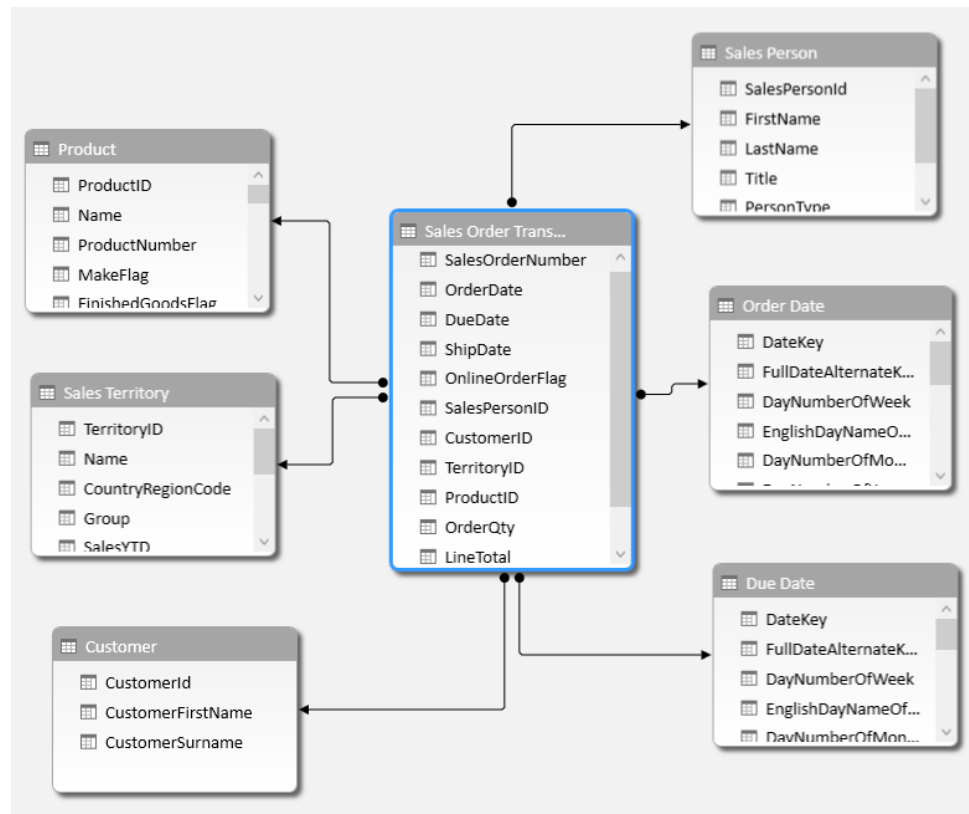
Tarefas envolvidas na construção de um Data Warehouse



- Extração de dados
- Limpeza
- Transformação
- Carga
- Atualização



Esquema de um banco de dados transacional



<https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>



Data Warehouse

OLTP		Data Warehouse
Complex data structures (3NF databases)		Multidimensional data structures
Few	Indexes	Many
Many	Joins	Some
Normalized DBMS	Duplicated Data	Denormalized DBMS
Rare	Derived Data and Aggregates	Common

https://docs.oracle.com/cd/A97630_01/server.920/a96520/concept.htm



Cubo de Dados

- Representação dos dados em múltiplas dimensões, definidos pelas tabelas de dimensões e pela tabela de fato.

Products ▾										
Quantity		Category Name ▾								
Country ▾	City ▾	Beverages	Condiments	Confections	Dairy Products	Grains/Cereals	Meat/Poultry	Produce	Seafood	Gr Total
▼ Argentina	Buenos Aires	82	45	57	54	20		33	48	
▸ Austria		982	720	575	1027	580	362	388	533	5
▸ Belgium		272	147	270	295	145	89	98	76	1
▼ Brazil	Campinas	50	70	15	48	60		31	41	
	Resende	72	12	54	15	28	35		51	
	Rio de Janeiro	384	317	362	261	127	48	52	342	1
	São Paulo	462	169	291	359	100	140	99	201	1
Brazil Total		968	568	722	683	315	223	182	635	4
▸ Canada		303	256	418	381	207	141	74	204	1
▸ Denmark		195	210	185	89	15	146	100	230	1
Grand Total		9532	5298	7906	9149	4562	4199	2990	7681	51

Page 1 of 3 (25 items) < 1 2 3 >

Demo: <https://demos.devexpress.com/aspxpivotgriddemos/OLAP/Browser.aspx>



Azure SQL Data Warehouse

- Data Warehouse na nuvem
- Disponibilidade de 99,9%
- Arquitetura de Processamento adequada
- Escalabilidade
- Disponibilidade global (mais de 30 regiões)

<https://azure.microsoft.com/pt-br/services/sql-data-warehouse>



Azure SQL Data Warehouse

- Processamento de Dados não-estruturados
- Integração com linguagens de programação e ferramentas líderes de mercado
- Separação da camada de armazenamento e computação
- Backups automático
- Uso de T-SQL (Transact SQL) e outras ferramentas conhecidas

<https://azure.microsoft.com/pt-br/services/sql-data-warehouse>



SQL (Structured Query Language)

DDL (Data Definition Language)

- CREATE, ALTER DROP

DML (Data Manipulation Language)

- INSERT, UPDATE, DELETE

DQL (Data Query Language)

- SELECT

DCL (Data Control Language)

- GRANT, REVOKE

DTL (Data Transaction Language)

- BEGIN TRANSACTION, COMMIT, ROLLBACK



T-SQL (Transact Structured Query Language)

- Extensão proprietária da Microsoft para a linguagem SQL

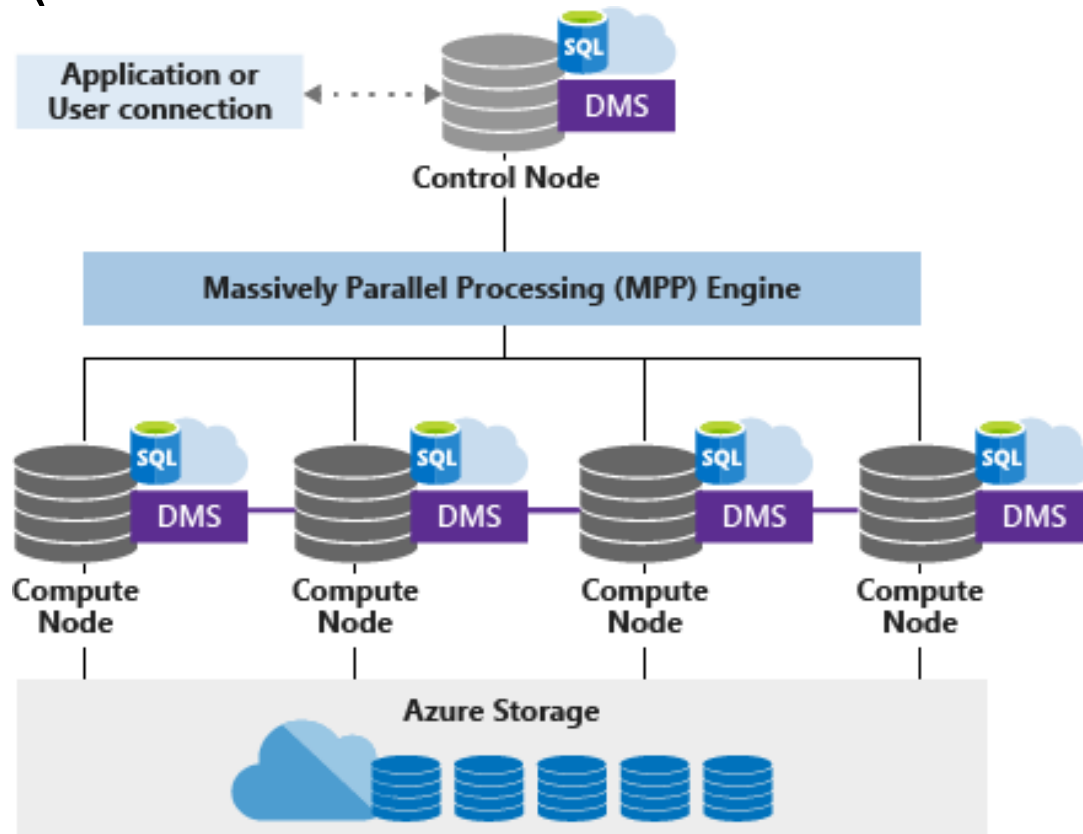
Suporte à:

- Controle de fluxo
- Variáveis
- Funções para processamento de strings, datas, etc

```
USE AdventureWorks2012;
GO
WHILE (SELECT AVG(ListPrice) FROM Production.Product) < $300
BEGIN
    UPDATE Production.Product
        SET ListPrice = ListPrice * 2
    SELECT MAX(ListPrice) FROM Production.Product
    IF (SELECT MAX(ListPrice) FROM Production.Product) > $500
        BREAK
    ELSE
        CONTINUE
END
PRINT 'Too much for the market to bear';
```



Arquitetura MPP (Processamento Paralelo Massivo)



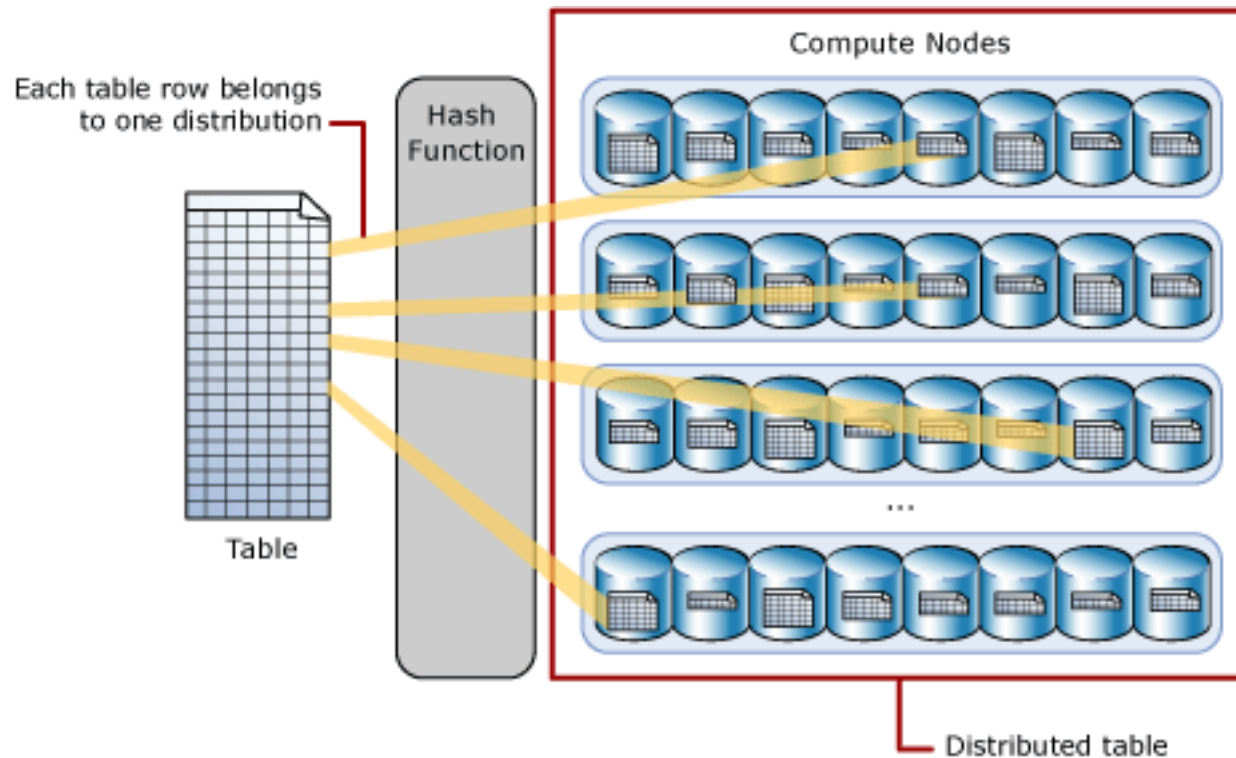
Conceitos

- Nó de Controle
- Nós de computação
- Serviço de Movimentação de Dados (DMS)

<https://docs.microsoft.com/pt-br/azure/sql-data-warehouse/massively-parallel-processing-mpp-architecture>



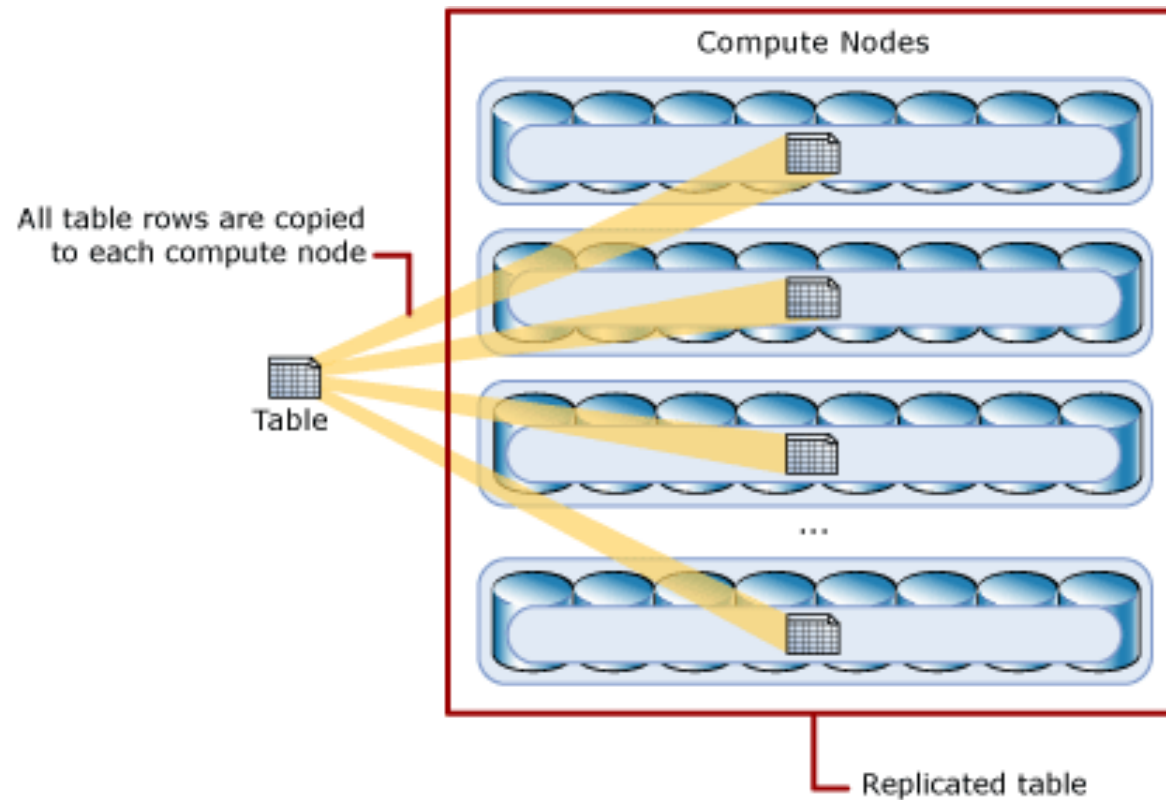
Tabelas distribuídas



<https://docs.microsoft.com/pt-br/azure/sql-data-warehouse/massively-parallel-processing-mpp-architecture>



Tabelas replicadas



<https://docs.microsoft.com/pt-br/azure/sql-data-warehouse/massively-parallel-processing-mpp-architecture>



Coluna de Distribuição

```
CREATE TABLE [dbo].[FactInternetSales]
(
    [ProductKey]          int          NOT NULL
,   [OrderDateKey]       int          NOT NULL
,   [CustomerKey]        int          NOT NULL
,   [PromotionKey]       int          NOT NULL
,   [SalesOrderNumber]   nvarchar(20) NOT NULL
,   [OrderQuantity]     smallint     NOT NULL
,   [UnitPrice]          money        NOT NULL
,   [SalesAmount]        money        NOT NULL
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX
,   DISTRIBUTION = HASH([ProductKey])
)
;
```



Coluna de Distribuição

Recomendações para escolha da coluna de distribuição:

- Coluna que não necessite de atualizações
- Coluna que distribua os dados uniformemente
- Tem muitos valores exclusivos
- Não tem valores nulos (ou com poucos valores nulos)
- Não é uma coluna de dados
- **COMO EU SEI SE ESCOLHI UMA BOA COLUNA DE DISTRIBUIÇÃO?**



Níveis de Desempenho

Desempenho otimizado para Elasticidade

- Separa camadas de armazenamento e computação na arquitetura
- Menor preço
- Unidade: DWU

```
--Optimized for Elasticity  
CREATE DATABASE myElasticSQLDW  
WITH  
(    SERVICE_OBJECTIVE = 'DW1000'  
)  
;
```

<https://docs.microsoft.com/pt-br/azure/sql-data-warehouse/performance-tiers>



Níveis de Desempenho

Desempenho otimizado para Computação

```
--Optimized for Compute  
CREATE DATABASE myComputeSQLDW  
WITH  
(    SERVICE_OBJECTIVE = 'DW1000c'  
)  
;
```

- Hardware com cache de disco de estado sólido NVMe mantendo dados frequentemente utilizados próximos às CPUs
- 2.5 vezes mais memória por consulta comparado ao nível de desempenho otimizado para Elasticidade
- Maior desempenho
- Unidade: cDWU

<https://docs.microsoft.com/pt-br/azure/sql-data-warehouse/performance-tiers>



Unidades de Data Warehouse

- DWU e cDWU são medidas abstrata de recursos de computação e desempenho (CPU, rede e E/S).
- Quanto mais DWUs, maior a performance do seu Data Warehouse, logo, maior o preço cobrado.

PARA MEU PROJETO, PRECISAREI DE QUANTAS DWUs?

<https://docs.microsoft.com/pt-br/azure/sql-data-warehouse/performance-tiers>



Criando um Data Warehouse no Microsoft Azure



Conectando ao Azure SQL Data Warehouse utilizando C# e ADO.NET



Conectando ao Azure SQL Data Warehouse utilizando Tableau



Limpendo Recursos

Importante para evitar
ser cobrado desnecessariamente

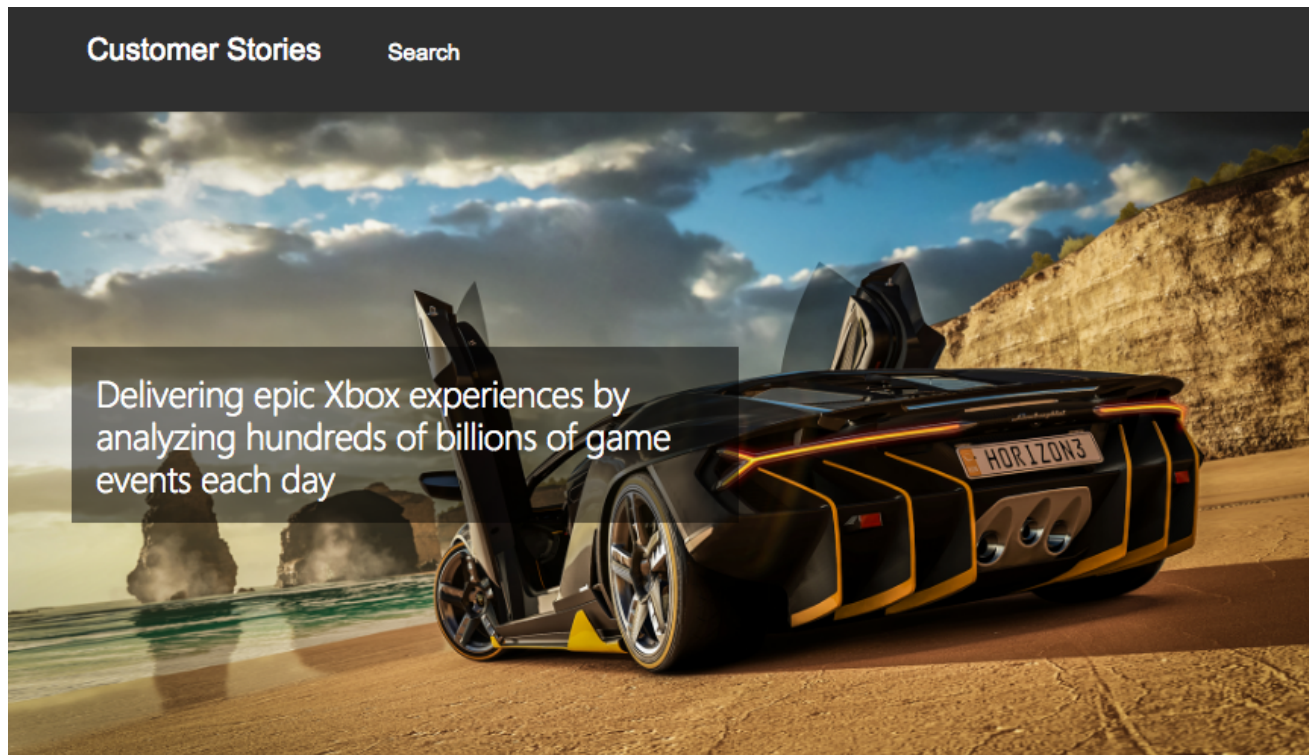


Backup e restauração

- Azure SQL Data Warehouse é um sistema distribuído:
 - Vários arquivos localizados no serviço de armazenamento do Azure.
- Serviço gerenciado e transparente
 - Backup de instantâneos locais
 - Backup geográfico
- Como restaurar um datawarehouse removido?



Estudo de Caso



<https://azure.microsoft.com/en-us/case-studies/?service=sql-data-warehouse>



FIM

Analytics, Visualização, Relatórios e Tomada de Decisões com Big Data

