BUCHANAN & EDWARDS

Event: Data Science On-Ramp

Presenter: Jon Tupitza, CTO Architect

BEOP.CTO.TP4
Owner: OCTO
Revision: 0001
Approved by: JAT
Effective: 08/30/2018

Data Science *On-Ramp*

SQL Machine Learning Services

Operationalizing Python & R using SQL Machine Learning

Jon Tupitza
Practice Director, Data Platform & Predictive Analytics

Take-Aways

- Understand Machine Learning and SQL Server ML Services
- Recognize the Categories of Machine Learning Algorithms along with Practical Applications
- Learn the Different Ways and Means to Implement SQL Server Machine Learning Services

Agenda

- SQL Server 2017 ML Services
  - Value Proposition
  - Getting Started
- Machine Learning Tasks
  - Classification
  - Regression
  - Clustering
  - Principal Component Analysis

SQL Server ML Services: Value Proposition

- The First Commercial Database Server with Built-In Artificial Intelligence
- Enables Developers to Train, Evaluate and Deploy Machine Learning Models Inside of SQL Server Databases for Enterprise Production

Overcomes Some Major Limitations Inherent to Statistical Software

- System Memory has been limited to the capacity of client workstations
- Data Movement has been saturating networks between remote storage and the development environment
- Performance and Scale have been limited by a lack of multi-threading and parallel processing capabilities

Provides a Convenient Way to Operationalize Machine Learning

- Access ML Algorithms using familiar T-SQL stored procedures
- Manage Machine Learning Models in SQL Server database tables
- Store Predictive Outcomes in SQL Server database tables
- Leverage database mechanisms like security, governance and monitoring

SQL Server 2017 Machine Learning Services

- **Adds Python Support:**
  - Grants access to deep learning tools like CNTK, TensorFlow and Keras.
- **RevoScalePy** in addition to **RevoScaleR:**
  - Enables scaling ML to arbitrarily large datasets; beyond available memory.
  - Linear & Logistic Regression, Decision Trees, Boosted Trees, Random Forests.
  - API's for ETL processing, remote compute contexts and data sources.
- **MicrosoftML:**
  - Simplifies complex AI scenarios involving Image and Text data sources
  - Features Deep Neural Networks, SVM, Fast Tree, Forests, etc.
  - Includes pre-trained models like ResNet for Image and Sentiment analysis
- **Python Remote Compute in SQL Server:**
  - Enables data scientists to execute Python code remotely from their development machines to explore data and develop models without moving data around.

Where to Process: Transact-SQL versus R or Python

Transact-SQL

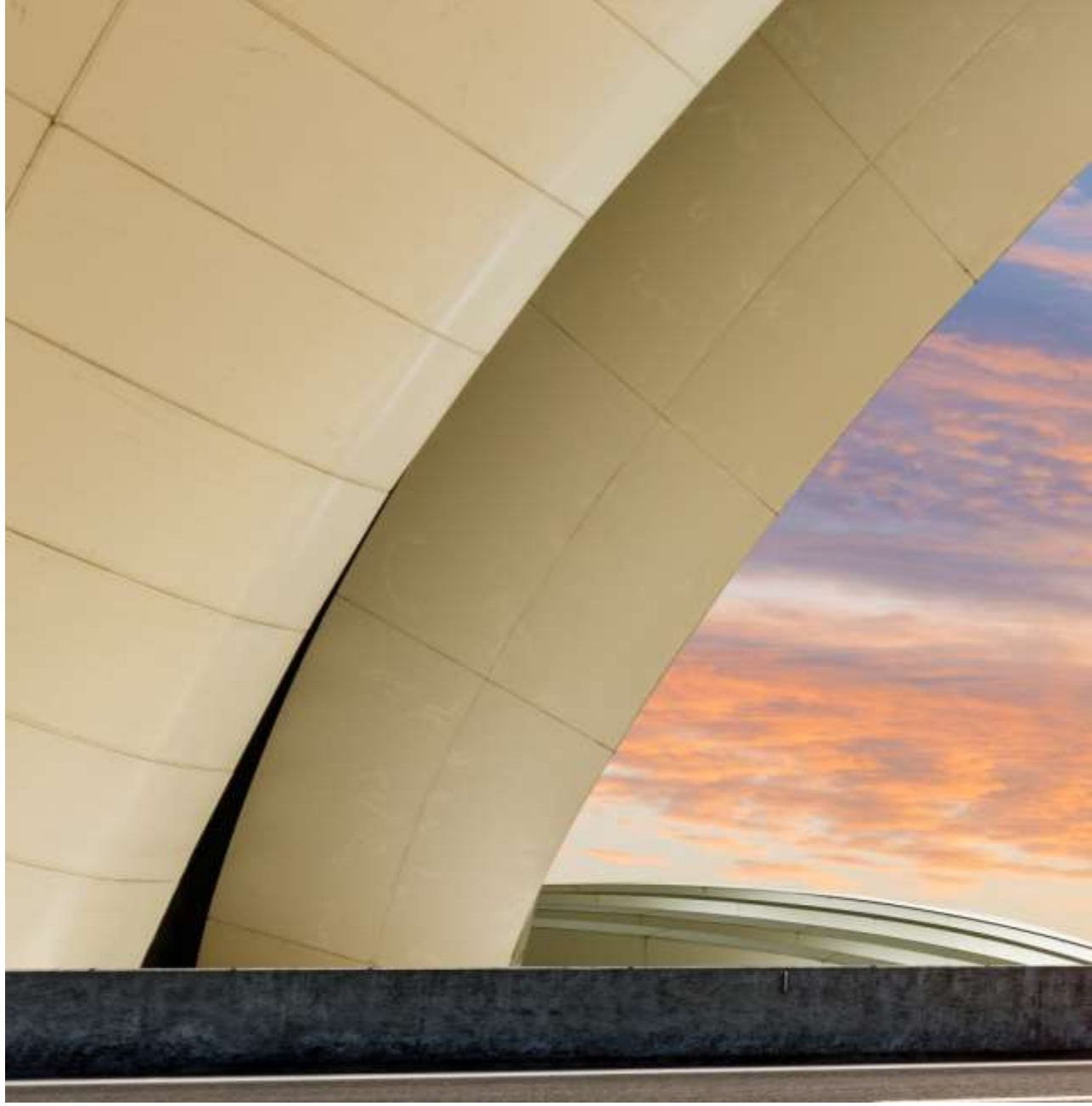
- Table Joins and Filters
- Set Operations (Union, Except, Intersect)
- Sorts & Aggregations

R or Python

- Statistical Calculations
- Predictive Algorithms
- Data Visualizations

Demonstration

- **Getting Started**
  with SQL Server 2017
  Machine Learning Services



Agenda

- SQL Server 2017 ML Services
  - Advantage and Value
  - Getting Started
- Machine Learning Tasks
  - Classification
  - Regression
  - Clustering
  - Principal Component Analysis

Demonstrations

- **Classification:**
  - Exploratory Data Analysis using Jupyter Notebooks with the RevoScalePy library's Remote Execution API's
  - Deploy a Machine Learning Model to Predict IRIS' Species using In-Database Scripting

Demonstration

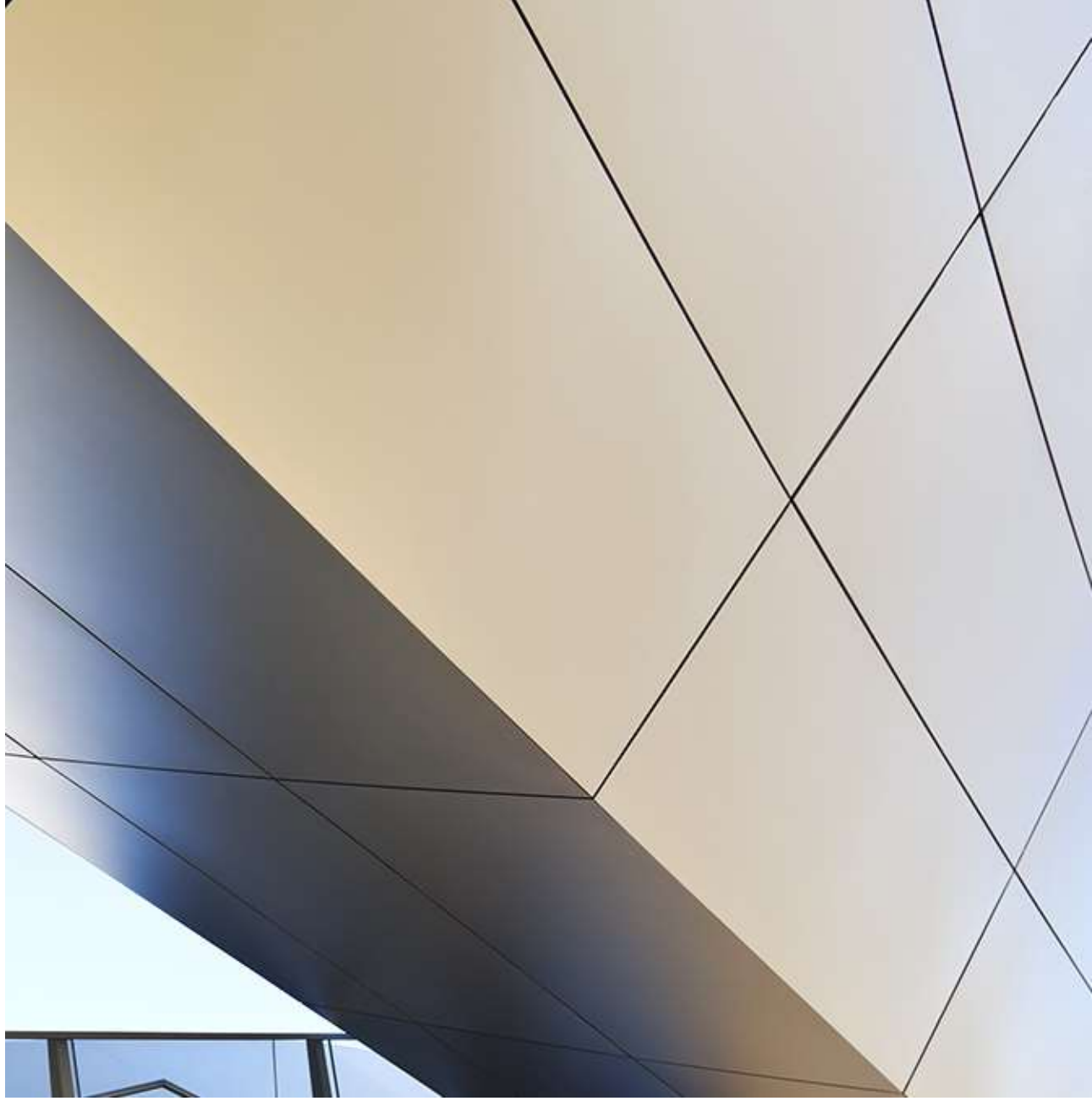
- **Linear Regression:**
  - Train and Evaluate a Model using Jupyter Notebooks with the RevoScalePy library's Remote Execution API's
  - Operationalize a Machine Learning Model that Predicts Future Ski Rentals using In-Database Script Execution

Demonstration

- **K-Means Clustering:**
  - Develop a ML Experiment using Jupyter Notebooks with the RevoScalePy library's Remote Execution API's
  - Operationalize a Model that Clusters Customers by their Purchasing Habits using In-Database Script Execution

Demonstration

- **Principal Component Analysis (PCA):**
  - Perform Dimension Reduction using Principal Component Analysis in Jupyter Notebooks with the RevoScalePy library's Remote Execution API's



Questions

Resources

- Microsoft Docs:
  - Tutorials for SQL Server Machine Learning Services
- Microsoft Machine Learning Server Blog:
  - Basics of R and Python Execution in SQL Server