Microsoft 365

How to leverage Machine Learning Services in SQL Server

Tejas Shah – Senior Program Manager
tejasaks@microsoft.com

Machine Learning Introduction

Predict properties of new data by learning from a sample

- Predict sales of stores in a region based on historical sales
- Predict probability of fraud on a new credit card transaction
- Predict default of a new loan based on loan / transaction history
- Predict sentiment of a new tweet or review
- Classify new image(s) based on sample images & attributes
- Classify data into groups or clusters

Popular ML technologies

- **R & Python**

Popular programming languages

- **Java**

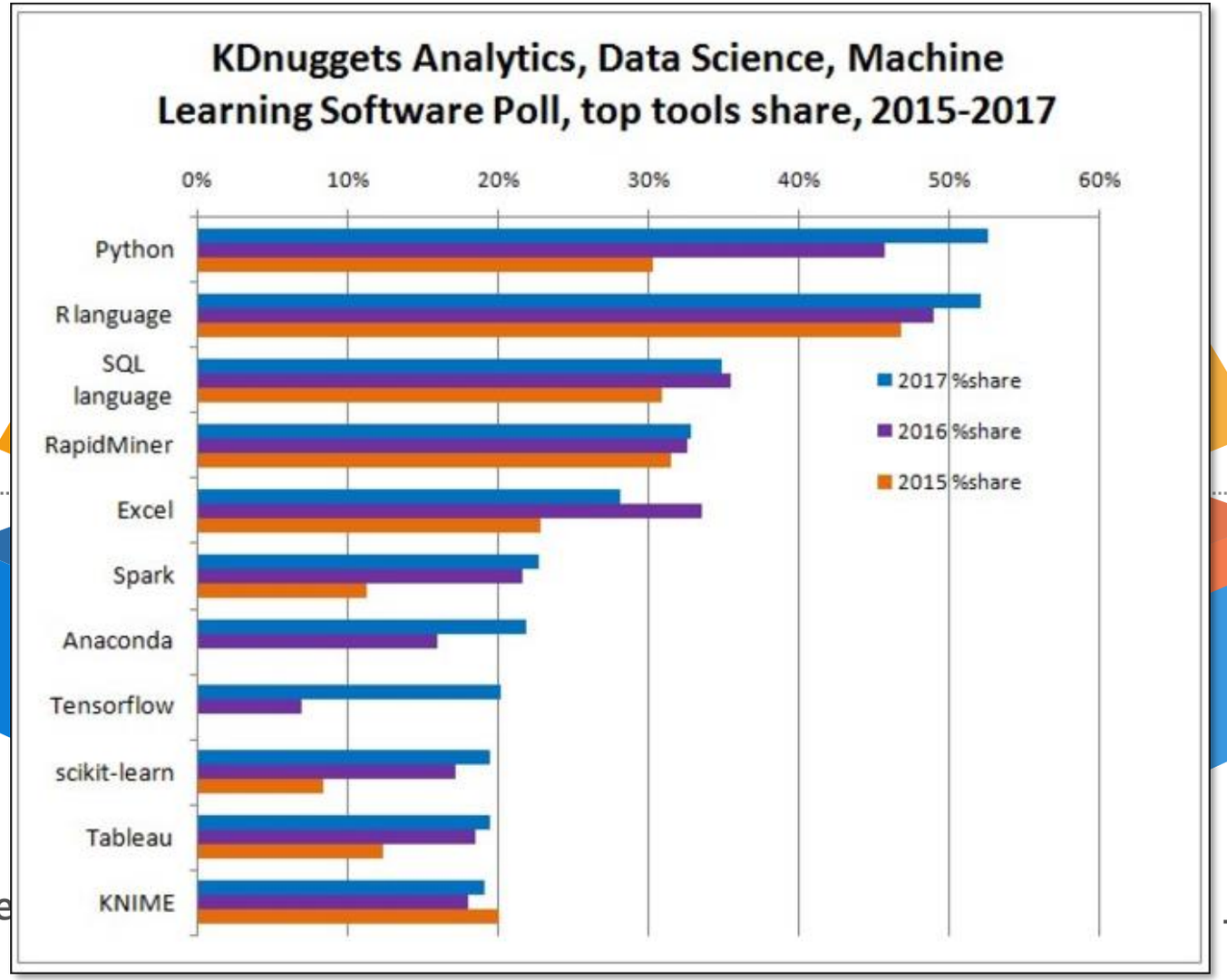
Bringing intelligence to where data lives

Application + Intelligence

Application

Database

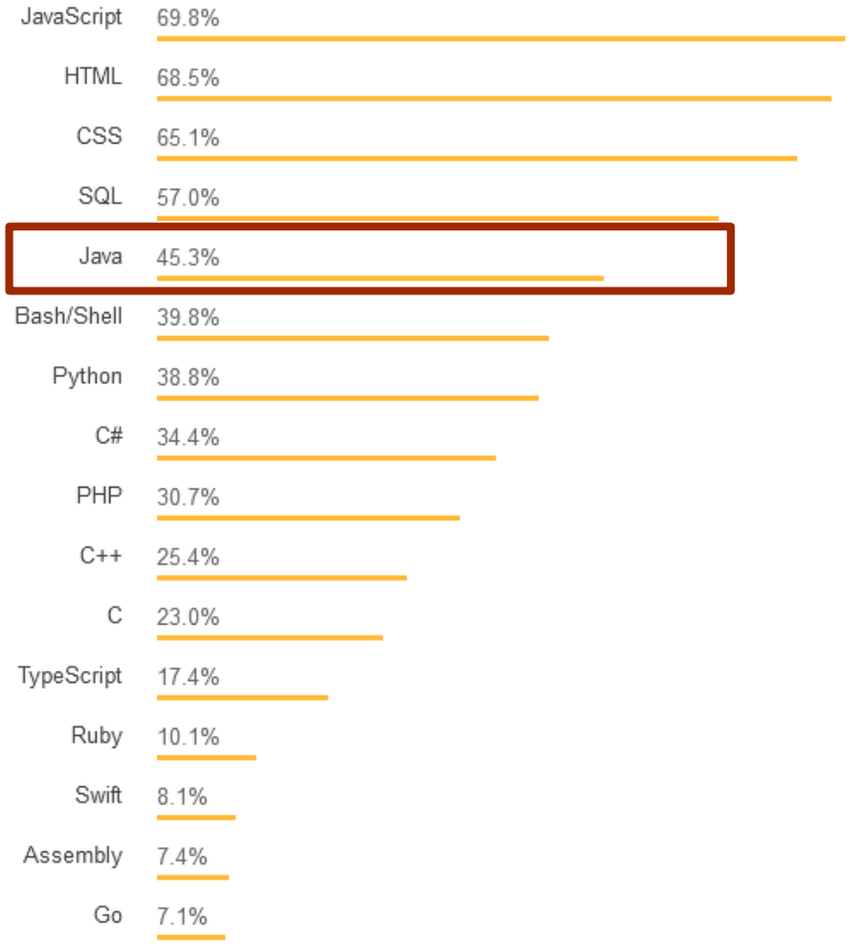
Intelligence + Database



Re

+ App

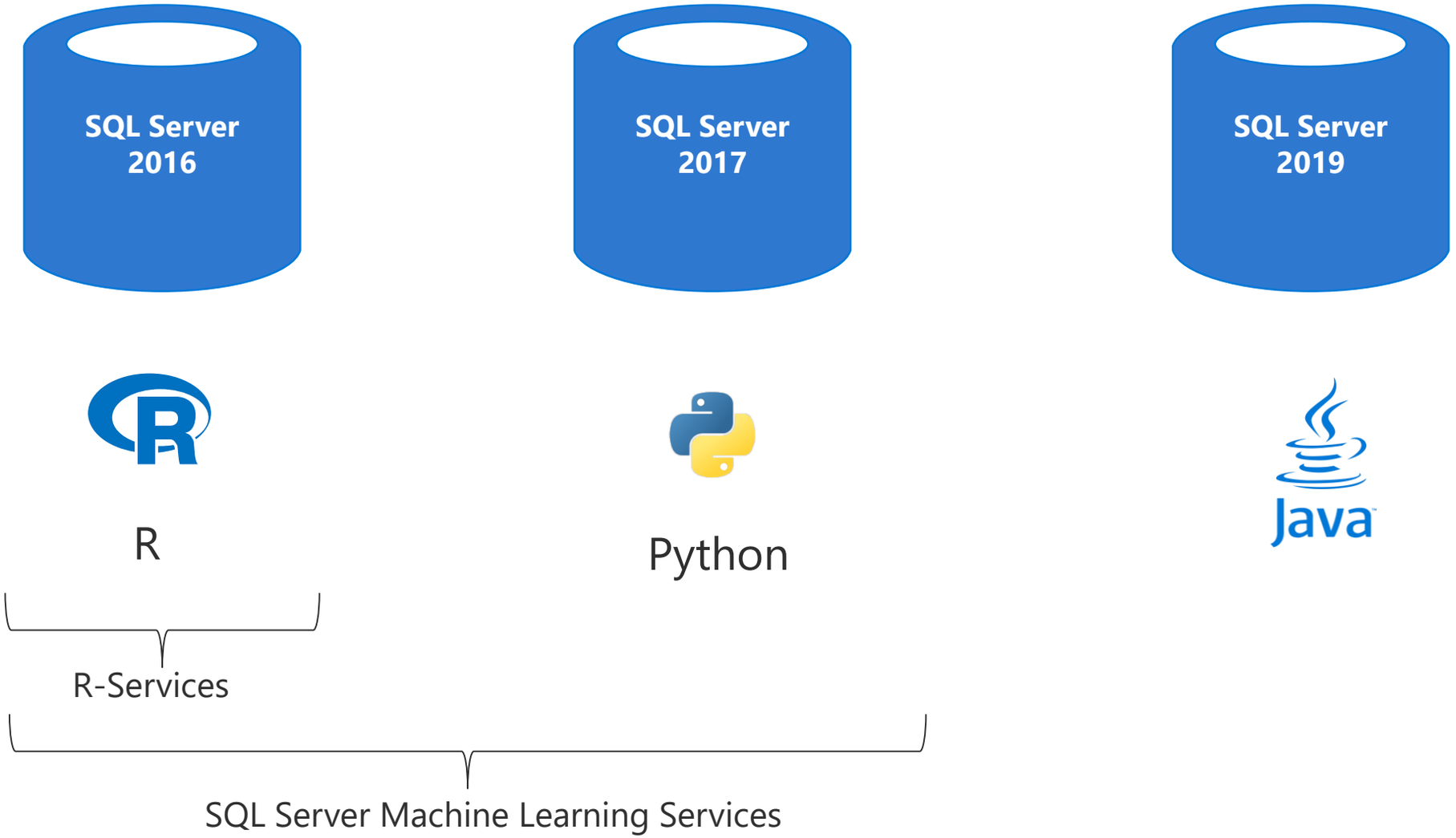
Popularity of Java



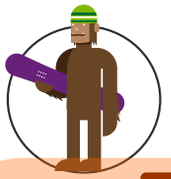
Source: Stack Overflow survey 2018

Tiobe programming language index

SQL Server external language integrations

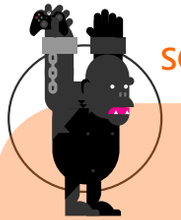


SQL Server - An end-to-end platform for Machine Learning



App Developer

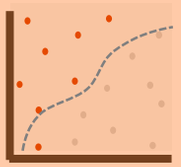
SQL Server's **rich ecosystem** integrates well with most technologies. **Share insights** and **make your apps more intelligent** by consuming machine learning from any app with a simple **simple stored proc call**



SQL Developer/DBA

Deploy and schedule training and prediction scripts in-DB and automatically make your in-DB models and predictions **accessible to any application**

Data Scientists



Working against SQL Server means that you can still **explore and experiment in your favorite IDE**. Transform data using TSQL, R or Python.

Data Scientists

Interact and collaborate directly with data

Train models from your IDE and **leverage the power of the server machine**. Or train directly in the SQL Server and **save the models and manage versions right in the DB**

Why Machine Learning in SQL Server?

Eliminate data movement

Leverage database security

Push ML compute to the database

Operationalize ML scripts and models

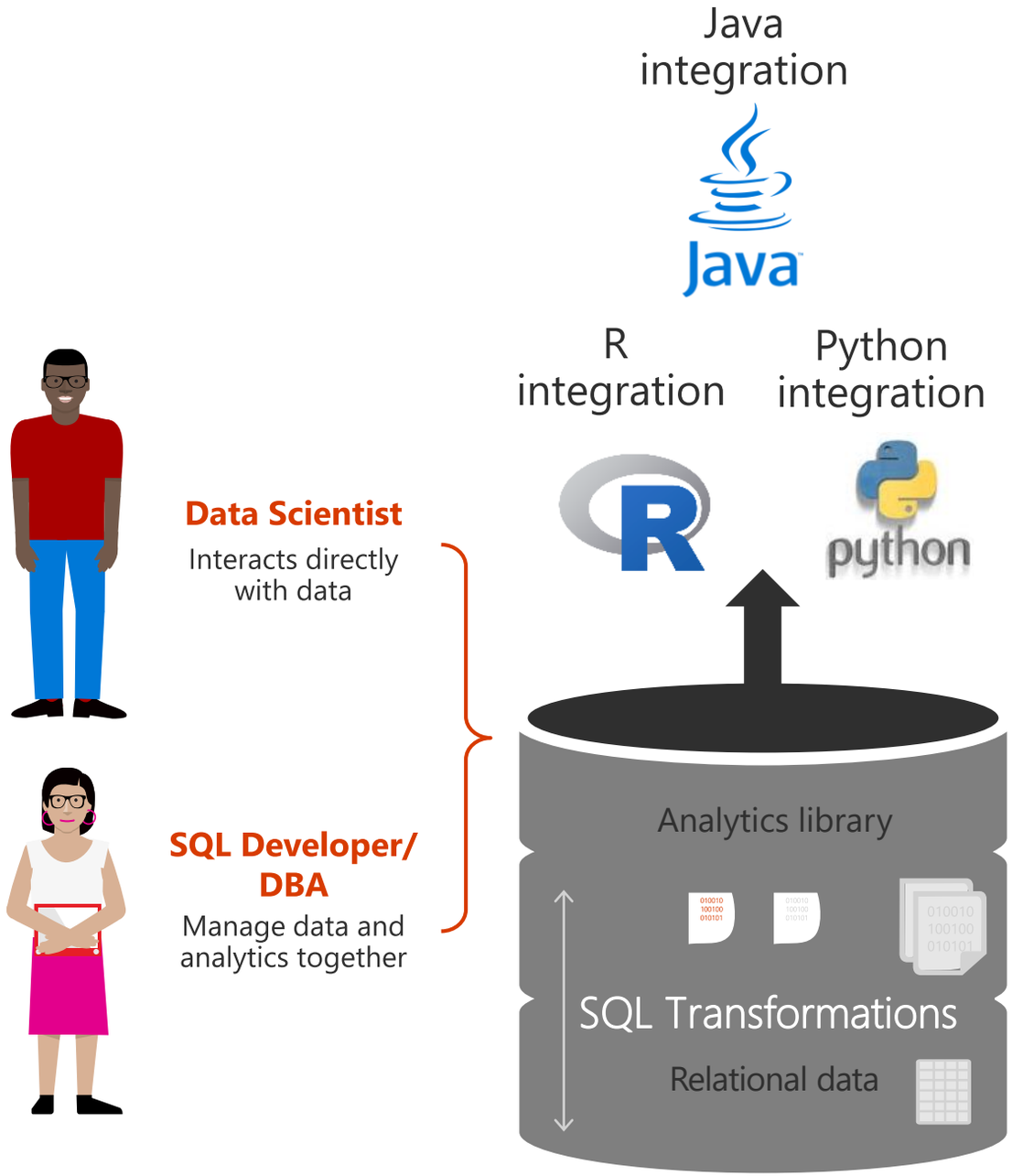
Calling familiar T-SQL stored procedures

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



Run R and Python in-db

```
1  EXEC sp_execute_external_script
2  @language =N'R',
3  @script=N'
4  OutputDataSet <- InputDataSet;
5  ',
6  @input_data_1 =N'SELECT 1 AS hello'
7  WITH RESULT SETS (([hello] int not null));
8  GO
9
```

RESULTS

	hello
1	1

```
1  EXEC sp_execute_external_script
2  @language =N'Python',
3  @script=N'
4  OutputDataSet = InputDataSet;
5  ',
6  @input_data_1 =N'SELECT 1 AS hello'
7  WITH RESULT SETS (([hello] int not null));
8  GO
9
```

RESULTS

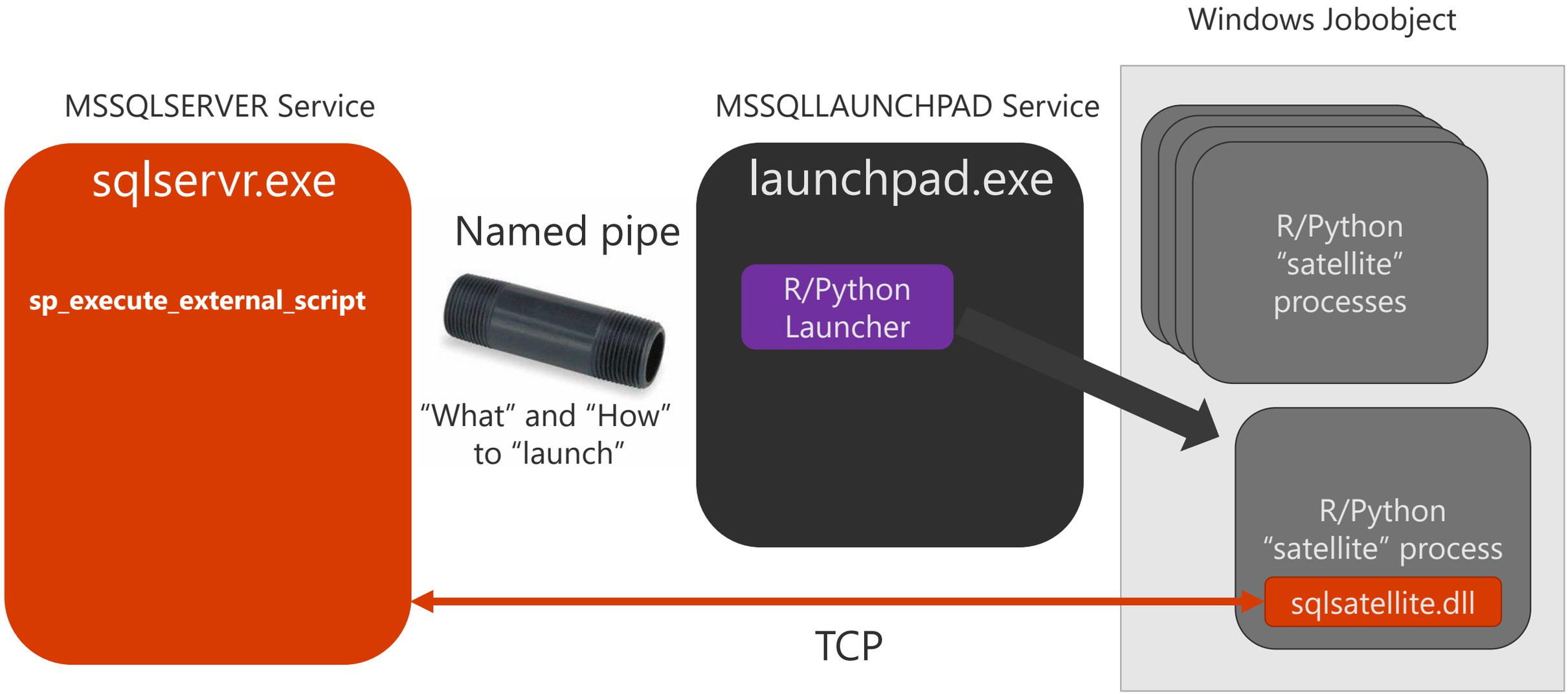
	hello
1	1

Microsoft

Example use cases for Java extensibility

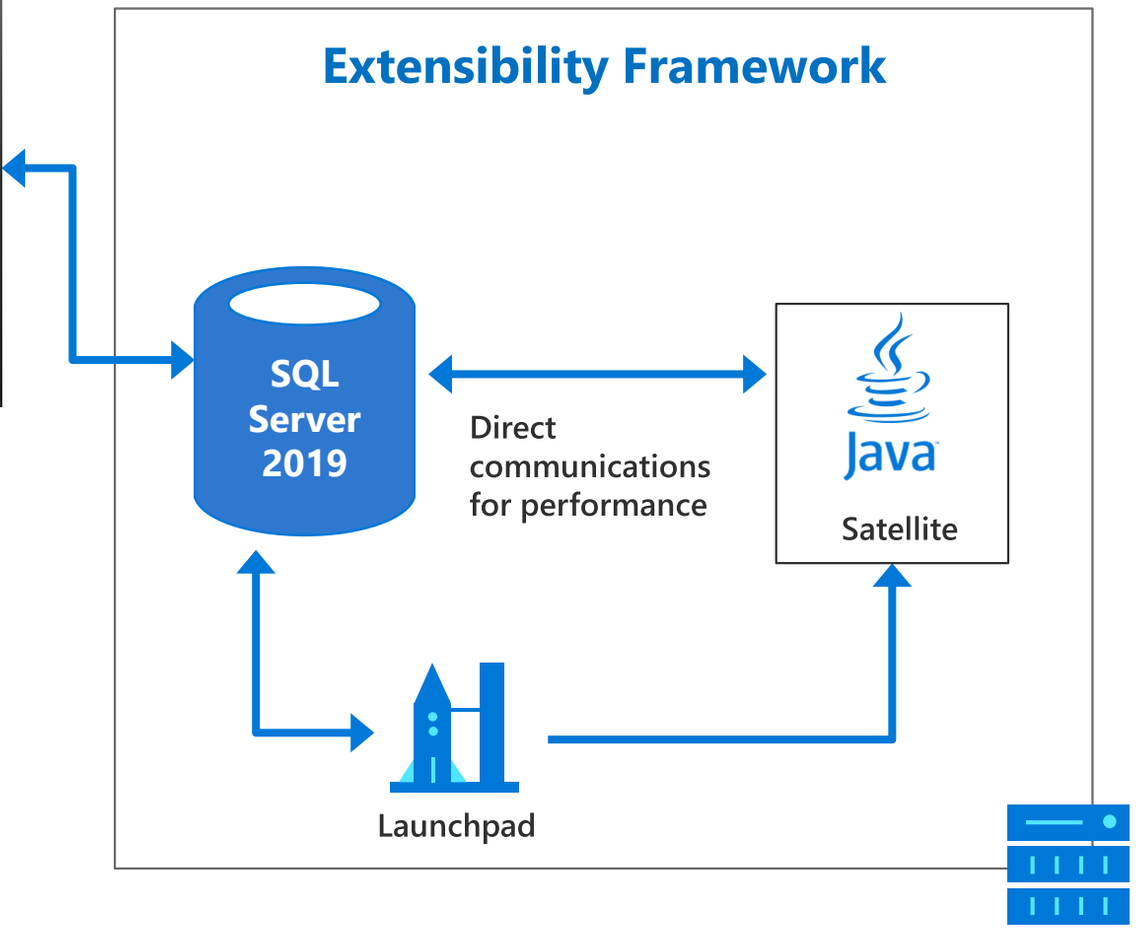
- Move middle-tier (application server) logic closer to the data
- Oracle Java Procs migration
- Natural language processing
- Regular expressions (Regex)

The SQL Extensibility Architecture – SQL 2017

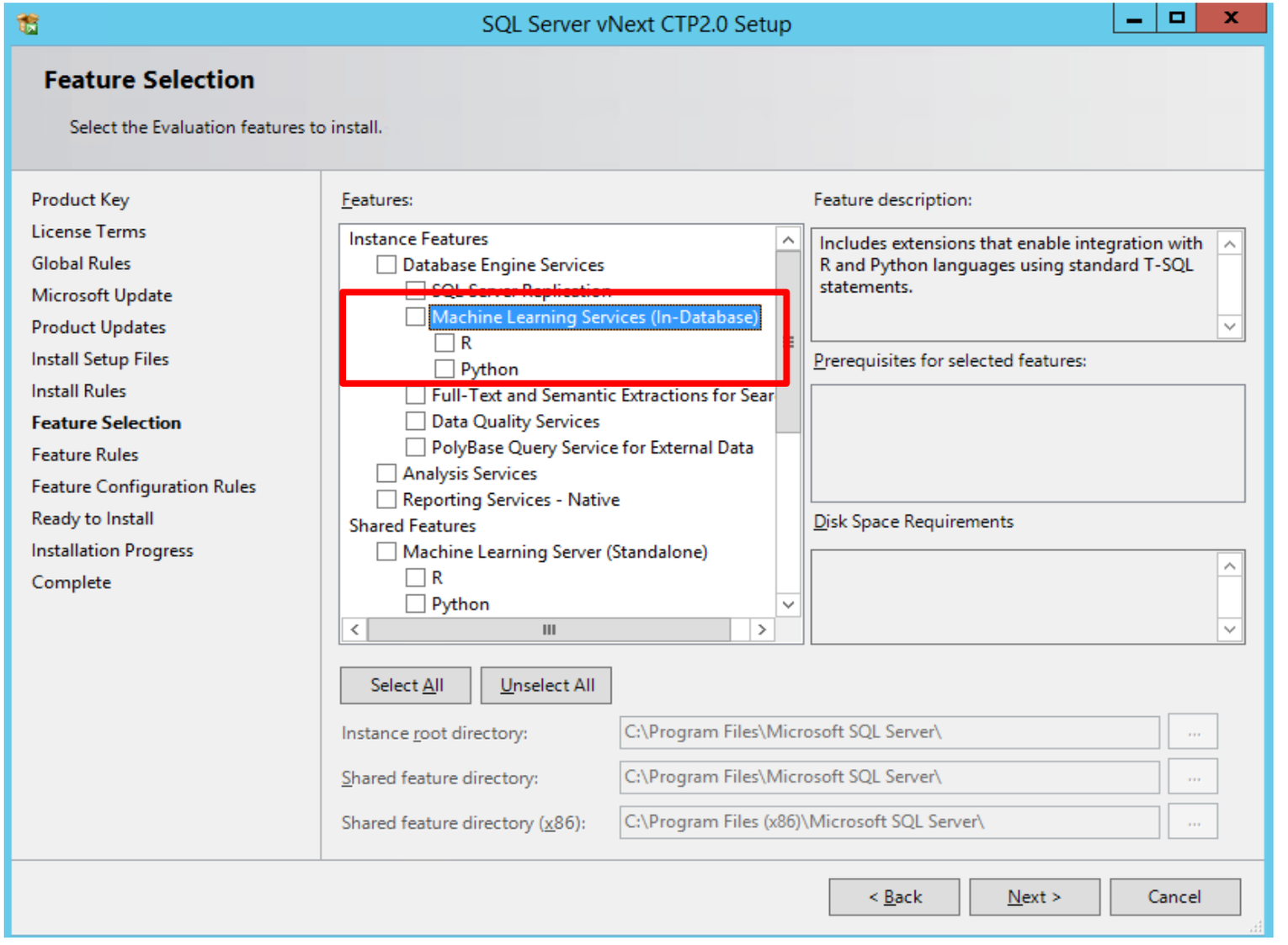


SQL Server 2019 - Java Language Extension

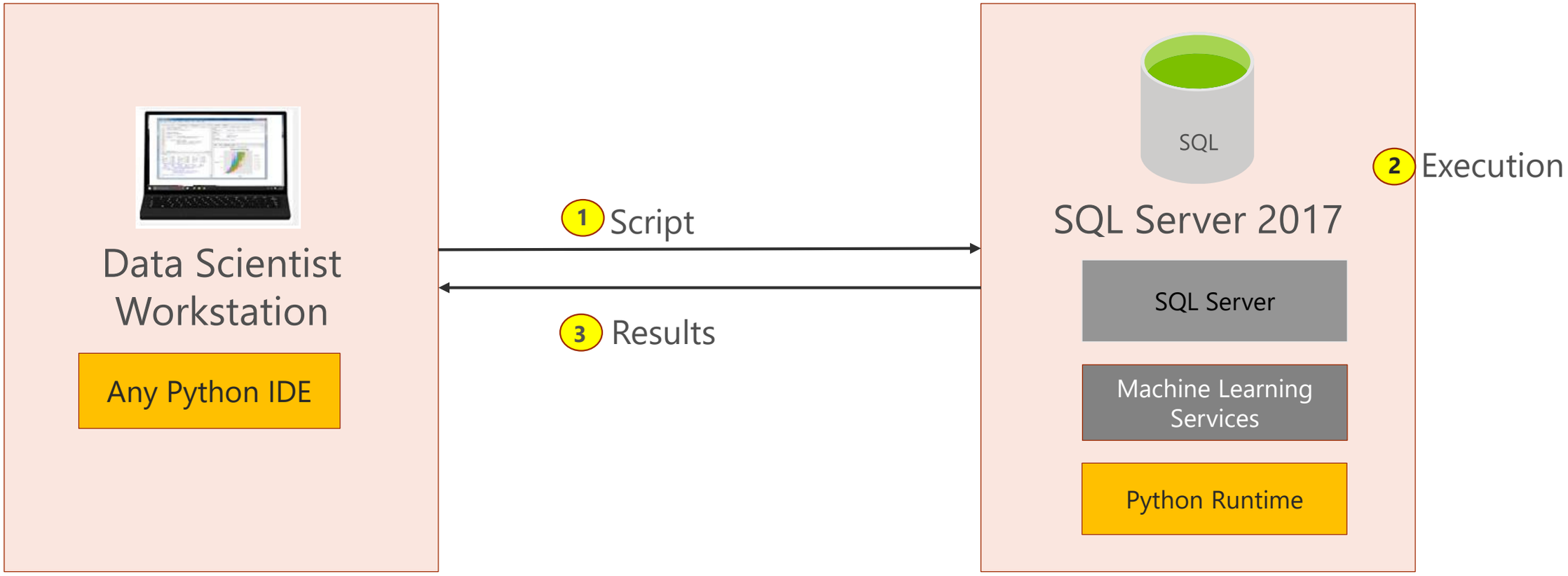
```
DECLARE @myClassPath nvarchar(30)
SET @myClassPath = N'/<my path>/program.jar'
SET @param1 = 3

EXEC sp_execute_external_script
  @language = N'Java'
, @script = N'package.ClassName.MethodName'
, @input_data_1 = N'<Input Query>
, @params = N'@CLASSPATH nvarchar(30), @param1 INT'
, @CLASSPATH = @myClassPath
, @param1 = @param1
with result sets ((outputcol int, outputcol2 int))
```



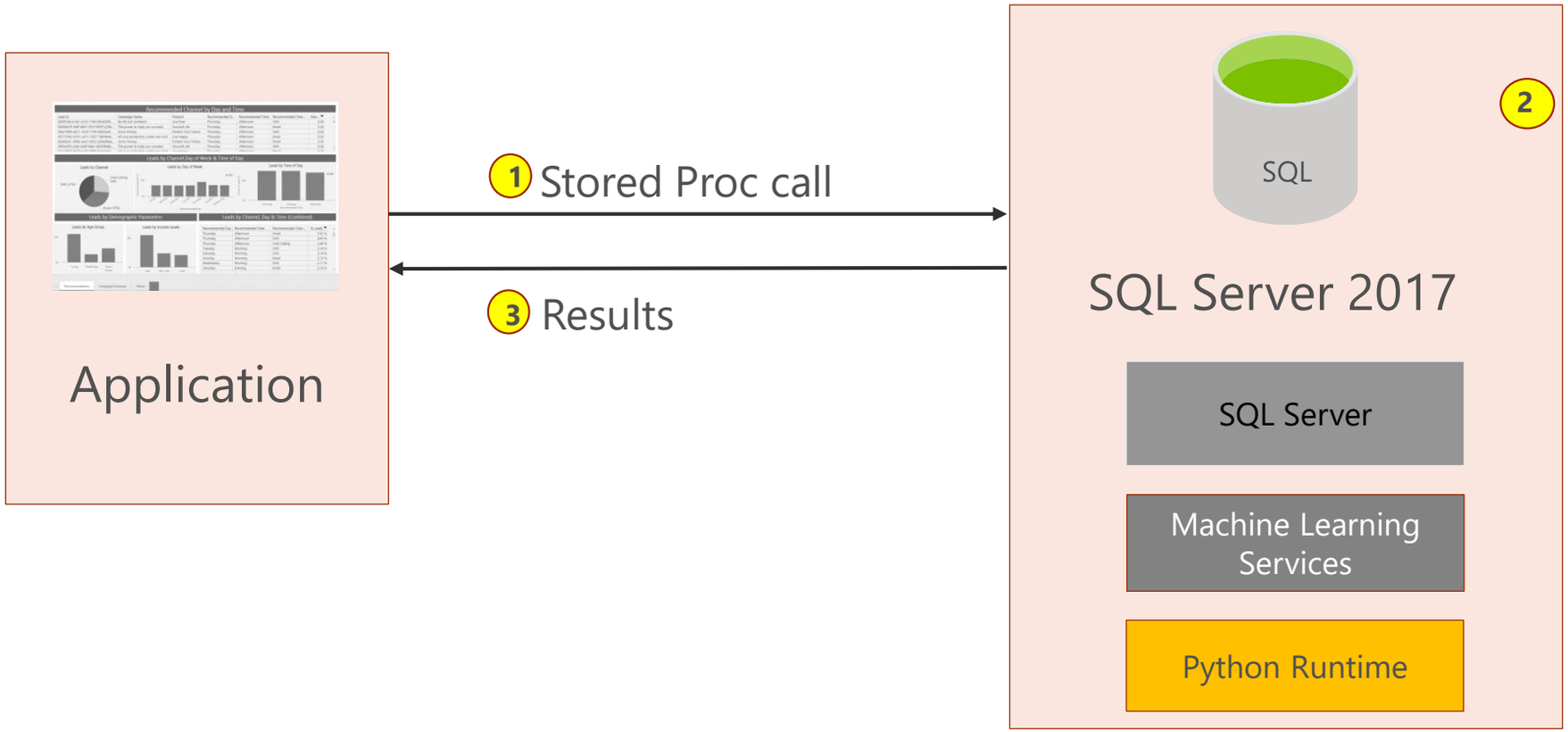
How to install Server in-database ML Services and language extensibility



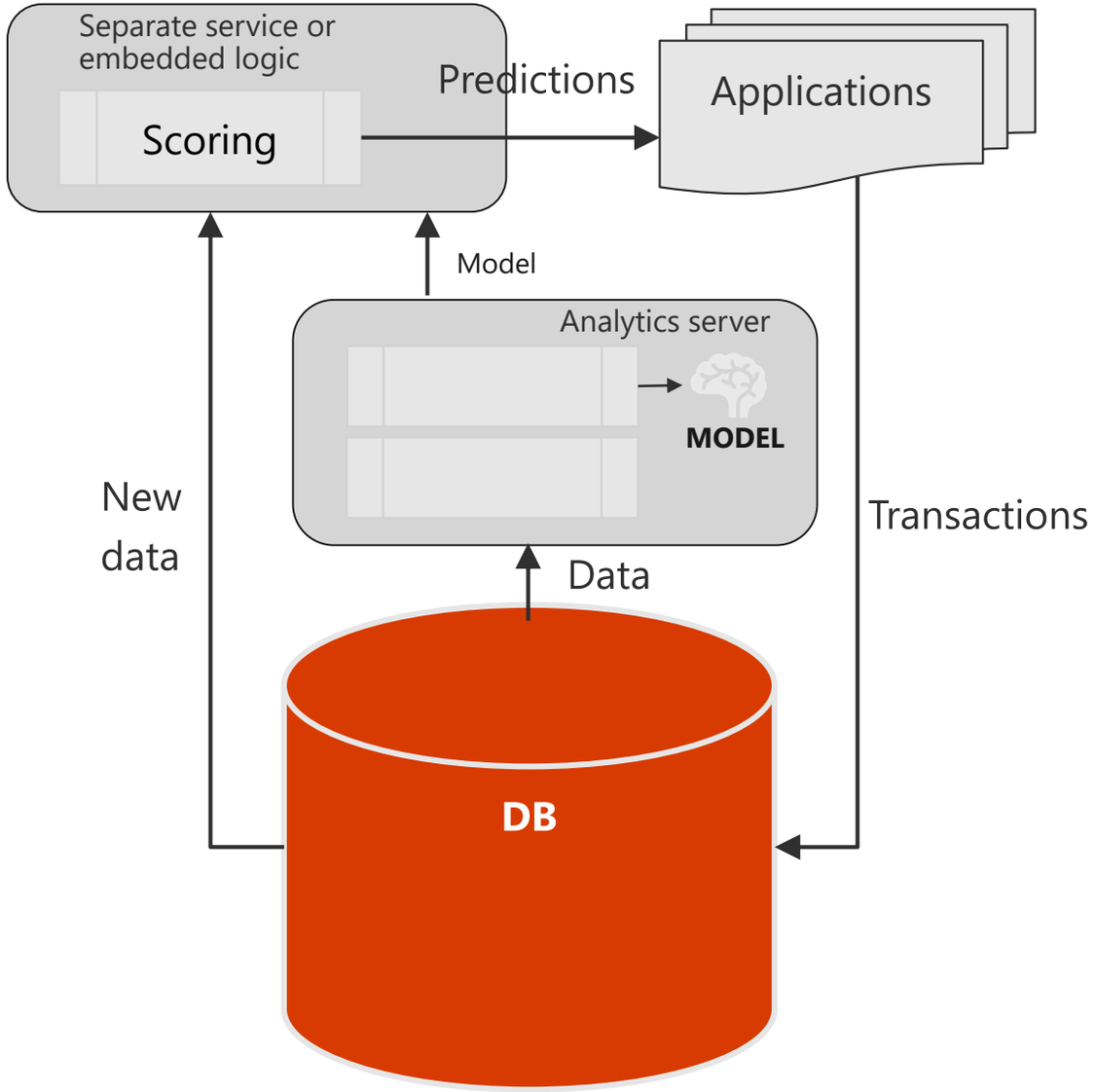
Data Scientists - Data Exploration and Model Development

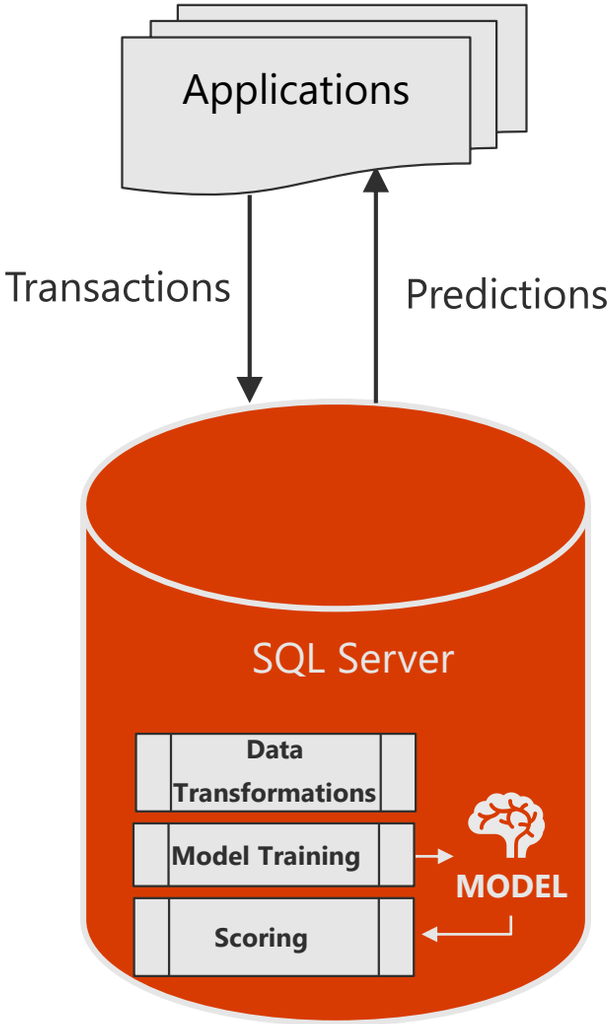


Application Developer - Model Operationalization



Streamline Productivity and Simplify Deployment





Machine Learning outside of DB

In-DB Machine Learning

Demo

Text Classification

Predictions for New Text

PredictedLabelName	Subject	Text
talk.religion.misc	Re: Who's next? Mormons and Jews?	"In article <1r7prg$hpq@agate.berkeley.edu> isaackuo@spam.berkeley.edu (Isaac Kuo) write kennejs@a.cs.okstate.edu (KENNEDY JAMES SCOT) writes: [...] > >You are speculating that the ridiculously extraordinary claim, and have no evidence >whatsoever, at least give us a specula possibility that the fire was started accidentally is much more reasonable, >but we don't have >"Isaac Kuo (isaackuo@math.berkeley.edu) * ___ Mr. Kuo: I don't recall seeing your byline aro newsgroup. t.p.m is populated largely by people whose hatred for the U.S. government, espe people, the suggestion that agents of the U.S. government would carry out a spectacular mas 'extraordinary claim'. It's just another day in the life of these United States. They don't have to citizens. Have you seen 'The Crying Game'? ""It's in their nature"". Don Pajerek Standard discl
talk.politics.misc	Re: solvent for duct-tape adhesive?	Another alternative is to cover the adhesive with another piece of duct tape then remove the <C67stK.FoK@athena.cs.uga.edu>, mcovingt@aisun3.ai.uga.edu (Michael Covington) writes: (Dave Stephenson) writes: |> >mcovingt@aisun3.ai.uga.edu (Michael Covington) writes: |> >> try MEK (methyl |> >>>ethyl keytone?). It worked getting the stickum left over from shelf pap better still |> >industrial goggles. The small est drop in the eye will casue [etc.] |> |> Please w |> -- |> :- Michael A. Covington, Associate Research Scientist : ***** |> :- Artificial Intelligence Georgia phone 706 542-0358 : * * * |> :- Athens, Georgia 30602-7415 U.S.A. amateur radio N4
talk.politics.mideast	Emergency Brake on '86 Nissan Maxima	Just a quick, simple question really... How many wheels are affected by the emergency brake would seem unusual to me. I thought the emergency brake on most cars only affected the re Enough to lock wheels at 30mph? Hmmmm... I just have to wonder about some of the things
talk.politics.guns	Re: Nature of the Waco gas	"cash@convex.com (Peter Cash) writes: >In article <1r6170INNdlu@cronkite.Central.Sun.COM was that the use causes extreme nausea, >>blindness, disorientation, total irrationality, raging the results all the >>earlier. >If we are indeed talking about CS, then this is not quite accurate

Text

Count of Id by PredictedLabelName and Subject

Microsoft

SQL Server Extensibility Framework

Host external runtimes securely on SQL Server machine

Resource governance on external processes

- EXTERNAL RESOURCE POOL to control CPU, Memory, CPU Affinity

Integrate with SQL query execution

- New external script operator to exchange data / parameters
- Parallel query pushing data to multiple external processes / threads
- Streaming mode execution
- Batch mode execution (in SQL Server 2017)

Implied Authentication

- Impersonation for loopback connections from external scripts
- Just use trusted connection in connection string

Faster Time to Insight

Boost performance of your machine learning processing with the following features:

- **Leverage the power of the server in remote execution (Compute Context)**
- **Integration with SQL query execution**
  - Parallel query pushing data to multiple external processes / threads
    - Use in-memory technology and Columnstore Indexes alongside your ML scripts
- **Streaming mode execution**
  - Stream data in batches to the R/Python process to scale beyond
  - available memory
- **Train and Predict using parallelism**
  - Leverage RevoScaleR/revoscalepy and scale your R and Python scripts using ale multi-threading and parallel processing
  - Leverage trivial parallelism for scoring any arbitrary R or Python script
- **Native scoring for faster real-time predictions (New in 2017)**

Native Scoring using PREDICT function

Requirements:

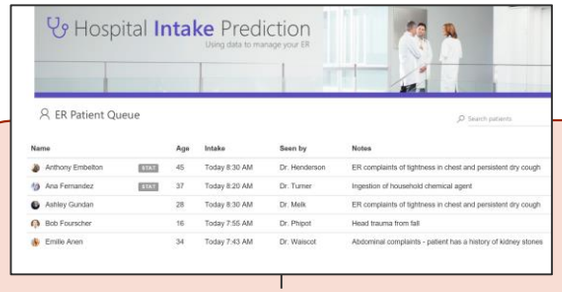
- rx* models
- Serialized model from rxSerializeModel (R)
- Serialized model from rx_serialize_model (Python)

Key Benefits:

- Runs natively in SQL Server (No R / Python dependency)
- Low latency for execution
- Ideal for highly concurrent scoring of few rows
- Can be used in INSERT/UPDATE/MERGE statement directly

Hybrid Scenario

Application



Anonymized patient data

Train model using Python or R

SQL Server 2017 with ML Services

On premises

Deploy model to Azure SQL DB

Azure SQL DB

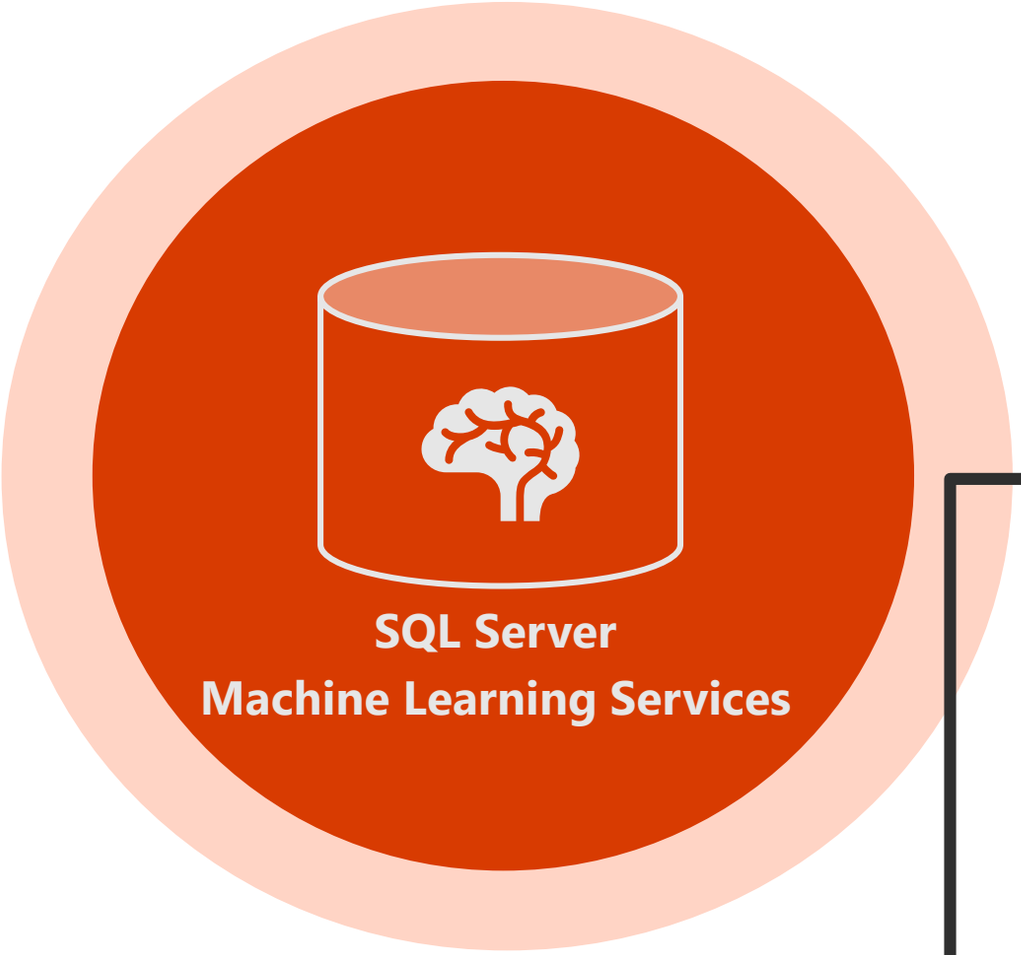
Score using deployed model

Cloud

Why *in-database* machine learning?

Better Collaboration & Sharing Insights

Faster Time to Insight



Streamline Productivity and Deployment

Better Security & Compliance

Better Security & Compliance

Reduced surface area and isolation

- 'external scripts enabled' required
- R/Python script execution outside of SQL Server process space

Script execution requires explicit permission

- sp_execute_external_script requires EXECUTE ANY EXTERNAL SCRIPT for non-admins
- SQL Server login/user required and db/table access

R/Python processes have limited privileges

- R/Python processes run under local low privileged accounts
- User isolation
- Default Windows firewall rules to block outbound traffic

Better Security & Compliance

Reduce need for data movement

- Remote execution capabilities using SQL compute context

Easy to share data but still secure

- Row level security to granularly protect sensitive data

Secure connection strings

- Use trusted connection in connection strings to avoid revealing passwords

Leverage Resource Governance to isolate R and Python scripts

- EXTERNAL RESOURCE POOL lets you control CPU, Memory and CPU Affinity

Customer Learnings

Customer stories

- ZEN3 Infosolutions (SayInt)
  - Customer call transcript in SQL Server
  - Use Python in-Db for text analysis
- ATTOM Data Solutions
  - Real estate property insights
- Acxiom
  - Migrating from SAS
  - Targeted marketing campaigns
- Heartland Bank
  - Migrating from SAS
  - Build financial models

ML Services with Python helps to process customer call transcripts to extract n-grams using Python to better analyze calls.

ATTOM will be able to take advantage of a host of new capabilities, including machine learning and predictive analytics.

With R Services, Acxiom is building better, more accurate models using larger data sets, which results in more precisely targeted marketing for customers.

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

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

Leverage best of T-SQL & R / Python. Use the right tool!

Roadmap

- Azure SQL Database
  - R support is coming
  - Python support is coming
- Machine Learning Services in SQL Server on Linux **in public preview**
- Additional algorithms and pre-trained models
- Failover cluster support
- Partitioning support for input data
- Native Scoring for more models
- Updated Java developer experience
- Expand data type support
- External library support

Resources

- SQL Server Samples on GitHub – R Services & ML Services
- Getting started tutorials: AKA.MS/MLSQLDEV
- Configure instance: SSMS Reports for ML Services
- ML cheat sheet
- Text classification scripts:
  - https://microsoft.github.io/ml-server-text-classification/
- Hospital length of Stay demo scripts
- Microsoft documentation: SQL Server Machine Learning Services

Microsoft 365

Thank you