



DATA TOBOGGAN

# What's New and Cool in Azure Synapse Analytics

**Pawel Potasinski**

Senior Program Manager



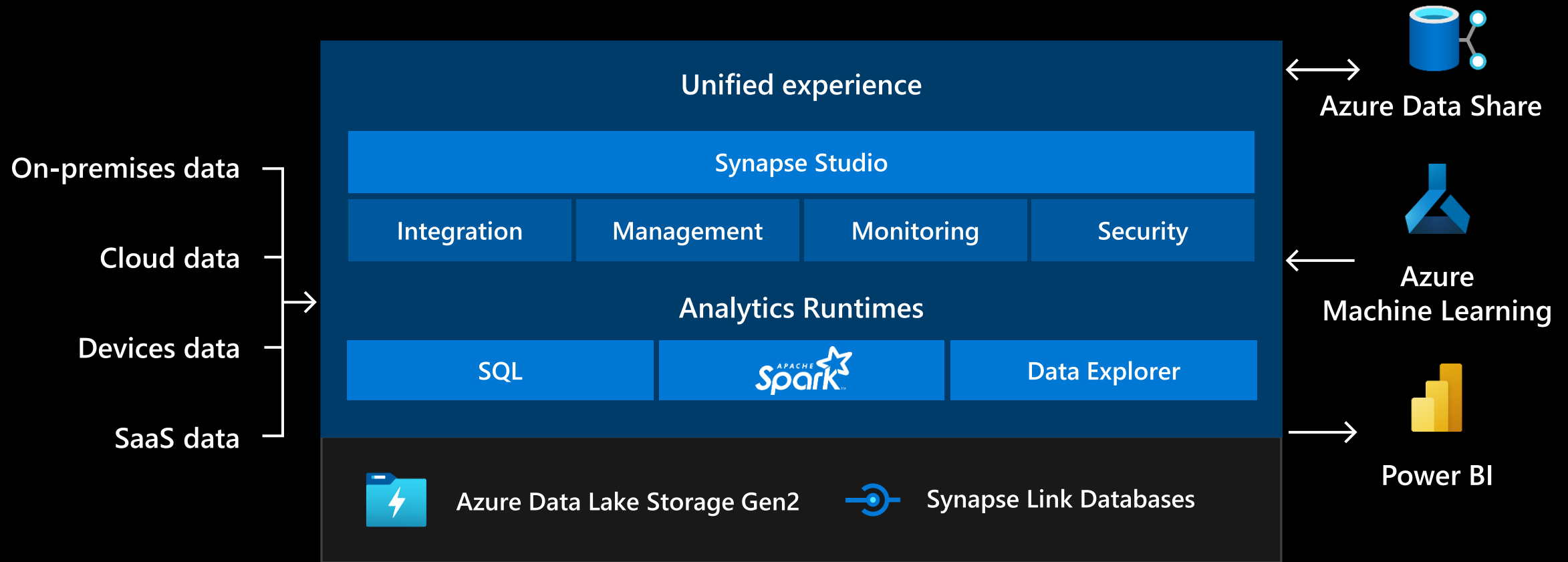
@pawelpotasinski



/in/pawelpotasinski



# Azure Synapse Analytics

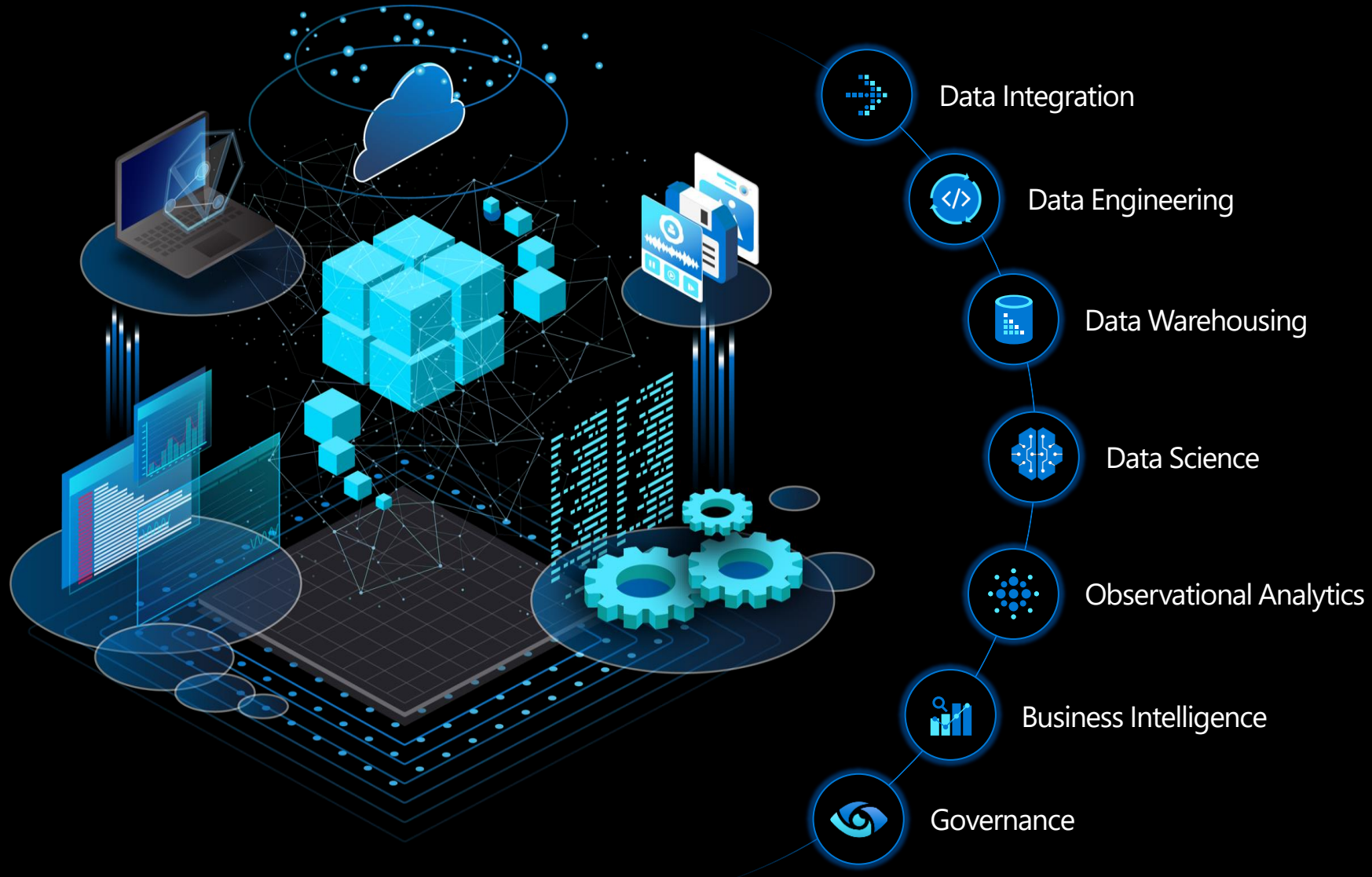




Stay up to date with Azure Synapse Analytics:

[aka.ms/SynapseMonthlyUpdate](https://aka.ms/SynapseMonthlyUpdate)

# Synapse + Power BI







# Data Integration

July-December 2022

- [Time-To-Live in managed VNET \[Public Preview\]](#)
- [Mapping data flows now support visual Cast transformation](#)
- [Change default activity time out from 7 days to 12 hours](#)
- [Expression builder UI update](#)
- [New UI for mapping dataflows inline datasets](#)
- [\*\*Gantt Chart view supported in Integration Runtime Monitoring\*\*](#)
- [Monitoring Improvements](#)
- [Maximum column optimization in mapping dataflow](#)
- [Support NUMBER to integer conversion in Oracle source connector](#)
- [Support for sending a body with HTTP request DELETE method in Web activity](#)
- [New Microsoft 365 Pipeline Template Dataflows in Preview](#)
- [\*\*SAP Change Data Capture \(CDC\) Connector is now General Available\*\*](#)
- [Reduce cost with batch mode in Synapse link for SQL](#)

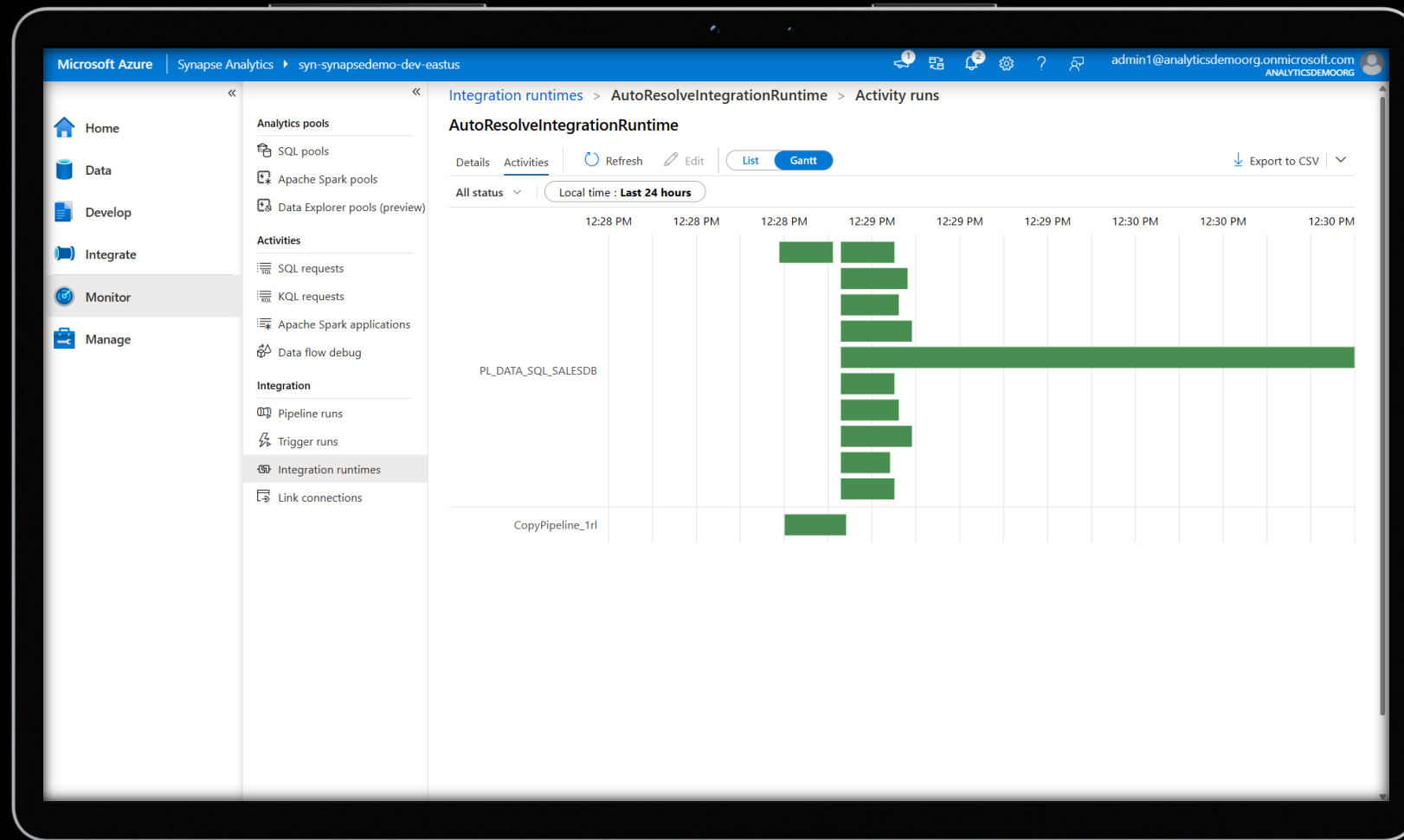


# Gantt View for IR Monitoring

Gantt chart of pipeline runs  
for specific IR

Easy to find overlapping  
pipeline runs/activities

Easy to find longest activity  
runs (with additional info on  
duration deviation)





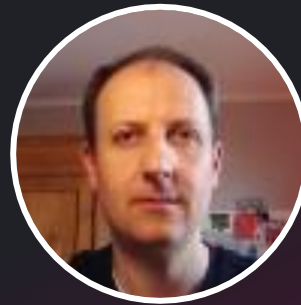
Generally Available

## SAP Change Data Capture (CDC) Connector

Incremental data ingestion from SAP applications

Access to a wide range of sources across all major SAP applications

Leverages SAP Operational Data Provisioning (ODP) framework



Credits go to:

**Ulrich Christ**

Principal Program Manager



# Data Engineering

# Data Engineering

July-December 2022

- **Synapse runtime for Spark 3.2 is now generally available**
- Intelligent Cache is now generally available
- Synapse runtime for Apache Spark 2.4 officially enters its retirement lifecycle
- Synapse Notebooks is now fully compatible with IPython
- Mssparkutils now has spark.stop() method
- Azure Synapse Analytics Spark elastic pool storage is available for public preview
- Azure Synapse Analytics Spark Optimized Write is available for public preview
- New informative Livy errors codes
- **R language support with key library management capabilities is now in preview**
- Increasing Spark performance
- **Azure Synapse Runtime for Apache Spark 3.3 is now in Public Preview**

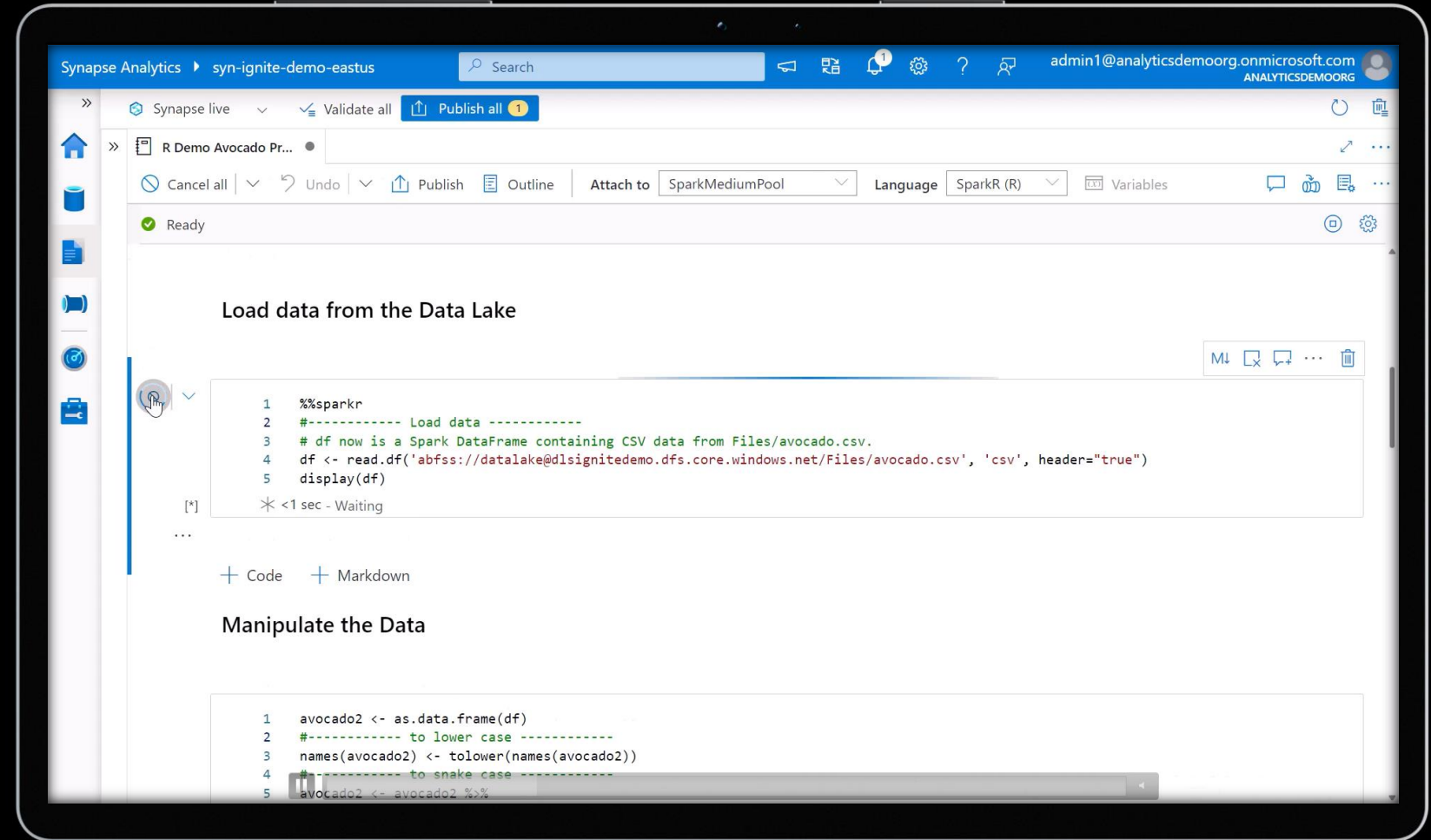
## Public Preview



# R Language Support

R language support with key library management capabilities

Enables data scientists to apply the industry standard R language to process data and develop ML models



## Generally Available

# Spark 3.2

Enables developers  
can leverage the latest  
innovations in the  
Spark ecosystem

### Pandas (Koalas) integration

A highly popular and flexible library with broad industry adoption

### Adaptive Query Execution (AQE) enabled by default

Significant improvements in query performance out-of-the-box

### Small Query execution improvements

Small queries run faster due to reduced initialization overhead

## Public Preview

# Spark 3.3

---

Enables developers can leverage the latest innovations in the Spark ecosystem

## Refreshed and upgraded runtime

Delta 2.1, Python 3.9, Scala 2.12.15, Java 1.8.0\_272, .NET 3.2, .NET for Apache Spark 1.0, R 4.2.1

## Performance improvements

Bloom Filter Joins, Query Execution improvements, supporting in Spark's vectorized Parquet reader for complex types such as lists, maps, and arrays.

## Pandas improvements

Default index from 'sequence' to 'distributed-sequence' what results in 2x faster data scanning in Pandas API on Spark. Pandas API enhancements.



# Data Warehousing



# Data Warehousing

July-December 2022

- **Add SQL objects and users in Lake databases**
- **Spark Delta Lake tables are available in serverless SQL pool for public preview**
- Multi-column Distribution is available for public preview in dedicated SQL pools
- Distribution Advisor is available for public preview in dedicated SQL pools
- Accelerate your Oracle data warehouse migration to dedicated SQL pools in Azure Synapse Analytics
- Auto-statistics for OPENROWSET in CSV datasets
- MERGE T-SQL is now Generally Available

## Public Preview

# Add SQL Objects and Users in Lake Databases

Views, stored procedures, inline table functions

Custom SQL objects require custom schema

Add lake database scoped users from AAD

The screenshot displays the Microsoft Azure Synapse Analytics interface. The top navigation bar shows 'Microsoft Azure | Synapse Analytics | syn-analytics-dev-eastus'. The left sidebar shows the 'Data' workspace with a tree view of resources. The main pane shows a SQL query being executed in a Lake database. The query is a stored procedure that calculates aggregated data for a specific year. The results are displayed in a table format.

**SQL Objects in Lake ...**

```
19
20 CREATE PROC SQL.procAggFactTrip
21 @year int = 2010
22 AS
23 SELECT
24     d.Year,
25     d.Month,
26     d.Date,
27     SUM(f.PassengerCount) AS PassengerCount,
28     SUM(f.Trips) AS Trips
29 FROM dbo.FactTripAgg AS f
30 INNER JOIN SQL.vwDimDate AS d
31 ON f.DateKey = d.DateKey
32 WHERE d.Year = @year
33 GROUP BY d.Year, d.Month, d.Date;
34 GO
35
36 EXEC SQL.procAggFactTrip @year = 2011;
37 GO
```

**Results** | Messages

View: Table | Chart | Export results

Year	Month	Date	PassengerCount	Trips
2011	10	2011-10-13T00:00:00.0000000	876907	535782
2011	4	2011-04-19T00:00:00.0000000	804165	496251
2011	11	2011-11-25T00:00:00.0000000	666590	367964

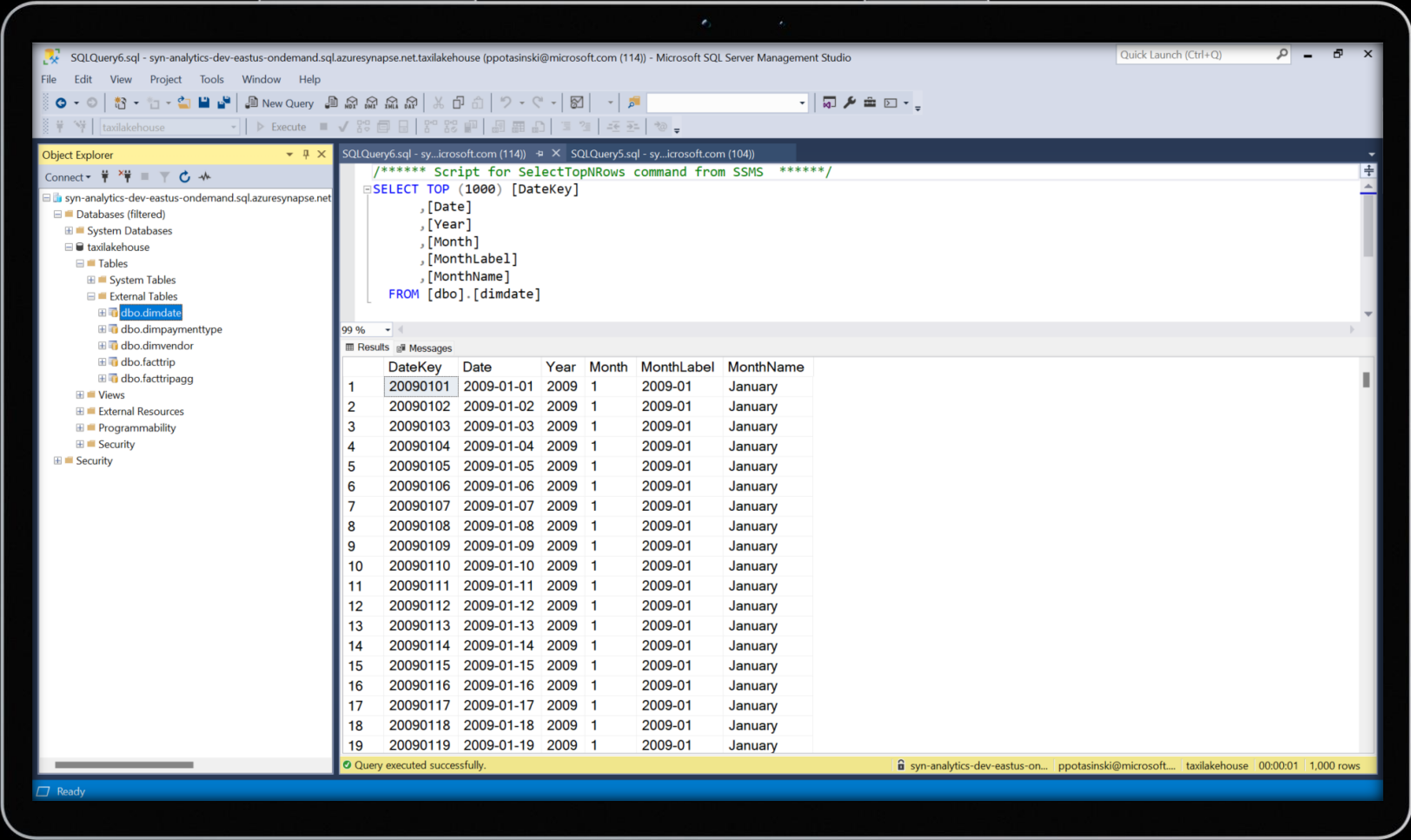
00:00:03 Query executed successfully.

Public Preview

# Spark Delta Lake Tables Available in Serverless SQL pool

Create Delta Lake tables in  
Spark and read their data  
using serverless SQL

External tables for serverless  
SQL pool are created  
automatically



Generally Available

# MERGE SQL Statement

Improve readability and ease migration by executing INSERT, UPDATE, and DELETE functionality in a single statement

INSERT ...  
UPDATE ...  
DELETE ...



MERGE ...



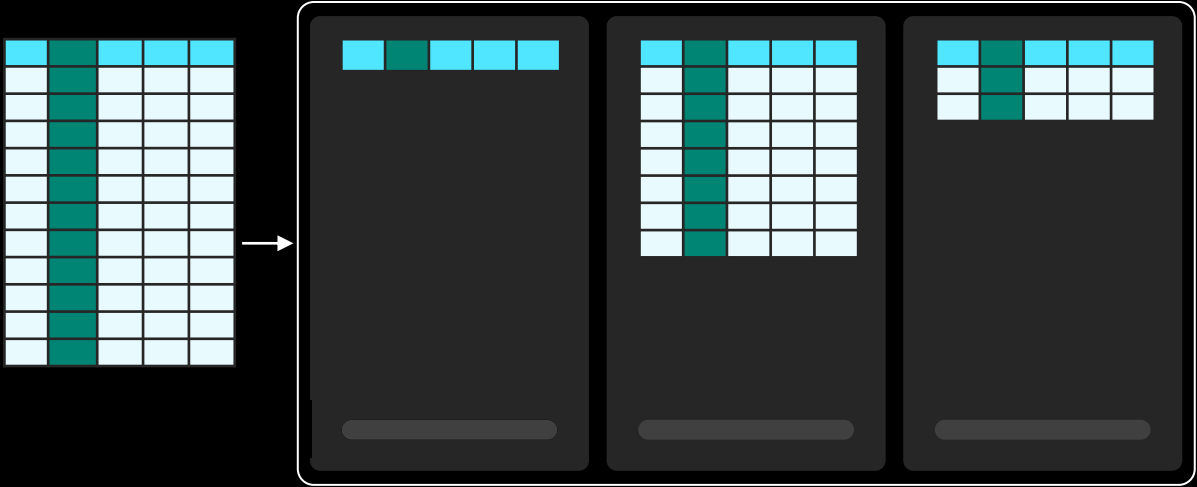
Public Preview

# Multi-Column Distribution

Faster query performance, less  
data skew and  
easier migrations

## Uneven Distribution (Skew)

```
CREATE TABLE SalesTransactions (  
  WITH DISTRIBUTION =  
  (HASH(ProductKey))
```



## Balanced Distribution

```
CREATE TABLE SalesTransactions (  
  WITH DISTRIBUTION =  
  (HASH(ProductKey, RegionKey))
```



Balanced Distribution: Queries execute faster

# Public Preview

# Distribution Advisor

Recommendation system that analyses the chosen queries or past run query data to provide suggestions on data distribution which can improve performance

[aka.ms/SynapseToolbox](https://aka.ms/SynapseToolbox)

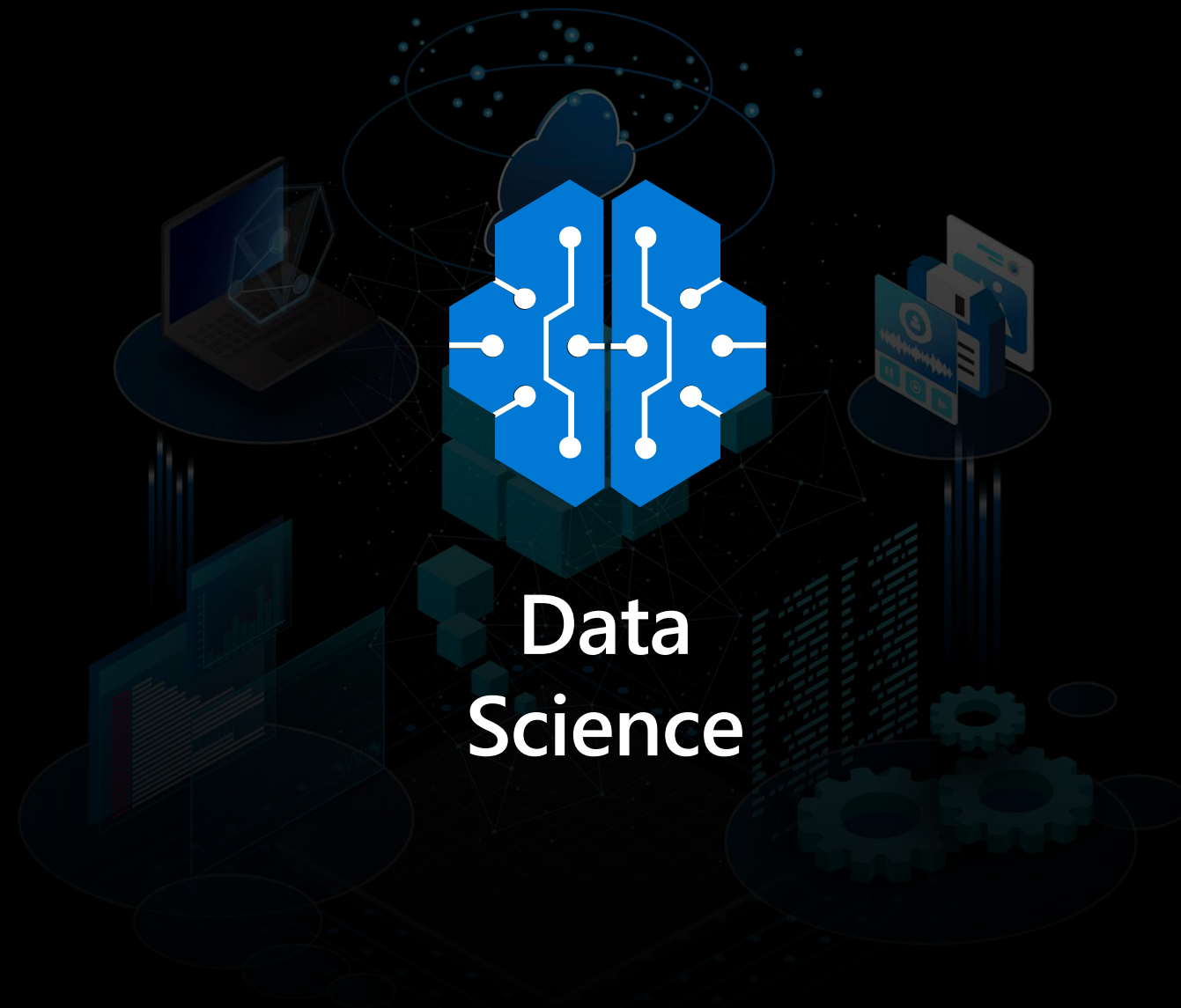
```
SELECT
    location,
    product
FROM sales
GROUP BY location, product
ORDER BY location, product;
```

```
SELECT
    location,
    COUNT(*) AS number_of_sales
FROM sales
WHERE product = 'coffee'
```

[illegible]

**Advisor recommends redistributing table on the Product column**

[illegible]





# Data Science

July-December 2022

- [.NET, C#, and F# Support](#)
- [\*\*OpenAI language models\*\*](#)
- [Full support for MLflow](#)
- [No cluster, no problem: Experiment in browser with Binder](#)
- [Other updates](#)

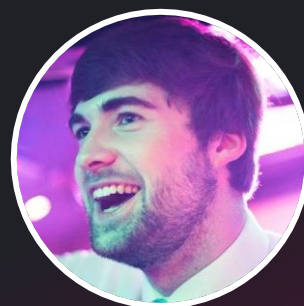
Generally Available

## SynapseML 0.10.0

Open-Source Spark Library  
Distributed ML algorithms for  
Apache Spark

Cognitive Services Integration  
for Spark

Language Binding Generators:  
Python, R, Java (, .NET)



Credits go to:

Mark Hamilton

Senior Software Engineer



# Observational Analytics

# Observational Analytics

July-December 2022

- [Synapse Web Data Explorer now supports rendering charts for each y column](#)
- [Create tables graphically in Synapse Web Data Explorer](#)
- [Free cluster upgrade option](#)
- [Analyzing fresh ADX data from Excel pivot table](#)
- [ADX Web Explorer- Query results exploration - Color by value](#)
- [ADX Web Explorer- Crosshair support for charts \(ysplit=panels\)](#)
- [Data mapping transformations](#)
- [Logstash connector HTTP/HTTPS proxy configuration](#)
- [Kafka Connect support of Protobuf format](#)
- [Embed Azure Data Explorer dashboards in third-party apps](#)
- [Funnel visuals](#)
- [.NET and Node.js support in Sample App Generator](#)
- [ADX Emulator](#)
- [Ingesting files from AWS S3](#)
- [Azure Stream Analytics ADX output \[Generally Available\]](#)
- [Open Telemetry exporter](#)
- [Streaming support in Telegraf connector](#)
- [Protobuf support in Kafka sink](#)
- [Leader follower discoverability](#)
- [Aliasing follower databases](#)
- [\*\*Parse-kv operator\*\*](#)
- [\*\*Scan operator\*\*](#)

## KQL – parse-kv

Extracts structured information from a string expression and represents the information in a key/value form

Supports specified delimiter, non-specified delimiter and Regex

Kusto

Copy

```
print str="2021-01-01T10:00:34 [INFO] ThreadId:458745723, Machine:Node001, Text: Started"
| parse-kv str as (Text: string, ThreadId:long, Machine: string)
| project-away str
```

Output

Text	ThreadId	Machine
Started	458745723	Node001

## KQL – scan

Scans data, matches, and builds sequences based on the predicates

Many use scenarios: process mining, cumulative calculations, fill forward, sequences of events

Kusto

Copy

```
range x from 1 to 5 step 1
| extend y = 2 * x
| scan declare (cumulative_x:long=0, cumulative_y:long=0) with
(
    step s1: true => cumulative_x = iff(s1.cumulative_x >= 10, x, x + s1.cumulative_x),
    cumulative_y = iff(s1.cumulative_y >= 10, y, y + s1.cumulative_y);
)
```

Output

x	y	cumulative_x	cumulative_y
1	2	1	2
2	4	3	6
3	6	6	12
4	8	10	8
5	10	5	18





# Business Intelligence

July-December 2022

- [Improvements to Power BI Datamarts](#)
- [Analysis Services server properties in Power BI Premium](#)
- [\*\*“Horizontal Fusion,” a query performance optimization in Power BI and Analysis Services\*\*](#)
- [Cross-tenant Power BI Dataset Sharing](#)
- [Pushing Top N filter to DirectQuery sources](#)
- [\*\*Optimize ribbon in Power BI Desktop\*\*](#)

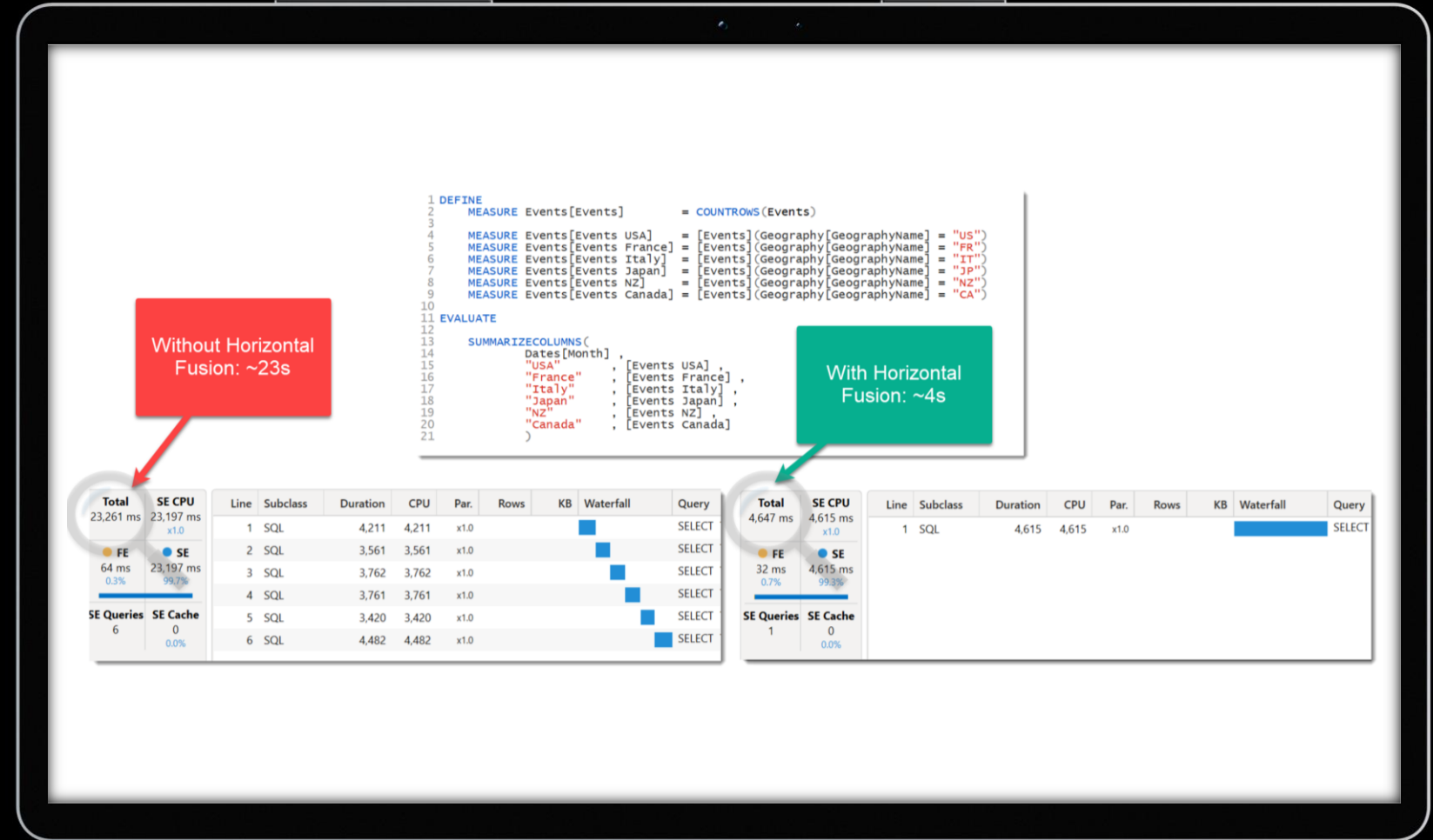
## Public Preview

# Horizontal Fusion

Reduced number of data source queries required to generate the results of DAX queries

Fewer data source roundtrips

Fewer expensive scans over large data sources



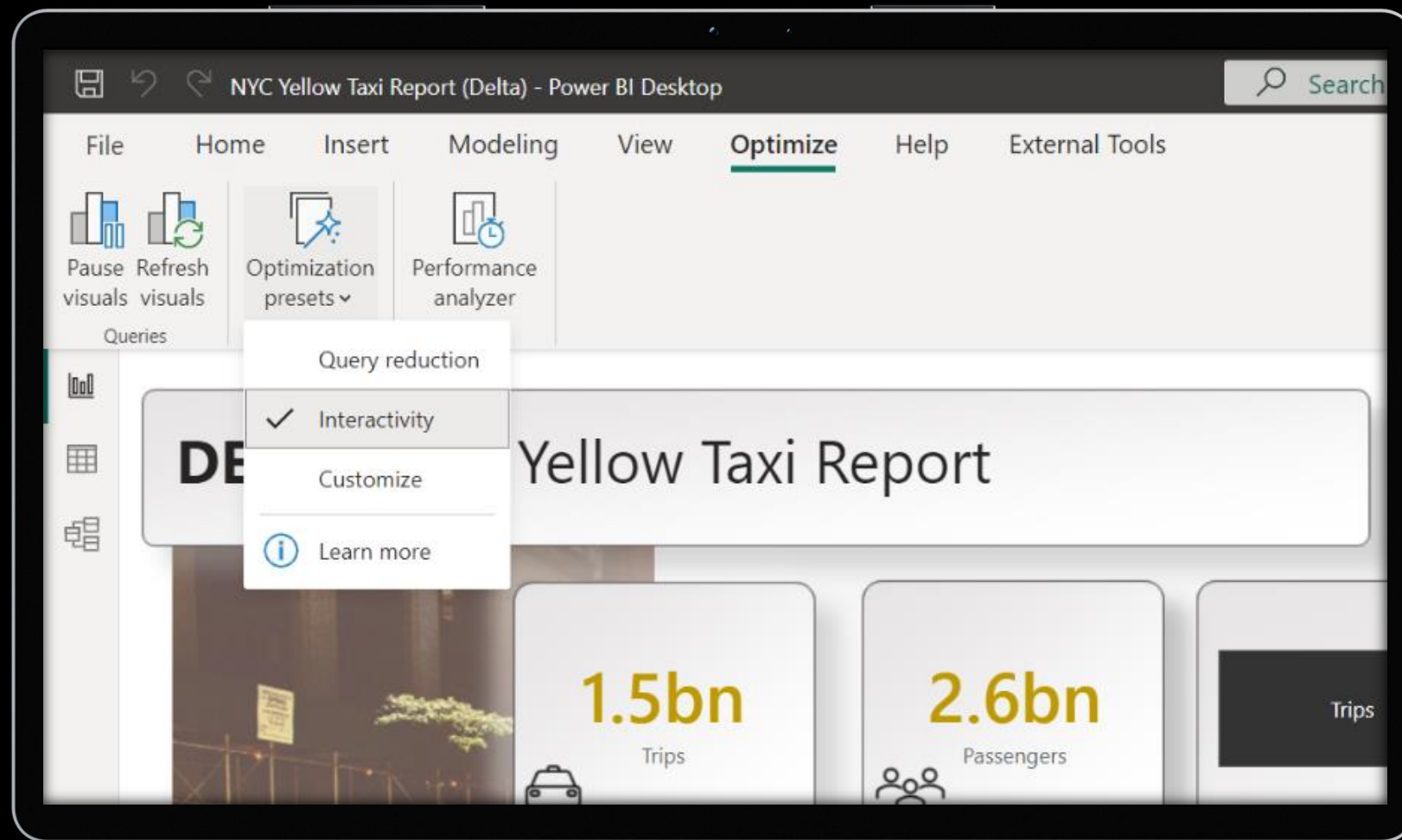
## Public Preview

# Optimize Ribbon in Power BI Desktop

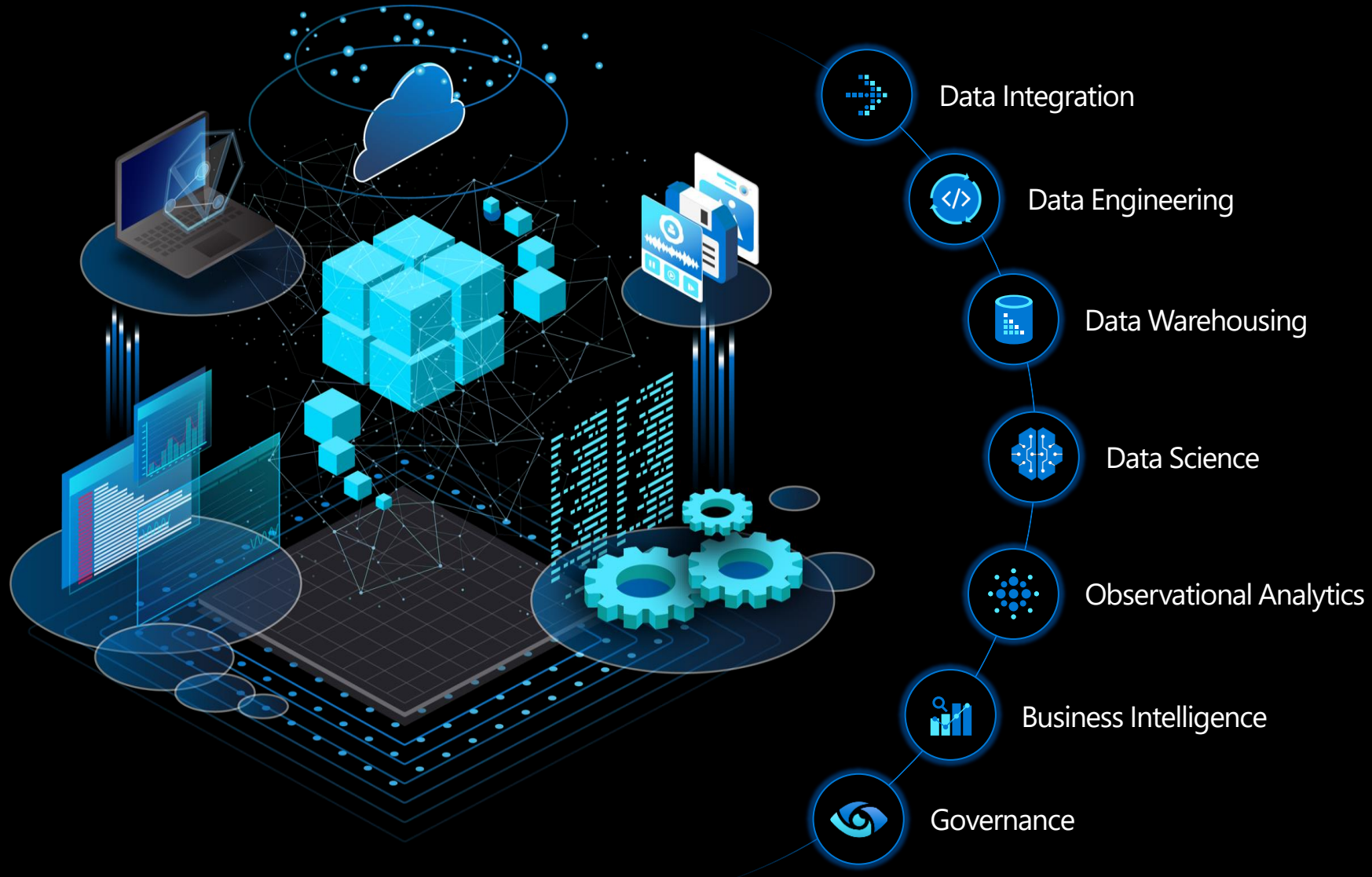
Edit DQ reports/models without running queries until you are ready

Optimization presets

Performance Analyzer



# Synapse + Power BI





# Azure Synapse Community Resources

## Community Call to Action

- ✓ Follow us on Twitter: [@Azure\\_Synapse](https://twitter.com/Azure_Synapse)
- ✓ Read and comment on our blog: <https://aka.ms/SynapseBlog>
- ✓ Check monthly updates blog: <https://aka.ms/SynapseMonthlyUpdate>
- ✓ Subscribe to YouTube channel: <https://aka.ms/SynapseYouTube>
- ✓ Share ideas and vote for them: <https://aka.ms/SynapseIdeas>
- ✓ Join Azure Synapse Influencers program: <https://aka.ms/SynapseInfluencers>

## Product

- Product page: <https://aka.ms/Synapse>
- Documentation: <https://aka.ms/SynapseDocs>
- Reference architectures: <https://aka.ms/SynapseArchitectures>
- Security whitepaper: <http://aka.ms/SynapseSecurity>

## Technical Discussions and Q&A

- Microsoft Q&A: <https://aka.ms/SynapseQuestions>
- StackOverflow: <https://aka.ms/SynapseStackOverflow>

## Quickstart

- Get started in 60 minutes: <https://aka.ms/SynapseGetStarted>
- Azure Synapse Analytics Toolkit: <https://aka.ms/SynapseToolkit>

## Free Learning

- MS Learn – learning paths: <https://aka.ms/SynapseLearningPaths>
- Synapse Practitioner: <https://aka.ms/SynapsePractitioner>
- Microsoft Virtual Training Days: <https://aka.ms/SynapseMVTD>
- 30-Day Cloud Skills Challenge: <https://aka.ms/SynapseSkillsChallenge>

## Samples & Accelerators

- Samples on GitHub: <https://aka.ms/SynapseSamples>
- Azure Synapse Toolkit: <https://aka.ms/SynapseToolkit>
- Accelerator – End-to-End Analytics: <https://aka.ms/azsynapse2e-git>

[aka.ms/SynapseCommunityResources](https://aka.ms/SynapseCommunityResources)







DATA TOBOGGAN

# What's New and Cool in Azure Synapse Analytics

**Pawel Potasinski**  
Senior Program Manager

 @pawelpotasinski  
 /in/pawelpotasinski

Get this slide deck from:  
<https://github.com/pawelpo/presentations>

