

# Azure + Python = Data Analytics

Snehith Allamraju

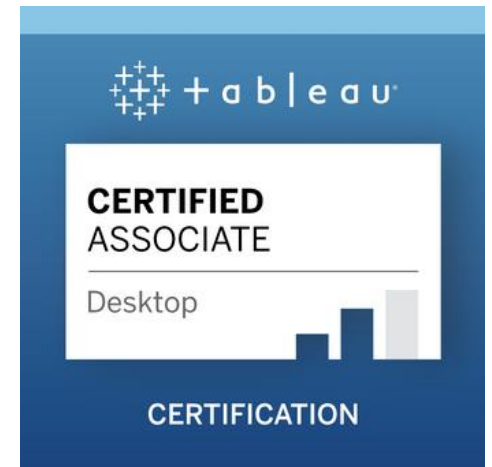
Azure Pyday

28<sup>th</sup> June 2020

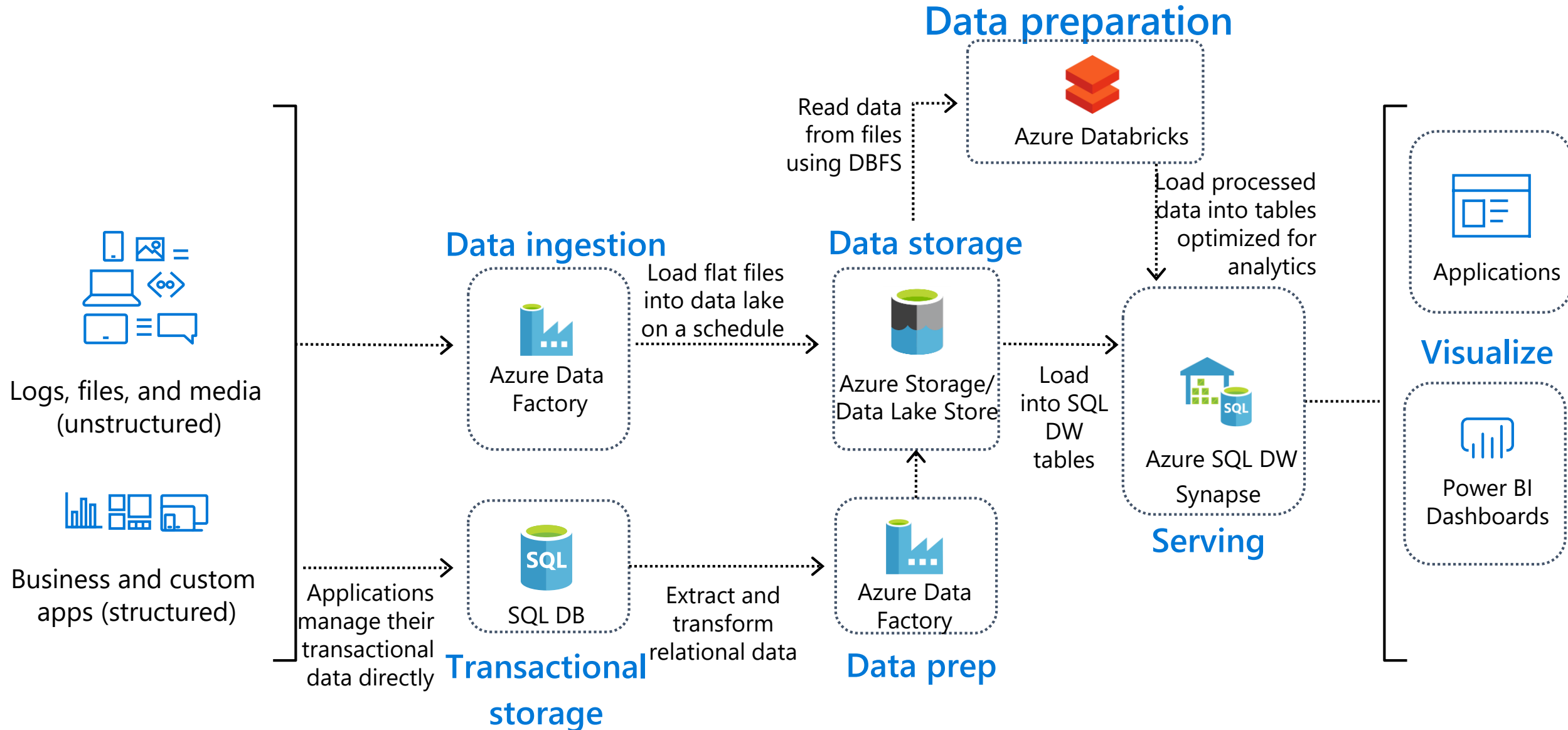
# About me

- BI & Data Analytics Manager
- Hyderabad Python user group
- ODSC Hyderabad
- Kaggle Days Hyderabad
- Hyderabad Tableau user group

**SAP<sup>®</sup>** Certified  
Associate



# Data warehousing pattern in Azure



# Big Data- Query a database in Azure SQL Database

You use Python to connect to Azure SQL Database or Azure SQL Managed Instance, and use T-SQL statements to query data.

- [Python SQL driver - pyodbc](#)
- [Python SQL driver - pymssql](#)

```
1 import pyodbc
2 server = 'your_server_name'
3 database = 'AdventureWorks2012'
4 username = 'your_username'
5 password = 'your_password'
6 driver= '{ODBC Driver 17 for SQL Server}'
7 cnxn = pyodbc.connect('DRIVER='+driver+';SERVER='+server+';PORT=1433;DATABASE='+database+';UID='+username+';PWD='+ password)
8 cursor = cnxn.cursor()
9 cursor.execute("SELECT TOP (10) SalesOrderNumber, PurchaseOrderNumber , OrderDate from [Sales].[SalesOrderHeader]")
10 row = cursor.fetchone()
11 while row:
12     print (str(row[0]) + " " + str(row[1]) + " " + str(row[2]))
13     row = cursor.fetchone()
```

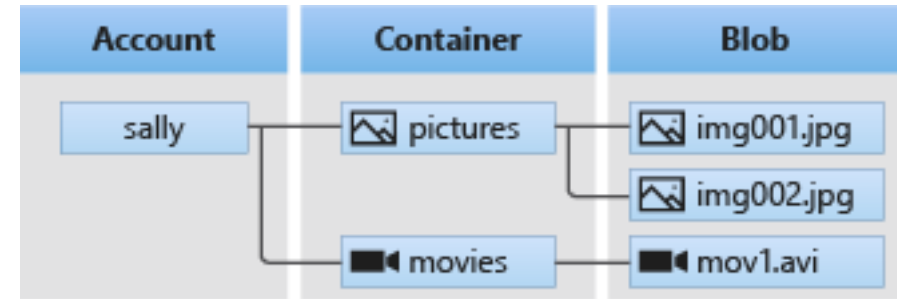
```
S01 P0522145787 2005-07-01 00:00:00
S02 P018850127500 2005-07-01 00:00:00
S03 P018473189620 2005-07-01 00:00:00
S04 P018444174044 2005-07-01 00:00:00
S05 P018009186470 2005-07-01 00:00:00
S06 P016617121983 2005-07-01 00:00:00
S07 P016588191572 2005-07-01 00:00:00
S08 P016008173883 2005-07-01 00:00:00
S09 P015428132599 2005-07-01 00:00:00
S010 P014732180295 2005-07-01 00:00:00
```

It is also possible to connect and query data in Azure Database for PostgreSQL ([psycopg2](#)), MySQL(mysql.connector)

# Big Data - Azure Storage library for Python

The following components make up a Azure Blob Service:

- The storage account itself
- A container within the storage account
- A blob within a container



The Azure Storage Blobs client library for Python allows you to interact with each of these components through the use of a dedicated client object.

<https://github.com/Azure/azure-sdk-for-python/tree/master/sdk/storage/azure-storage-blob/samples>

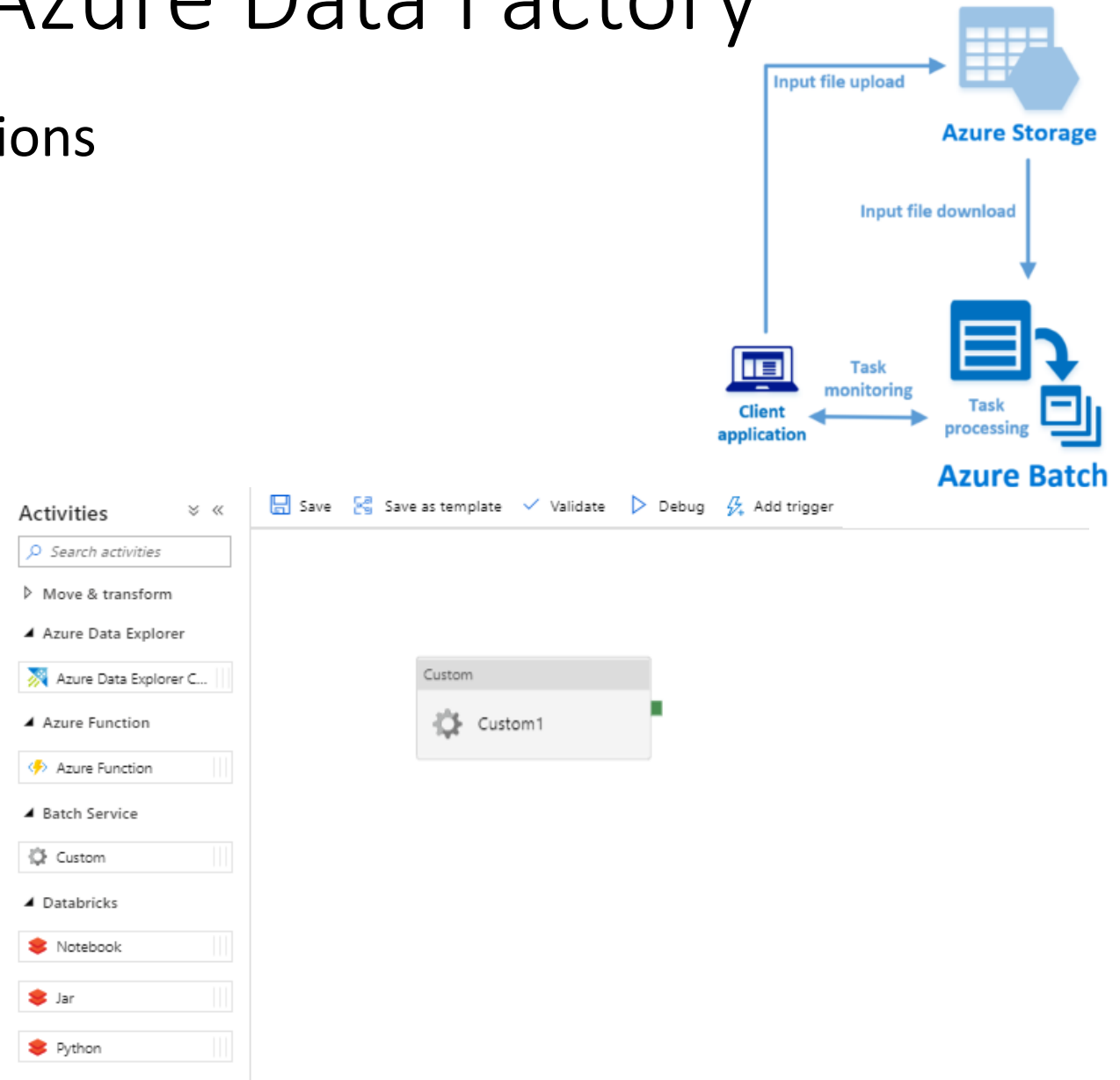
# Data Orchestration in Azure Data Factory

## Custom Scripting using Azure Batch functions

Azure Data Factory helps in Data movement and Data transformation activities

ADF supports Custom activity that helps define our own data movement or transformation logic and use the activity in a pipeline. The custom activity runs your customized code logic on an **Azure Batch** pool of virtual machines

You use the Python API to run an Azure Batch job from an app. The app uploads input data files to Azure Storage and creates a *pool* of Batch compute nodes (virtual machines). It then creates a job that runs *tasks* to process each input file in the pool using a basic command.



You can also process *Parallel Workloads* with Azure Batch

# Data Orchestration in Azure Data Factory

## ADF ETL Pipelines using Python SDK for Data Factory

Along with using the Azure portal UI, you can also use [Python SDK for Data Factory](https://docs.microsoft.com/en-us/azure/data-factory/quickstart-create-data-factory-python) to create a data factory by using Python

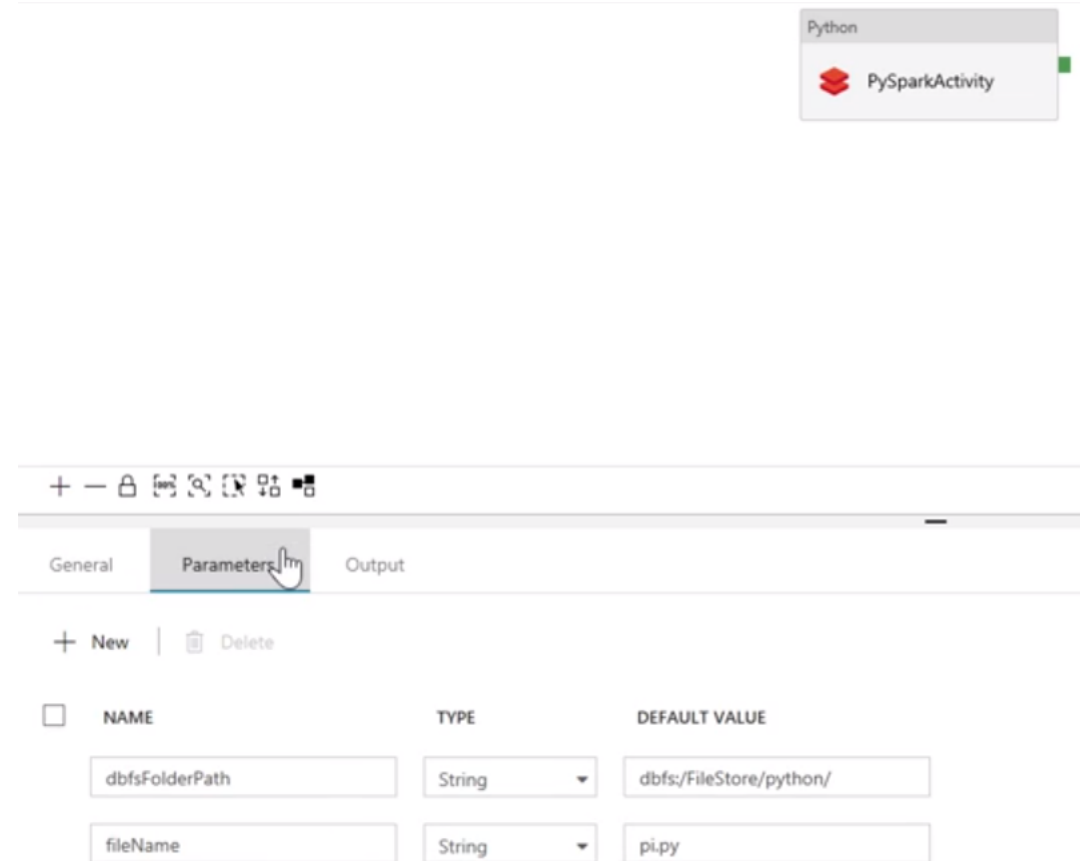
The Python package for Azure Data Factory is **azure-mgmt-datafactory**

<https://docs.microsoft.com/en-us/azure/data-factory/quickstart-create-data-factory-python>

## Consuming Databricks Python in ADF

The Azure Databricks Python Activity in a [Data Factory pipeline](#) runs a Python file in your Azure Databricks cluster.

<https://docs.microsoft.com/en-us/azure/data-factory/transform-data-databricks-python>



# Azure Machine Learning

Azure Machine Learning helps you build, train, deploy, and manage your models at cloud scale.



## Register data

Reference data from storage to easily access during model training and explore using summary statistics.



## Train models

Use machine learning algorithms with training data to create models.



## Evaluate models

Find the best model using test data.



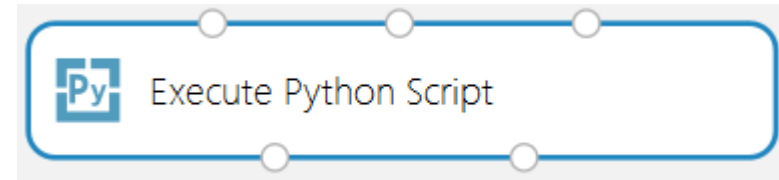
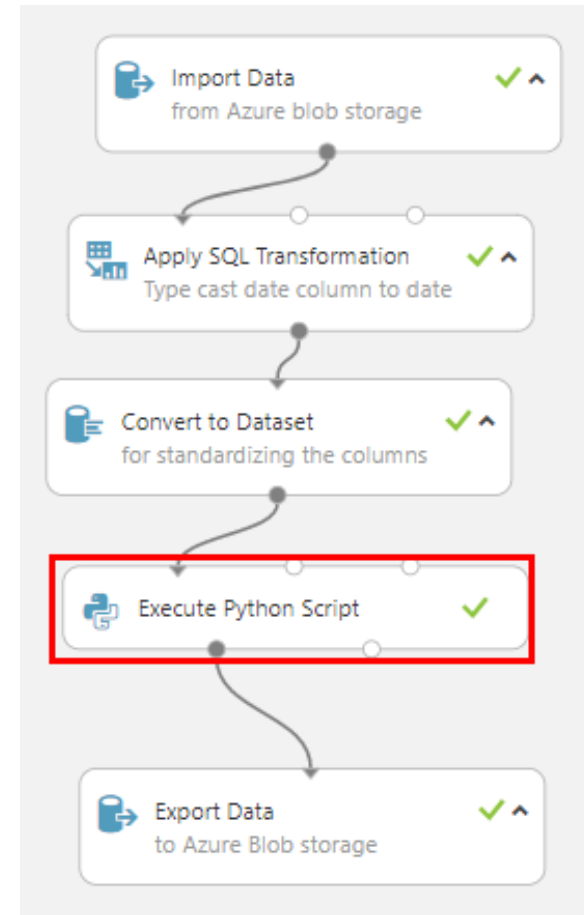
## Deploy models

Deploy model as a web service in the Azure cloud, or to IoT Edge devices.



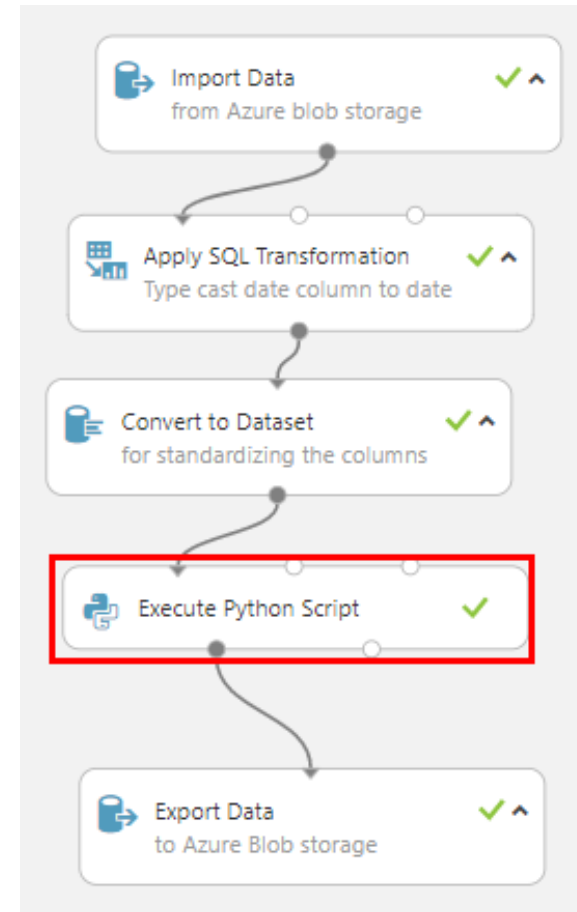
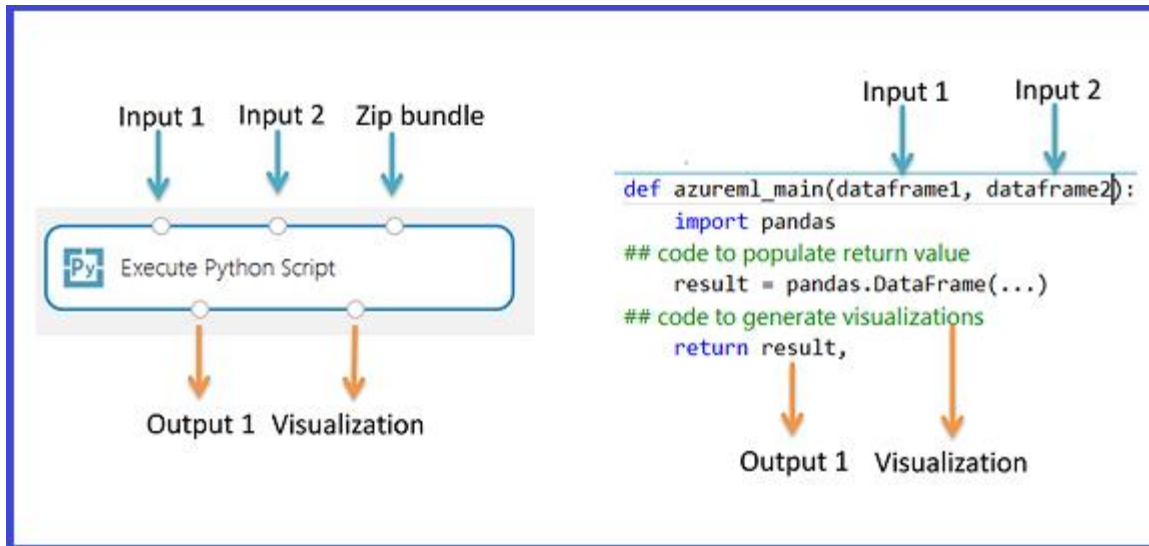
# Machine Learning using Azure ML Studio/workspace

- Azure Machine Learning Workspace
  - You can use the *Execute Python Script* module to use Python code in your Azure Machine Learning Studio (classic) experiments and web services
- Azure Machine Learning Python SDK running in Jupyter notebooks



# Machine Learning using Azure ML Studio/workspace

- Azure Machine Learning Workspace
  - You can use the *Execute Python Script* module to use Python code in your Azure Machine Learning Studio (classic) experiments and web services



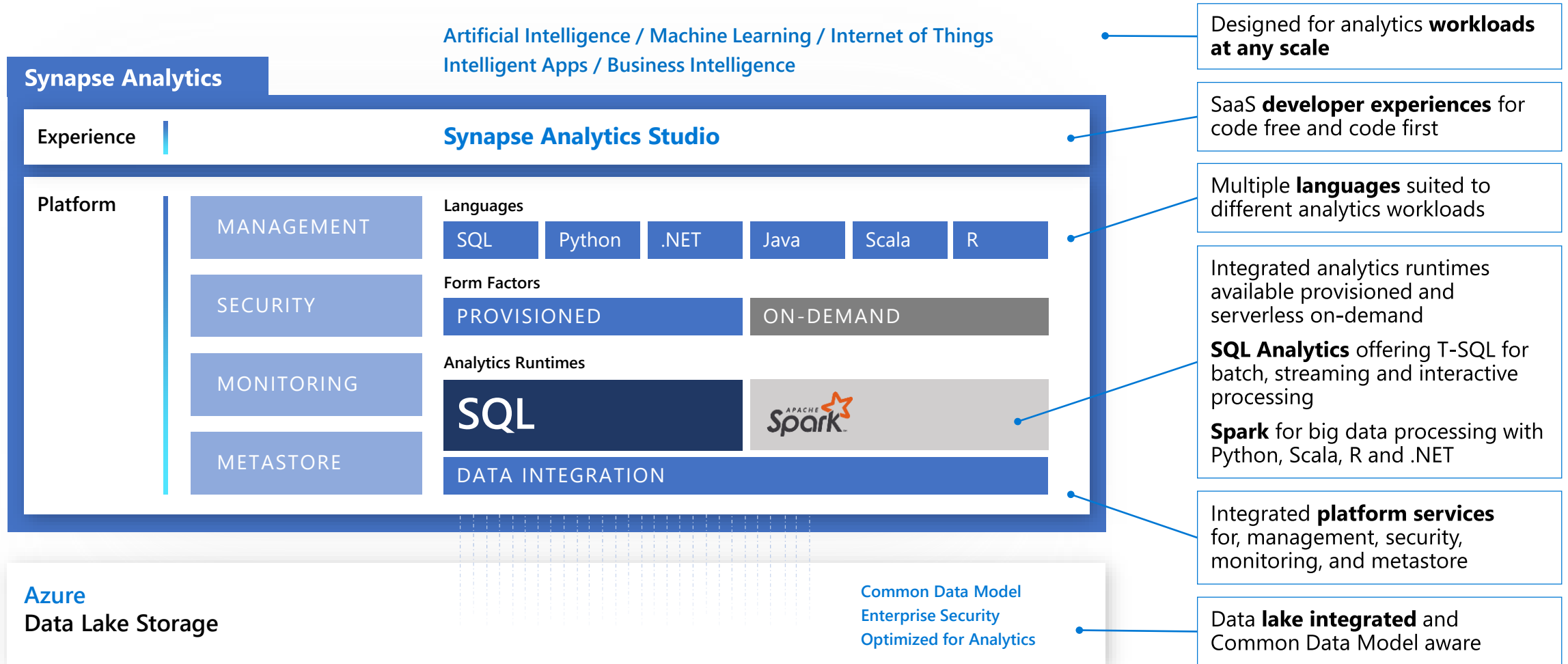
# Spark based Analytics using Azure Databricks

Ideal for running large-scale intensive machine learning workflows on the scalable Apache Spark platform.

You can access Azure Blob Storage using [Spark framework](#) on Python using Azure Databricks

# Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence

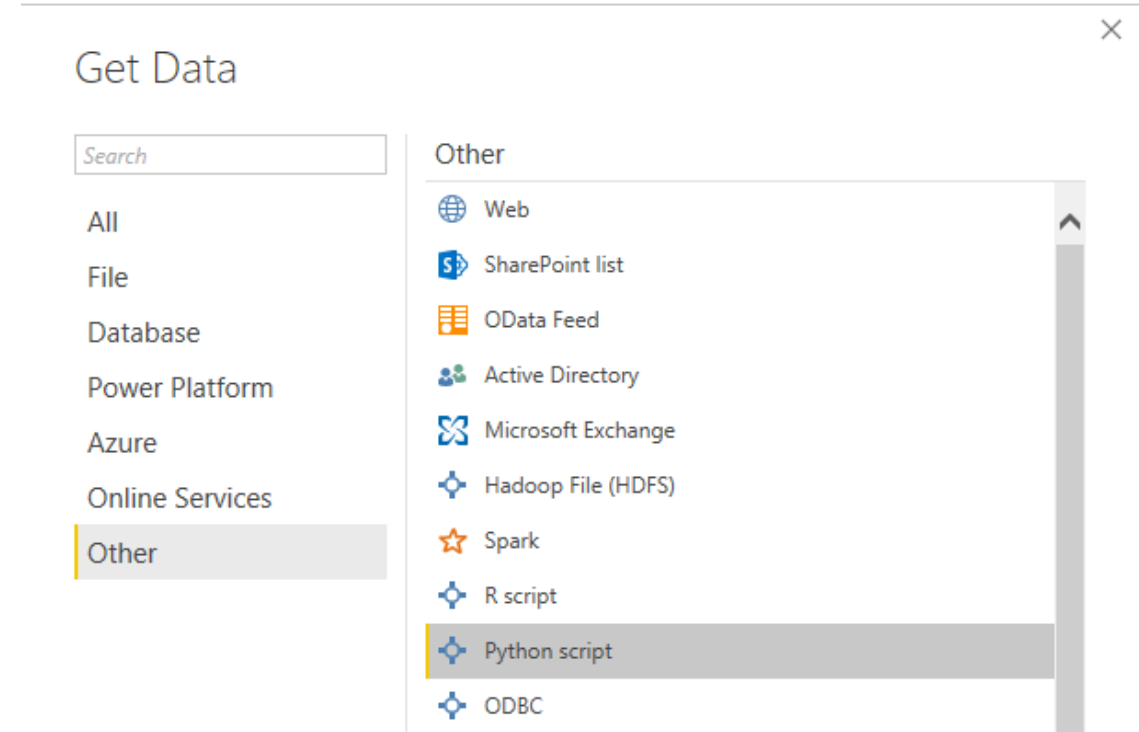


# Data Visualization – Power BI

You can run Python scripts directly in Power BI Desktop and import the resulting datasets into a Power BI Desktop data model

Pre requisites:

- Pandas
- Matplotlib



# Also Available ( AI & IoT )

- Working with Azure IoT Hub and Python for Telemetry data
- Azure IoT SDKs for Python that enable python developers to easily create IoT device solutions that seamlessly connect to the Azure IoT Hub ecosystem
- Python SDKs for various Cognitive Services services
  - Knowledge            QnA SDK
  - Language            LUIS/ Text Analytics SDKs
  - Vision                Face, Computer Vision, Ink recognizer SDKs



# CFP for Workshops is Open

**PyConIndia**  
2020, Online

<https://in.pycon.org/blog/2020/2020-workshop-cfp-announcement.html>

# Thank you!

Do connect and reach out at:

Twitter: @watrutalkin & @HydTechEvents

Linkedin: <https://www.linkedin.com/in/snehithallamraju/>