

# DATA SCIENCE







Por que  
aprender SQL?

A linguagem também pode ser  
utilizada com Big Data (Hadoop)








Por que  
SQL Server?



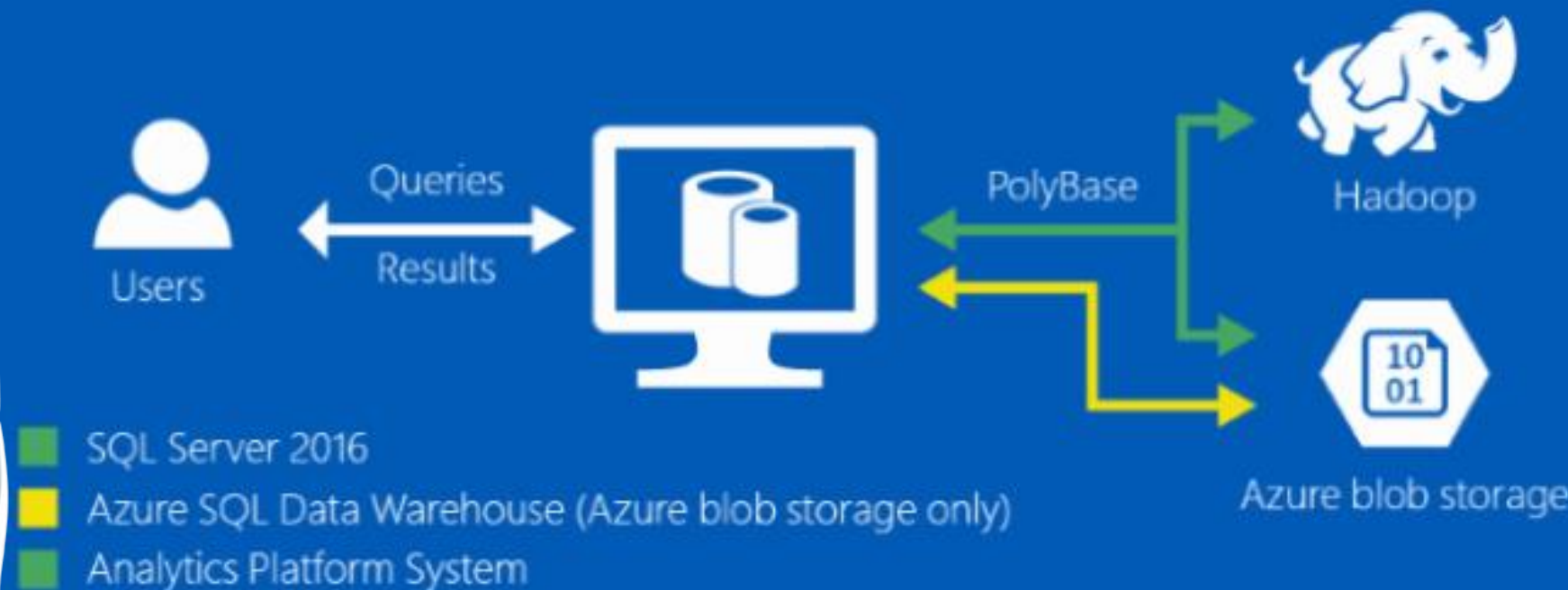


Polybase





O PolyBase é uma tecnologia que acessa dados fora do banco de dados por meio da linguagem TSQL. No SQL Server 2016, ela permite executar consultas em dados externos no Hadoop ou importar/exportar dados do Armazenamento de Blobs do Azure. As consultas são otimizadas para o Hadoop.







**Consultar dados armazenados no Hadoop do SQL Server ou do PDW.** Os usuários estão armazenando dados em sistemas escalonáveis e distribuídos mais econômicos, como o Hadoop. O PolyBase facilita a consulta dos dados com T-SQL.

**Consulte os dados armazenados no Armazenamento de Blobs do Azure.** O armazenamento de blobs do Azure é um local conveniente para armazenar dados para uso dos serviços do Azure. O PolyBase facilita o acesso aos dados com T-SQL.

**Importar dados do Hadoop, do Armazenamento de Blobs do Azure ou do Azure Data Lake Store** Aproveite a velocidade das funcionalidades de análise e tecnologia de columnstore do Microsoft SQL importando dados do Hadoop, do Armazenamento de Blobs do Azure ou do Azure Data Lake Store para tabelas relacionais. Não é necessário ter uma ferramenta separada de ETL ou importação.

**Exporte dados para o Hadoop, o Armazenamento de Blobs do Azure ou o Azure Data Lake Store.** Archive dados no Hadoop, no Armazenamento de Blobs do Azure ou no Azure Data Lake Store para obter um armazenamento econômico e mantê-los online para fácil acesso.

**Integre com ferramentas de BI.** Use o PolyBase com a business intelligence e pilha de análise da Microsoft, ou use as ferramentas de terceiros compatíveis com o SQL Server.





Além de ser um SGBD extremamente amigável, o SQL Server 2017 já é preparado para utilizar ferramentas de Data Science nativamente, como o R e a linguagem Python. Além disso, podemos tirar proveito das ferramentas de BI da suíte SQL Server.





Integration Services –  
Ferramenta de ETL, responsável  
pela extração, transformação e  
carga dos dados.

Analysis Services – Ferramenta  
de análise de DW através de  
Cubos OLAP.

Reporting Services – Disponibiliza  
relatórios via web.



# DATA SCIENCE



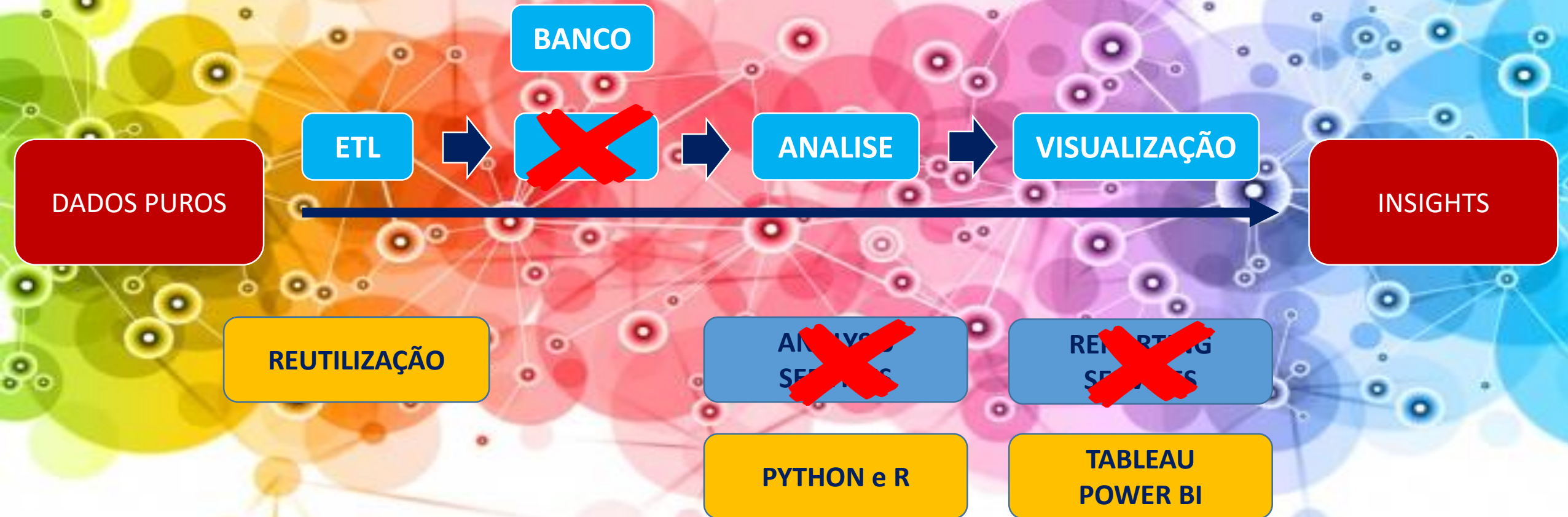




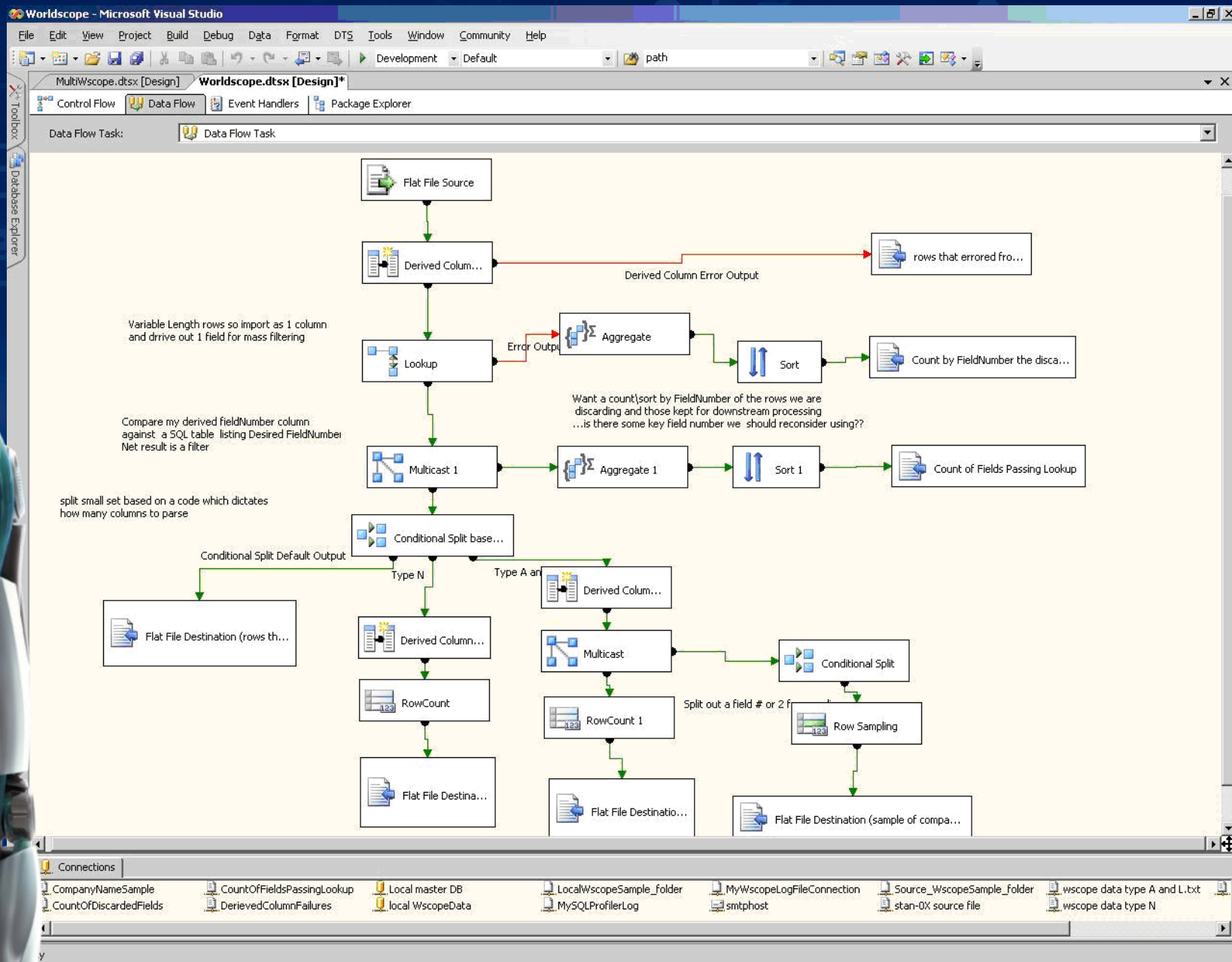
# Ciclo de Vida de B.I



# O Ciclo de Vida de Business Intelligence









# Ponto de Vista de Análise



BI TRADICIONAL

DATA SCIENCE

DESCRITIVO

O que aconteceu?

DIAGNÓSTICO

Por que aconteceu?

PREDITIVO

O que vai acontecer?

PRESCRITIVO

O que posso fazer?



Outras Ferramentas

Power Center

Oracle ODI

Talend

PDI (Spoon)







# LINKS PARA DOWNLOAD

<https://www.microsoft.com/pt-br/evalcenter/evaluate-windows-server-2016/>

<https://www.microsoft.com/en-us/sql-server/sql-server-2017>

<https://docs.microsoft.com/pt-br/sql/ssdt/download-sql-server-data-tools-ssdt>

<http://www.oracle.com/technetwork/pt/java/javase/downloads/jre8-downloads-2133155.html>

<https://docs.microsoft.com/pt-br/sql/ssms/download-sql-server-management-studio-ssms>

<https://www.visualstudio.com/thank-you-downloading-visual-studio/?sku=Community&rel=15#>

<https://imagine.microsoft.com/en-us/Catalog/Product/101>



Área	Analista de BI	Cientista de Dados
<b>Foco</b>	Relatórios, KPI's, Tendências	Padrões, Correlações, Modelos Preditivos
<b>Processo</b>	Estático, Comparativo	Exploratório, Experimental, Visual
<b>Fontes de Dados</b>	Data Warehouses, Bancos Transacionais	Big Data, Dados Não-Estruturados, Bancos Transacionais e NoSQL, Dados Gerados em Tempo Real
<b>Qualidade dos Dados na Fonte</b>	Alta	Baixa ou Média (requer processo de limpeza e transformação)
<b>Modelo de Dados</b>	Esquema de dados bem definido na fonte	Esquema de dados definido no momento da consulta
<b>Transformações nos Dados</b>	Pouca ou nenhuma (dados já organizados na fonte)	Transformação sob demanda, necessidade de complementar os dados
<b>Análise</b>	Descritiva, Retrospectiva	Preditiva, Prescritiva
<b>Responde à pergunta:</b>	O que aconteceu?	O que pode acontecer?