

Raiane Lins
Sidney Cirqueira

Carregando dados para Azure SQL Data Warehouse com Datafactory e Polybase

#bsbtechconf

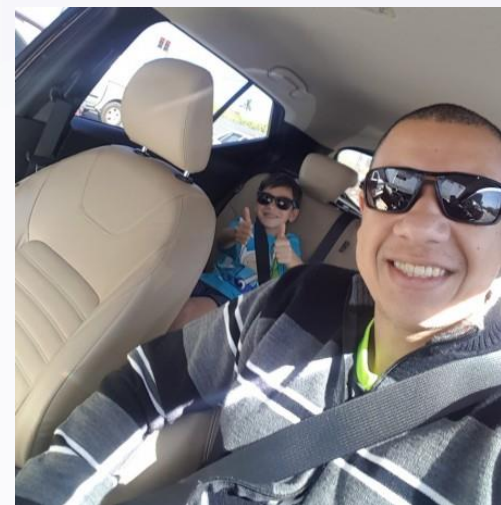
Apresentação

Raiane Lins



Há 7 anos trabalhando com Banco de dados;
DBA na empresa SICOOB Confederação;
Colaboradora do Blog Comunidade SQL Server

Sidney Cirqueira



Há 5 anos trabalhando com Banco de Dados e há 1 ano focado em Engenharia de Dados no Azure;
Analista de Banco de Dados na Wiz Soluções;
Comunidades: SQLServerDF, BrasiliaDataGroup e BSB IA, Data Science, BigData & Analytics.

Agenda

Azure SQL Data Warehouse;

ELT (Extract, Load and Transform);

Polybase;

Datafactory;

Demo: Carregar informações das fontes do SQL Server e do Blob Storage para o SQL Data Warehouse.

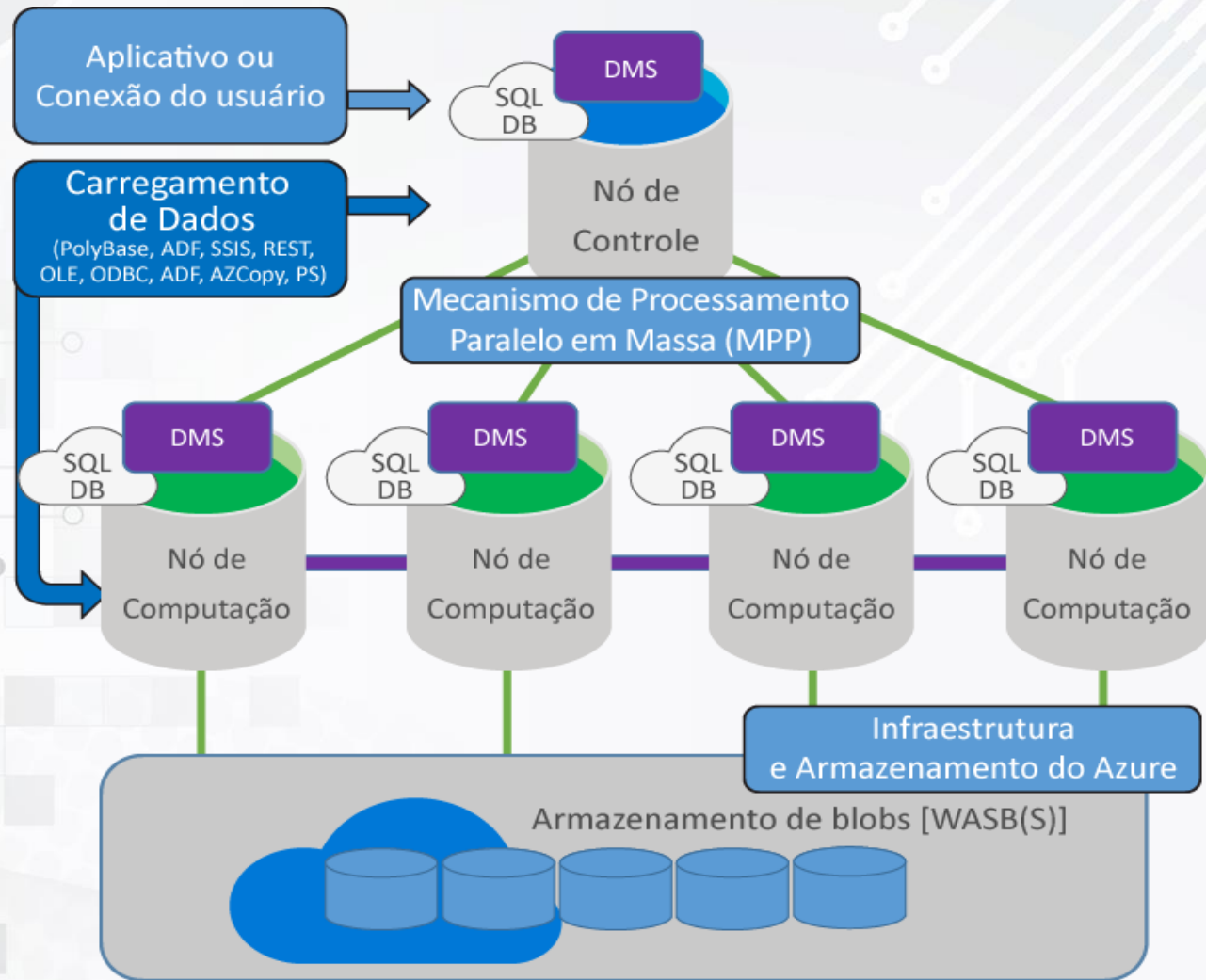


Azure SQL Data Warehouse



Azure SQL Data Warehouse

Disponível no azure no formato de plataforma como serviço (PaaS).

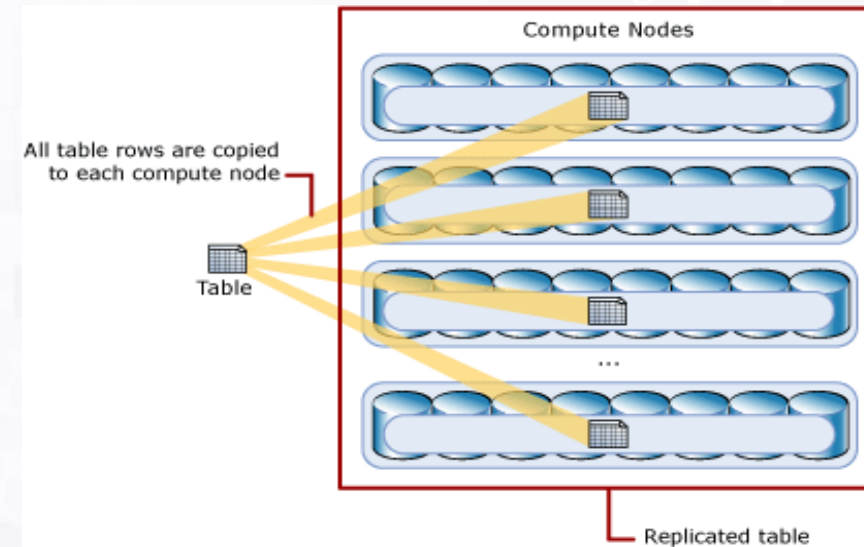
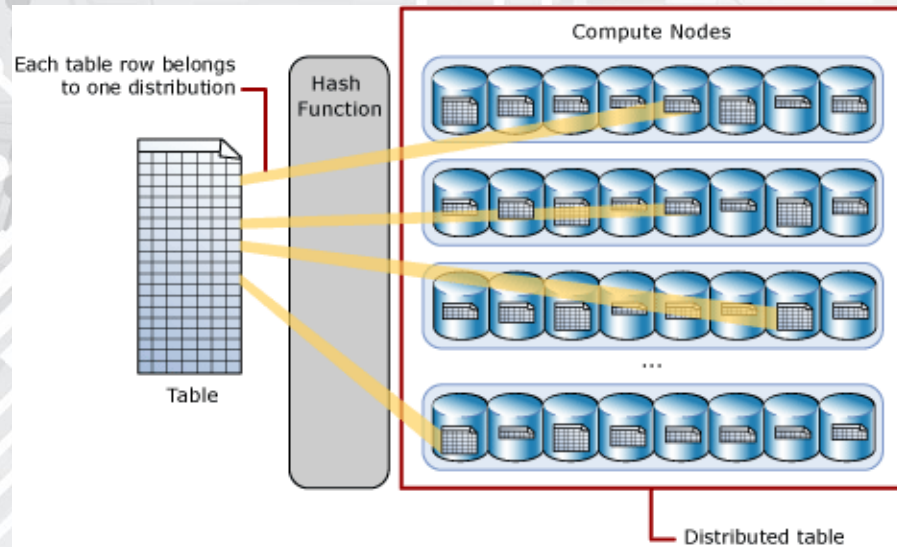


Azure Storage

Hash – Melhor desempenho para consulta e agregações.

Round Robin – Melhor desempenho para tabelas de carga.

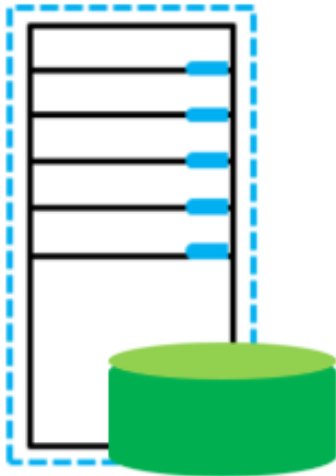
Replicate – Melhor desempenho de consulta para tabelas pequenas.



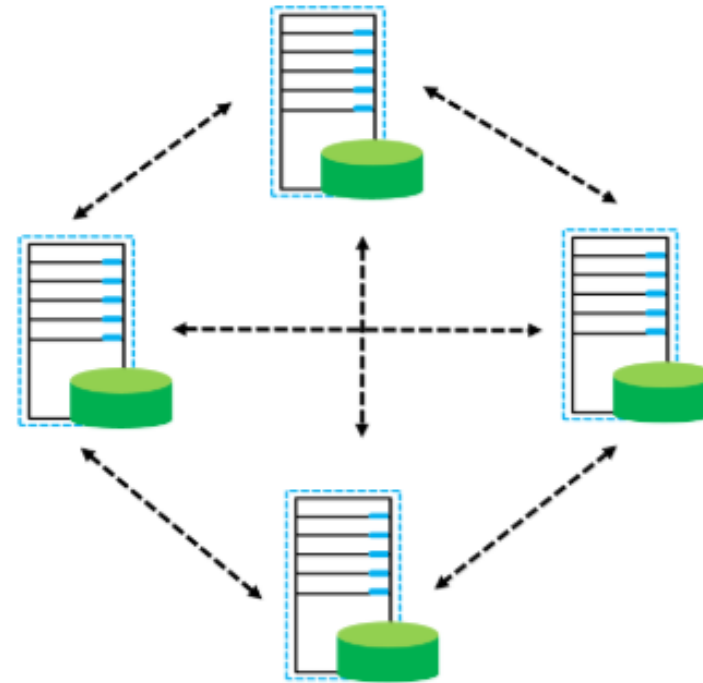
DW convencional x Azure SQL Data Warehouse

SMP vs. MPP

Symmetric Multiprocessing

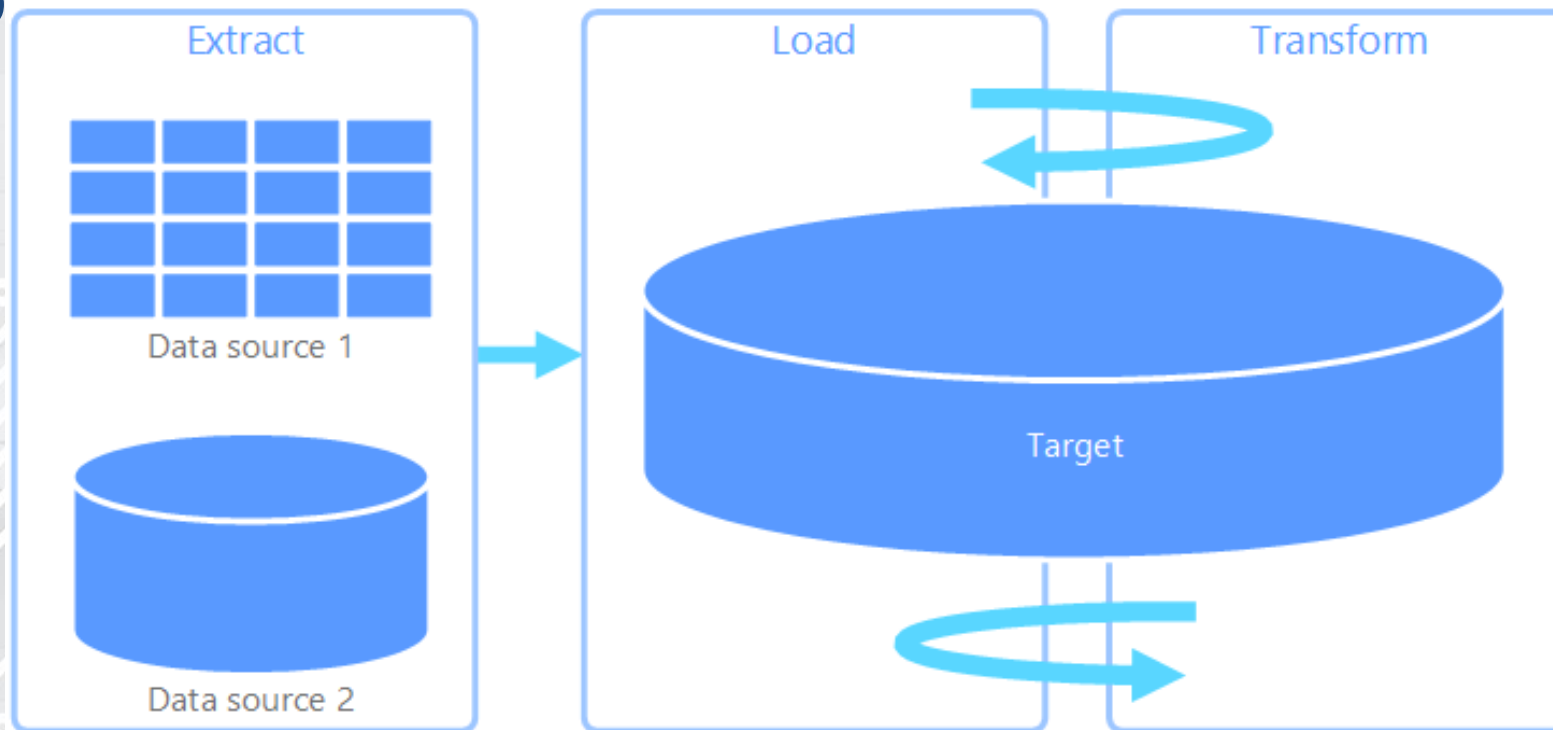


Massively Parallel Processing



ELT com Azure SQL DW e Polybase

- É o processo pelo qual os dados são carregados de um Sistema de origem para um Data Warehouse de destino



O que é o Polybase?

É uma tecnologia que acessa dados fora da base de dados do SQL Server via T-SQL;

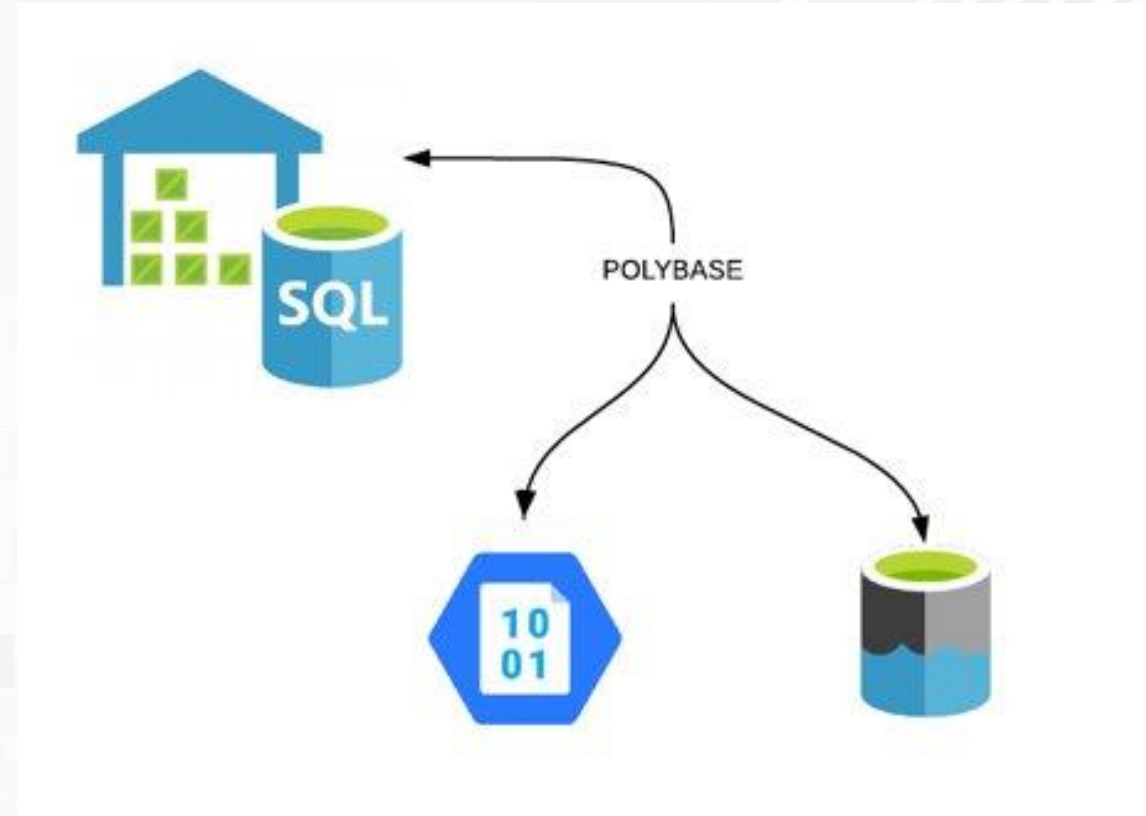
A partir do SQL Server 2016 isso permite você rodar consultas em dados externos no Hadoop ou importar/exportar dados de um Azure Blob Storage;

No SQL Server 2019 CTP 2.0 é possível acessar dados externos no SQL Server, Oracle, Teradata e MongoDB;

Opções para carregamento de dados: TSQL, SSIS, DataFactory e Databricks.

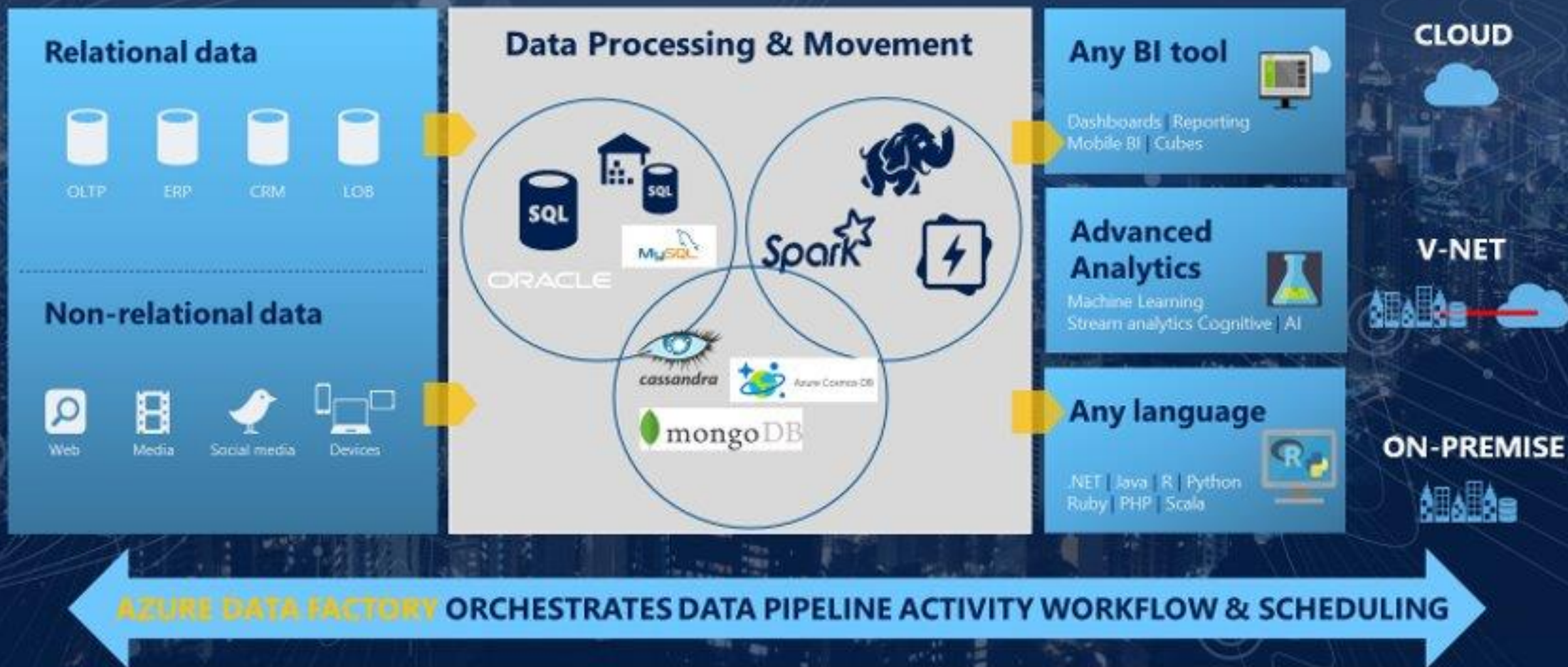
Integração Polybase com Azure

Permite que o SQL Data Warehouse importar e exportar dados do Azure Blob Storage e Azure Data Lake.





HYBRID DATA INTEGRATION AT SCALE



Azure Data Factory

Let's get started



Create pipeline



Copy Data



Configure SSIS
Integration Runtime



Set up Code
Repository

Demo:



Referências:

Criar conta gratuita <https://azure.microsoft.com/en-us/free/>

Documentação SQLDW: <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/>

Arquitetura MPP: <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/massively-parallel-processing-mpp-architecture>

Treinamento Azure SQLDW: <https://docs.microsoft.com/en-us/learn/paths/implement-sql-data-warehouse/>

Documentação Polybase: <https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide?view=sql-server-2017>

Documentação Datafactory: <https://docs.microsoft.com/en-us/azure/data-factory/>

Forum Datafactory: <https://feedback.azure.com/forums/270578-data-factory>

Rayane Lins

<https://www.linkedin.com/in/raiane-flores-borba-lins-07567463/>

nanelp@gmail.com

Sidney Cirqueira

<https://github.com/sidneycirqueira/brasiliatechconf2019>

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

sidneycirqueira@gmail.com

Obrigado!