



Luis Beltrán



# SQL Server Machine Learning Services with Python and R

Prague 2018

# Luis Beltrán



- Microsoft MVP in Developer Technologies
- Researcher & PhD student at Tomas Bata University in Zlín
- Lecturer at Tecnológico Nacional de México in Celaya



 @darkicebeam

 luis@luisbeltran.mx

 **Tomas Bata University in Zlín**  
Faculty of Applied Informatics



TECNOLÓGICO  
NACIONAL DE MÉXICO





# Agenda

# Agenda

- Machine Learning
- Why ML in SQL Server?
- SQL Server Machine Learning Services
- Demo





# Machine Learning

# What is Machine Learning?

Algorithms + Mathematical models →  
Computer systems to become **smarter  
with experience**.

“Experience” = past data + human input.

It uses data mining & learning techniques  
to **train** models using **historical data** to  
**predict future** outcomes.



# Past Data

- **Training Data:** A set of samples.
- **Sample:** A single observation.
- **Features:** Individual columns in our data set.
- **Label:** Historical results related to a set of samples.

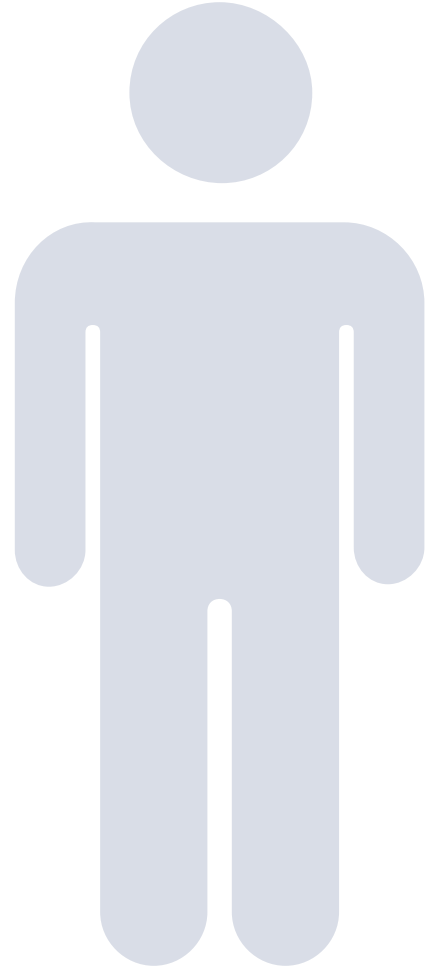




# Human Input

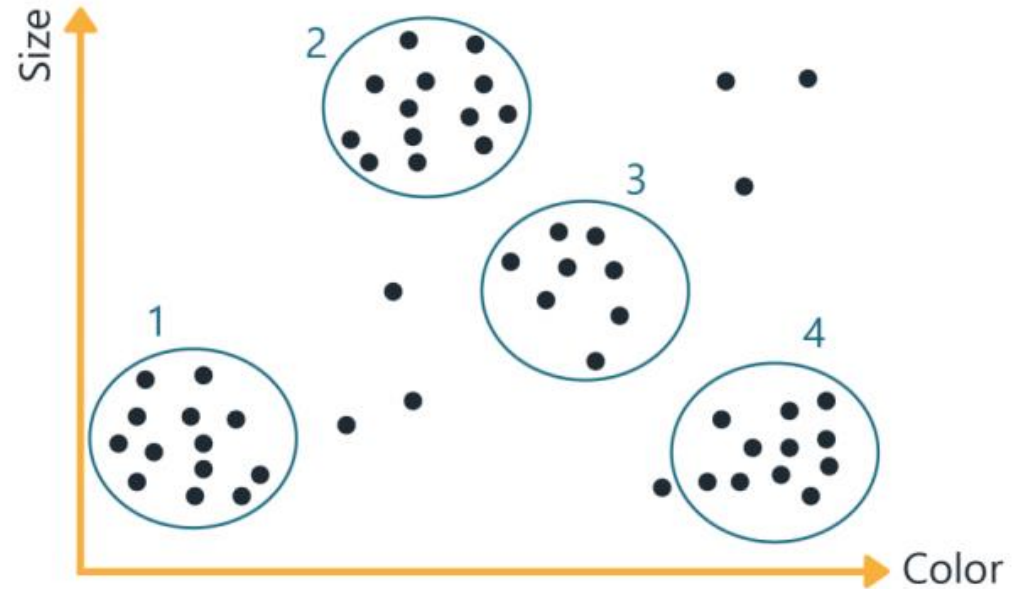
Domain of the Data Scientist and Subject Matter Expert to **determine if the data meets these criteria:**

- ✓ Relevant
- ✓ Connected
- ✓ Accurate
- ✓ Enough of it



# Machine Learning Problems

- Prediction
- Classification
- Anomaly Detection
- Clustering



# Examples

- Predict store sales based on historical performance.
- Predict default of a loan based on transaction history.
- Predict the sentiment of a new review.
- Classify customers into groups based on transaction characteristics.
- Classify images and extract features from them.



A large, teal-colored abstract graphic on the left side of the slide. It consists of several thick, curved lines that sweep from the top left towards the bottom right, creating a sense of motion and flow.

# **Why ML in SQL Server?**

# Why Machine Learning in SQL Server?

Eliminate data movement

Leverage database security

Push ML compute to the database

Operationalize ML scripts and models

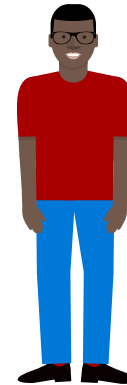
Calling familiar T-SQL stored procedures

Model management in SQL Server

Enterprise grade performance and scale

Scale your R and Python analytics with multi-threading and parallel processing

SQL Server security, compliance, resource governance, query performance, always on secondaries



**Data Scientist**

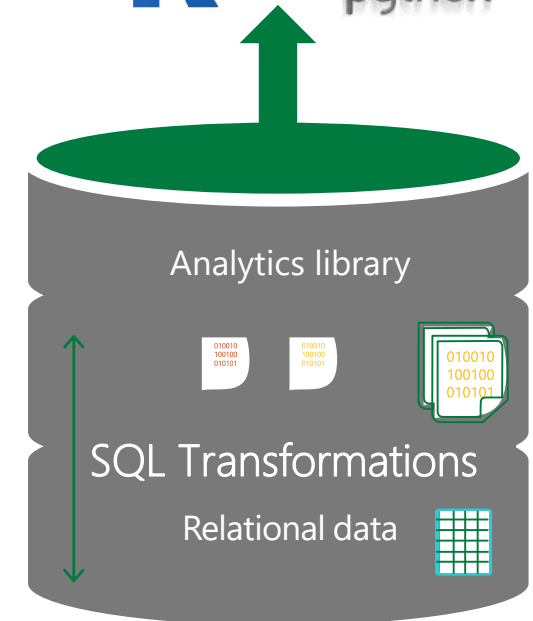
Interacts directly with data



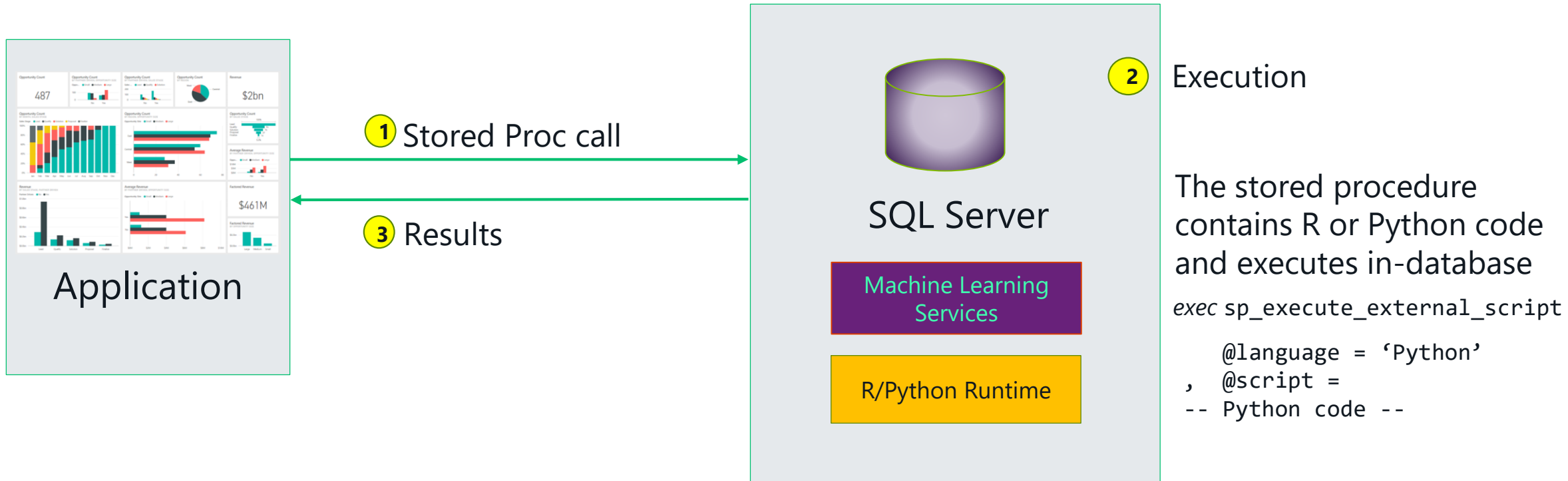
**SQL Developer/  
DBA**

Manage data and analytics together

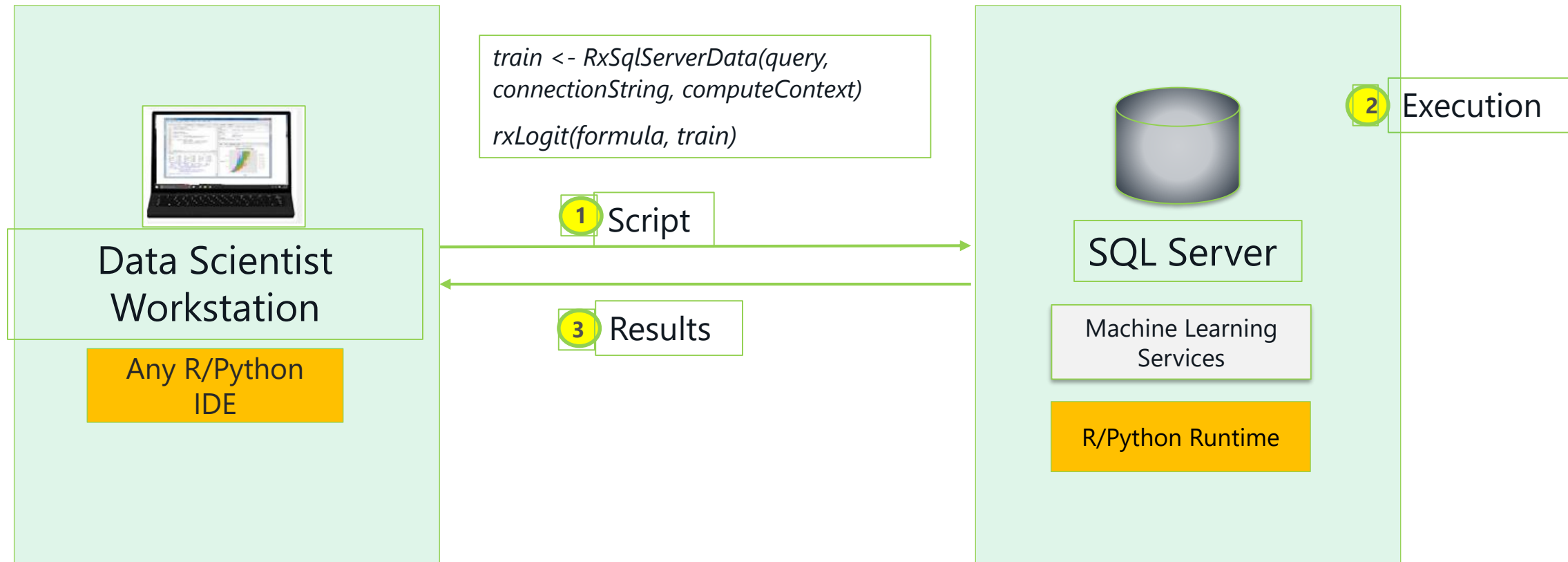
R integration      Python integration



# Developer's Model Operationalization



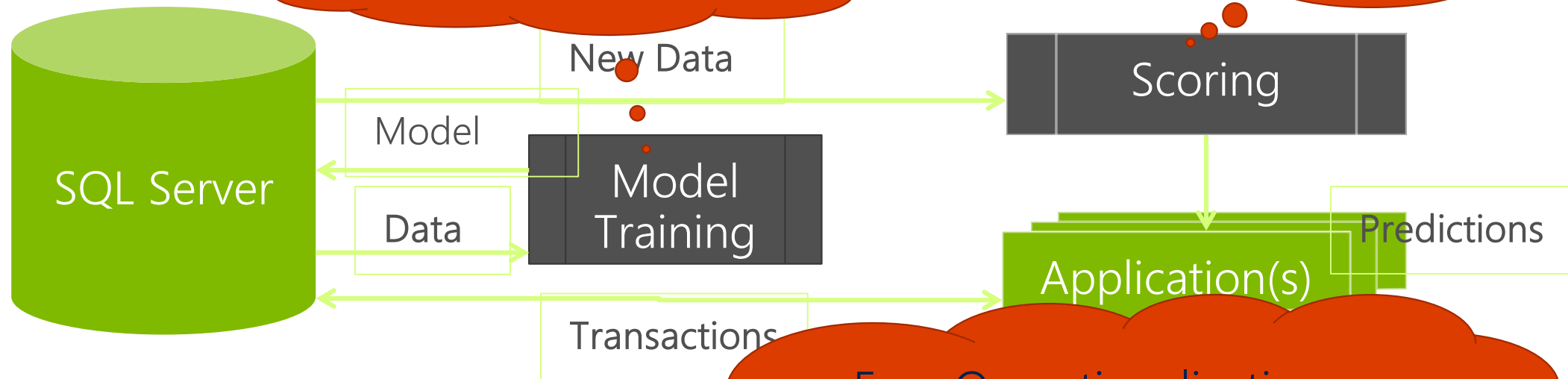
# Data Scientists' Data Exploration and Development



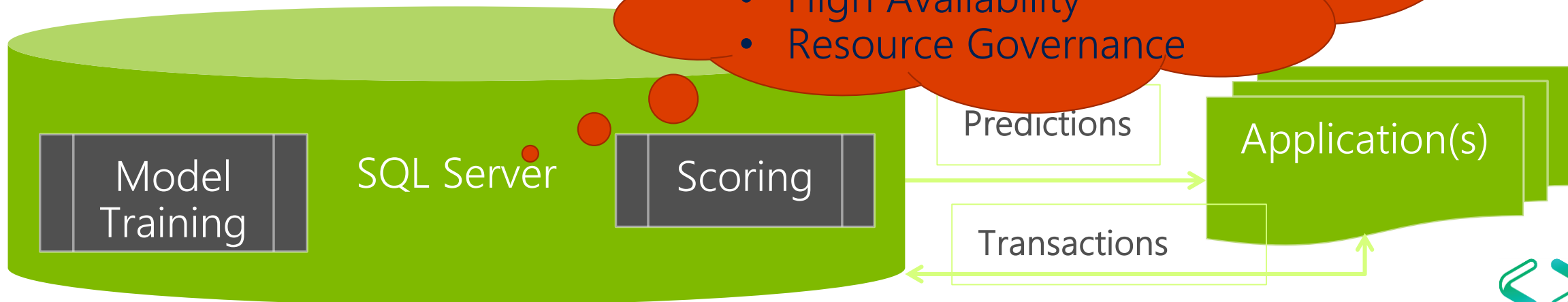
# Machin

Analytic Server

Separate Service or  
Embedded Logic



- Easy Operationalization
- Performance
- High Availability
- Resource Governance





A large, stylized teal graphic on the left side of the slide, resembling a thick, curved arrow or a stylized letter 'R' pointing towards the right.

# SQL Server Machine Learning Services

# SQL Server Machine Learning Services

## SQL Server 2017

- Batch mode execution support for input data
- **R** support (3.3.3 version)
- **Python** support (Anaconda 3.5.2)
- Native scoring using **PREDICT** function (+Linux support)



# R

Step 1.1 Install SQL Server with in-database R / Machine Learning Services

Step 1.2 Install SQL Server Management Studio (SSMS)

Step 1.3 Enable external script execution

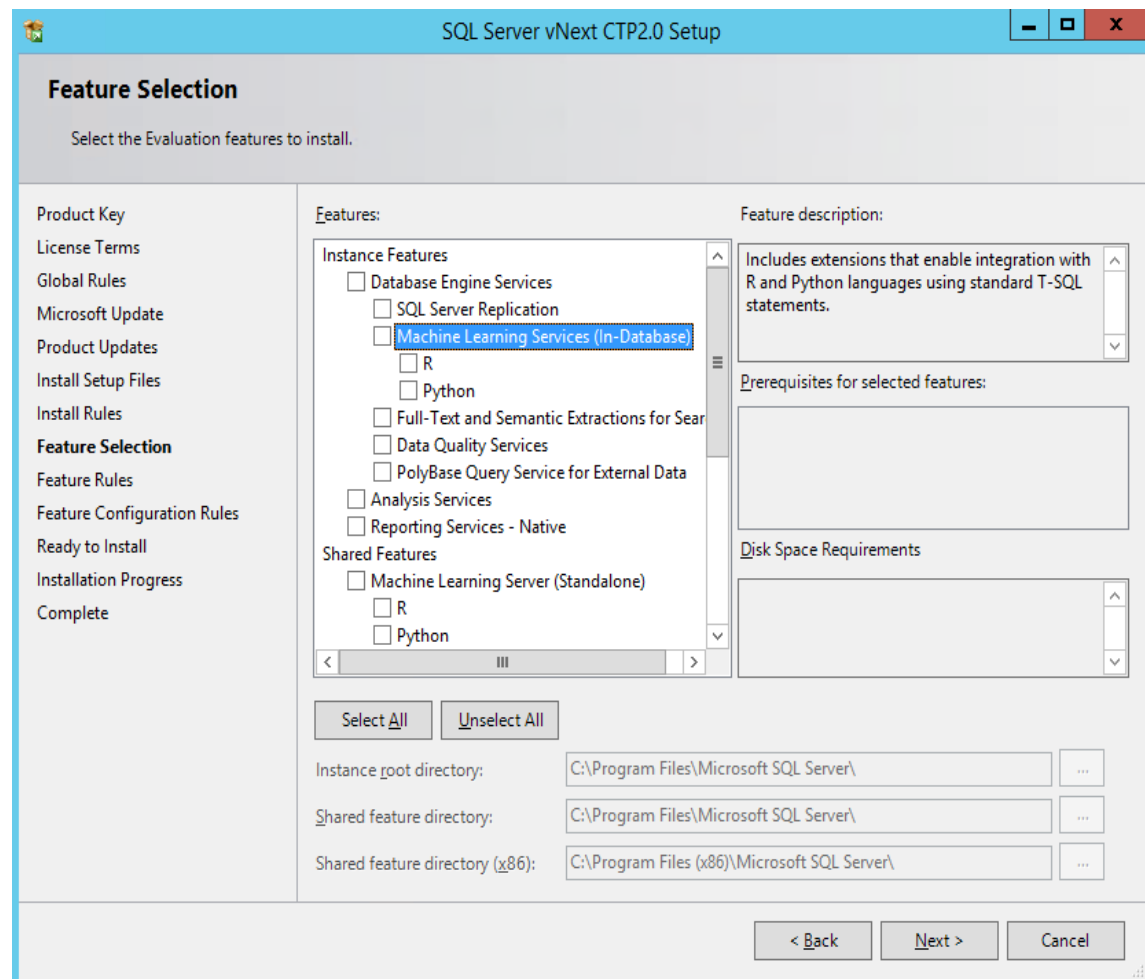
```
EXEC sp_configure 'external scripts enabled', 1;  
RECONFIGURE WITH OVERRIDE
```

**restart your SQL Server Instance after the configuration!**

Install and configure your R development environment

[Install Microsoft R Client](#)

Set up RStudio for R Client



# Python

Step 1.1 Install SQL Server with in-database R / Machine Learning Services

Step 1.2 Install SQL Server Management Studio (SSMS)

Step 1.3 Enable external script execution

```
EXEC sp_configure 'external scripts enabled', 1;  
RECONFIGURE WITH OVERRIDE
```

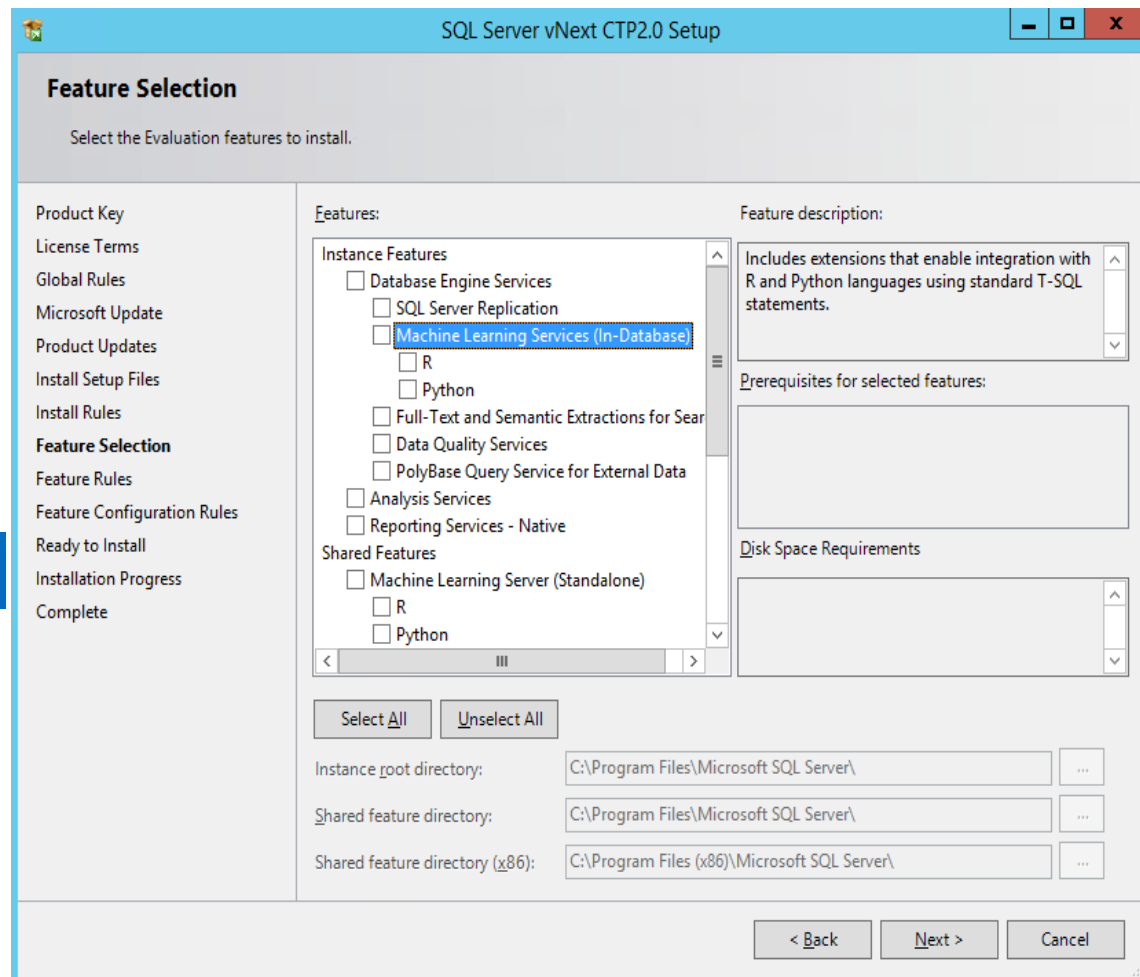
Step 1.4 Install and configure your Python development environment

install a Python IDE.

\*VS Code ([download](#)) with the [Python Extension](#) and the [mssql extension](#)

Step 1.5 Install remote Python client libraries

[install Python client libraries](#)



# Reference

Package	Language	Description
<a href="#">RevoScaleR</a>	R	Distributed and parallel processing for R tasks: data transformation, exploration, visualization, statistical and predictive analytics.
<a href="#">MicrosoftML</a>	R	Functions based on Microsoft's AI algorithms, adapted for R.
<a href="#">olapR</a>	R	Imports data from OLAP cube.s
<a href="#">sqlRUtils</a>	R	Helper functions for encapsulating R and T-SQL.
<a href="#">revoscalepy</a>	Python	Distributed and parallel processing for Python tasks: data transformation, exploration, visualization, statistical and predictive analytics.
<a href="#">microsoftml</a>	Python	Functions based on Microsoft's AI algorithms, adapted for Python.

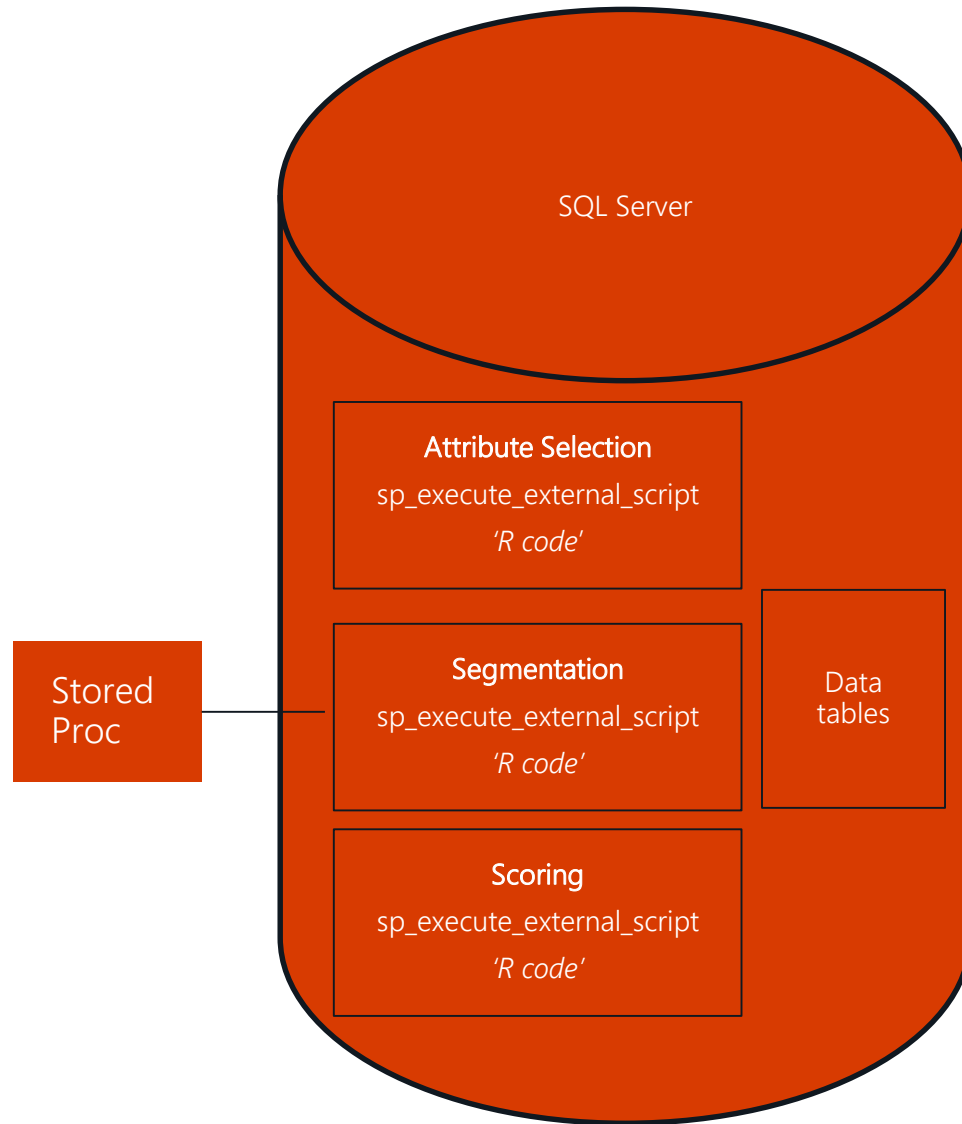


RxSqlServerData	Creates a SQL Server data source object.
rxSqlServerDropTable	Execute an SQL statement that drops a table.
rxSqlServerTableExists	Execute an SQL statement that checks for a table's existence.
rxHistogram	Creates a histogram from data.
rxLinePlot	Creates a line plot from data.
rxLorenz	Computes a Lorenz curve which can be plotted.
rxRocCurve	Computes and plots ROC curves from actual and predicted data.
rxPredict *	Calculates predictions for fitted models. Output must be an XDF data source.
rxKmeans *	Performs k-means clustering.
rxNaiveBayes *	Performs Naive Bayes classification.



RxInSqlServer		Creates a compute context for running revoscalepy analyses inside a remote Microsoft SQL Server.
RxLocalSeq		This is the default but you can call it switch back to a local compute context if your script runs in multiple. Computations using rx_exec will be processed sequentially.
rx_get_compute_context		Returns the current compute context.
rx_set_compute_context		Change the compute context to a different one.
RxSqlServerData		Local, RxInSqlServer Generates a data source object from a SQL table or query.
rx_import	All	Import data into an .xdf file or data frame.
rx_data_step	All	Transform data from an input data set to an output data set.
rx_predict_default	All	Compute predicted values and residuals using rx_lin_mod and rx_logit objects.
rx_predict_rx_dforest	All	Calculate predicted or fitted values for a data set from an rx_dforest or rx_btrees object.







A large, teal-colored abstract graphic on the left side of the slide, consisting of several thick, curved lines that form a stylized, open shape.

# Demo

<https://github.com/icebeam7/SQLSaturdayPrague2018>

# Recommendations

## Don't

- × Run R / Python script as-is
- × Embed secrets in scripts
- × Do data transformations that can be achieved in SQL
- × Access network resources
- × Process/transform files as part of the stored procedure call
- × Embed the R/Python code directly in applications

## Do

- ✓ Develop/Test from RTVS, PTVS, RStudio or other IDE
- ✓ SQL Compute Context from client
- ✓ Data processing & transformations in SQL Server
- ✓ Data integration using SQL Server features
- ✓ Model management in database



# Applications

- Real estate property insights

Take advantage of a host of new capabilities, including machine learning and predictive analytics.

- Targeted marketing campaigns

With R Services, it is possible to build better, more accurate models using larger data sets, which results in more precisely targeted marketing for customers.

- Build financial models

The bank has moved its credit scorecard development, arrears analysis, investment forecasting, and analysis of intermediary and broker performance to the R Server platform.



# Resources

- ❑ SQL Server Samples – [R Services](#) & [ML Services](#)
- ❑ [SSMS Reports for ML Services](#)
- ❑ [SQL Server Machine Learning Services](#) (aka.ms/mlsqldev)
  - ❑ Getting started ML tutorials
- ❑ [SQL Server Developer Tutorials](#) (aka.ms/sqldev)
  - ❑ \*\*Getting started SQL tutorials
- ❑ [Performance Patterns for SQL Server Machine Learning Services](#)



# Luis Beltrán



 @darkicebeam

 luis@luisbeltran.mx

## THANK YOU!

## Questions?



