Microsoft

Azure Synapse Analytics

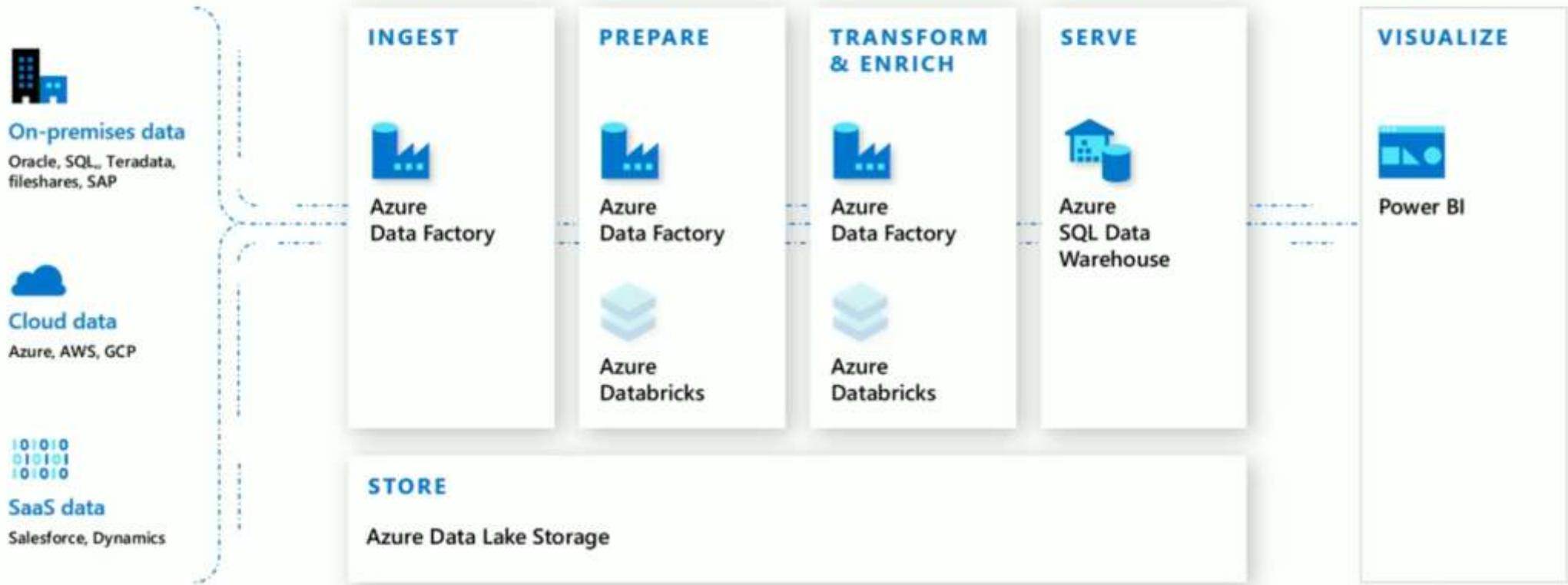
- ***Adrián J. Fernández Zenteno***
- *Sr. Cloud Solution Architect - Azure Advanced Analytics*
- Email: adrian.fernandez@microsoft.com
- Twitter: @AdrianFZ10

Microsoft

Azure Synapse Analytics

- Introduction
- Studio
- Data Integration
- Execution Pools:
  - SQL Analytics *(formerly SQL Datawarehouse)*
  - SQL On-Demand
  - Spark
- Foundation
- Connected Services

Modern Data Warehouse



Azure Synapse Analytics - *Data Lakehouse*

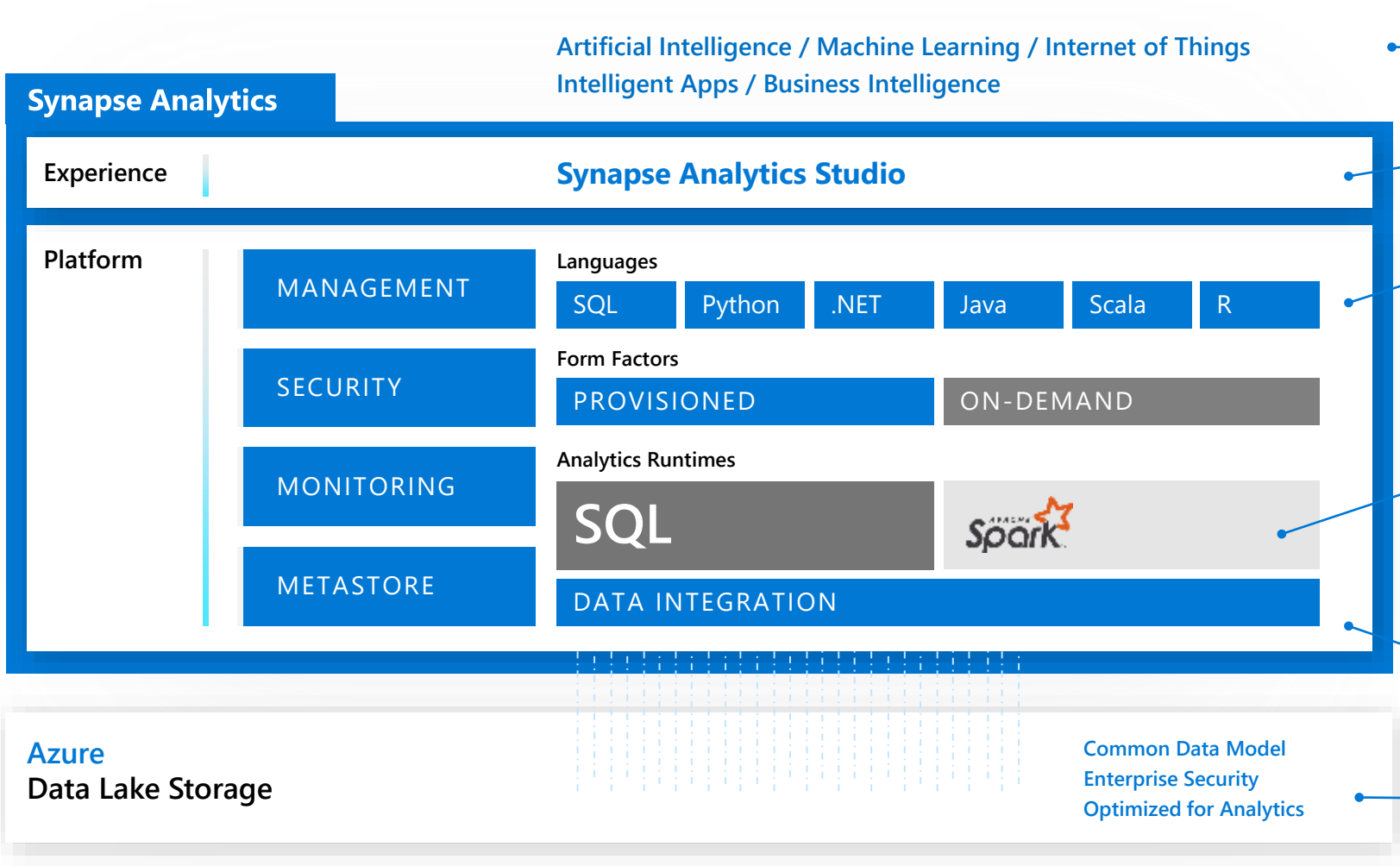
Microsoft

Azure Synapse Analytics

Azure Synapse Analytics is a limitless analytics service, that brings together enterprise data warehousing and Big Data analytics. It gives you the freedom to query data on your terms, using either serverless on-demand or provisioned resources, at scale. Azure Synapse brings these two worlds together with a unified experience to ingest, prepare, manage, and serve data for immediate business intelligence and machine learning needs.

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

Provisioning Synapse workspace

Providing Synapse is easy

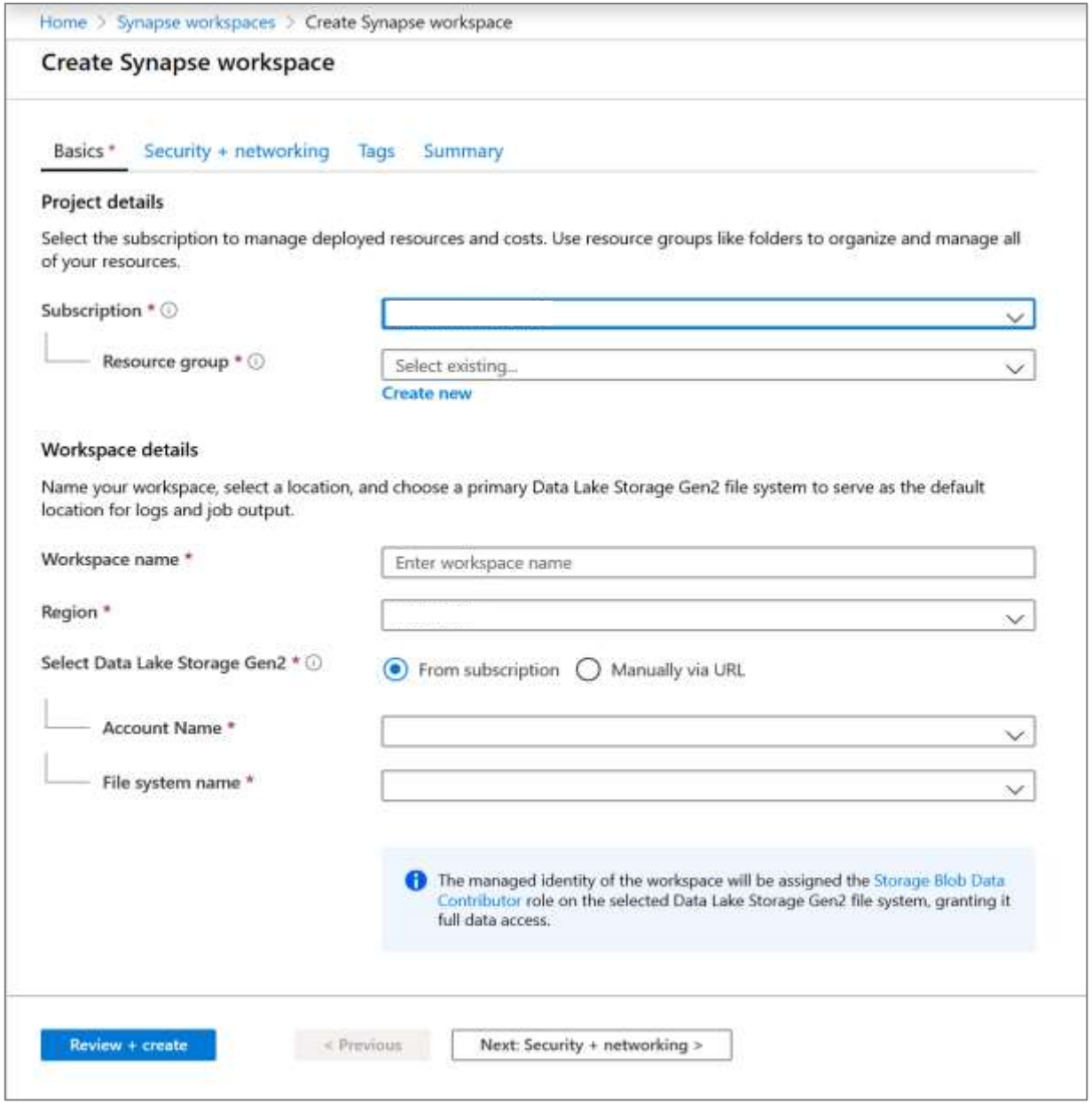
Subscription

Resource Group

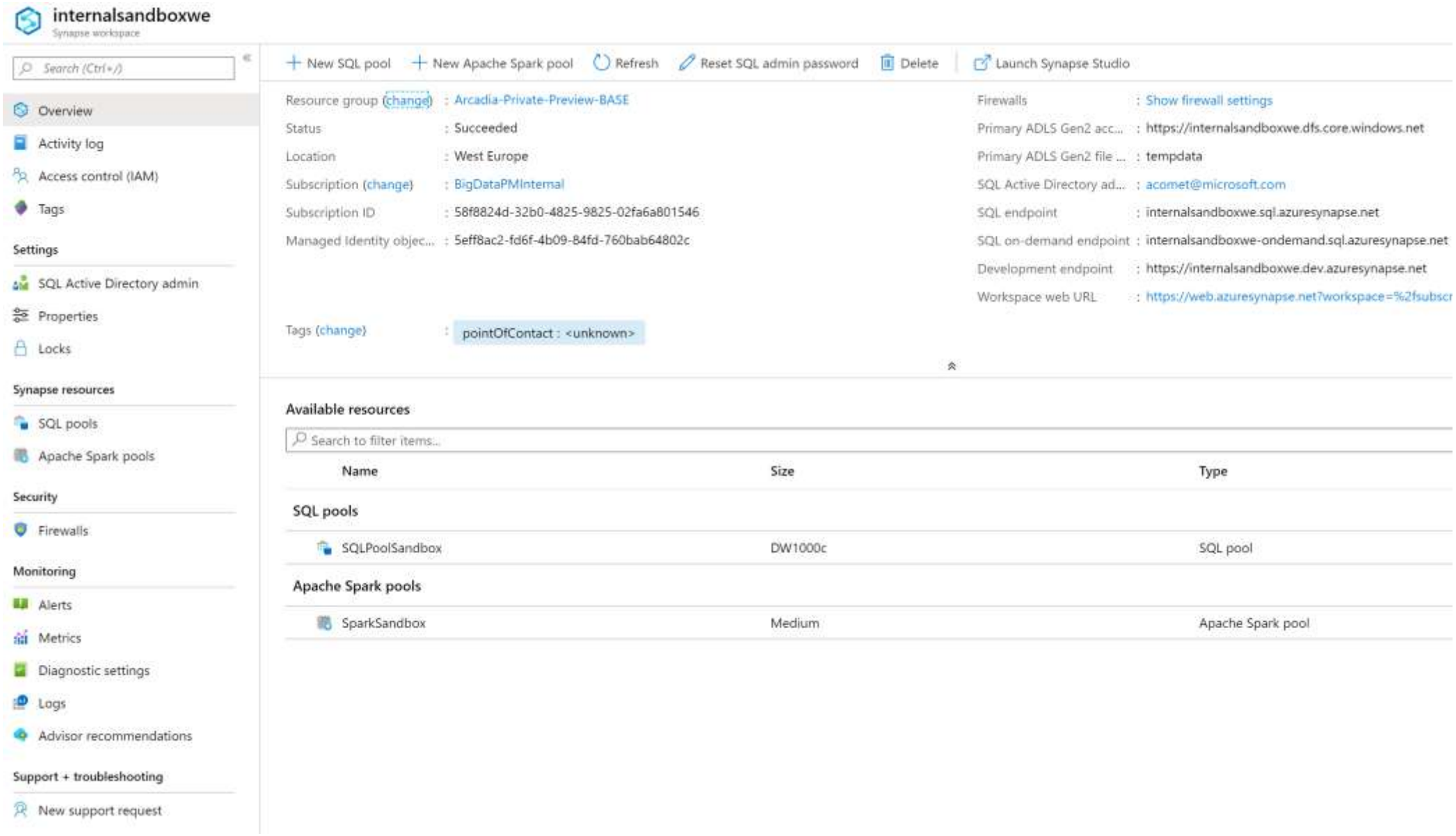
Workspace Name

Region

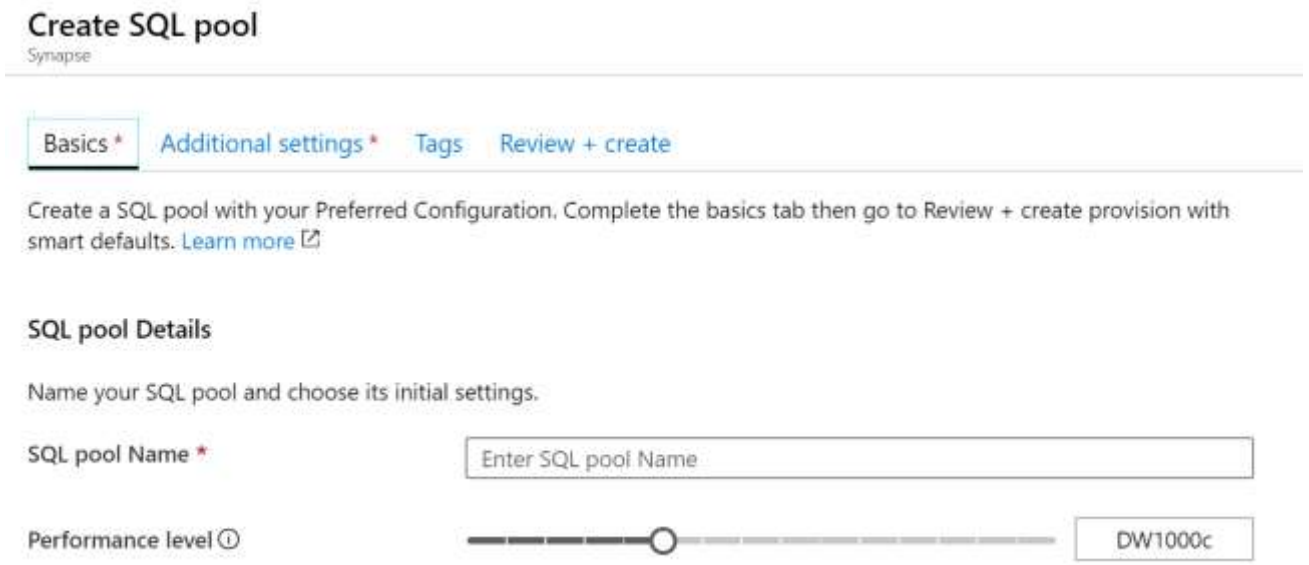
Data Lake Storage Account

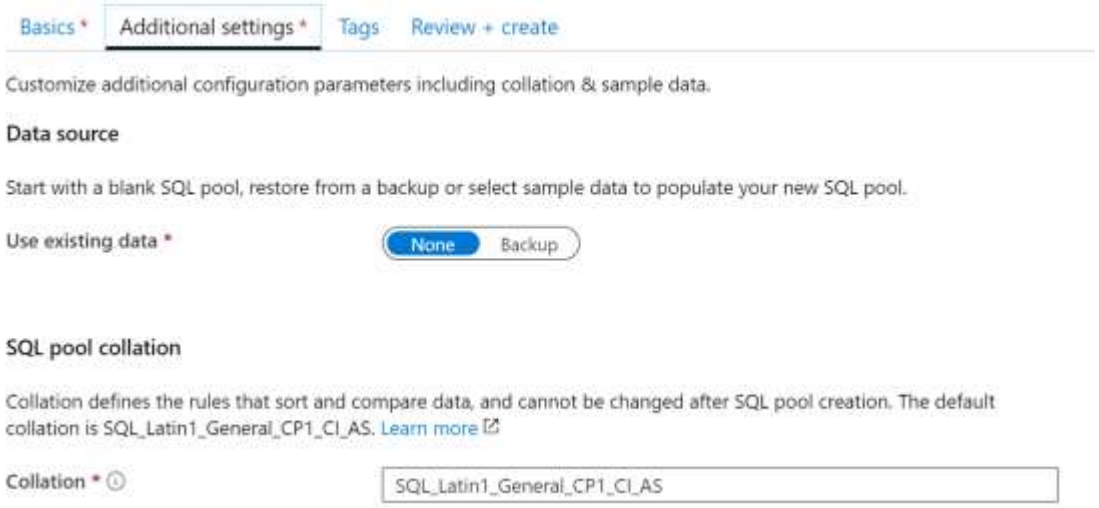


Synapse workspace



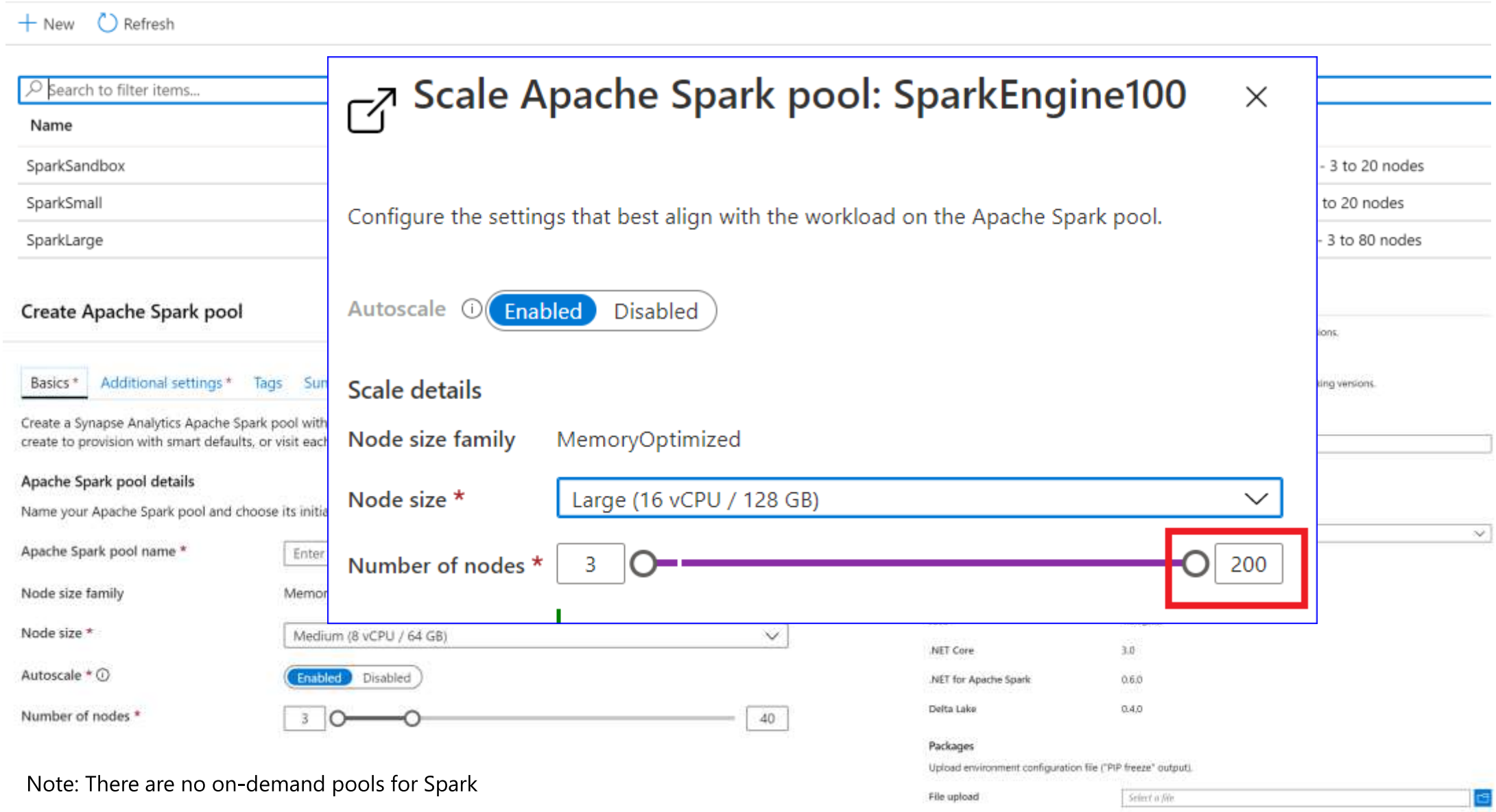
SQL pools





SQL Analytics pool = SQL Data Warehouse

Apache Spark pools



Note: There are no on-demand pools for Spark

Synapse workspace

supplychaindataanalytics

New

Ingest
Use the copy data tool to import data once or on a schedule.

Explore
Learn how to navigate and interact with your data.

Analyze
Learn how to use SQL or Spark to get insights from your data.

Visualize
Build interactive reports with integrated Power BI capabilities.

Resources

Recent | Pinned

No recent resources

Your recently opened resources will show up here.

Useful links

Synapse Analytics overview
Discover the capabilities offered by Synapse and learn how to make the most of them.

Pricing
Learn about pricing details for Synapse capabilities.

Documentation
Visit the documentation center for quickstarts, how-to guides, and references for PowerShell, APIs, etc.

Give feedback
Share your comments or suggestions with us to improve Synapse.

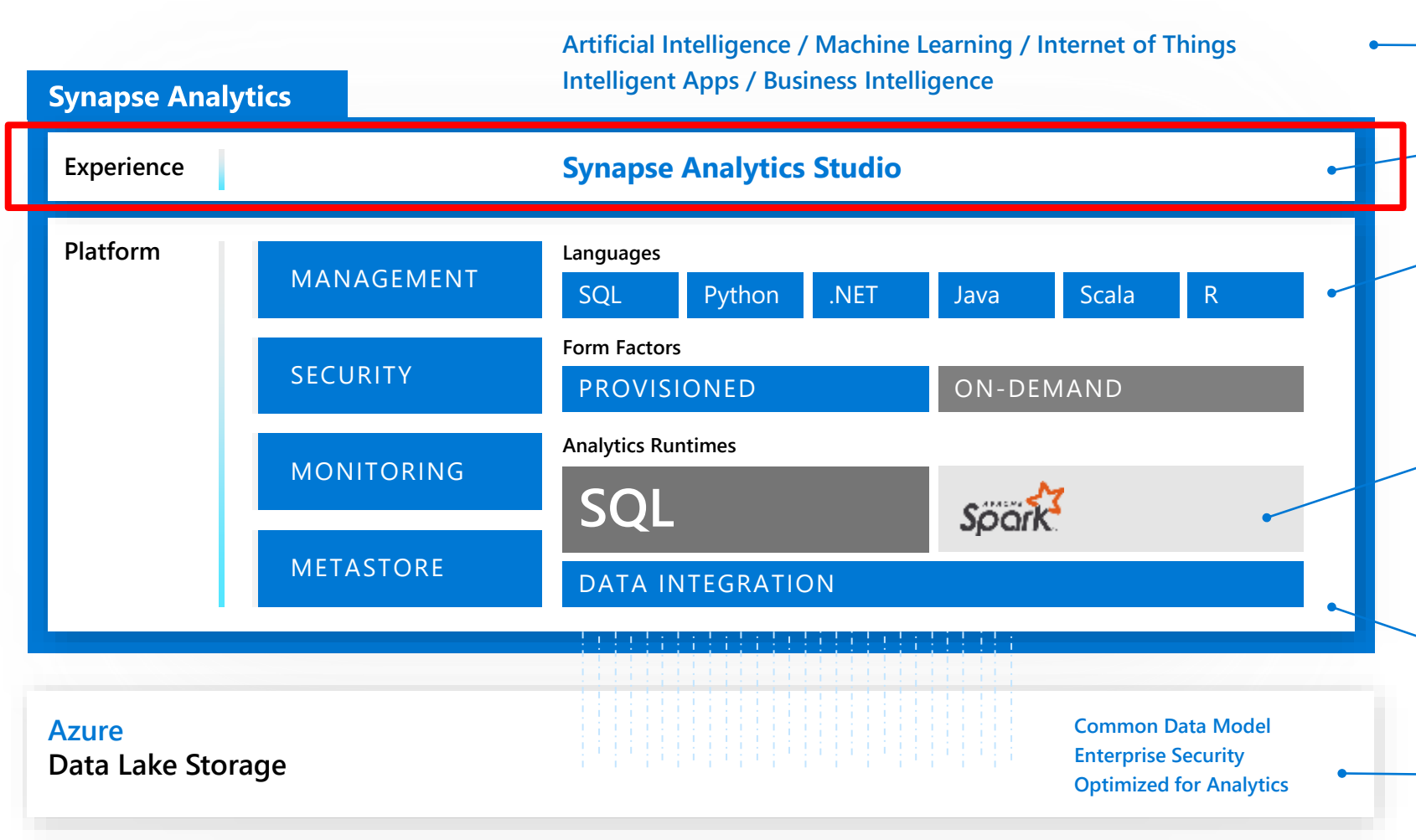
Microsoft

Azure Synapse Analytics

Studio

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

Studio

https://web.azuresyanapse.net

Microsoft Azure

Select workspace

Azure Synapse Analytics is a limitless cloud data warehouse with unmatched time-to-insight. Learn more

Azure Active Directory *

Microsoft

Subscription

Adrian Microsoft Azure Internal Consumption (80a07312-7e2c-49...

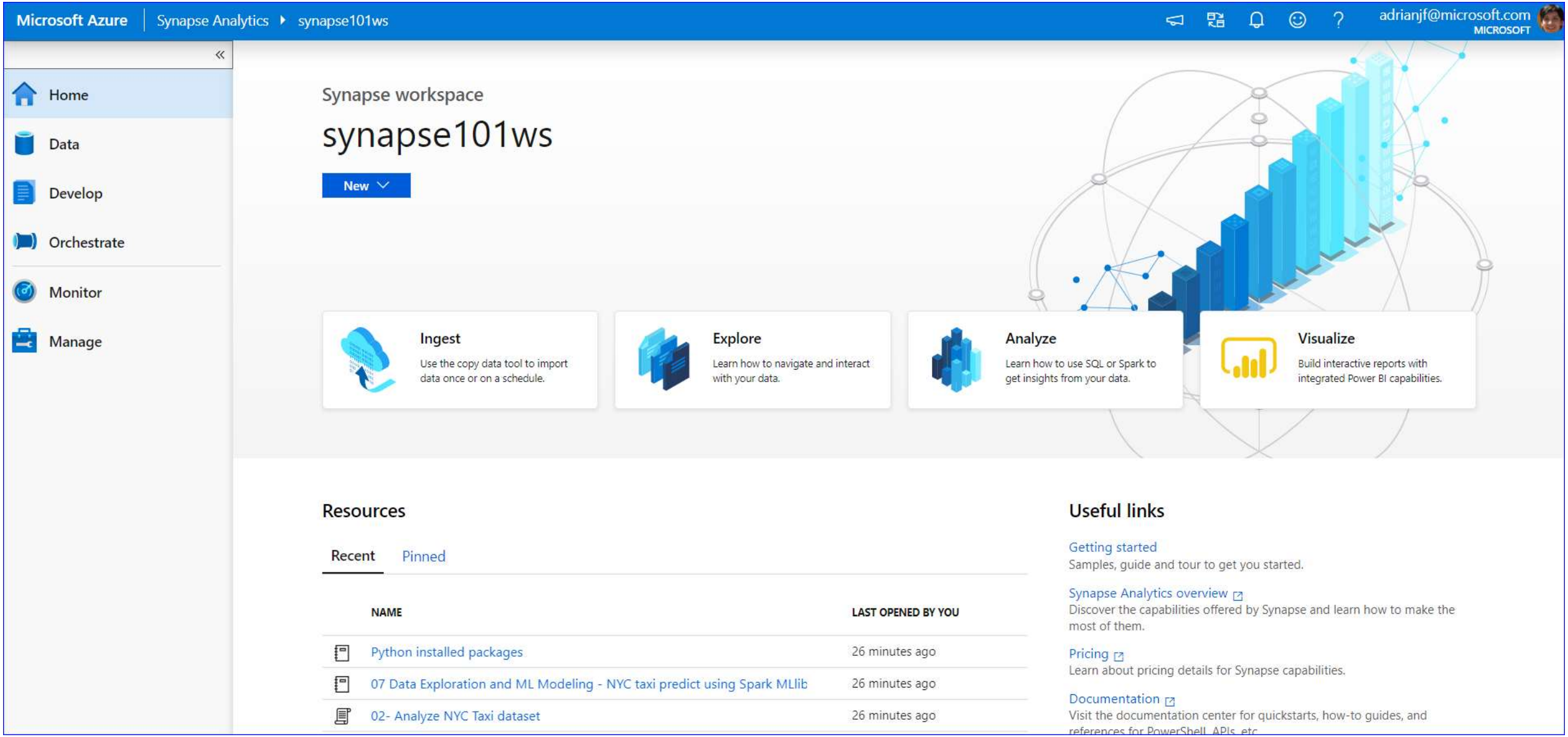
Workspace name *

synapse101ws

Continue

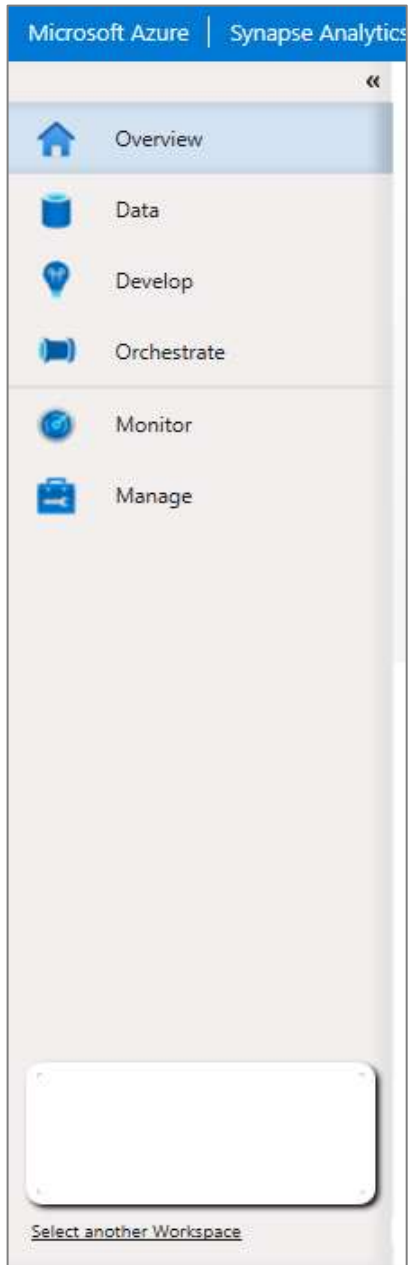
Studio

https://web.azuresyanapse.net



Studio

https://web.azuresynapse.net

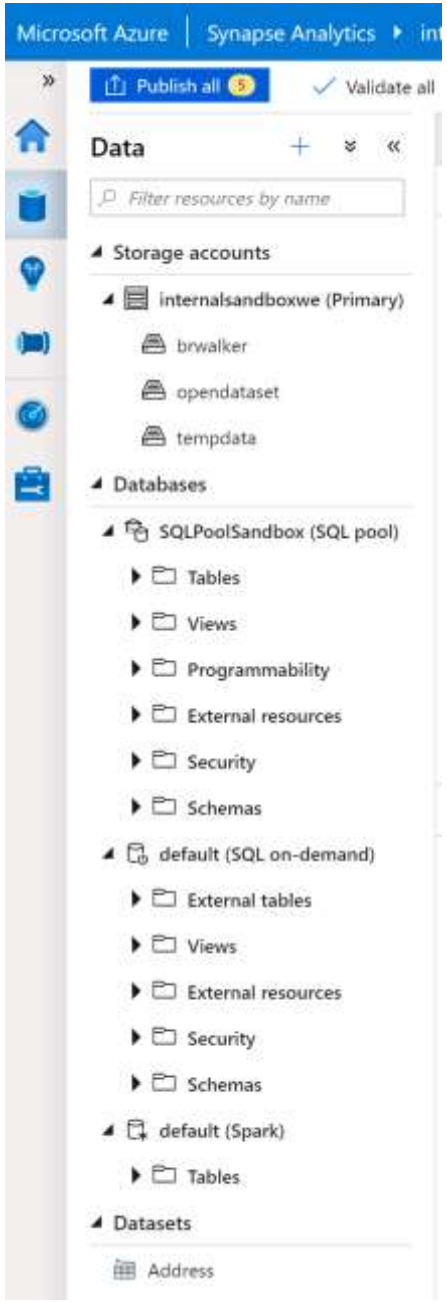


- **Data**: Shows the available data sources available to the workspace. These can exist internally in the workspace (such as a SQL compute database or a Spark database), or externally (such as a Data Lake Store Gen2, or Azure Blob Storage account)
- **Develop**: Shows the different objects used to query or operate with the data, such as SQL scripts, notebooks, data flows, Spark job definitions, Power BI, etc
- **Orchestrate**: Shows the objects used to automate analytics processes (such as pipelines, datasets, etc.)
- **Monitor**: Shows metrics for pipeline runs, trigger runs, integration runtimes, and spark applications
- **Manage**: Create linked services, pipeline triggers, integration runtimes, and manage access to Synapse

Data Hub

Overview

Access to storage accounts, databases, and datasets (for Data Integration)



Develop Hub

Overview

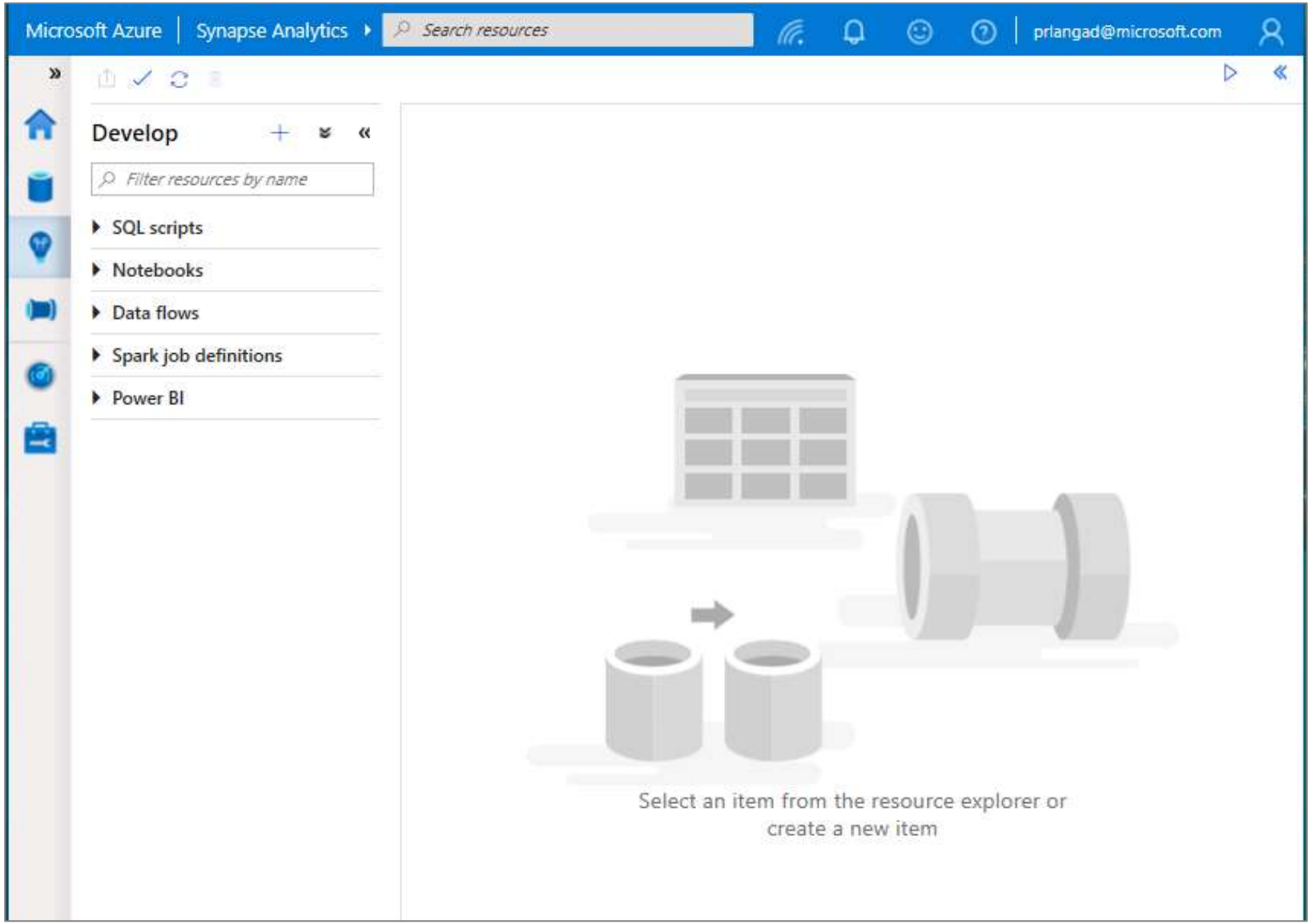
It provides development experience to query, analyze, model data

Benefits

Multiple languages to analyze data under one umbrella

Switch over notebooks and scripts without losing content

Code intellisense offers reliable code development



Develop Hub/SQL scripts

SQL Script

Authoring SQL Scripts

Execute SQL script on provisioned SQL Pool or SQL On-demand

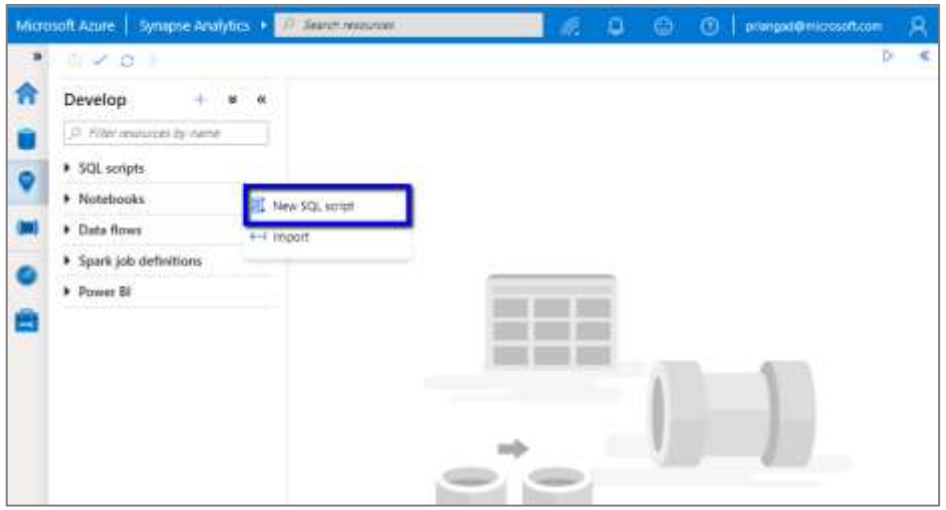
The key thing here is that a "pool" is a set of pre-provisioned resources for DW style workloads. In fact, it's SQLDW with net new enhancements

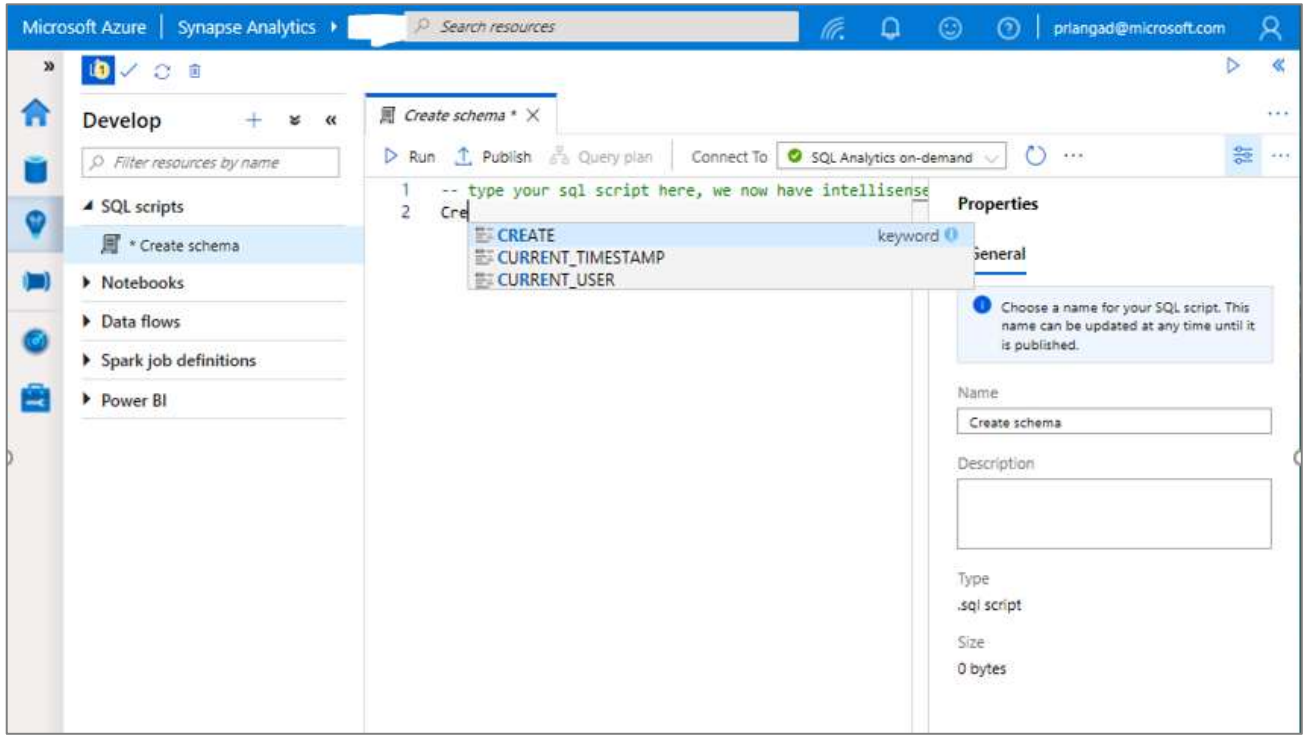
Publish individual SQL script or multiple SQL scripts through Publish all feature

Language support and intellisense

View result in tabular or chart format

Export result





Develop Hub/Notebooks

Notebooks

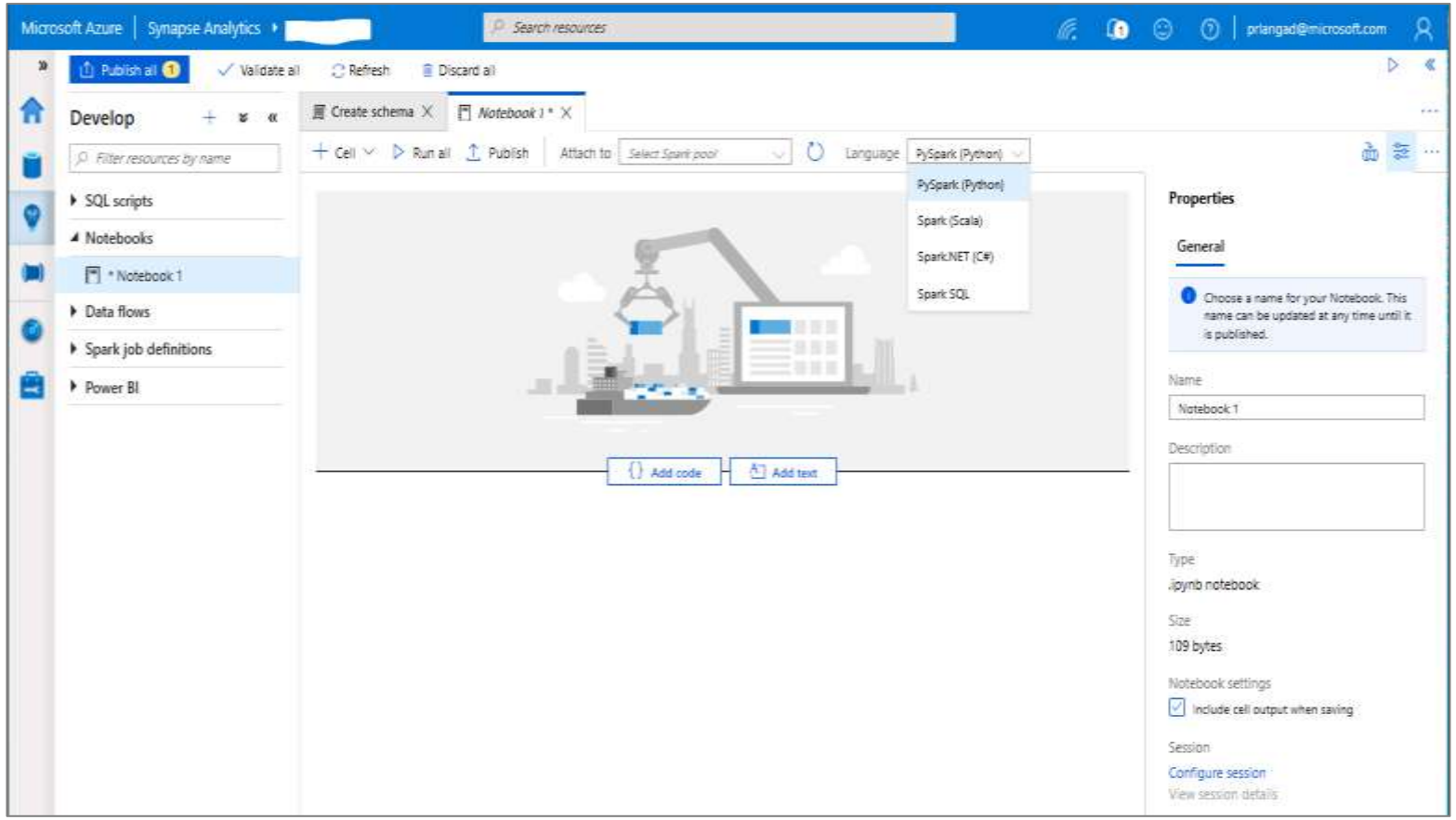
Allows to write multiple languages in one notebook

%%<Name of language>

Offers use of temporary tables across languages

Language support for Syntax highlight, syntax error, syntax code completion, smart indent, code folding

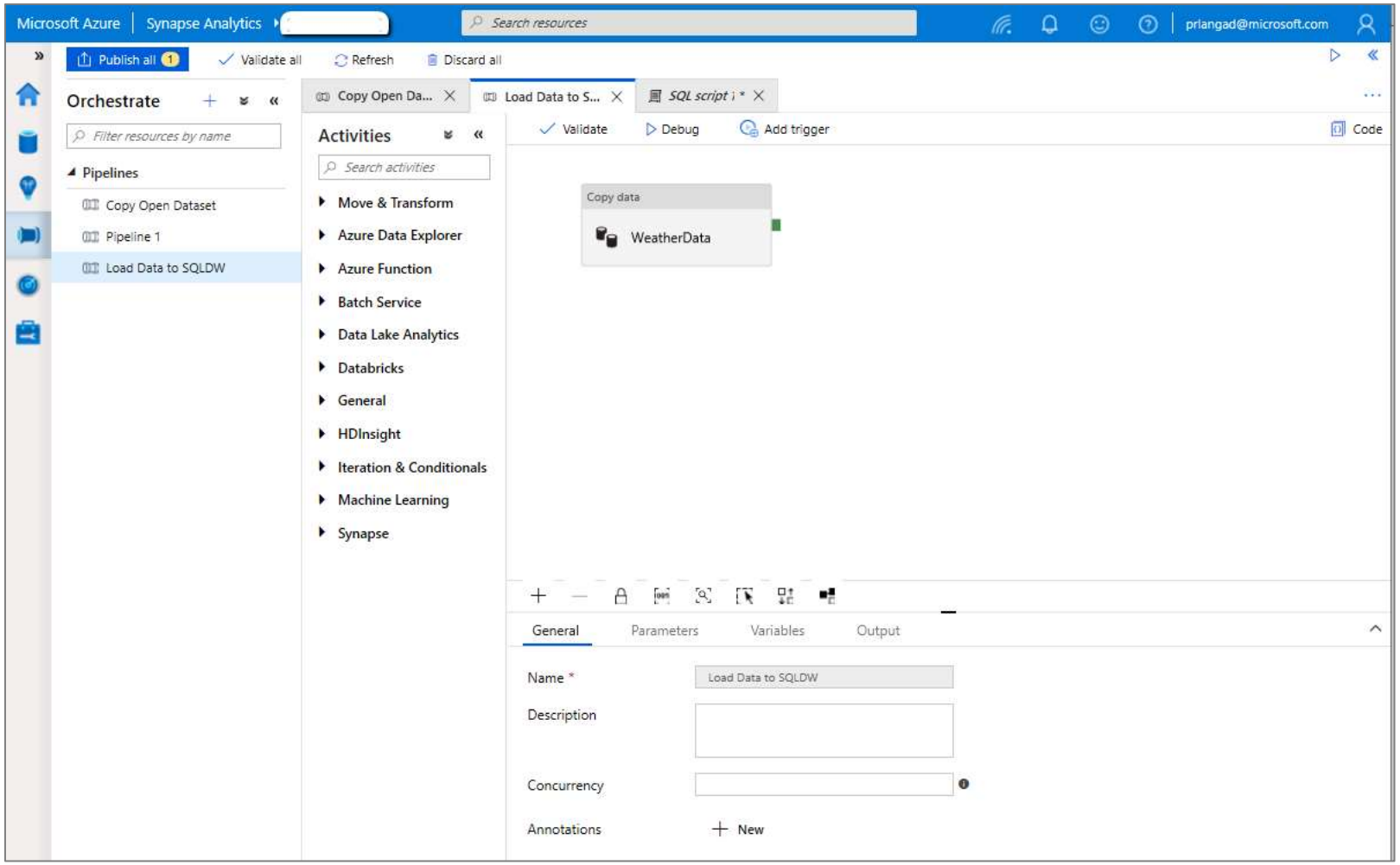
Export results



Orchestrate Hub

Overview

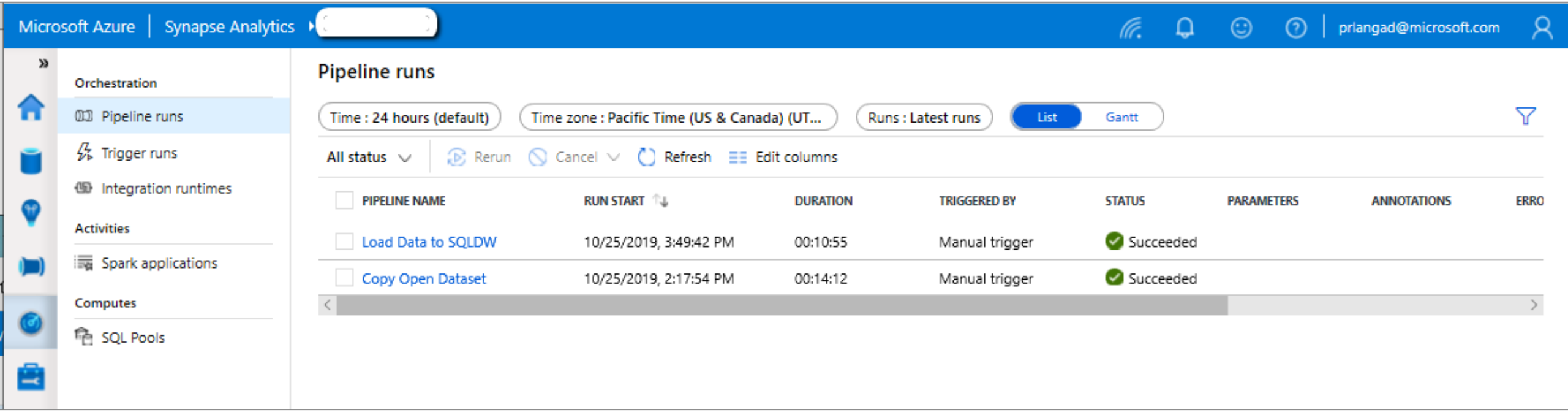
It provides ability to create pipelines to ingest, transform and load data with 90+ inbuilt connectors



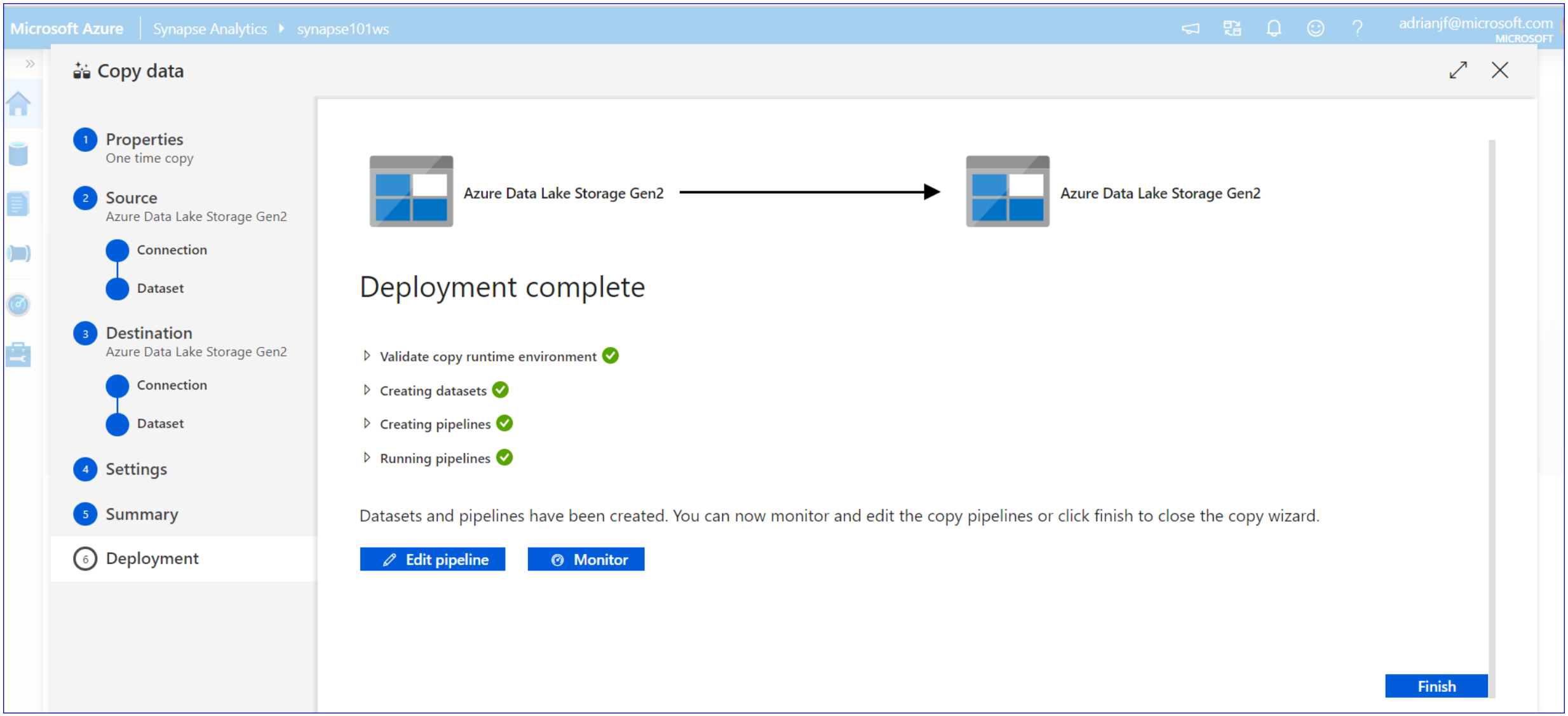
Monitor Hub

Overview

This feature provides ability to monitor orchestration, activities and compute resources.



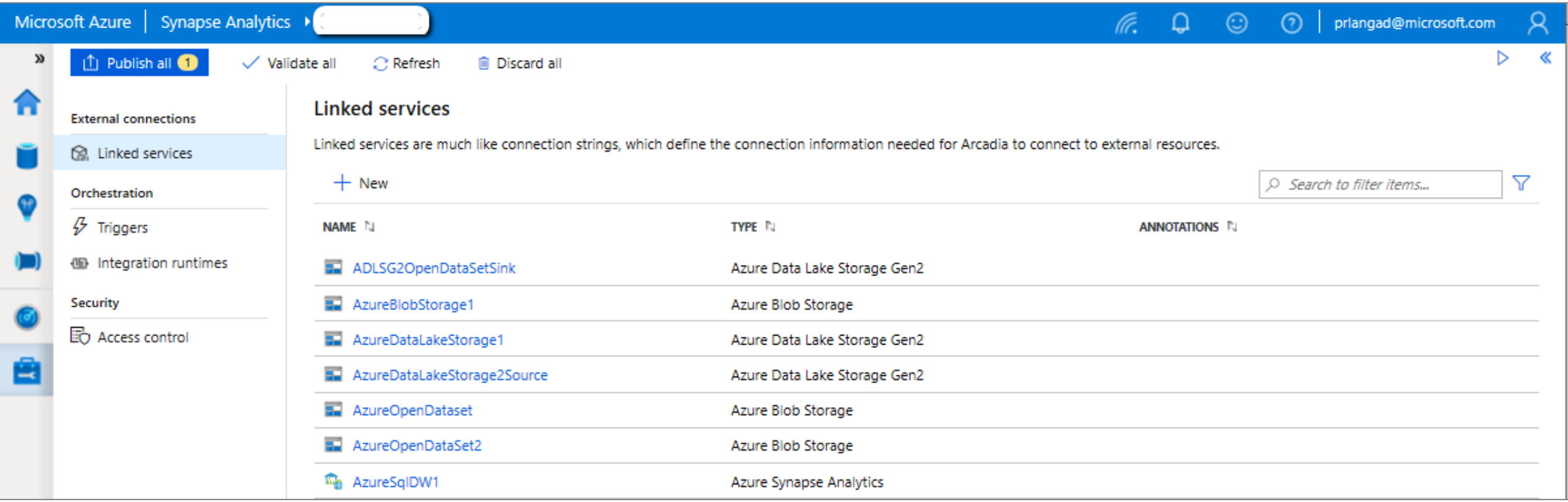
Monitor Hub


Microsoft Azure
Synapse Analytics
synapse101ws
adrianjf@microsoft.com
MICROSOFT
Copy data
1 Properties
One time copy
2 Source
Azure Data Lake Storage Gen2
Connection
Dataset
3 Destination
Azure Data Lake Storage Gen2
Connection
Dataset
4 Settings
5 Summary
6 Deployment
Azure Data Lake Storage Gen2
Azure Data Lake Storage Gen2
Deployment complete
Validate copy runtime environment
Creating datasets
Creating pipelines
Running pipelines
Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close the copy wizard.
Edit pipeline
Monitor
Finish

Manage Hub

Overview

This feature provides ability to manage Linked Services (in Data Integration), Orchestration and Security.



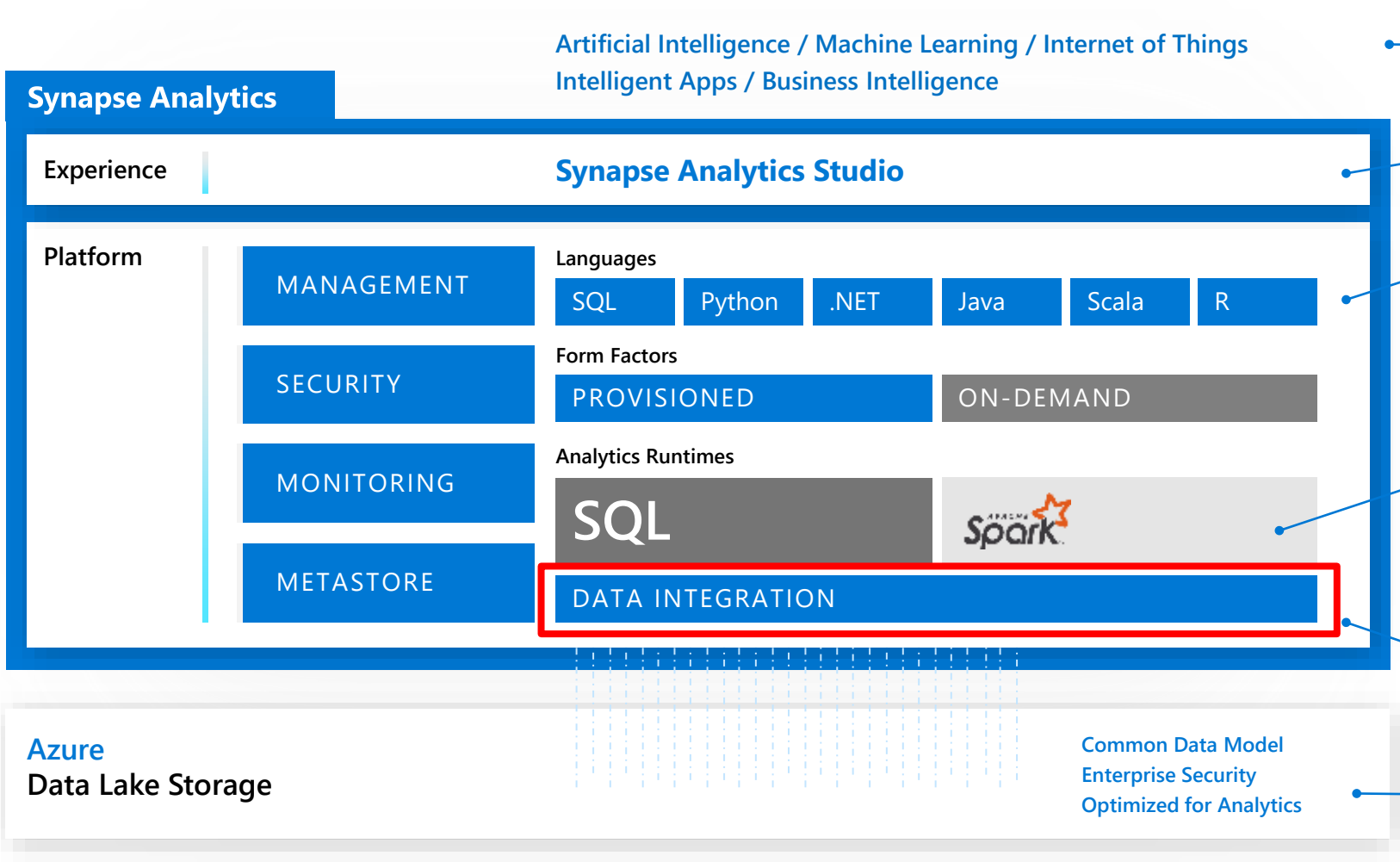
Microsoft

Azure Synapse Analytics

Data Integration

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

Data Integration = Separate version Azure Data Factory (ADF). Will have 1-click migration

Manage – Linked Services

Overview

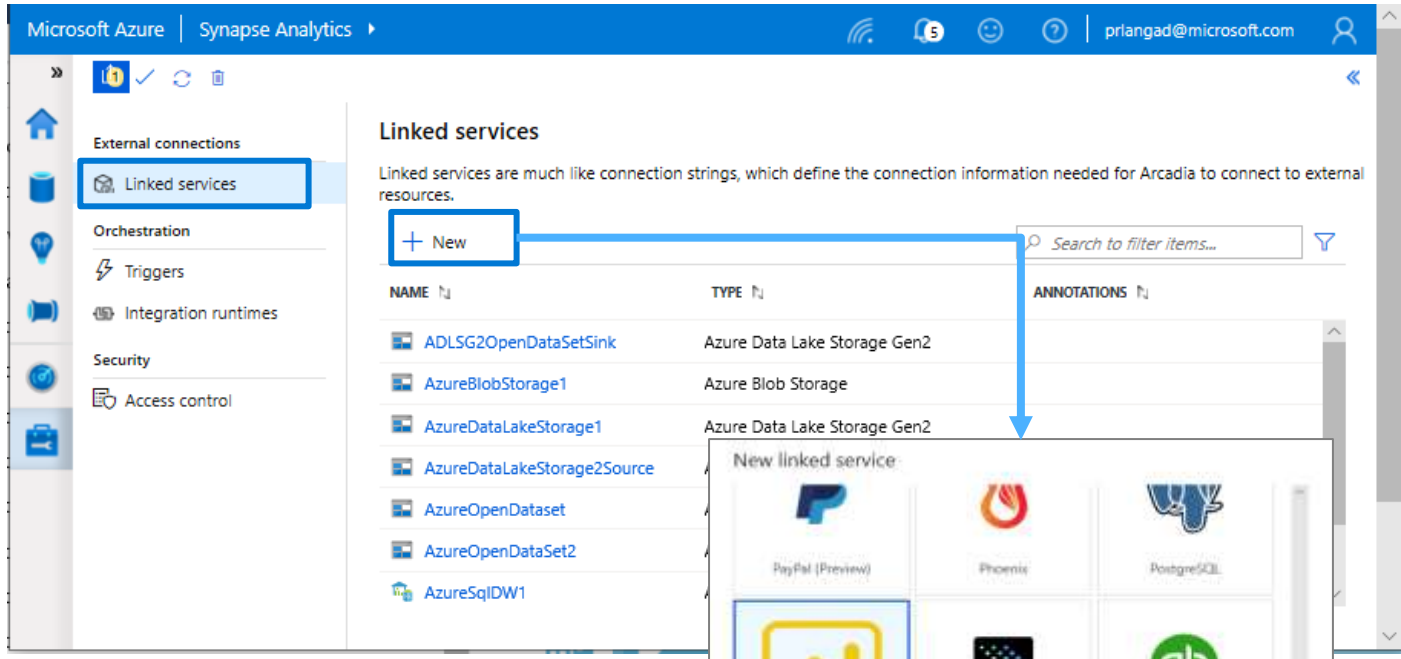
It defines the connection information needed for Pipeline to connect to external resources.

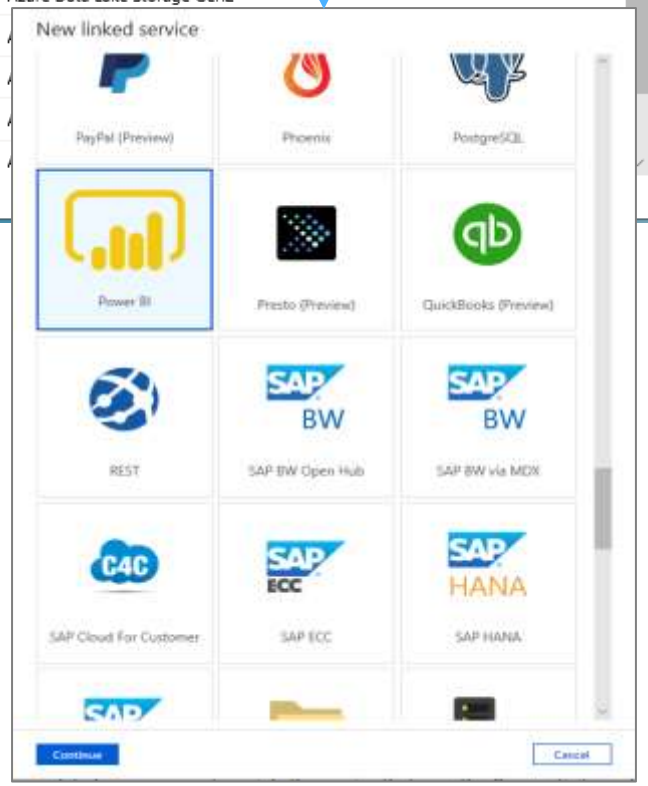
Benefits

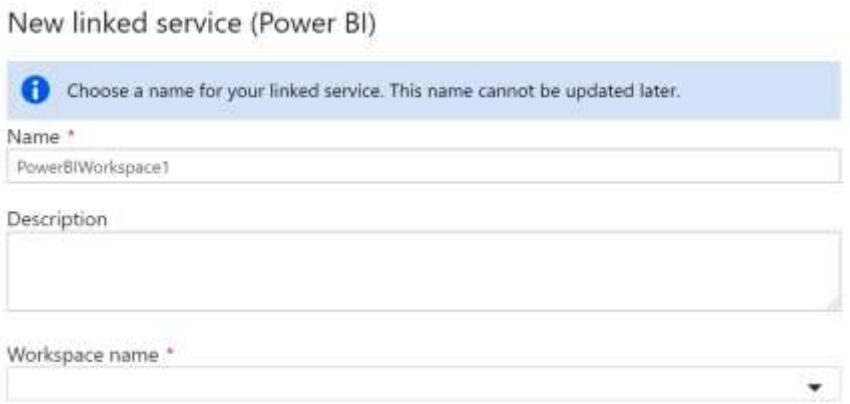
Offers pre-build 90+ connectors

Easy cross platform data migration

Represents data store or compute resources







Manage – Integration runtimes

Overview

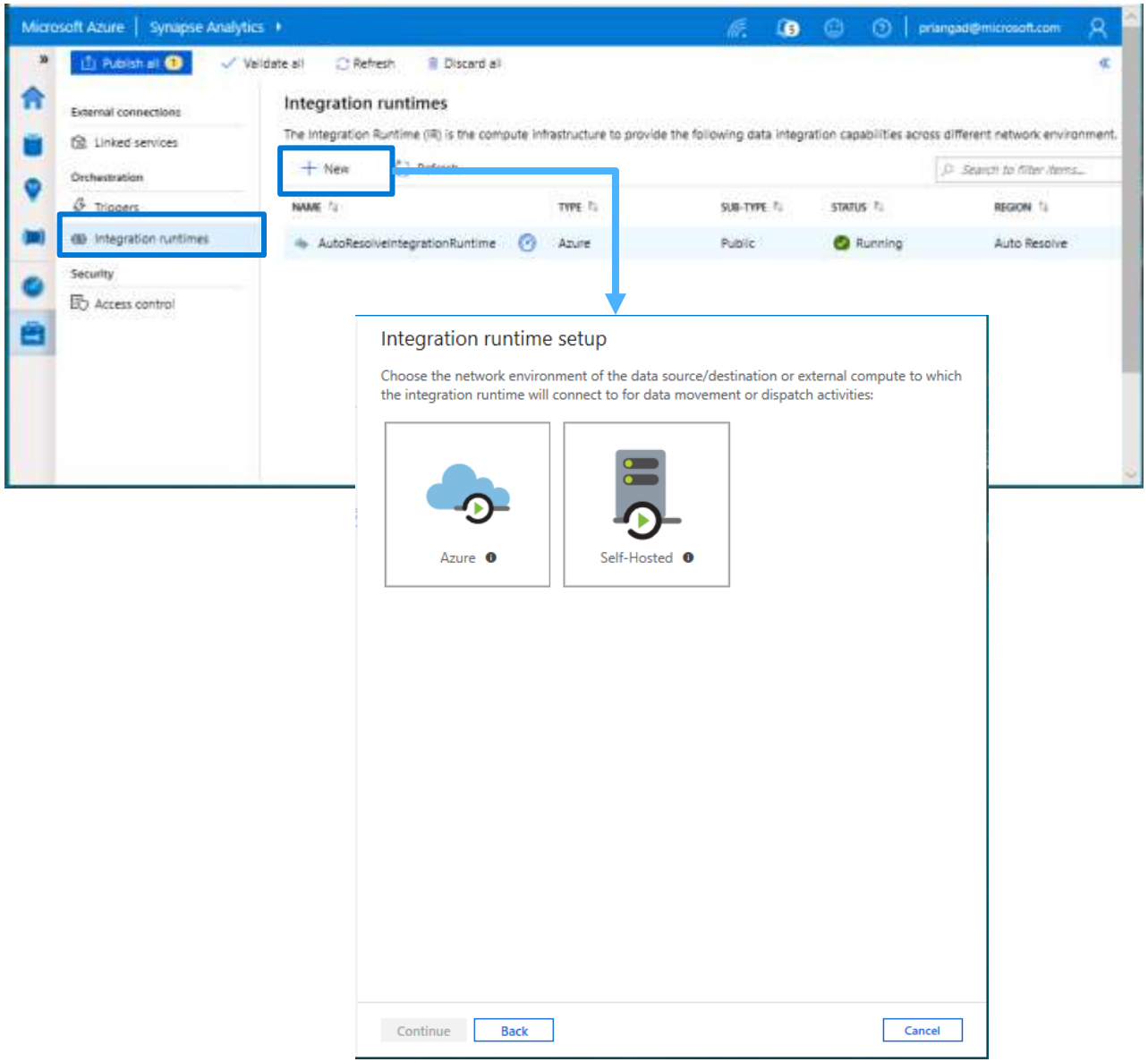
It is the compute infrastructure used by Pipelines to provide the data integration capabilities across different network environments. An integration runtime provides the bridge between the activity and linked Services.

Benefits

Offers Azure Integration Runtime or Self-Hosted Integration Runtime

Azure Integration Runtime – provides fully managed, serverless compute in Azure

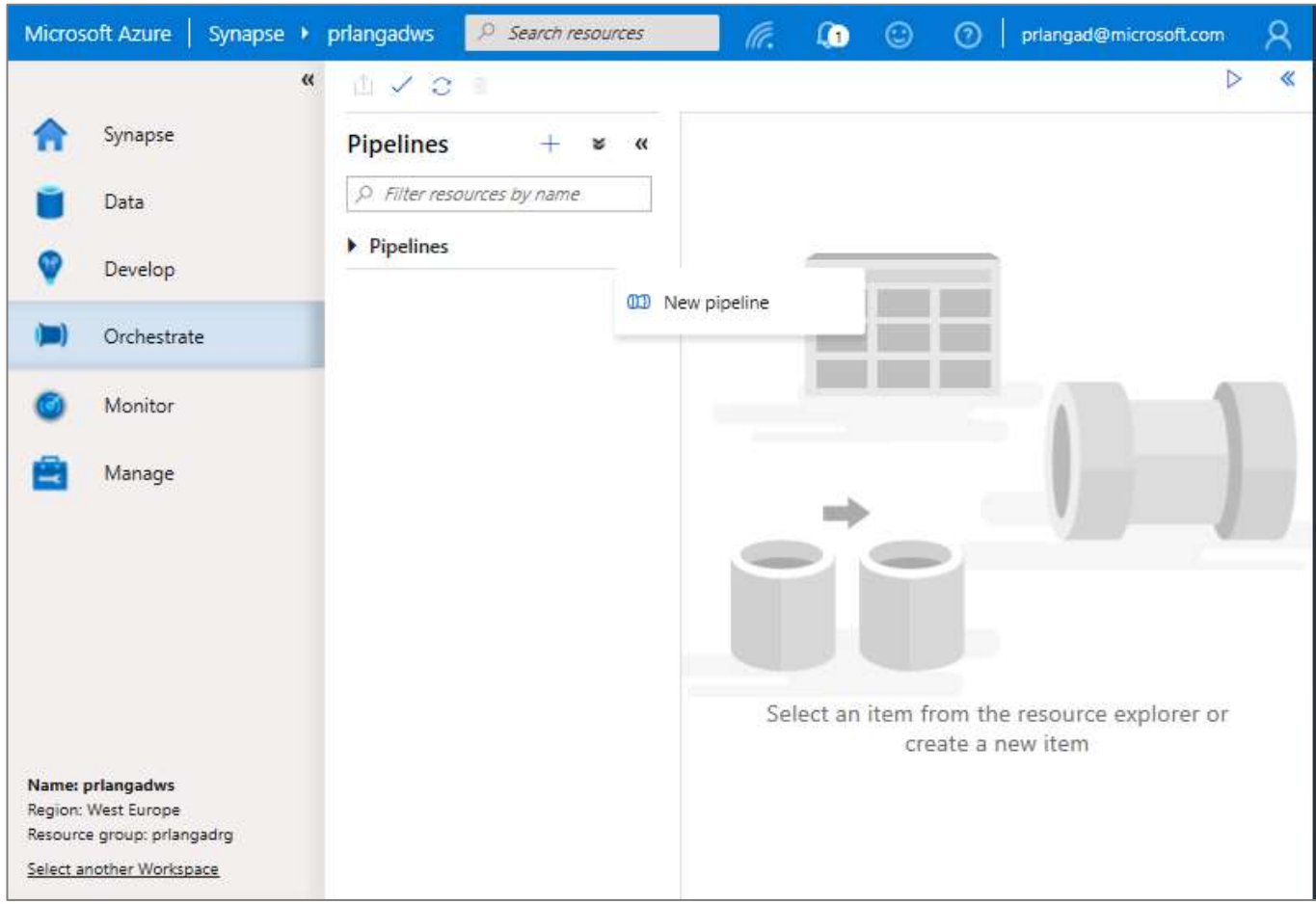
Self-Hosted Integration Runtime – use compute resources in on-premises machine or a VM inside private network



Pipelines

Overview

- Scalable – per job elasticity up to 4GB/s
- Visually author or via code (Python, Spark(Scala), Spark.NET,(C#) etc.)
- Serverless, no infrastructure to manage
- 24 points of presence worldwide
- Self-hostable Integration Runtime for hybrid movement
- Supported file formats: CSV, AVRO, ORC, Parquet, JSON



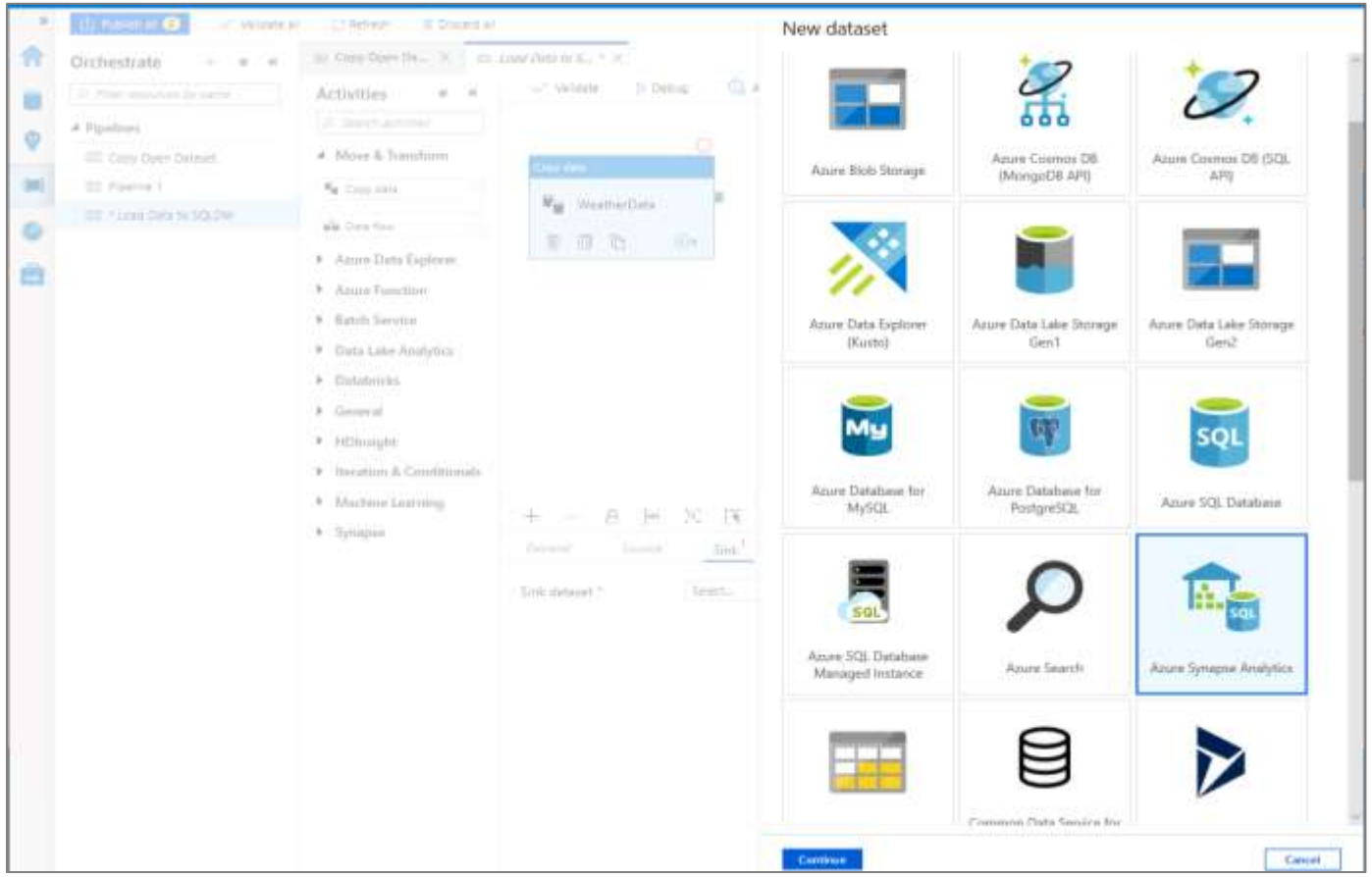
90+ Connectors out of the box

Azure (15)	Database & DW (26)		File Storage (6)	NoSQL (3)	Services and App (28)		Generic (4)
Blob storage	Amazon Redshift	Oracle	Amazon S3	Cassandra	Amazon MWS	Oracle Service Cloud	Generic HTTP
Cosmos DB - SQL API	DB2	Phoenix	File system	Couchbase	Common Data Service	PayPal	Generic OData
Cosmos DB - MongoDB API	Drill	PostgreSQL	FTP	MongoDB	Concur	QuickBooks	Generic ODBC
Data Explorer	Google BigQuery	Presto	Google Cloud Storage		Dynamics 365	Salesforce	Generic REST
Data Lake Storage Gen1	Greenplum	SAP BW Open Hub	HDFS		Dynamics AX	Salesforce Service Cloud	
Data Lake Storage Gen2	HBase	SAP BW via MDX	SFTP		Dynamics CRM	Salesforce Marketing Cloud	
Database for MariaDB	Hive	SAP HANA			Google AdWords	SAP Cloud for Customer (C4C)	
Database for MySQL	Apache Impala	SAP table			HubSpot	SAP ECC	
Database for PostgreSQL	Informix	Spark			Jira	ServiceNow	
File Storage	MariaDB	SQL Server			Magento	Shopify	
SQL Database	Microsoft Access	Sybase			Marketo	Square	
SQL Database MI	MySQL	Teradata			Office 365	Web table	
SQL Data Warehouse	Netezza	Vertica			Oracle Eloqua	Xero	
Search index					Oracle Responsys	Zoho	
Table storage							

Load Data

Overview

It provides the ability to load data from storage account to desired linked service. Load data by manual execution of a pipeline or by orchestration



Benefits

Supports common loading patterns

Polybase for data loading @ scale

Note there is a new COPY statement for loading data

Prep & Transform Data

Overview

It offers data cleansing, transformation, aggregation, conversion, etc

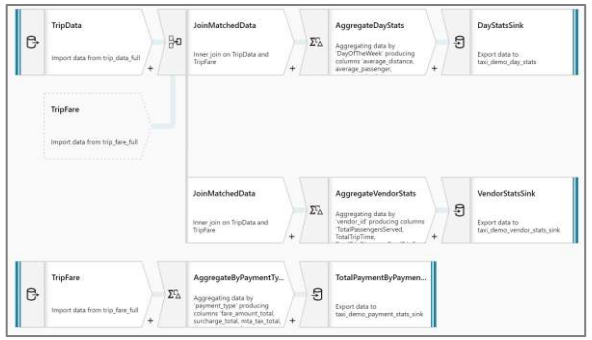
Benefits

Cloud scale via Spark execution

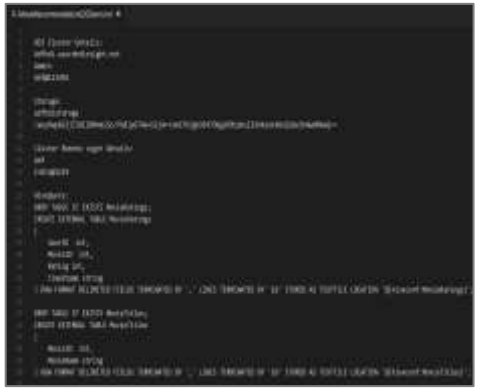
Guided experience to easily build resilient data flows

Flexibility to transform data per user's comfort

Monitor and manage dataflows from a single pane of glass



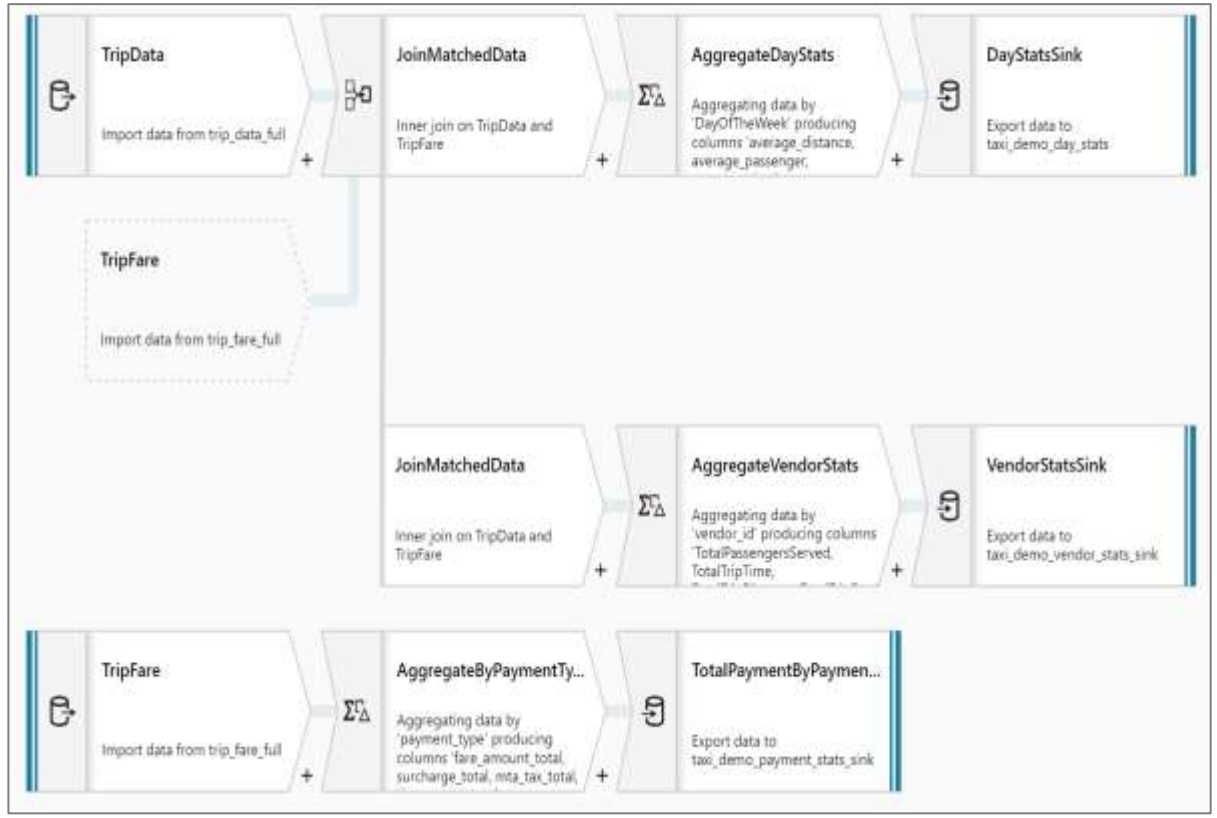
...
not



Prep & Transform Data

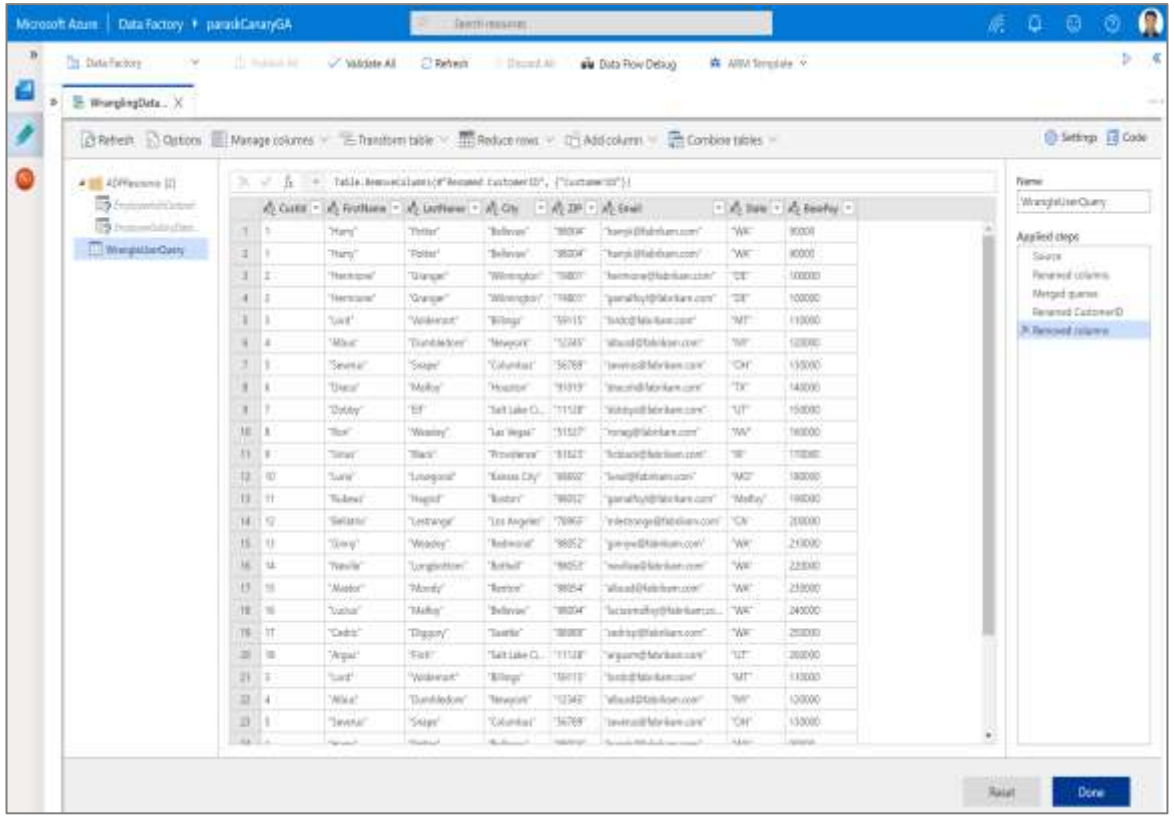
Mapping Dataflow

Code free data transformation @scale



Wrangling Dataflow

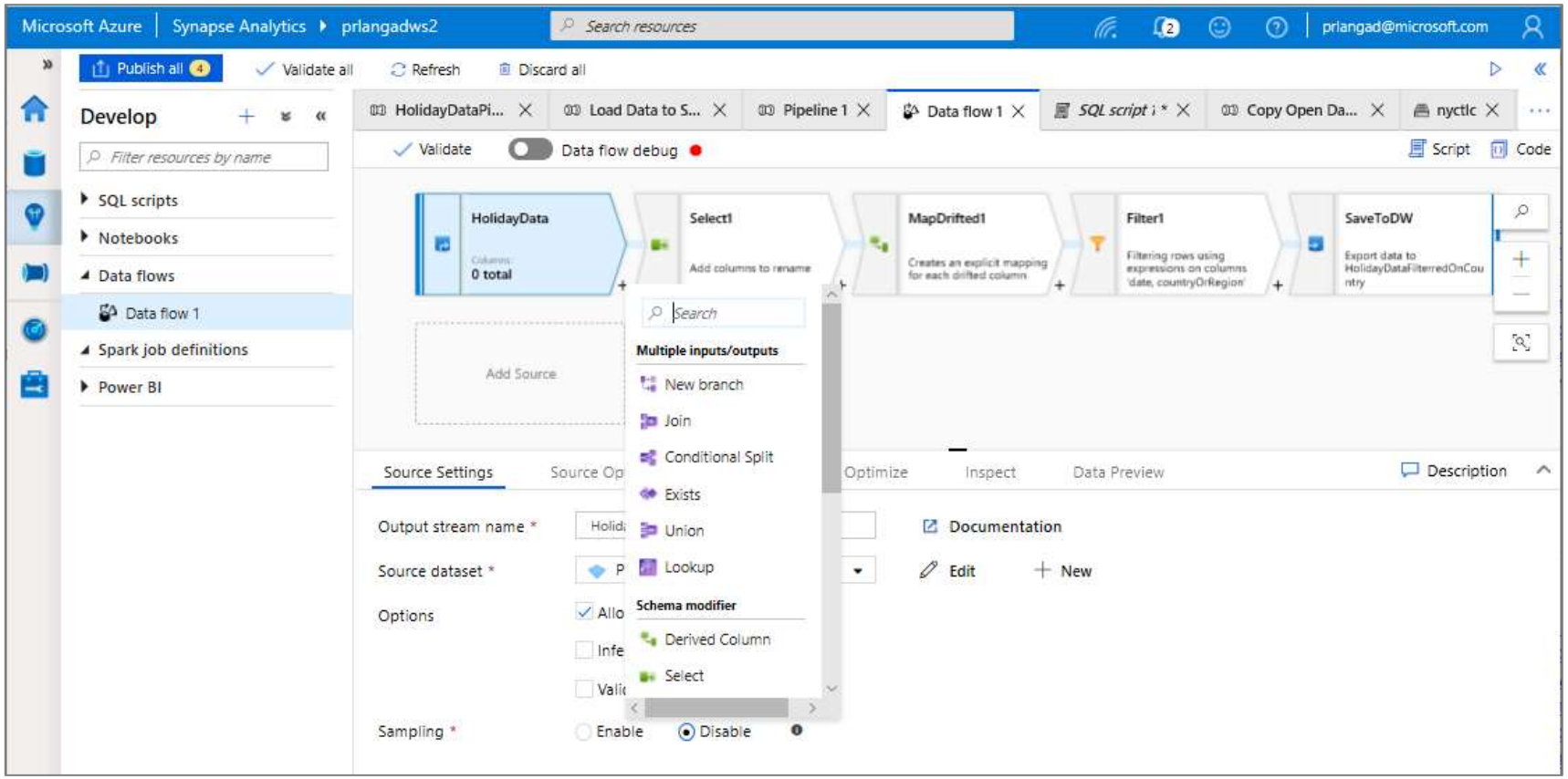
Code free data preparation @scale



Data Flows

Overview

It provides ability to transform data at scale without any coding required. Offers data manipulation operations such as schema modifier, row modifier, input/output – join, branch, union etc.



Benefits

Use data flows as part of pipeline to transform data.

Debug data flows offers data preview to verify applied transformation.

Orchestrate @ Scale

Overview

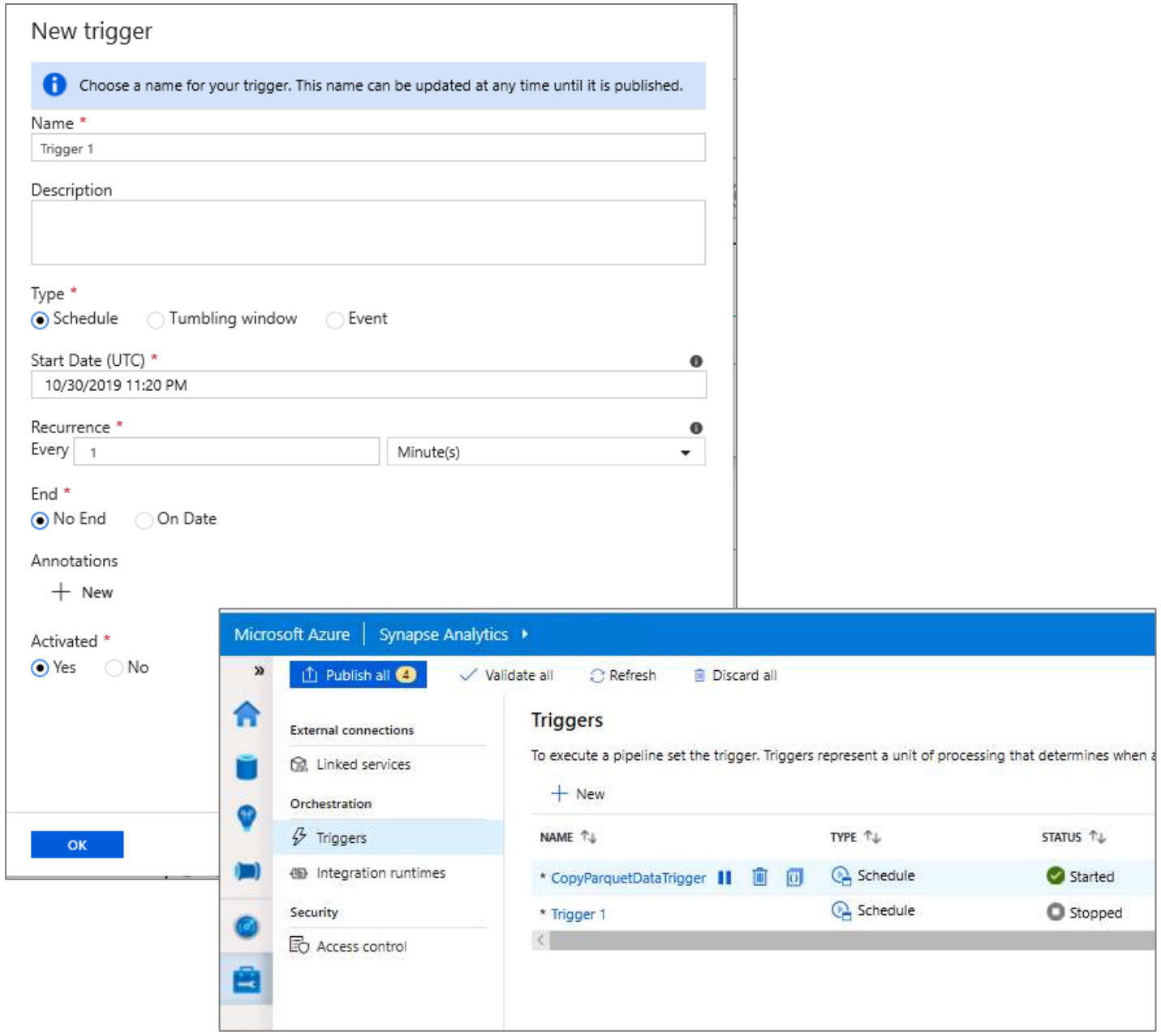
It offers trigger types as - schedule, event, tumbling window. Monitor pipeline runs, control trigger execution.

Benefits

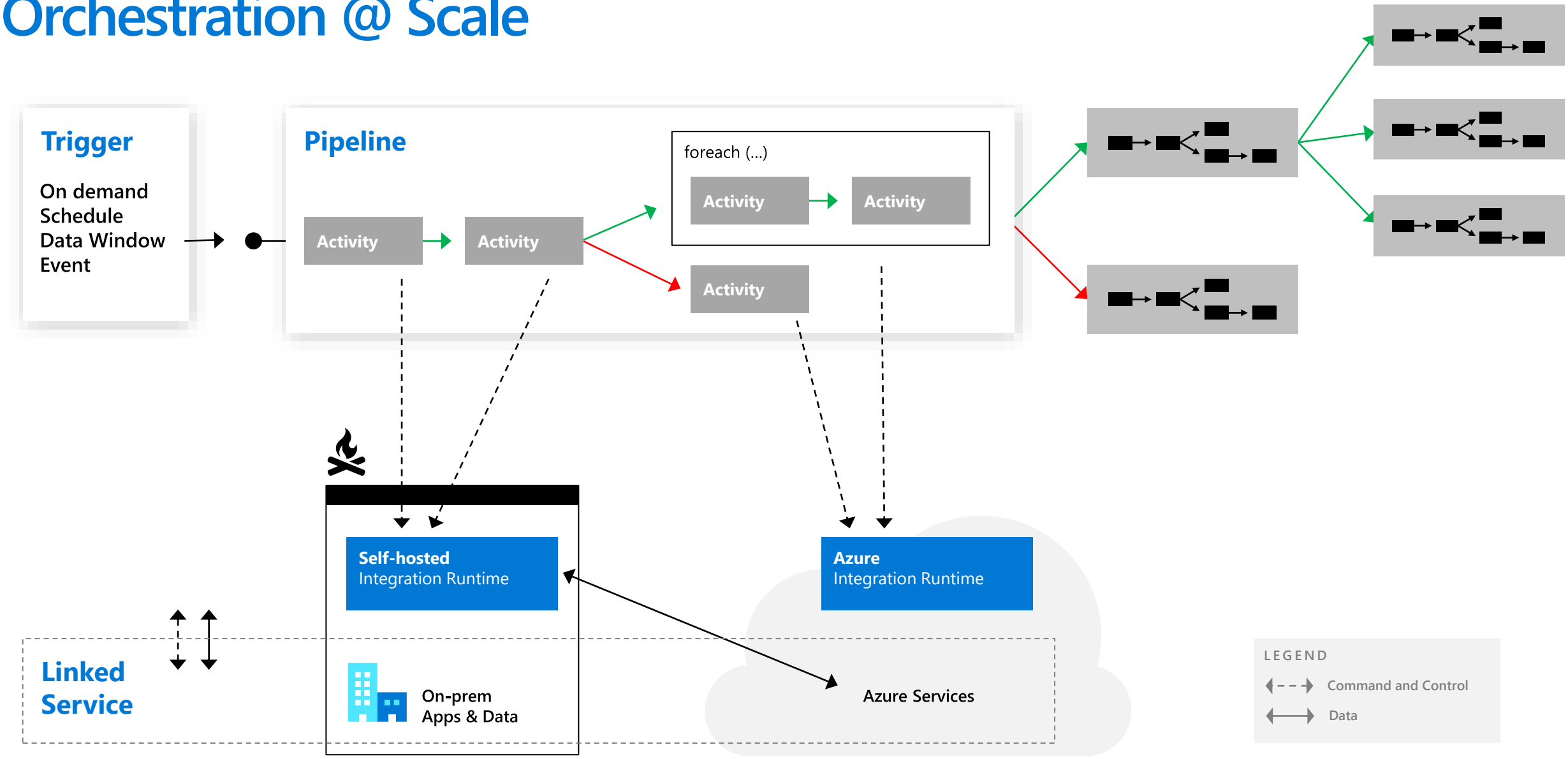
Flexible invocation

Control execution flows, conditional logic

Ability to monitor execution



Orchestration @ Scale



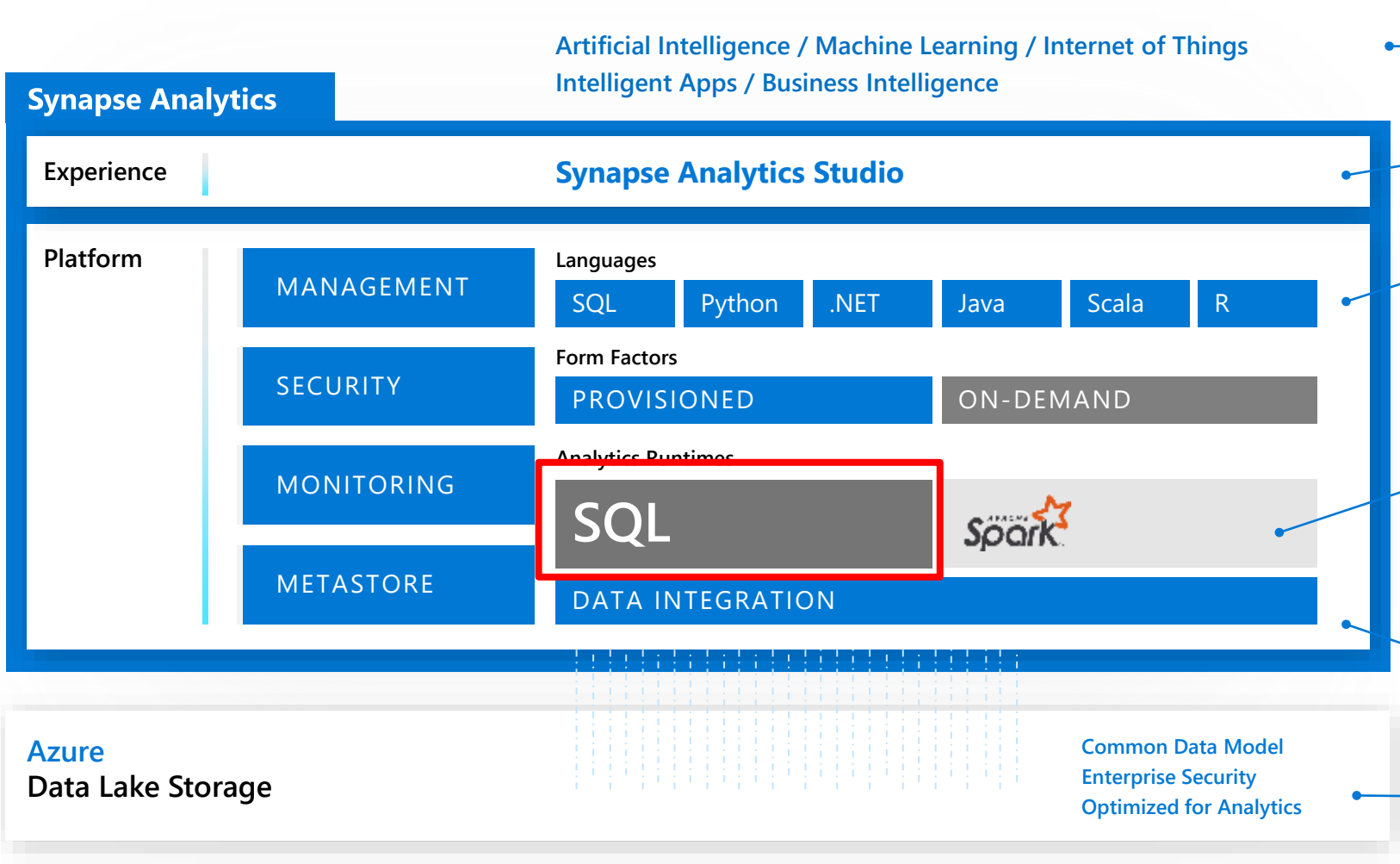
Microsoft

Azure Synapse Analytics

SQL Analytics *(formerly SQL Datawarehouse)*

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

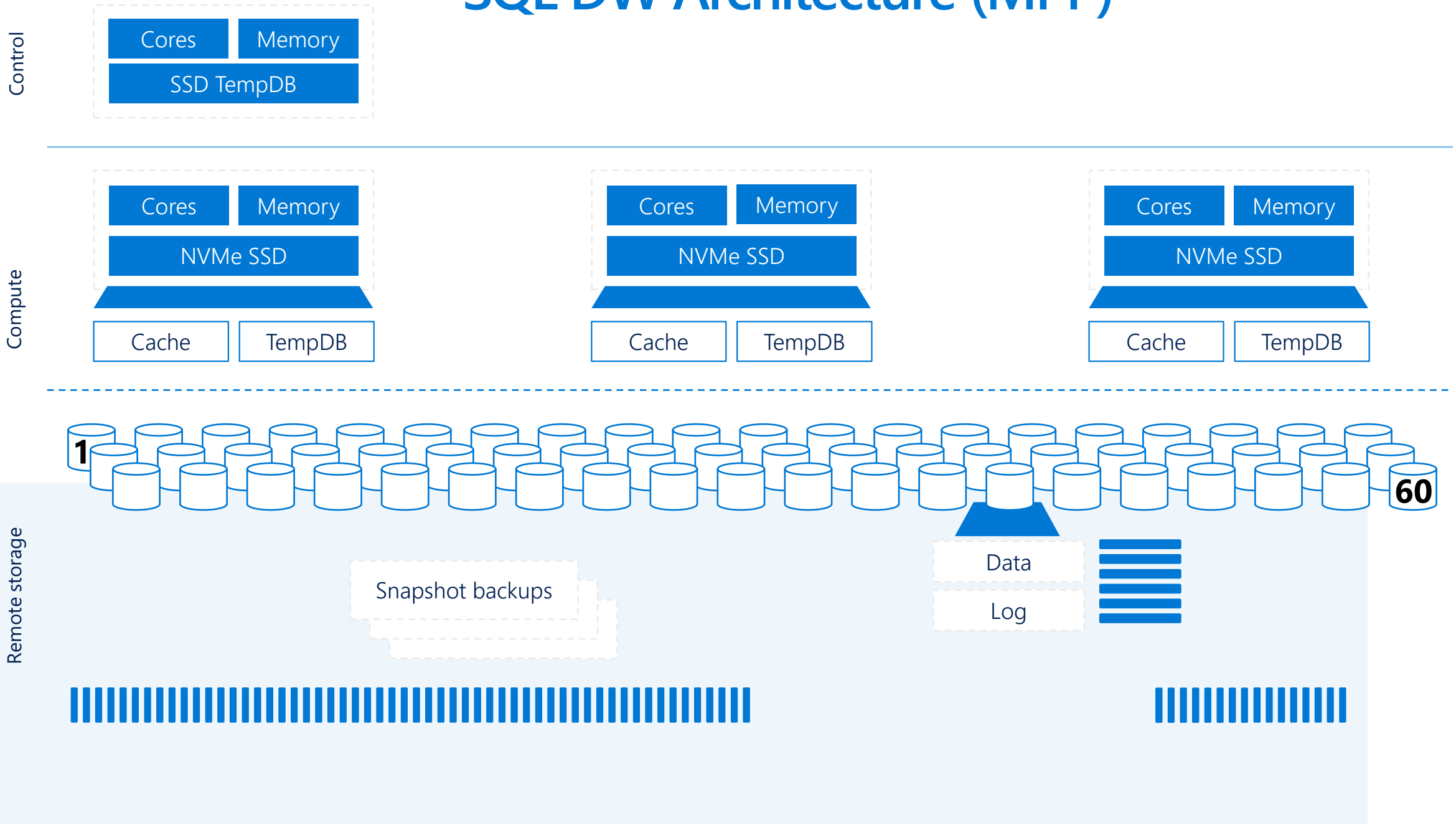
SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

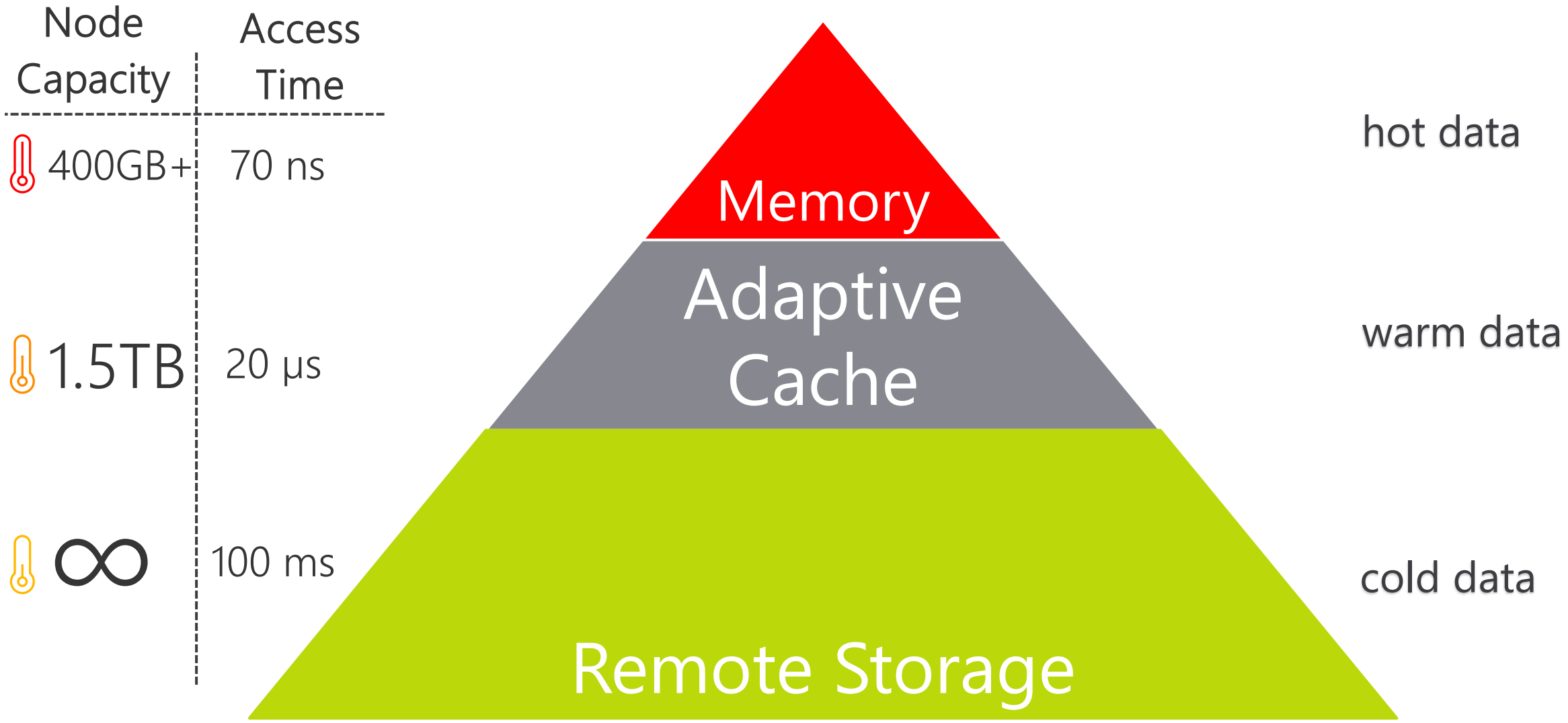
Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

SQL DW Architecture (MPP)



Automated Tiering Of Storage Layers



Data Warehouse Units (DWU)

Normalized amount of compute

Converts to billing units i.e. what you pay

DWUc	
100	
200	
300	
400	
500	→ 1 Compute Node
1000	→ 2 Compute Nodes
1500	→ 3 Compute Nodes
2000	
2500	
3000	
5000	
6000	
7500	
10000	
30000	→ 60 Compute Nodes

Reserved capacity pricing

Overview

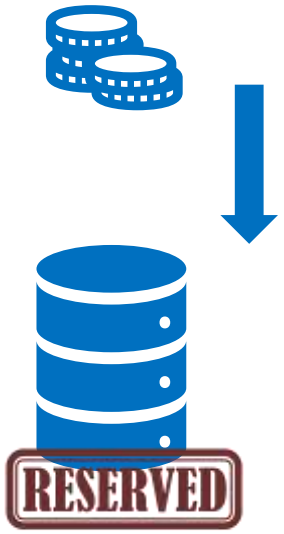
Provides customers with a significant discount (up to 69%) compared to on-demand instance pricing for signing up for an upfront monetary commitment in an Azure region.

Size flexibility: Reserved instance pricing applies to a purchased capacity amount, and is independent of instance size. So, a customer who reserves DW3000c can deploy 1x DW3000c, or 3x 1000c etc. within the same region.

Capacity returns: Paused DW's do not count towards a customer's reservation purchase. This capacity (and associated discount) can be re-used for additional data warehouses.

Pricing details: Reserved capacity discounts are based on full upfront payment and are determined as follows:

- 1 year: 37% discount
- 3 years : 65% discount



Customer buys 1 year's worth of DW10,000c capacity in US West2 region

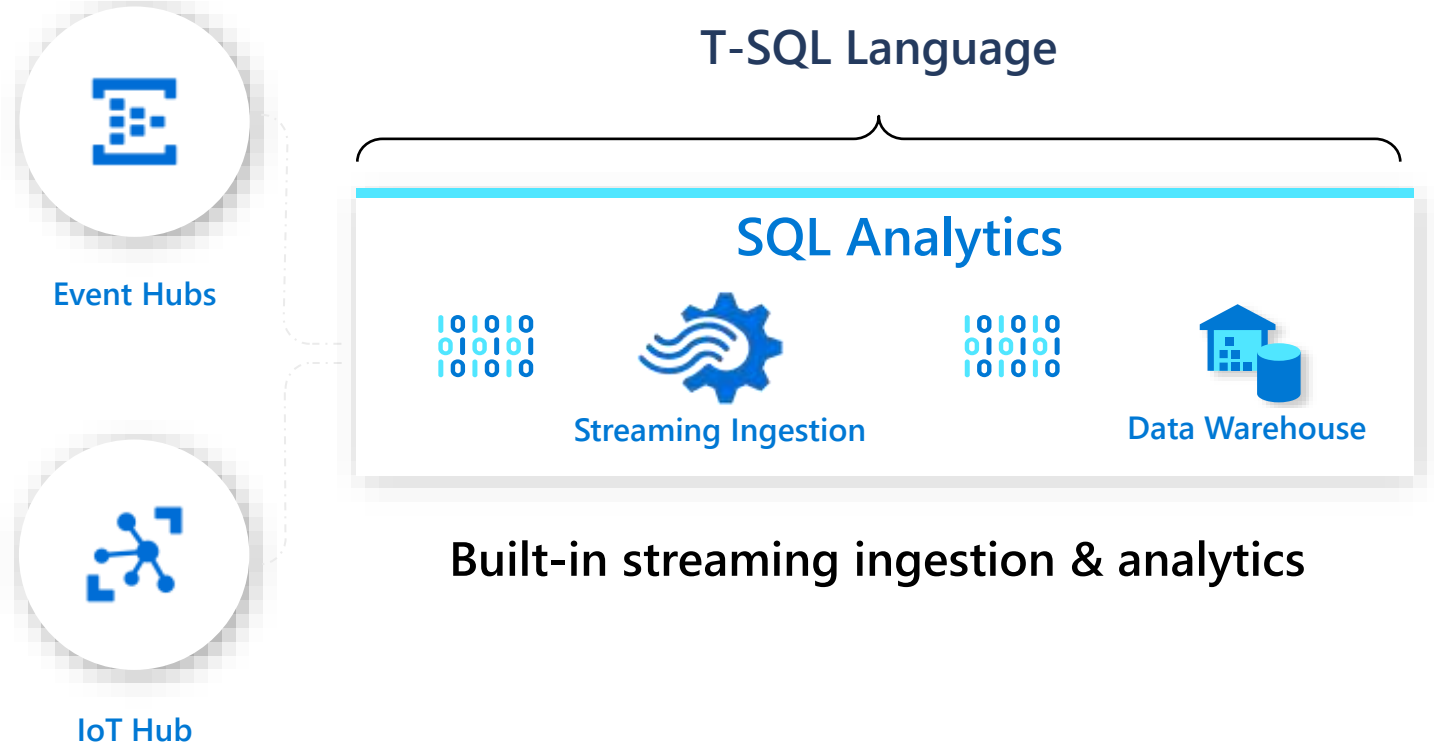
Customer deploys DW10,000c warehouse in US West2 region and receives reserved capacity discount (37% discount)

Scope: **Private Preview (whitelisting needed)**

Heterogenous Data Preparation & Ingestion

Native SQL Streaming

- High throughput ingestion (up to 200MB/sec)
- Delivery latencies in seconds
- Ingestion throughput scales with compute scale
- Analytics capabilities (SQL-based queries for joins, aggregations, filters)

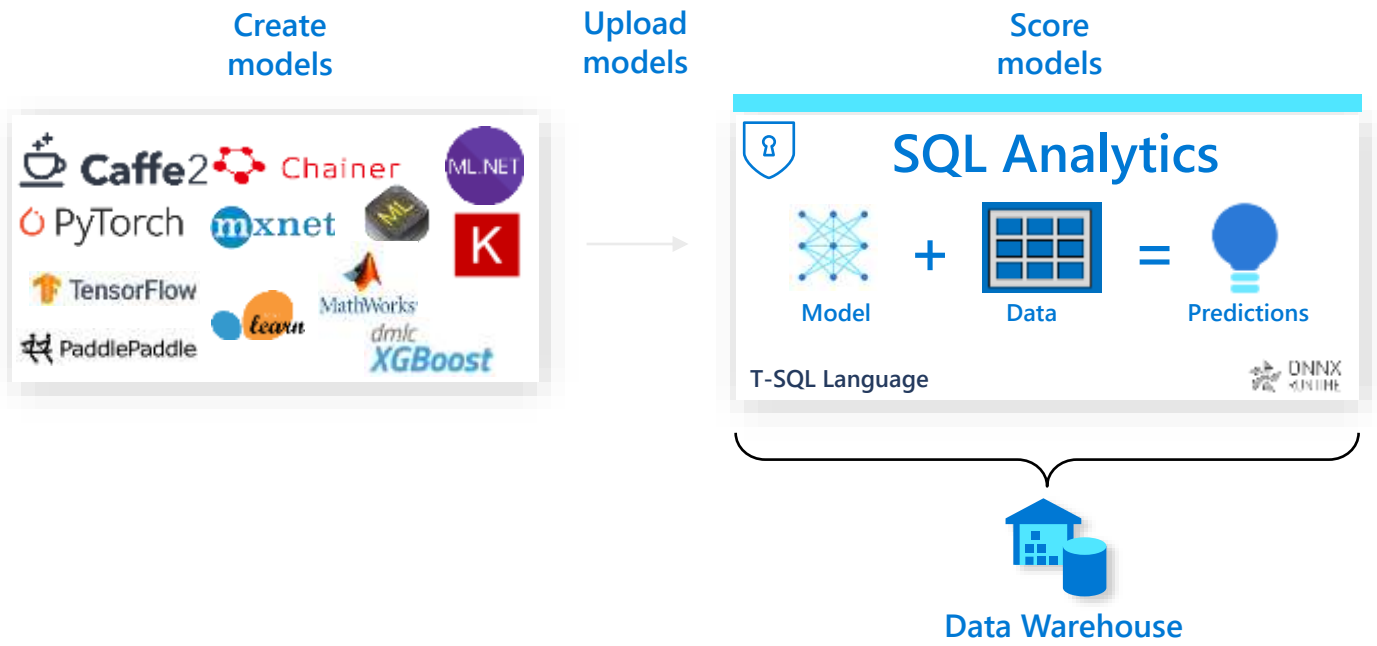


Scope: **Private Preview (whitelisting needed)**

Machine Learning enabled DW

Native PREDICT-ion

- T-SQL based experience (interactive./batch scoring)
- Interoperability with other models built elsewhere
- Execute scoring where the data lives



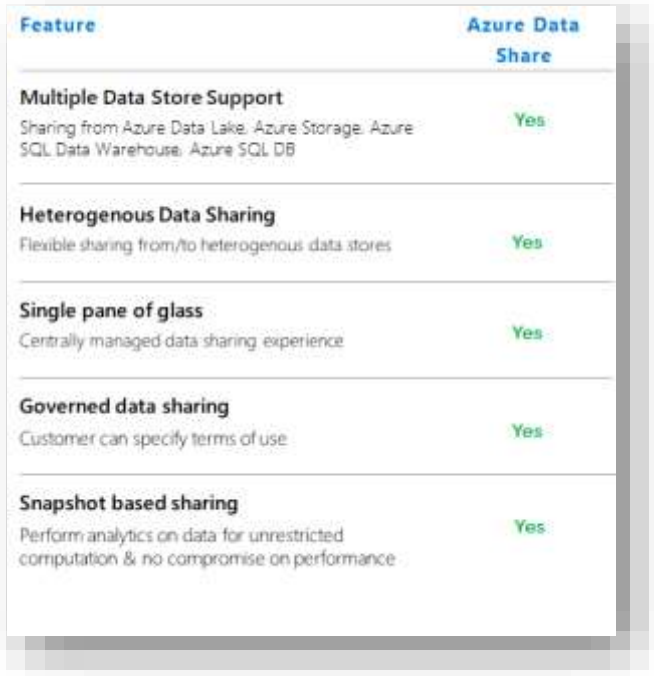

```
--T-SQL syntax for scoring data in SQL DW
SELECT d.*, p.Score
FROM PREDICT(MODEL = @onnx_model, DATA = dbo.mytable AS d)
WITH (Score float) AS p;
```

Scope: **Generally Available**

Azure Data Share

Enterprise data sharing

- Share from DW to DW/DB/other systems
- Choose data format to receive data in (CSV, Parquet)
- One to many data sharing
- Share a single or multiple datasets



Feature	Azure Data Share
Multiple Data Store Support Sharing from Azure Data Lake, Azure Storage, Azure SQL Data Warehouse, Azure SQL DB	Yes
Heterogenous Data Sharing Flexible sharing from/to heterogenous data stores	Yes
Single pane of glass Centrally managed data sharing experience	Yes
Governed data sharing Customer can specify terms of use	Yes
Snapshot based sharing Perform analytics on data for unrestricted computation & no compromise on performance	Yes

Any Azure Data Sources

Share data from any Azure regions and data stores

Single Pane of Glass

Manage and monitor data sharing with multiple organizations

Rich Analytics Tools

Use Azure analytics tools to prepare data and derive insights

Governance

Control data access governed by enterprise policies

Monetization

Charge for data or cost of data curation and access

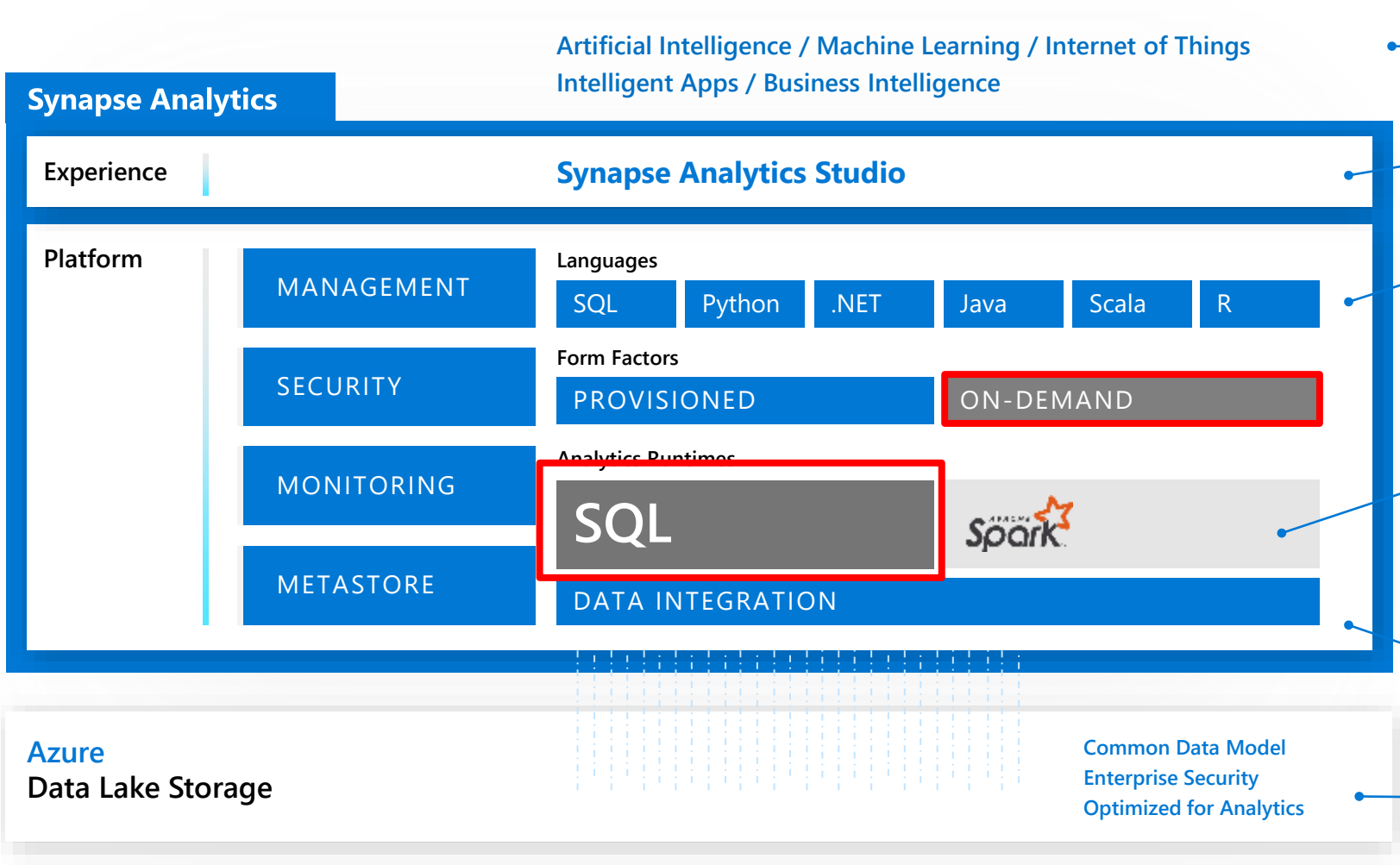
Microsoft

Azure Synapse Analytics

SQL On-Demand

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

SQL On-Demand

Overview

An interactive query service that provides T-SQL queries over high scale data in Azure Storage.

Benefits

Serverless

No infrastructure

Pay only for query execution

No ETL

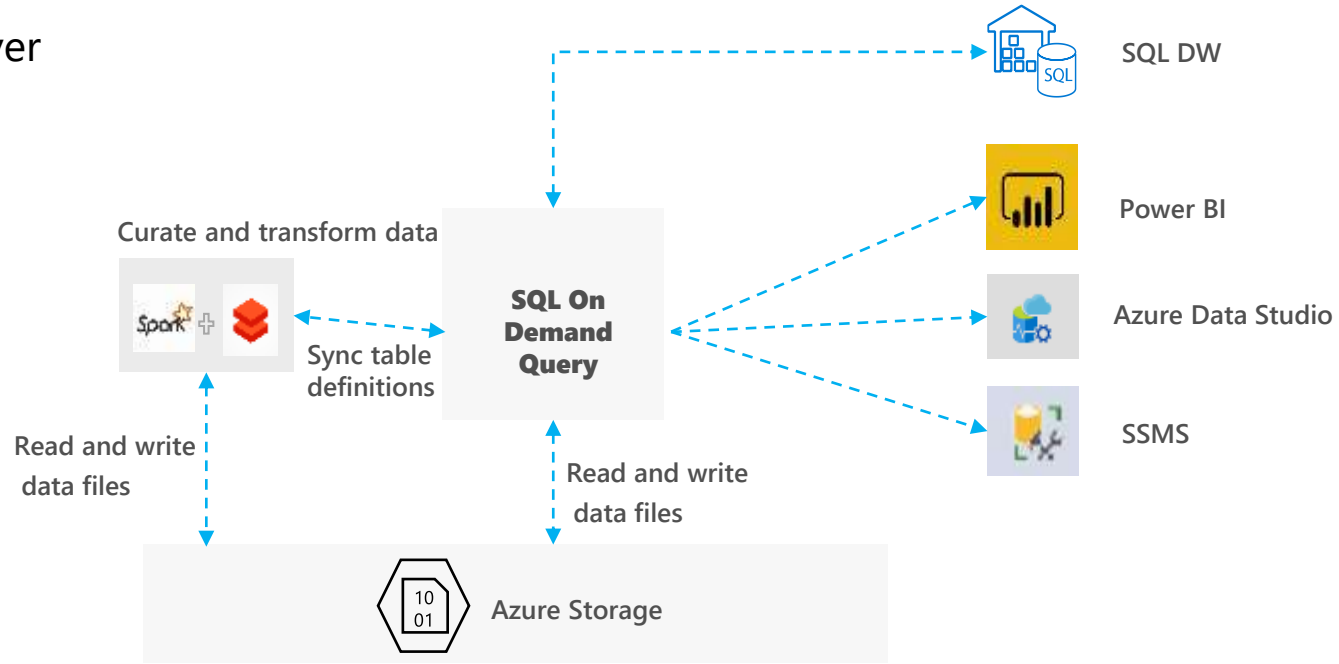
Offers security

Data integration with Databricks, HDInsight

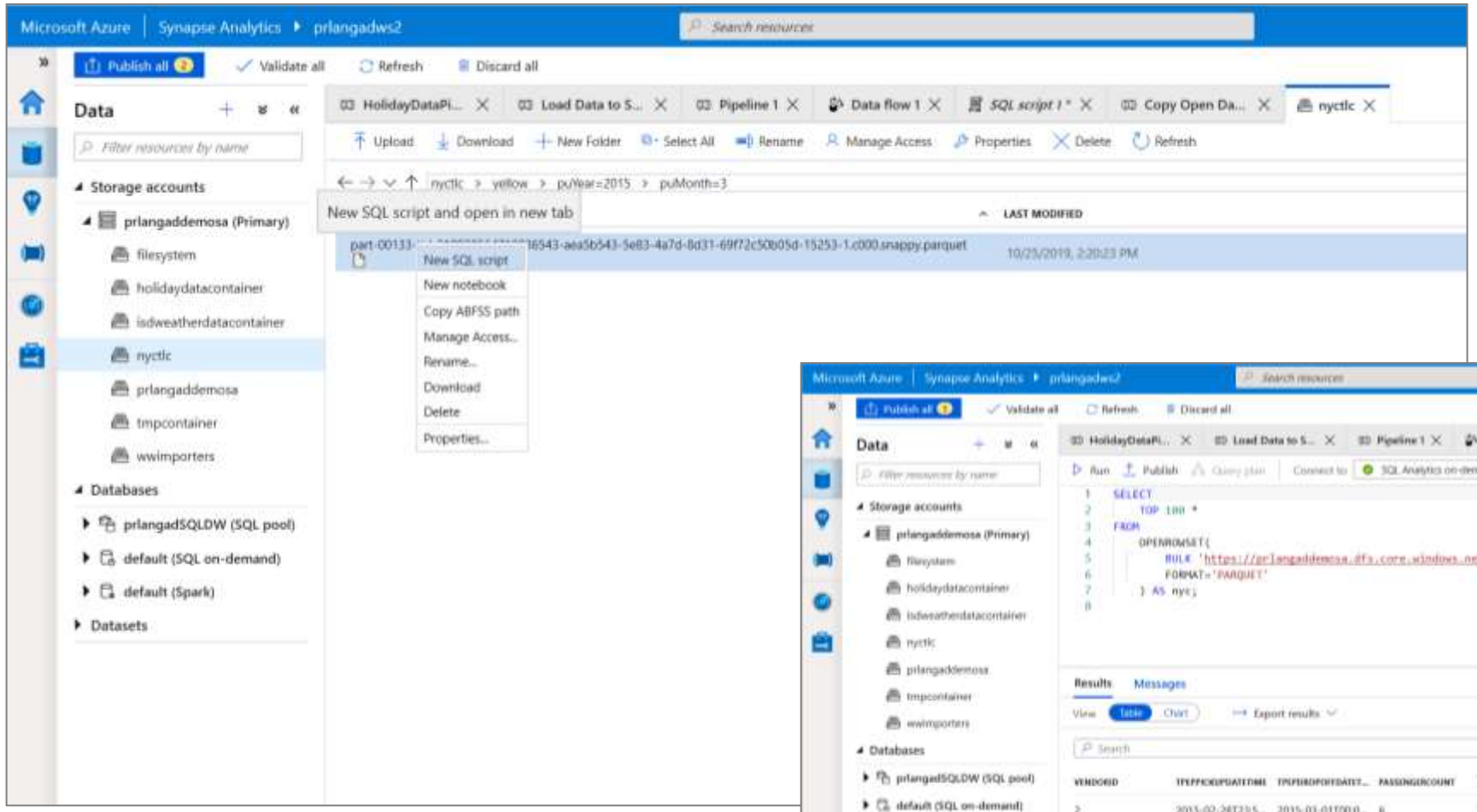
T-SQL syntax to query data

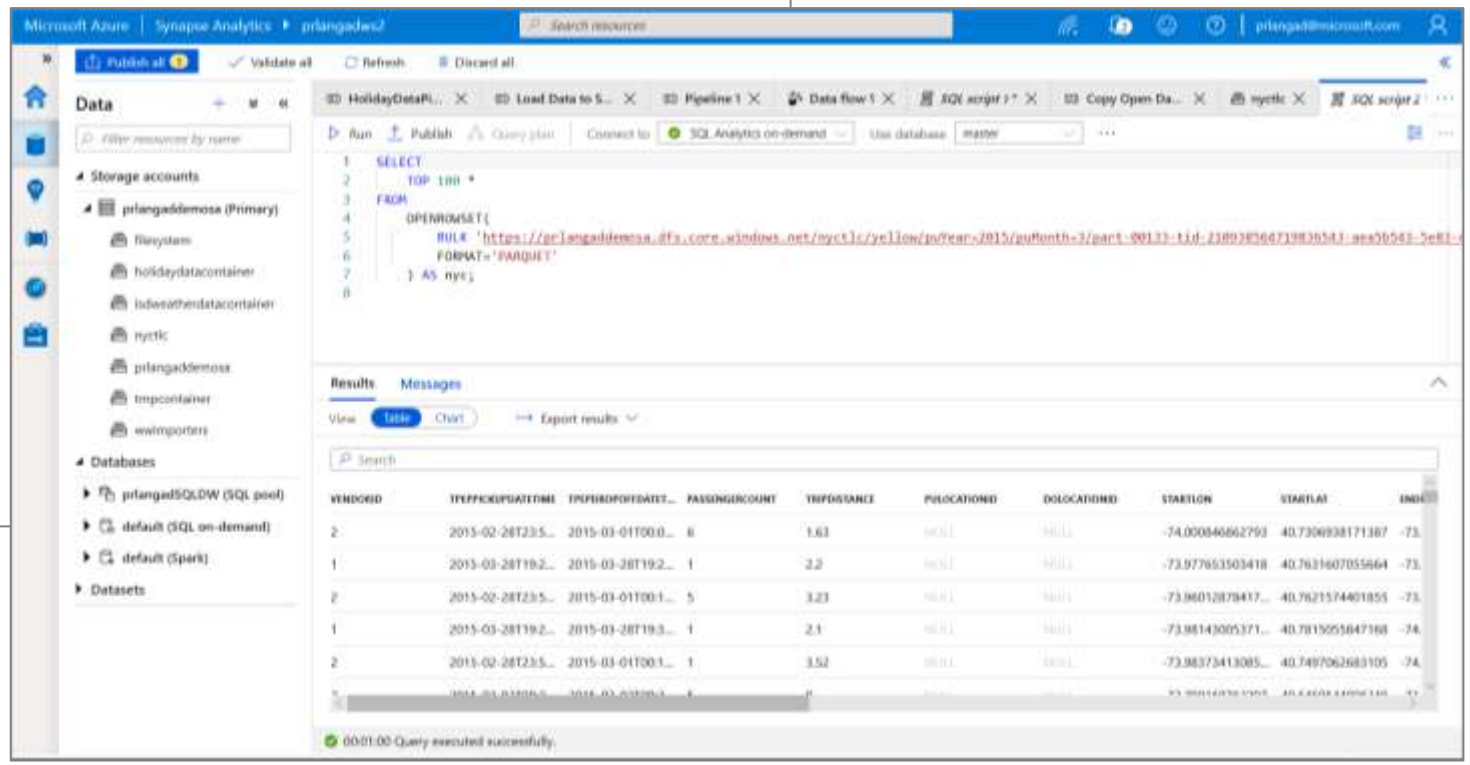
Supports data in various formats (Parquet, CSV, JSON)

Support for BI ecosystem



SQL On Demand – Querying on storage





SQL On Demand – Querying CSV File

Overview

Uses OPENROWSET function to access data

Benefits

Ability to read CSV File with

- no header row, Windows style new line
- no header row, Unix-style new line
- header row, Unix-style new line
- header row, Unix-style new line, quoted
- header row, Unix-style new line, escape
- header row, Unix-style new line, tab-delimited
- without specifying all columns

```
SELECT *
FROM OPENROWSET(
        BULK 'https://XXX.blob.core.windows.net/csv/population/population.csv',
        FORMAT = 'CSV',
        FIELDTERMINATOR =',',
        ROWTERMINATOR = '\n'
    )
WITH (
    [country_code] VARCHAR (5) COLLATE Latin1_General_BIN2,
    [country_name] VARCHAR (100) COLLATE Latin1_General_BIN2,
    [year] smallint,
    [population] bigint
) AS [r]
WHERE
    country_name = 'Luxembourg'
    AND year = 2017
```

	country_code	country_name	year	population
1	LU	Luxembourg	2017	594130

SQL On Demand – Querying JSON files

Overview

Read JSON files and provides data in tabular format

Benefits

Supports OPENJSON, JSON_VALUE and JSON_QUERY functions

```
SELECT *
FROM
    OPENROWSET(
        BULK 'https://XXX.blob.core.windows.net/json/books/book
1.json’,
        FORMAT='CSV',
        FIELDTERMINATOR ='0x0b',
        FIELDQUOTE = '0x0b',
        ROWTERMINATOR = '0x0b'
    )
    WITH (
        jsonContent varchar(8000)
    ) AS [r]
```

	jsonContent
1	{"_id": "kim95", "type": "Book", "title": "Modern Databas…

SQL On Demand – Querying JSON files

Example of JSON_VALUE function

```
SELECT
    JSON_VALUE(jsonContent, '$.title') AS title,
    JSON_VALUE(jsonContent, '$.publisher') as publisher,
    jsonContent
FROM
    OPENROWSET(
        BULK 'https://XXX.blob.core.windows.net/json/books/*.json',
        FORMAT='CSV',
        FIELDTERMINATOR ='0x0b',
        FIELDQUOTE = '0x0b',
        ROWTERMINATOR = '0x0b'
    )
    WITH (
        jsonContent varchar(8000)
    ) AS [r]
WHERE
    JSON_VALUE(jsonContent, '$.title') = 'Probabilistic and Statist
ical Methods in Cryptology, An Introduction by Selected Topics'
```

	title	publisher	jsonContent
1	Probabilistic and Statistical Methods in Cryptology, An Int…	Springer	{"_id": "neuen…

Example of JSON_QUERY function

```
SELECT
    JSON_QUERY(jsonContent, '$.authors') AS authors,
    jsonContent
FROM
    OPENROWSET(
        BULK 'https://XXX.blob.core.windows.net/json/books/*.json',
        FORMAT='CSV',
        FIELDTERMINATOR ='0x0b',
        FIELDQUOTE = '0x0b',
        ROWTERMINATOR = '0x0b'
    )
    WITH (
        jsonContent varchar(8000)
    ) AS [r]
WHERE
    JSON_VALUE(jsonContent, '$.title') = 'Probabilistic and Statist
ical Methods in Cryptology, An Introduction by Selected Topics'
```

	authors	jsonContent
1	["Daniel Neuenschwander"]	{"_id": "neuenschwander04", "type": "Book", "title": "Probabi…

Microsoft

Azure Synapse Analytics

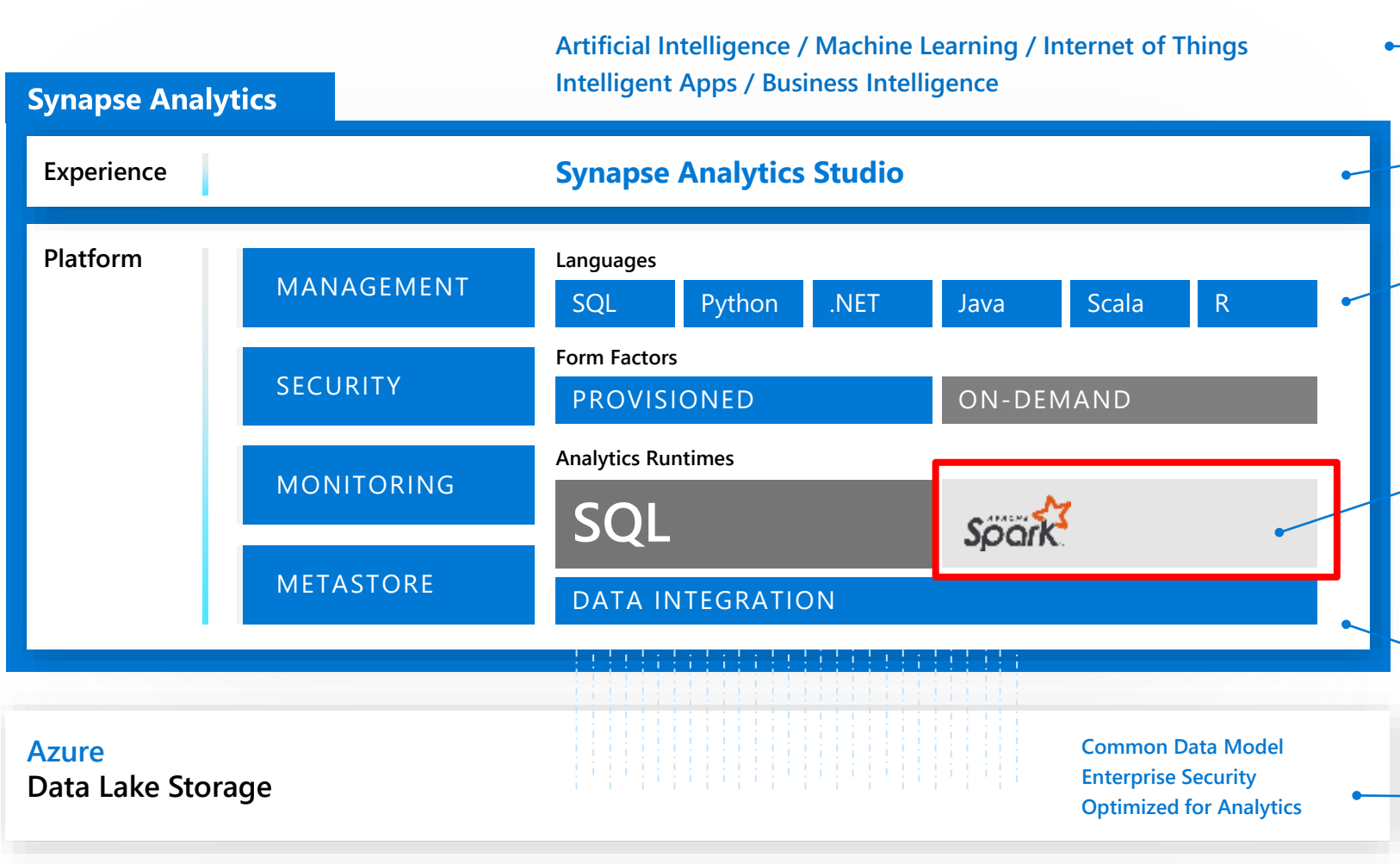
Spark

Azure Synapse Apache Spark - Summary

- Apache Spark 2.4 derivation
  - Linux Foundation Delta Lake 0.4 support
  - .Net Core 3.0 support
  - Python 3.6 + Anacondas support
- Tightly coupled to other Azure Synapse services
  - Integrated security and sign on
  - Integrated Metadata
  - Integrated and simplified provisioning
  - Integrated UX including nteract based notebooks
  - Fast load of SQL Analytics pools
- Core scenarios
  - Data Prep/Data Engineering/ETL
  - Machine Learning via Spark ML and Azure ML integration
  - Extensible through library management
- Efficient resource utilization
  - Fast Start
  - Auto scale (up and down)
  - Auto pause
  - Min cluster size of 3 nodes
- Multi Language Support
  - .Net (C#), PySpark, Scala, Spark SQL, Java

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

Spark for big data processing with Python, Scala, R and .NET

Integrated **platform services** for, management, security, monitoring, and metastore

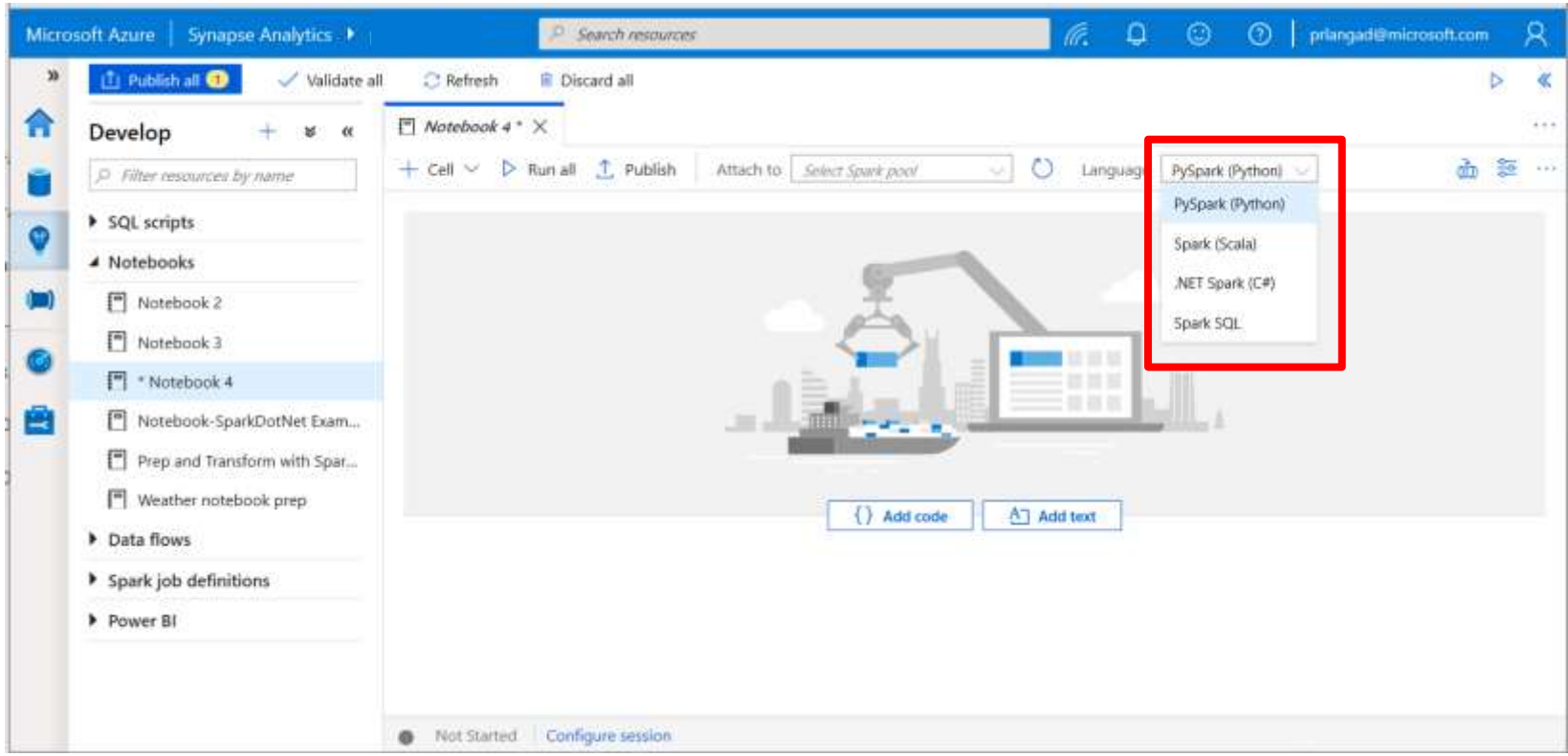
Data **lake integrated** and Common Data Model aware

Languages

Overview

Supports multiple languages to develop notebook

- PySpark (Python)
- Spark (Scala)
- .NET Spark (C#)
- Spark SQL



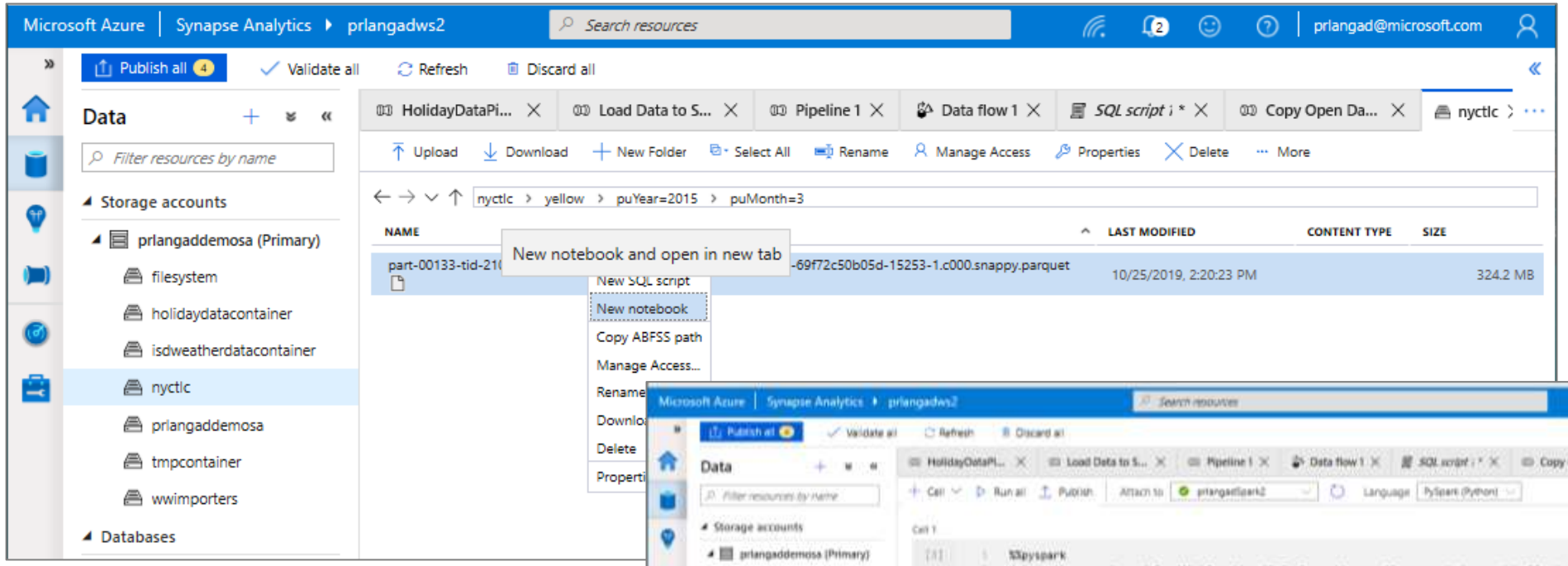
Benefits

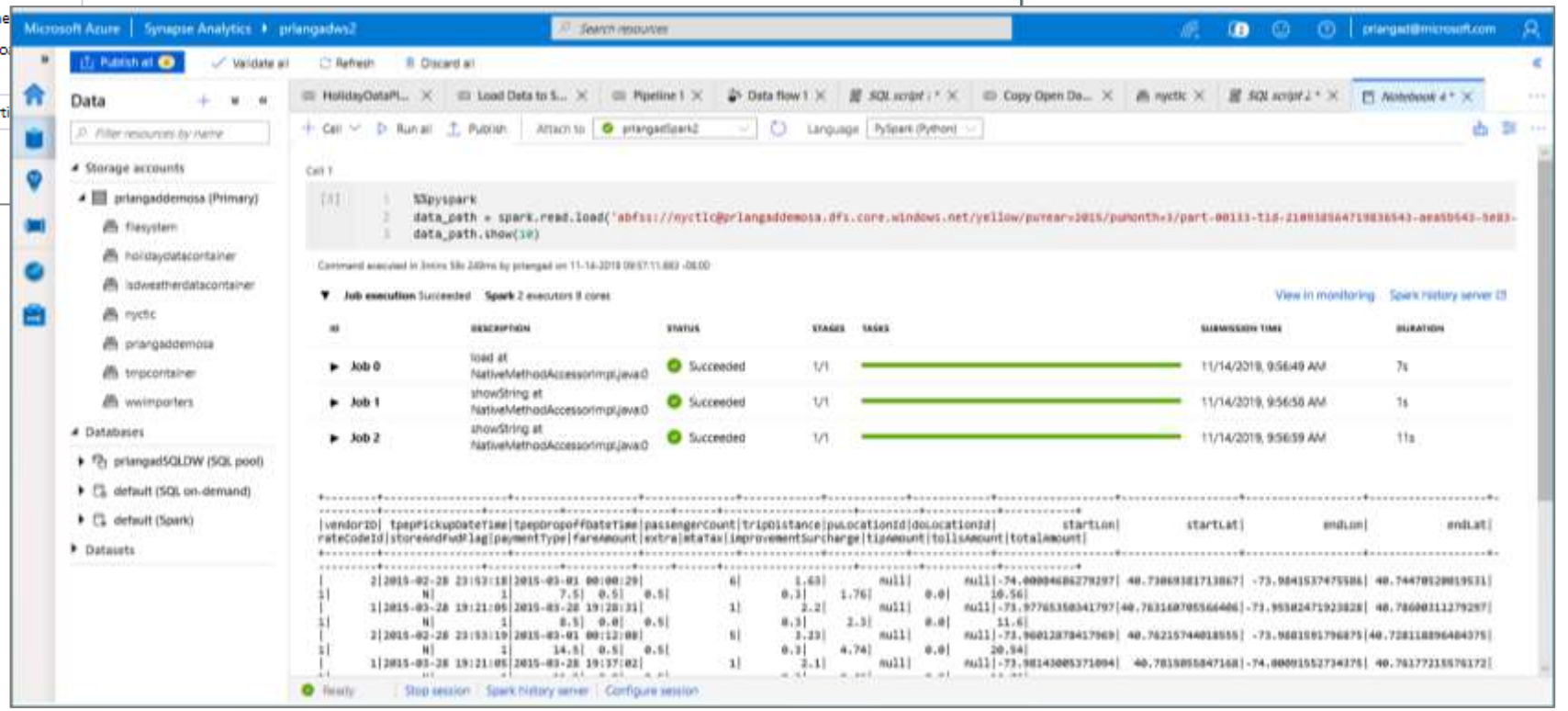
Allows to write multiple languages in one notebook

%%<Name of language>

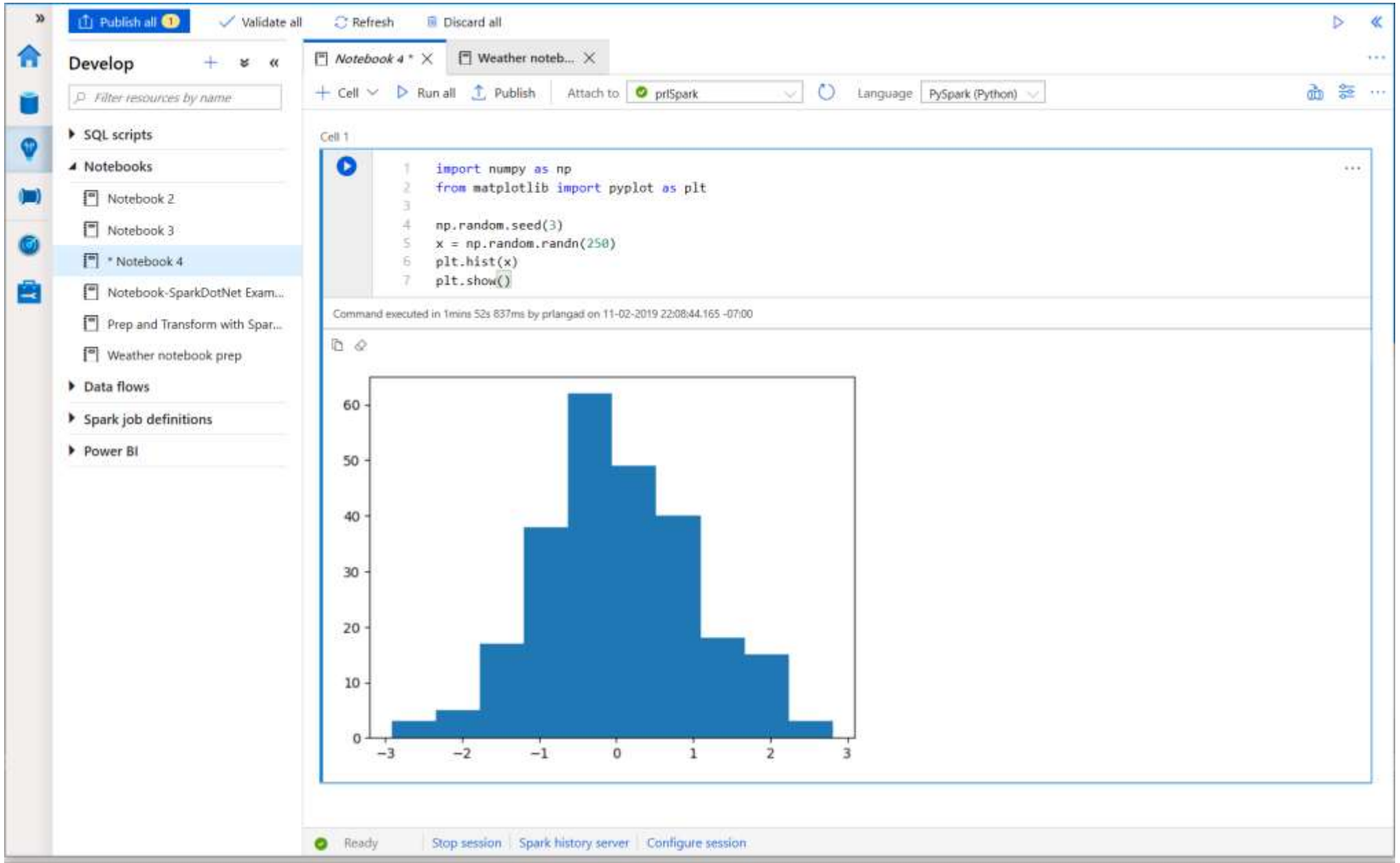
Offers use of temporary tables across languages

Create Notebook on files in storage

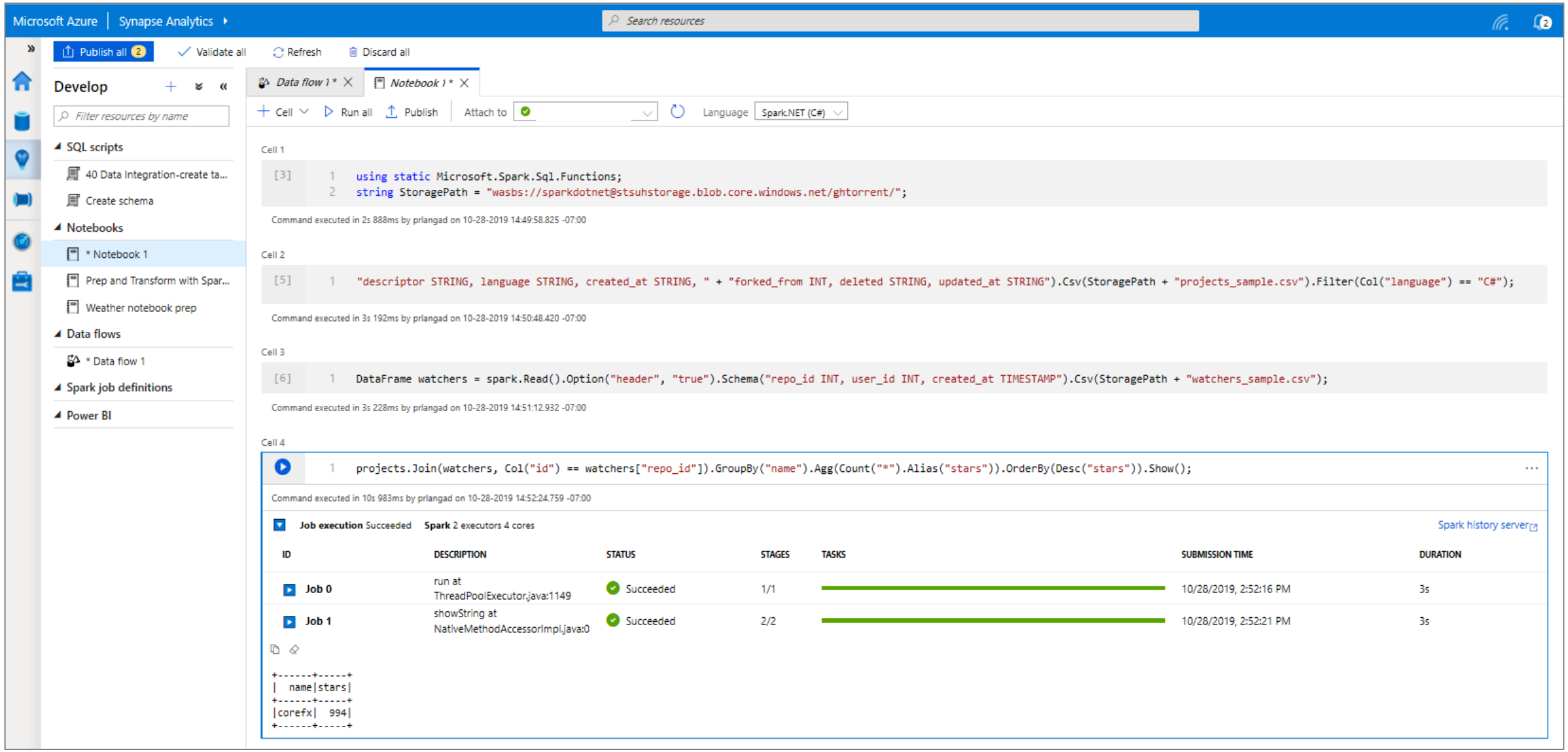




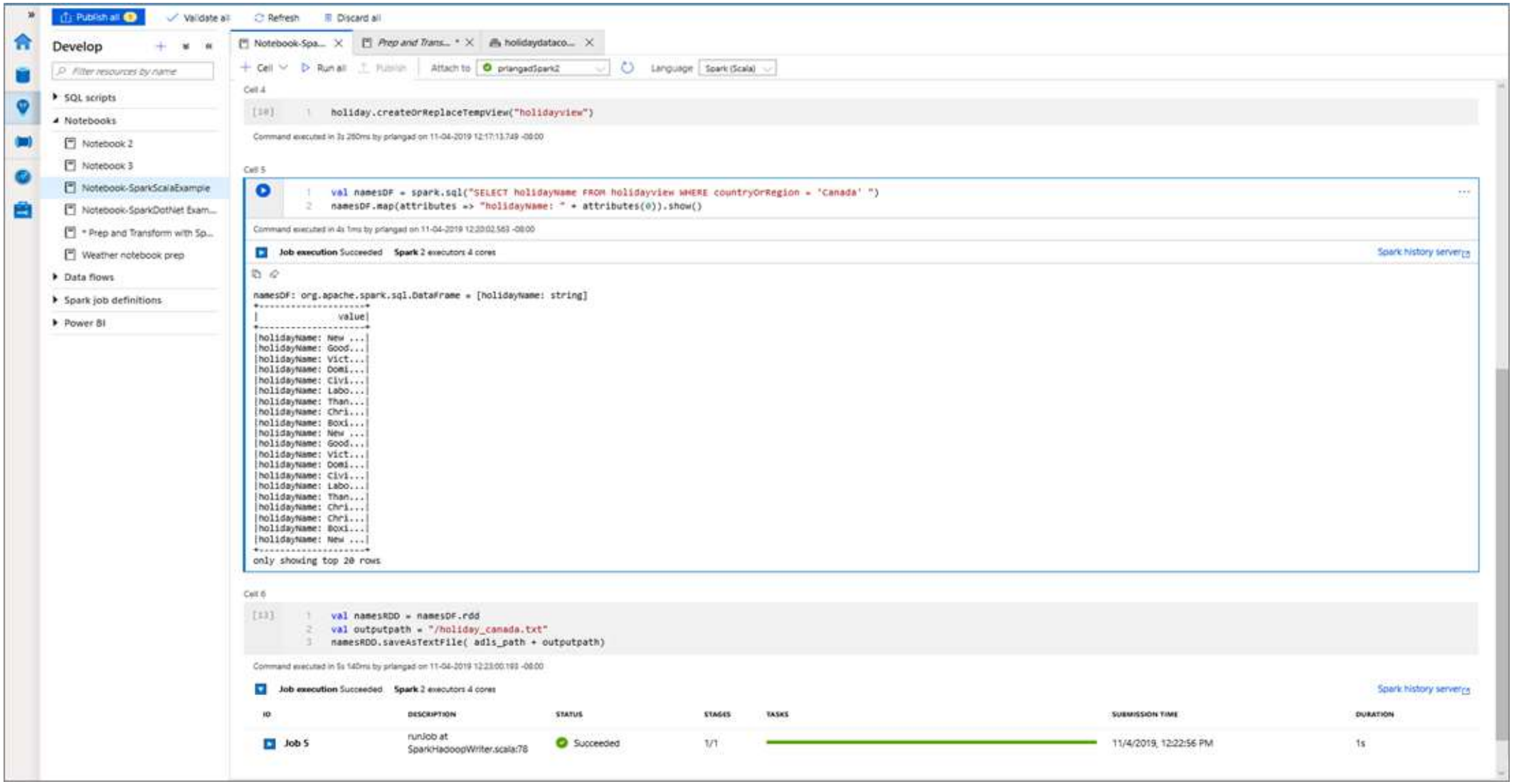
Languages – PySpark (Python)



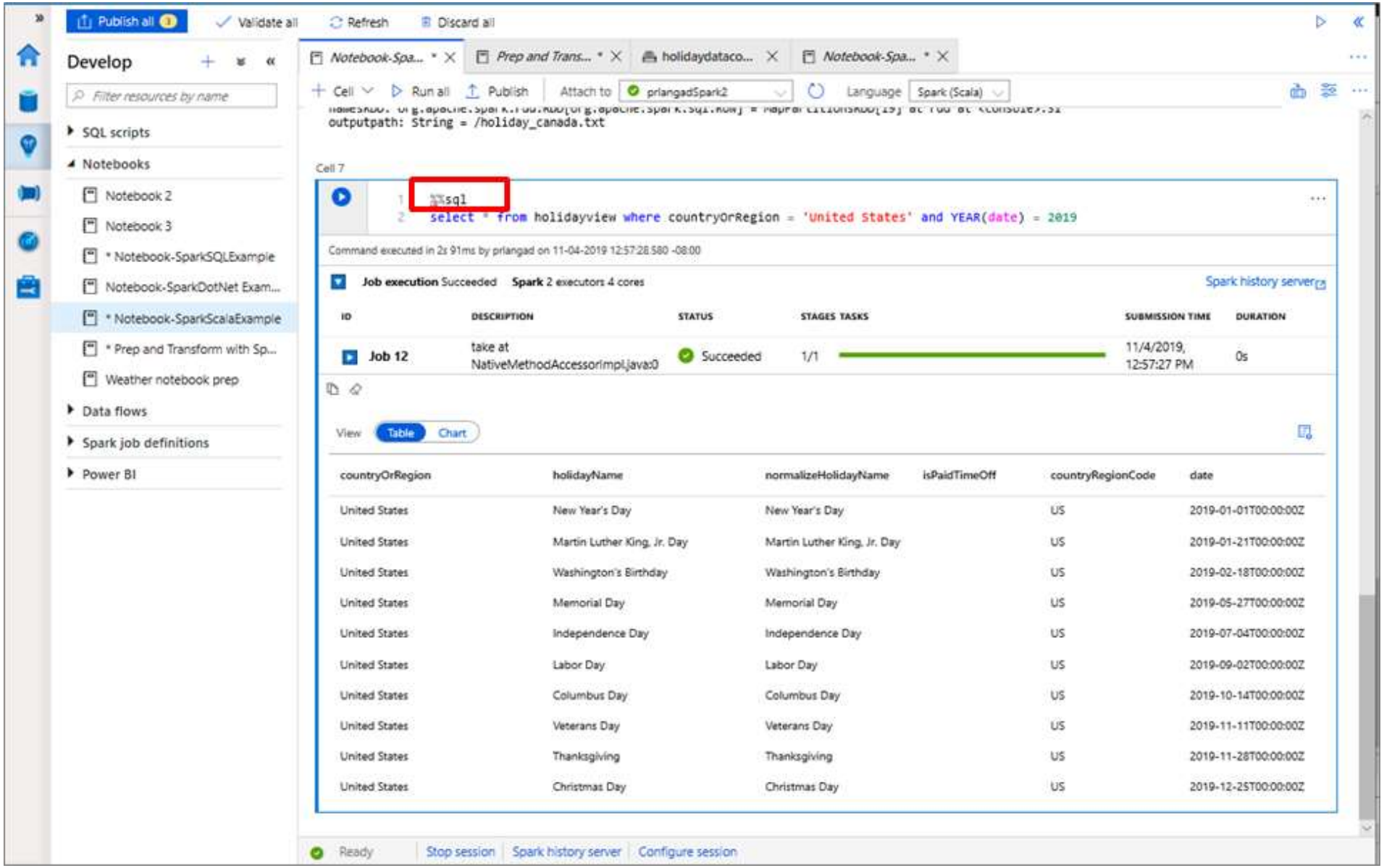
Languages – Spark.NET (C#)



Languages – Spark (Scala)



Languages – Spark SQL



Library Management - Python

Overview

Customers can add new python libraries at Spark pool level

Benefits

Input requirements.txt in simple pip freeze format

Add new libraries to your cluster

Update versions of existing libraries on your cluster

Libraries will get installed for your Spark pool during cluster creation

Ability to specify different requirements file for different pools within the same workspace

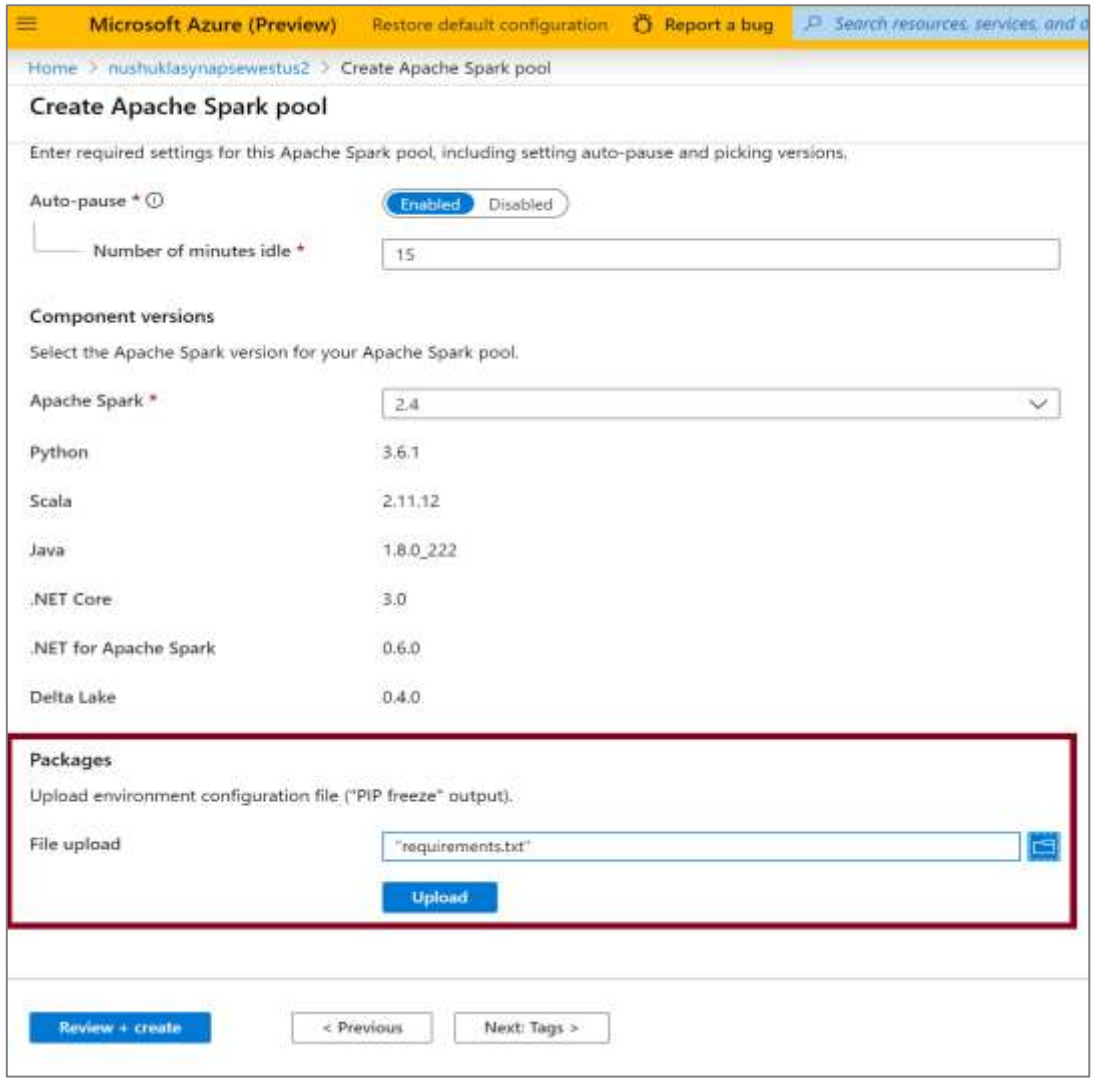
Constraints

The library version must exist on PyPI repository

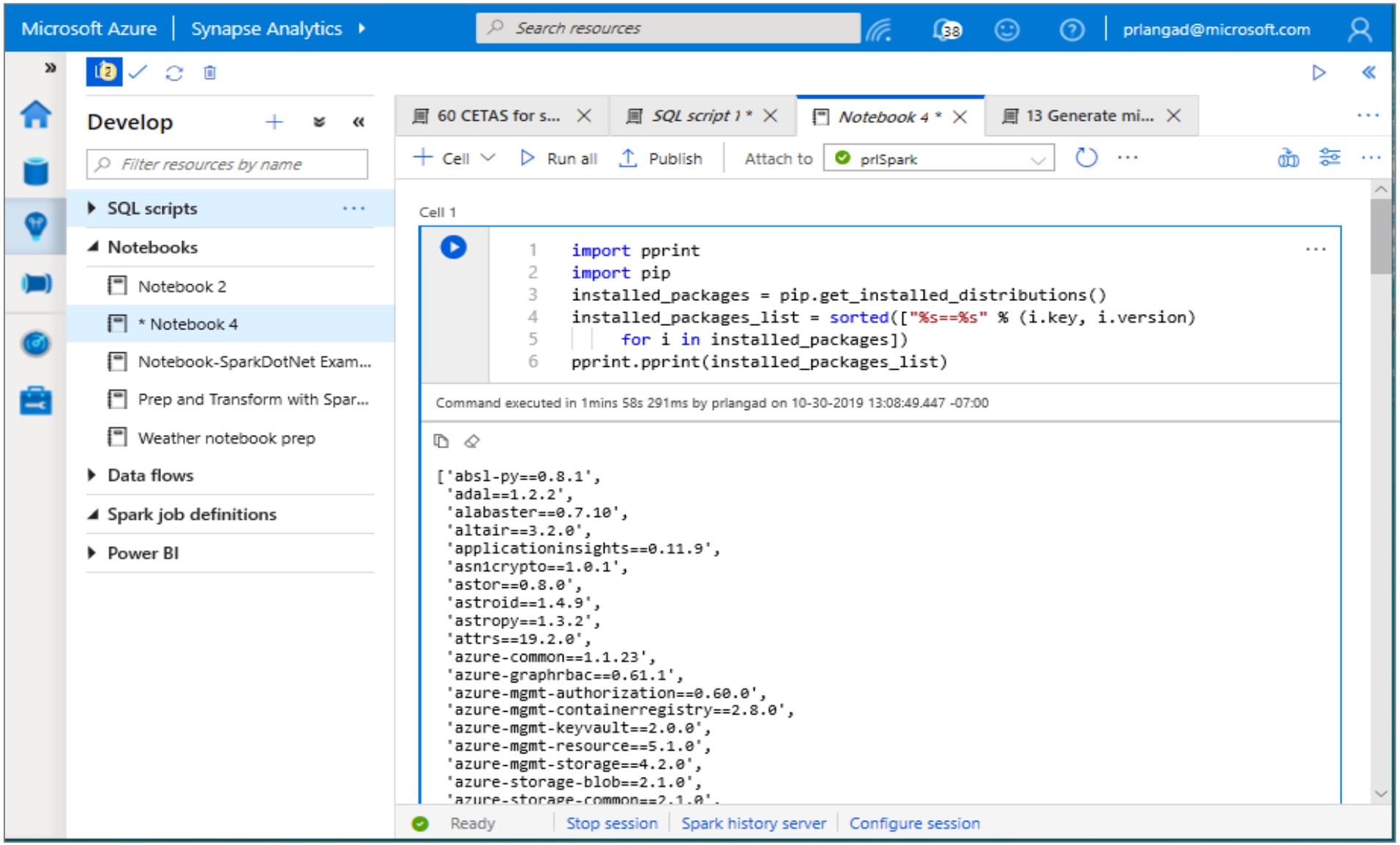
Version downgrade of an existing library not allowed

In the Portal

Specify the new requirements while creating Spark Pool in Additional Settings blade



Library Management - Python



Microsoft

Azure Synapse Analytics

Foundation

Manage – Access Control

Overview

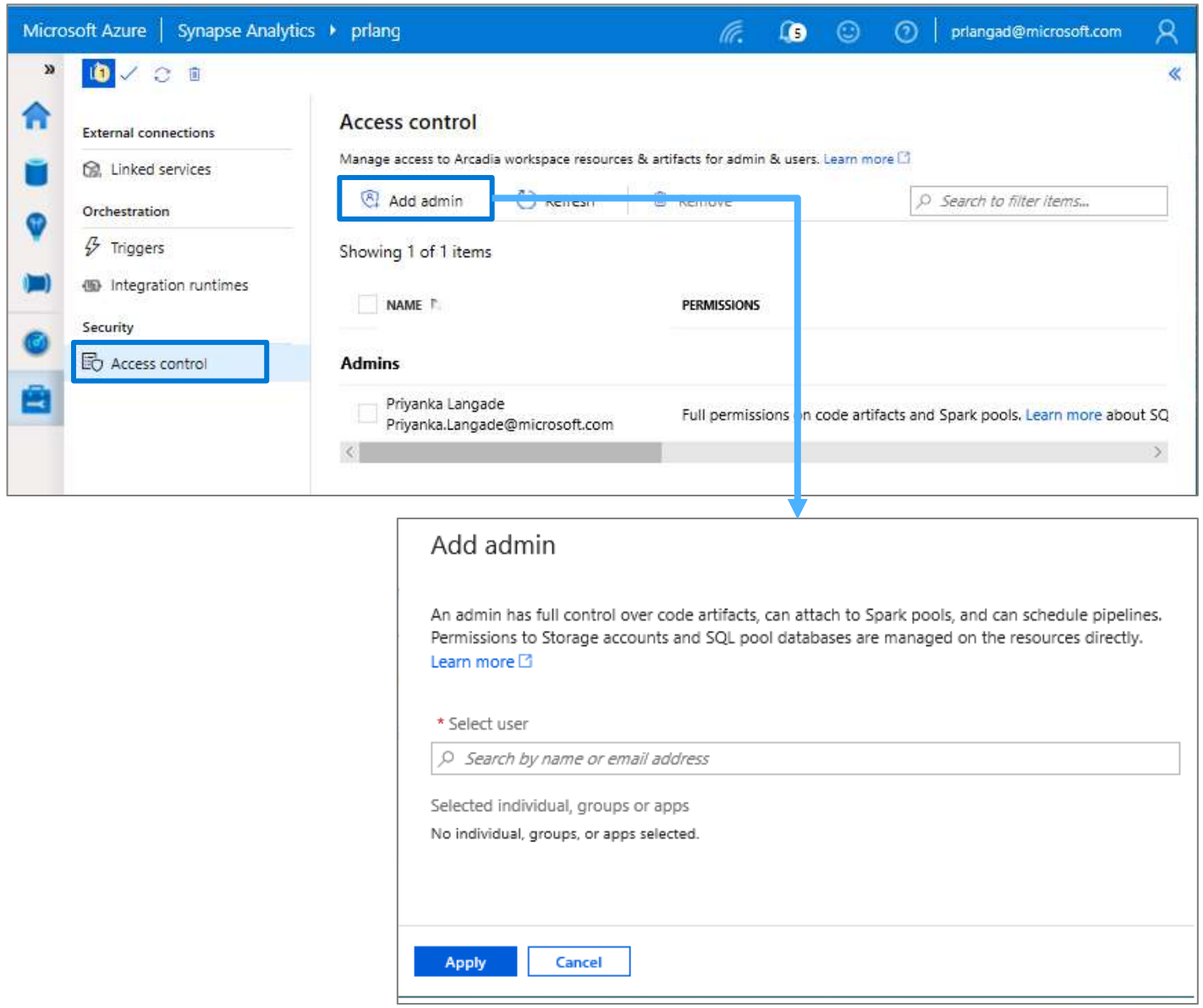
It provides access control management to workspace resources and artifacts for admin and users

Benefits

Share workspace with the team

Increases productivity

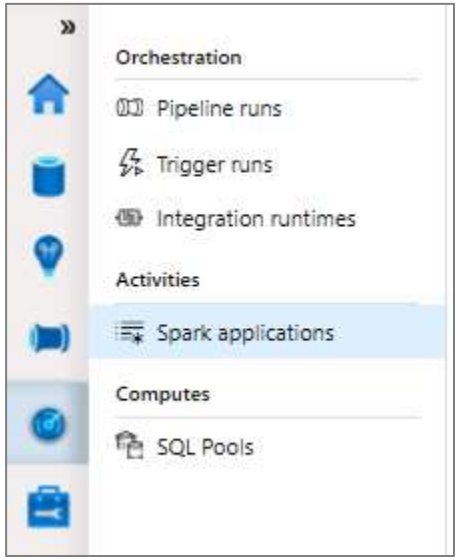
Manage permissions on code artifacts and Spark pools



Spark Monitoring

Overview

Monitor Spark pools, Spark applications for the progress and status of activities

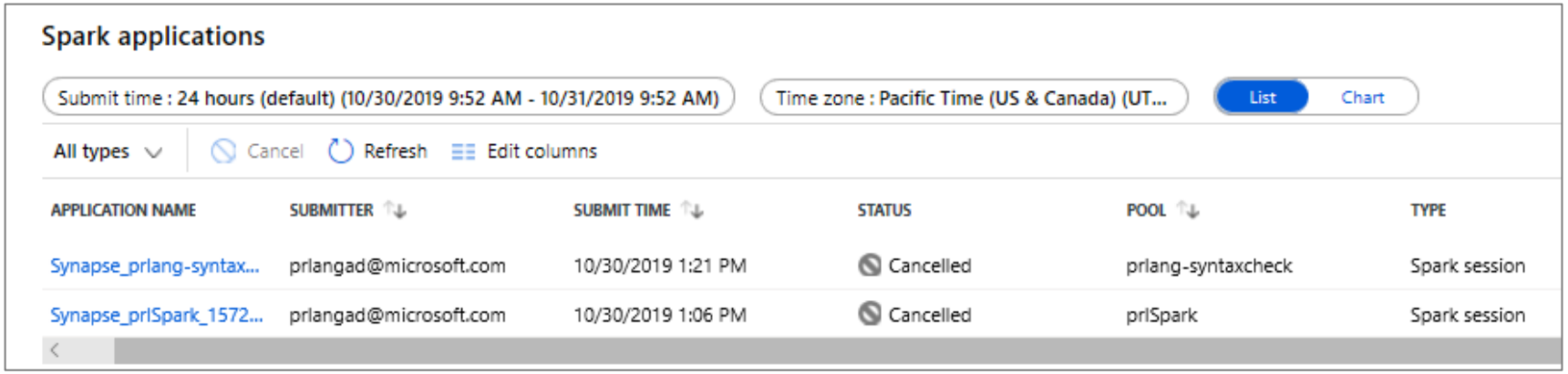


Benefits

Monitor Spark pools for the status as paused, active, resume, scaling and upgrading

Build a dashboard to monitor performance

Track the usage of resources



SQL Monitoring

Overview

Monitor SQL Pool in Azure Portal for overall usage and query activities.

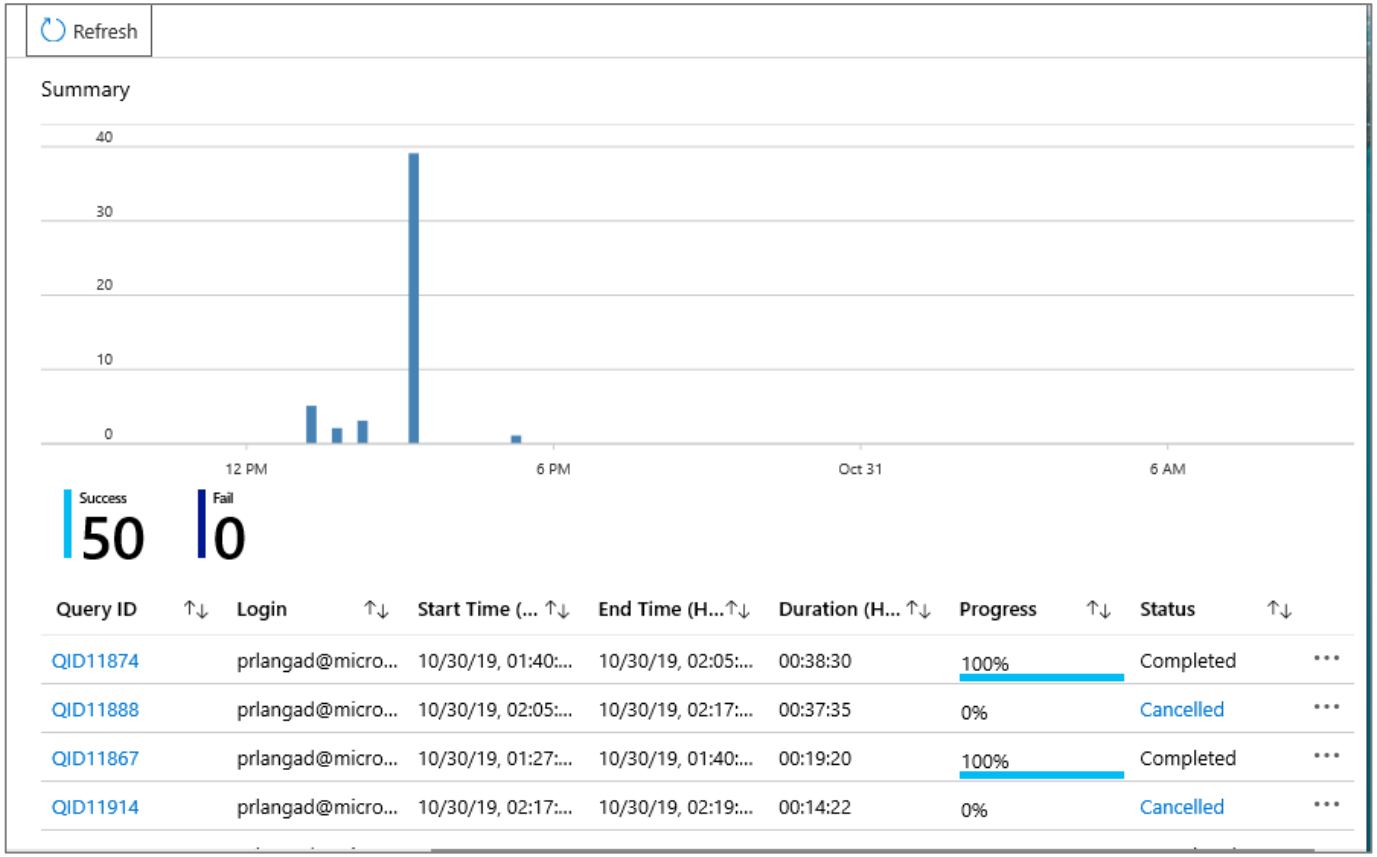
Benefits

Access SQL Audit Logs for my SQL computes

Monitor status and progress of all/specific activities

Dashboard view to monitor performance

Get to know scale of SQL compute resource



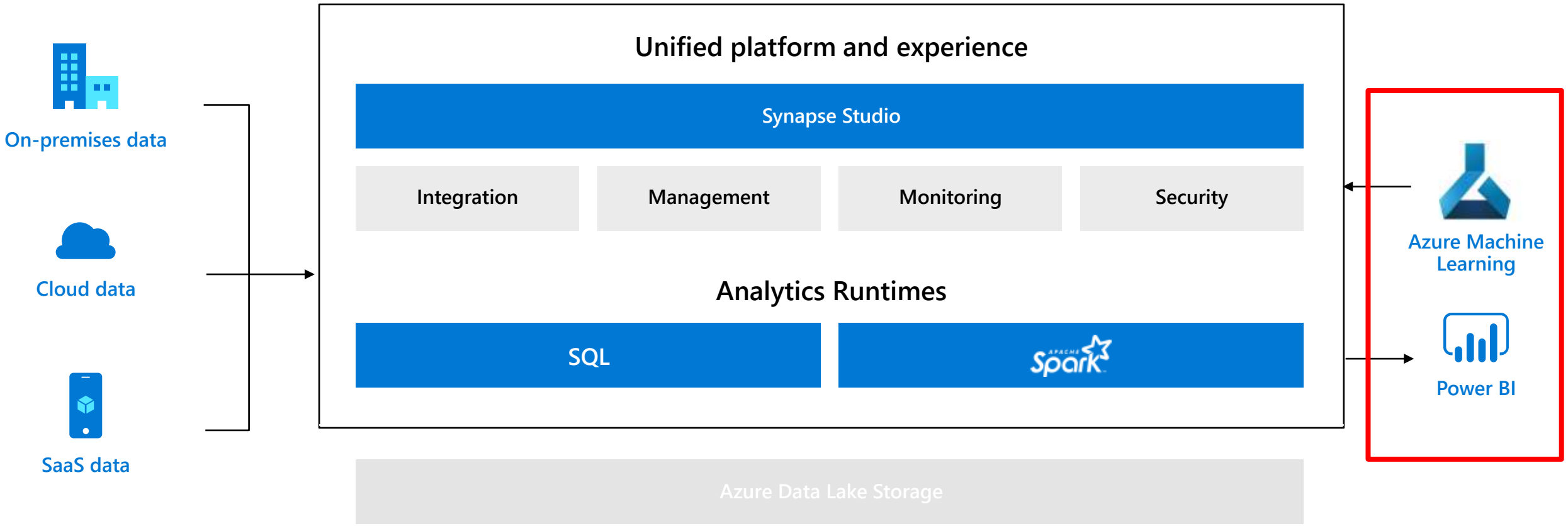
Microsoft

Azure Synapse Analytics

Connected Services

Azure Synapse Analytics

Limitless analytics service with unmatched time to insight



Azure Machine Learning

Overview

Data Scientists can use Azure ML notebooks to do (distributed) data preparation on Synapse Spark compute.

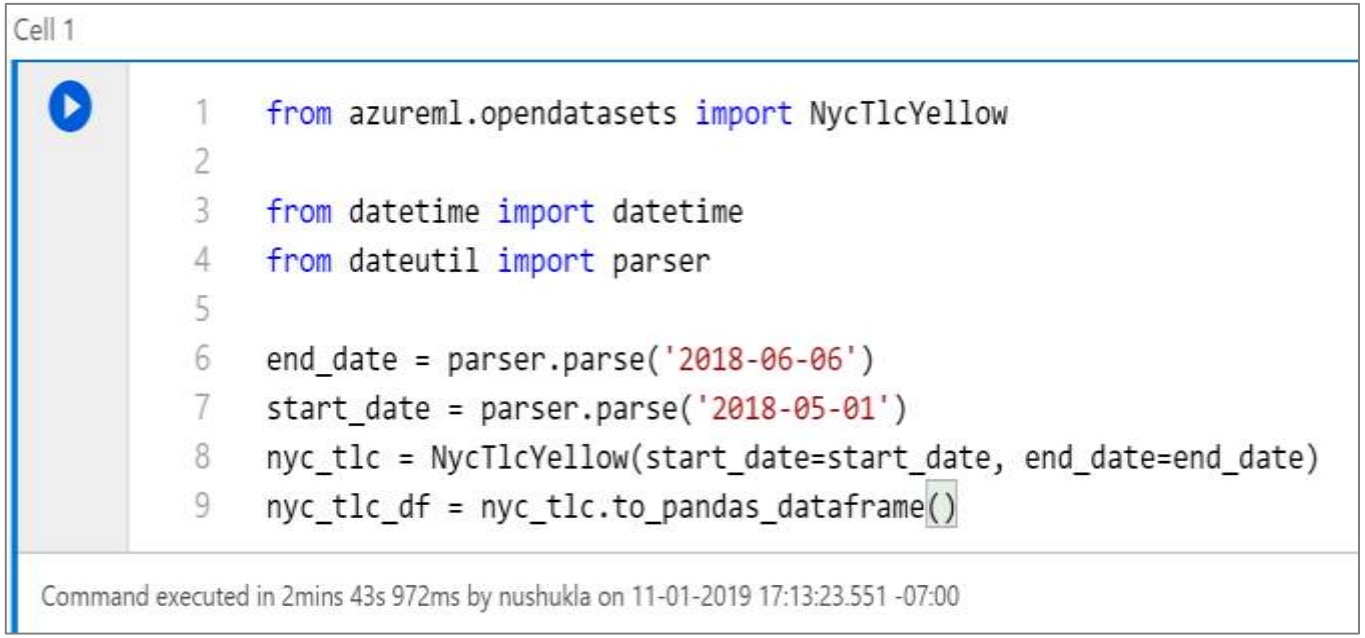
Benefits

Connect to your existing Azure ML workspace and project

Use the AutoML Classifier for classification or regression problem

Train the model

Access open datasets



Azure Machine Learning (continued)

Configure AutoML and Train the Models

Cell 9

```
l_config = AutoMLConfig(task = 'regression',debug_log = 'automl_errors.log',
                        primary_metric = 'normalized_root_mean_squared_error', iteration_timeout_minutes = 10,
                        iterations = 2, preprocess = True, n_cross_validations = 2,max_concurrent_iterations = 2,
                        verbosity = logging.INFO,spark_context=sc, enable_onnx_compatible_models=True, cache_store=Tru
```

Cell 10

```
local_run = experiment.submit(automl_config, show_output = True)
```

Best Model

Cell 12

```
best_run, fitted_model = local_run.get_output(return_onnx_model=True)
print(fitted_model)
```

Portal URL for Monitoring Runs

Cell 14

```
more Insights of experiment
displayHTML("<a href={} target='_blank'>Your experiment in Azure Portal: {}</a>".format(local_run.get_portal_url(), local_r
```

Synapse

Power BI

Power BI

Overview

Power BI is a business analytics service that delivers insights to enable fast, informed decisions

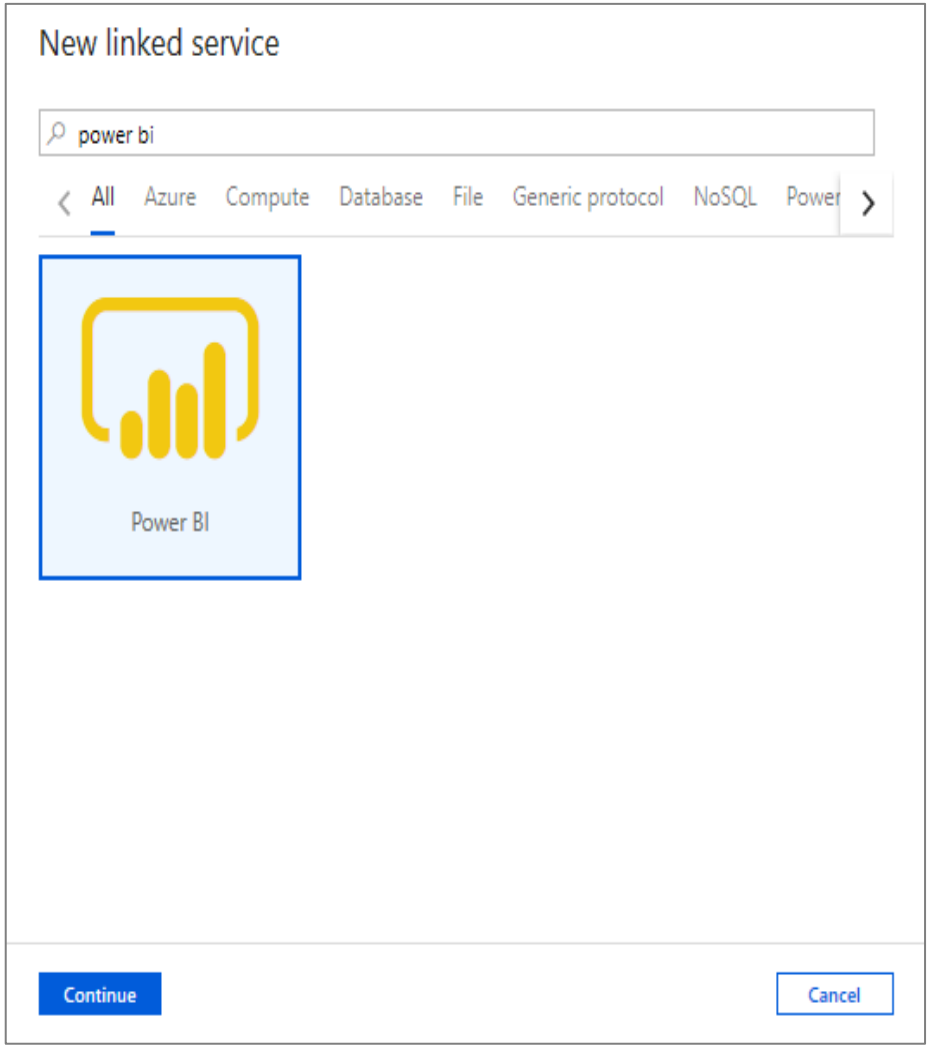
Benefits

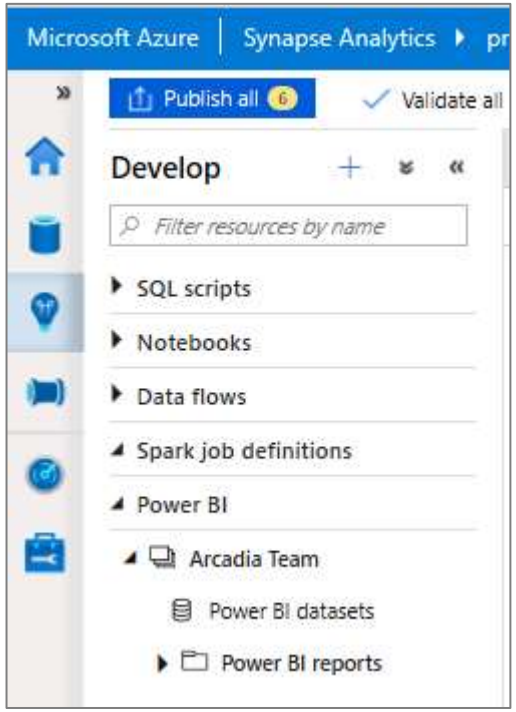
Create Power BI reports in the workspace

Have access to published reports in workspace

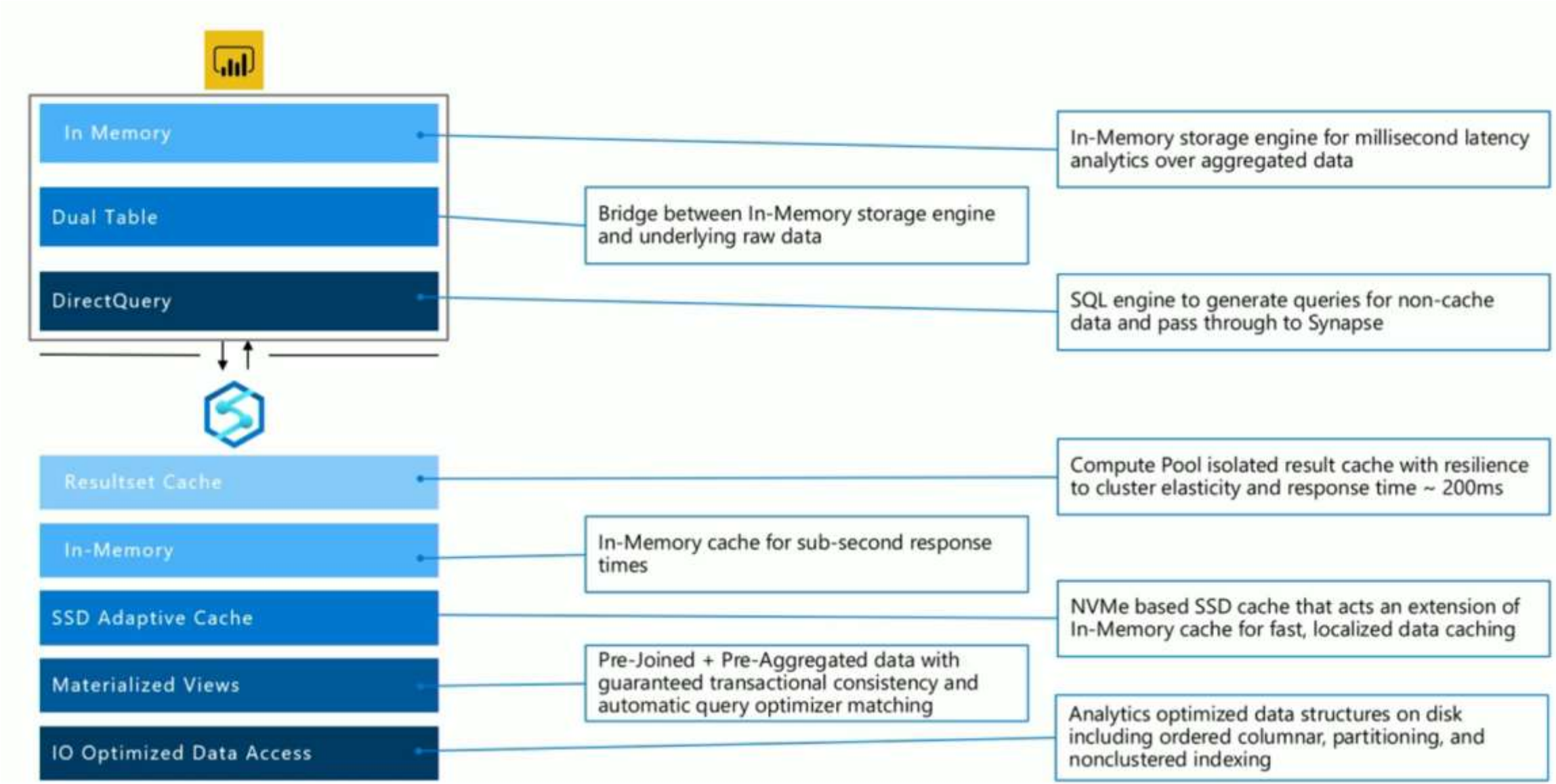
Update reports real time from Synapse workspace to get it reflected on Power BI service

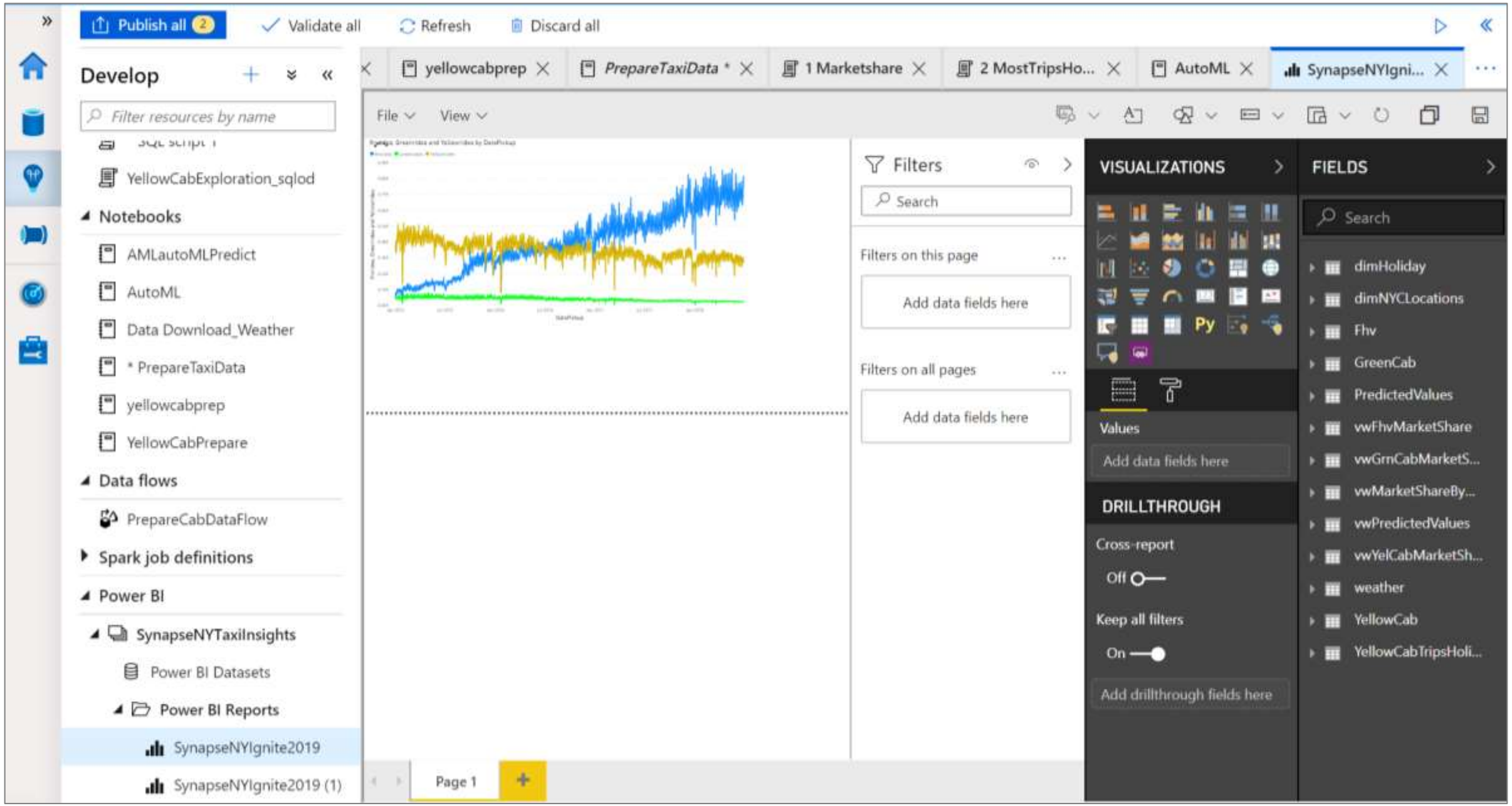
Visually explore and analyze data

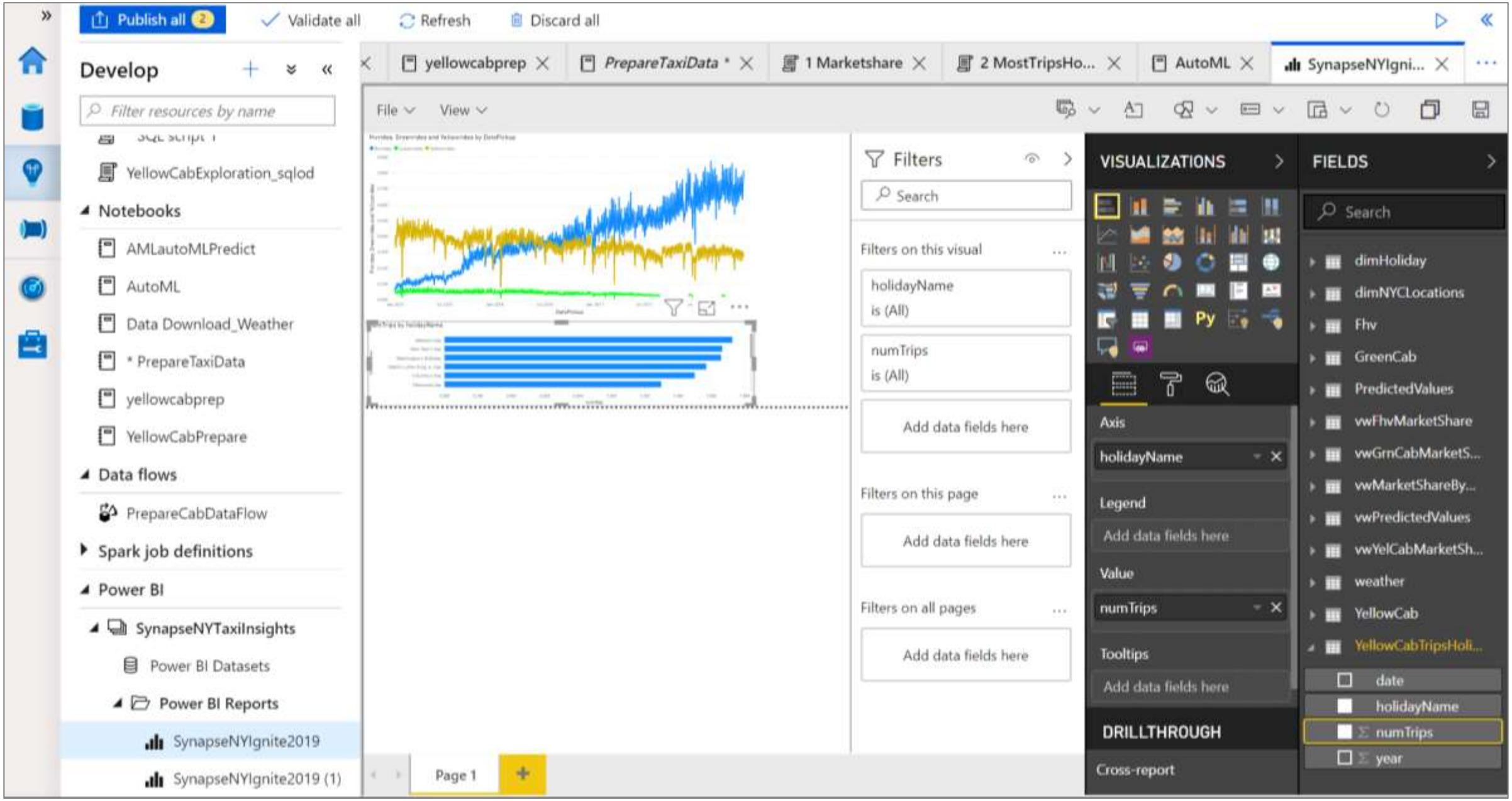


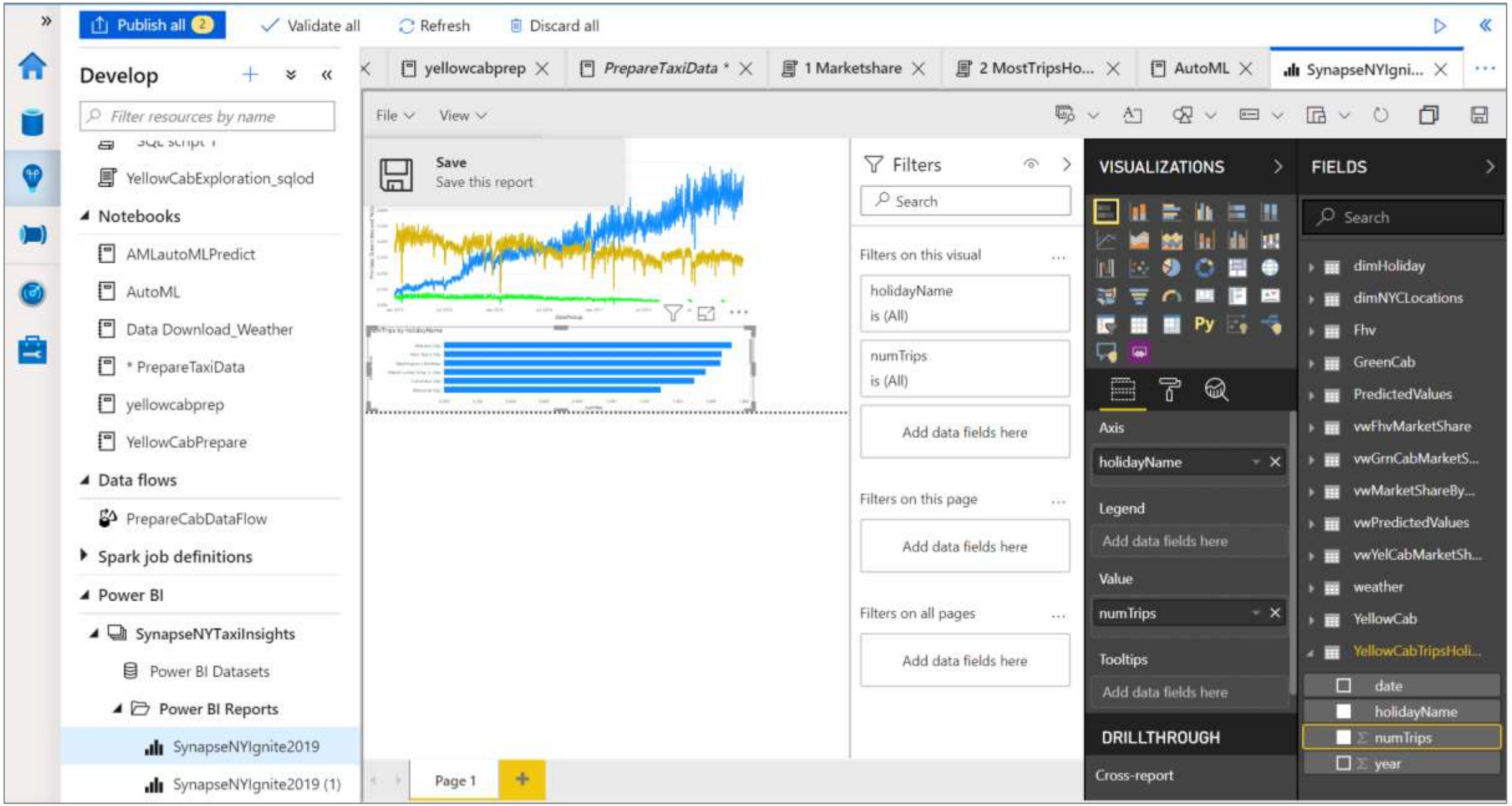


Power BI Aggregations and Synapse query performance



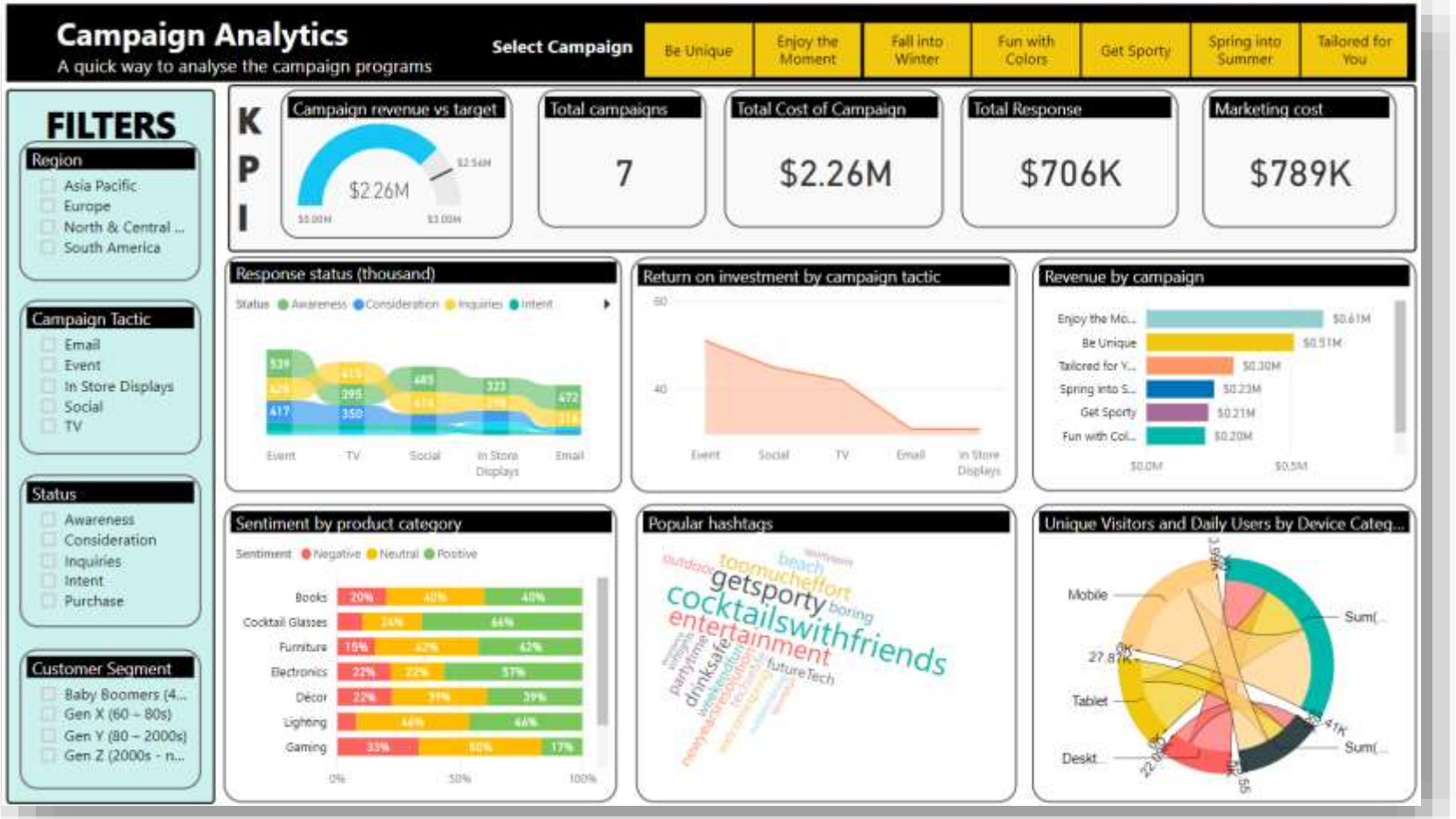

Publish all 2
Validate all
Refresh
Discard all
Develop
Filter resources by name
YellowCabExploration_sqlod
Notebooks
AMLautoMLPredict
AutoML
Data Download_Weather
* PrepareTaxiData
yellowcabprep
YellowCabPrepare
Data flows
PrepareCabDataFlow
Spark job definitions
Power BI
SynapseNYTaxiInsights
Power BI Datasets
Power BI Reports
SynapseNYIgnite2019
SynapseNYIgnite2019 (1)
yellowcabprep
PrepareTaxiData *
1 Marketshare
2 MostTripsHo...
AutoML
SynapseNYIgni...
File
View
Filters
Search
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
VISUALIZATIONS
Values
Add data fields here
DRILLTHROUGH
Cross-report
Off
Keep all filters
On
Add drillthrough fields here
FIELDS
Search
dimHoliday
dimNYCLocations
Fhv
GreenCab
PredictedValues
vwFhvMarketShare
vwGrnCabMarketS...
vwMarketShareBy...
vwPredictedValues
vwYelCabMarketSh...
weather
YellowCab
YellowCabTripsHoli...
Page 1


Publish all 2
Validate all
Refresh
Discard all
Develop
Filter resources by name
YellowCabExploration_sqlod
Notebooks
AMLautoMLPredict
AutoML
Data Download_Weather
* PrepareTaxiData
yellowcabprep
YellowCabPrepare
Data flows
PrepareCabDataFlow
Spark job definitions
Power BI
SynapseNYTaxiInsights
Power BI Datasets
Power BI Reports
SynapseNYIgnite2019
SynapseNYIgnite2019 (1)
yellowcabprep
PrepareTaxiData *
1 Marketshare
2 MostTripsHo...
AutoML
SynapseNYIgni...
File
View
Filters
Search
Filters on this visual
holidayName
is (All)
numTrips
is (All)
Add data fields here
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
VISUALIZATIONS
Axis
holidayName
Legend
Add data fields here
Value
numTrips
Tooltips
Add data fields here
DRILLTHROUGH
Cross-report
FIELDS
Search
dimHoliday
dimNYCLocations
Fhv
GreenCab
PredictedValues
vwFhvMarketShare
vwGrnCabMarketS...
vwMarketShareBy...
vwPredictedValues
vwYelCabMarketSh...
weather
YellowCab
YellowCabTripsHoli...
date
holidayName
numTrips
year
Page 1


Publish all 2
Validate all
Refresh
Discard all
Develop
Filter resources by name
YellowCabExploration_sqlod
Notebooks
AMLautoMLPredict
AutoML
Data Download_Weather
* PrepareTaxiData
yellowcabprep
YellowCabPrepare
Data flows
PrepareCabDataFlow
Spark job definitions
Power BI
SynapseNYTaxiInsights
Power BI Datasets
Power BI Reports
SynapseNYIgnite2019
SynapseNYIgnite2019 (1)
yellowcabprep
PrepareTaxiData *
1 Marketshare
2 MostTripsHo...
AutoML
SynapseNYIgni...
File
View
Save
Save this report
Filters
Search
Filters on this visual
holidayName
is (All)
numTrips
is (All)
Add data fields here
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
VISUALIZATIONS
Axis
holidayName
Legend
Add data fields here
Value
numTrips
Tooltips
Add data fields here
DRILLTHROUGH
Cross-report
FIELDS
Search
dimHoliday
dimNYCLocations
Fhv
GreenCab
PredictedValues
vwFhvMarketShare
vwGrnCabMarketS...
vwMarketShareBy...
vwPredictedValues
vwYelCabMarketSh...
weather
YellowCab
YellowCabTripsHoli...
date
holidayName
numTrips
year
Page 1

Amazing insights made possible by migration to Azure Synapse Analytics

Campaign analytics insights based on historical data



Top documentation links

- What is SQL on-demand?: link
- What is Apache Spark in Azure Synapse Analytics?: link
- Best practices for SQL pool in Azure Synapse Analytics: link
- Best practices for SQL on-demand in Azure Synapse Analytics: link
- Azure Synapse Analytics shared metadata: link
- Use maintenance schedules to manage service updates and maintenance: link
- Cheat sheet for Azure Synapse Analytics (formerly SQL DW): link
- Best practices for SQL Analytics in Azure Synapse Analytics (formerly SQL DW): link
- Synapse Analytics documentation is here: aka.ms/SynapseDocs

Microsoft

Q & A

Adrián J. Fernández Zenteno

Sr. Cloud Solution Architect - Azure Advanced Analytics

Email: adrian.fernandez@microsoft.com

Twitter: @AdrianFZ10