Getting started with building your Azure Synapse environment

Erwin de Kreuk
Principal Consultant – Lead Data & AI
InSpark

@erwindekreuk
linkedin.com/in/erwindekreuk
erwindekreuk.com
github.com/edkreuk

Microsoft® Most Valuable Professional

Erwin de Kreuk

Principal Consultant – Lead Data & AI
InSpark

Erwin de Kreuk
Principal Consultant - Lead Data & AI
InSpark

@erwindekreuk
linkedin.com/in/erwindekreuk
erwindekreuk.com
github.com/edkreuk
https://sessionize.com/erwin-de-kreuk/

Let's connect

Microsoft® Most Valuable Professional

We Are InSpark

We help organizations **accelerating** their **digital transformation** with impactful Microsoft solutions & expertise

What is Synapse Analytics

1. Unified Analytics Platform
2. Analytic Runtimes
3. Synapse Studio
4. Networking
5. Pipelines
6. Load Data

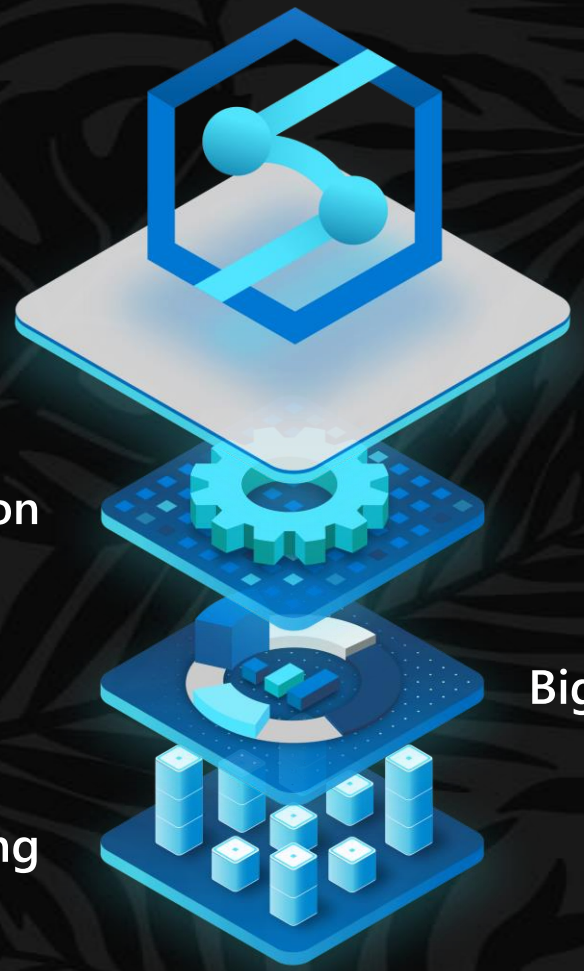
Each new technology creates another siloed operation

- Big data
- Data integration
- Machine learning
- Business intelligence
- Data governance
- Security

Azure Synapse Analytics

The first unified, cloud native platform for converged analytics

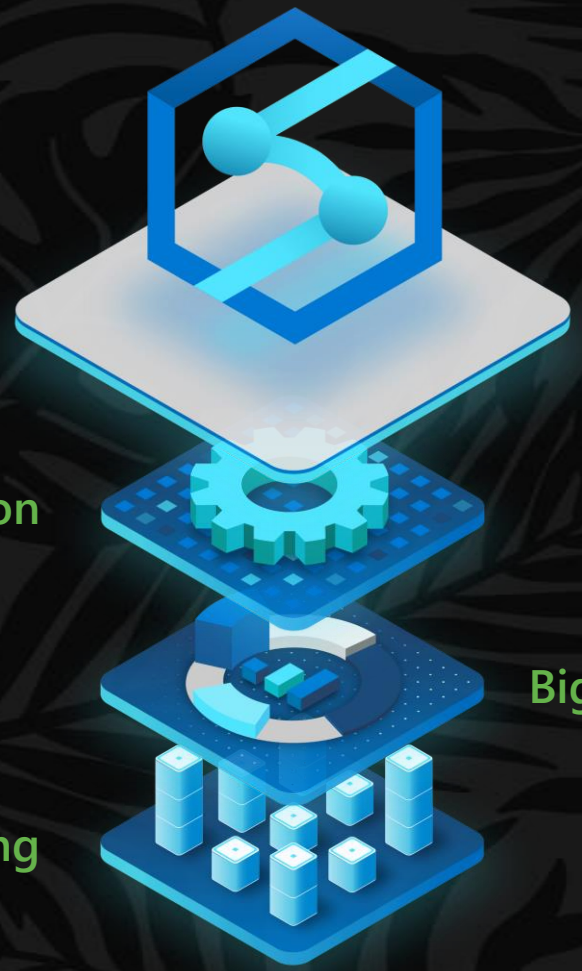
Azure Synapse is the only unified platform for analytics, blending big data, data warehousing, and data integration into a single cloud native service for end-to-end analytics at cloud scale.



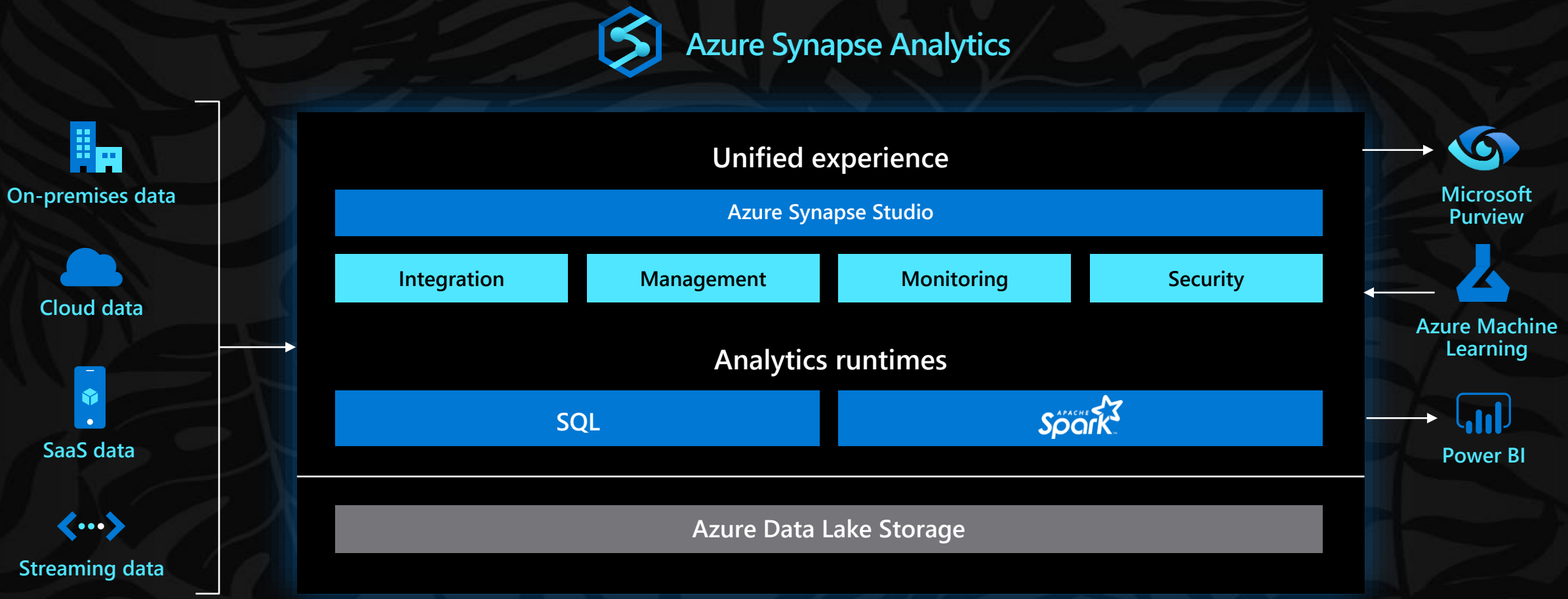
Azure Synapse Analytics

The first unified, cloud native platform for converged analytics

Azure Synapse is the only unified platform for analytics, blending big data, data warehousing, and data integration into a single cloud native service for end-to-end analytics at cloud scale.



Azure Synapse lies at the heart of business, AI, and BI



Azure Synapse Analytics

Synapse SQL

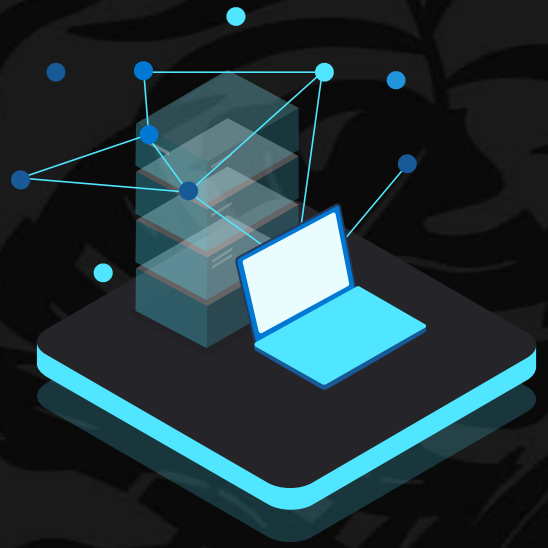
Synapse Spark

Synapse Pipelines

Synapse Studio

Synapse SQL

- Serverless pay-per-query ideal for ad-hoc data lake exploration and transformation
- Dedicated clusters optimized mission-critical data warehouse workloads

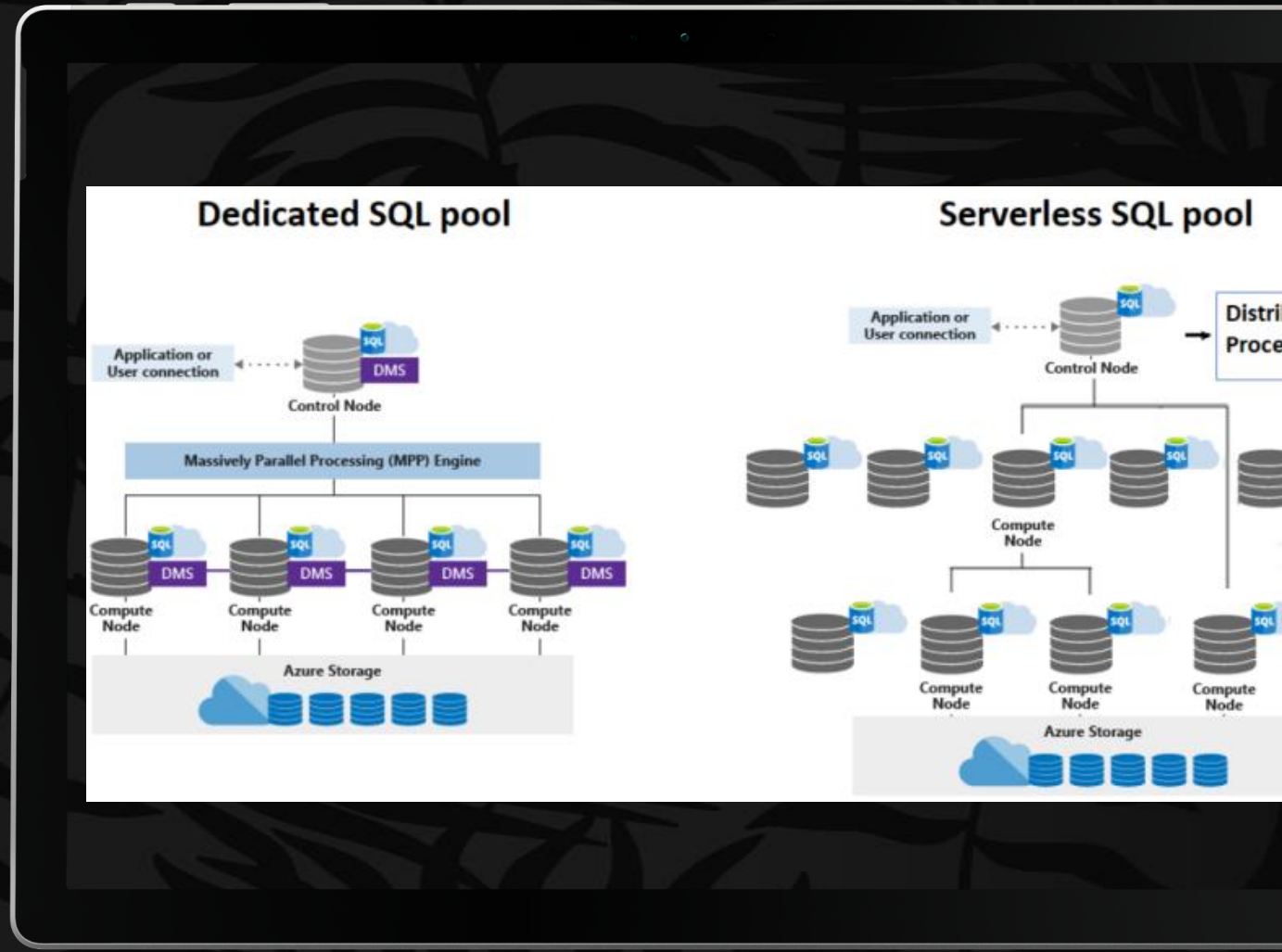


Serverless

Dedicated

Synapse SQL

- Serverless pay-per-query ideal for ad-hoc data lake exploration and transformation
- Dedicated clusters optimized mission-critical data warehouse workloads
  - Minimum 1 TB, 60 million rows
  - Consider serverless
  - No small updates



Synapse SQL

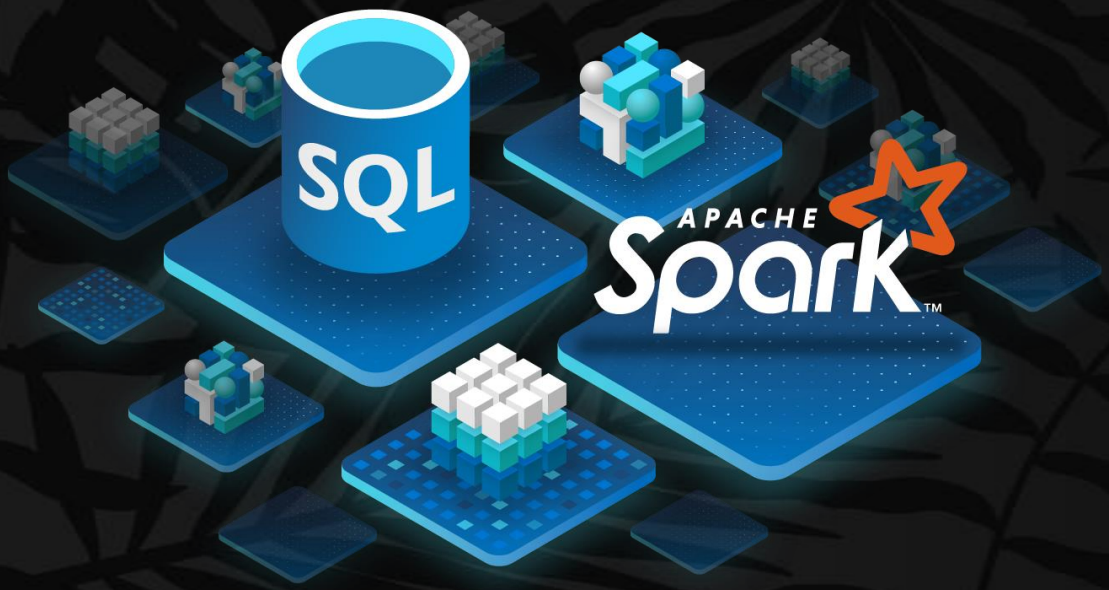
- Serverless pay-per-query ideal for ad-hoc data lake exploration and transformation
- Dedicated clusters optimized mission-critical data warehouse workloads
  - Minimum 1 TB, 60 million rows
  - Consider serverless
  - No small updates

Service Level	DWU	Pay as you go	1 year reserved
DW100c	100	€1,048.661/month	€660.6287/mo ~37% savings
DW200c	200	€2,097.322/month	€1,321.2573/mo ~37% savings
DW300c	300	€3,145.983/month	€1,981.8860/mo ~37% savings
DW400c	400	€4,194.644/month	€2,642.5146/mo ~37% savings
DW500c	500	€5,243.305/month	€3,303.1433/mo ~37% savings
DW1000c	1000	€10,486.610/month	€6,606.2865/mo ~37% savings

Serverless = **€4.757** per TB of data processed

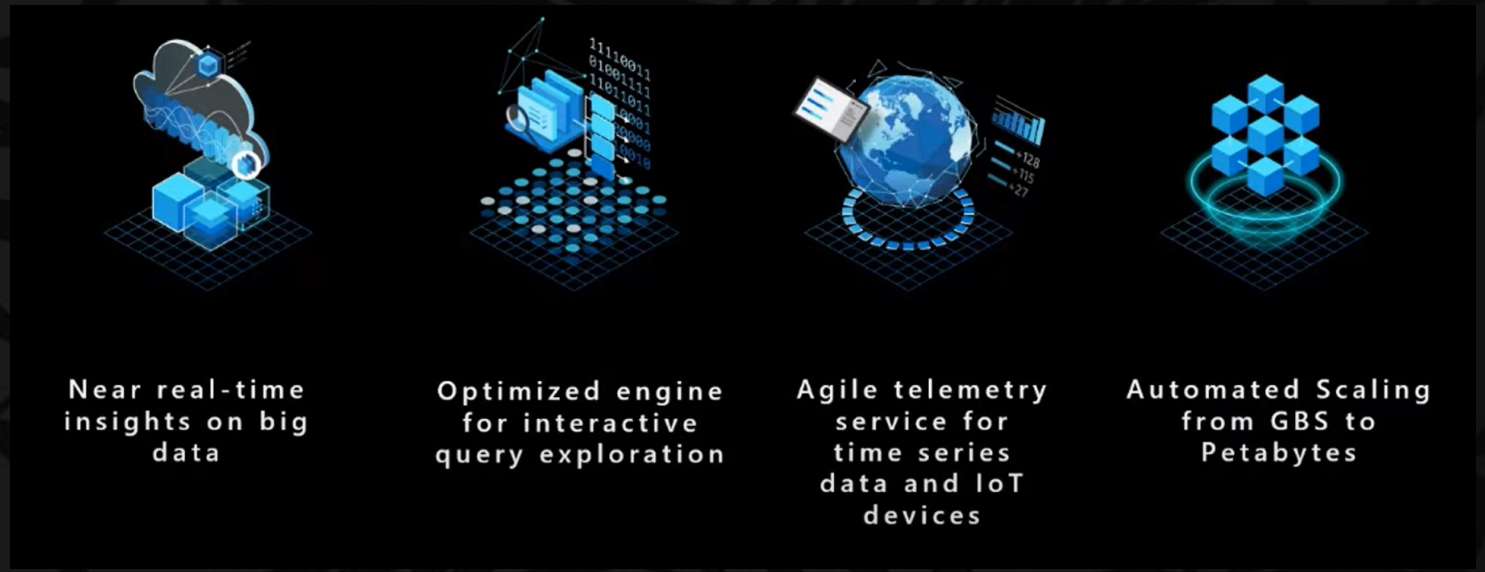
Synapse Spark

- New Cloud native Engine
- Integration with Spark
- Allows multiple languages in one notebook
- Offers use of temporary tables across languages
- Support for syntax highlight, syntax error, syntax code completion, smart indent, and code folding



Synapse Data Explorer (PREVIEW)

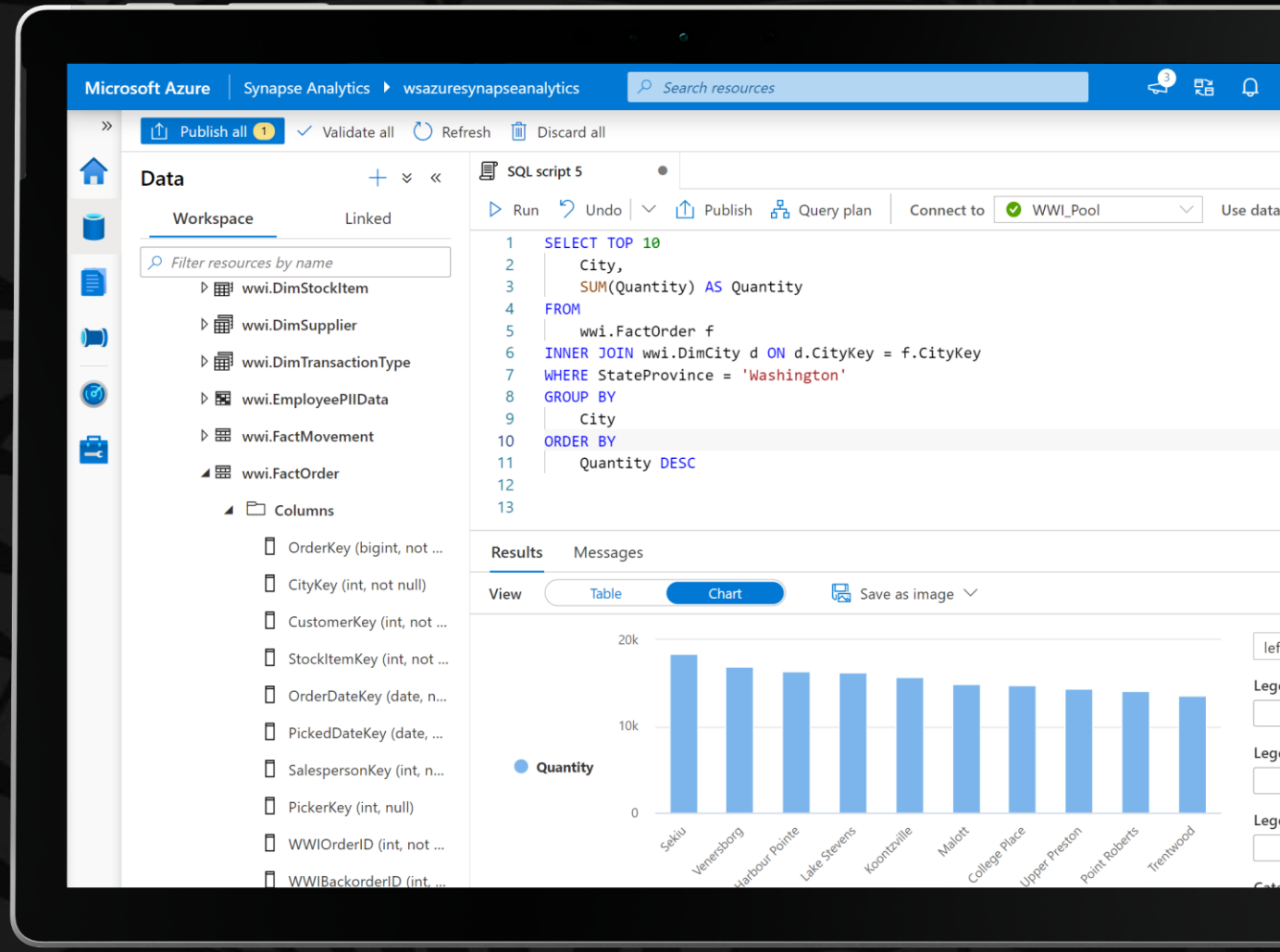
- Building near Realtime Big Data Analytics on custom event/log data
- Security log analytics
- IOT Solutions analytics



Synapse Studio

SQL Editor

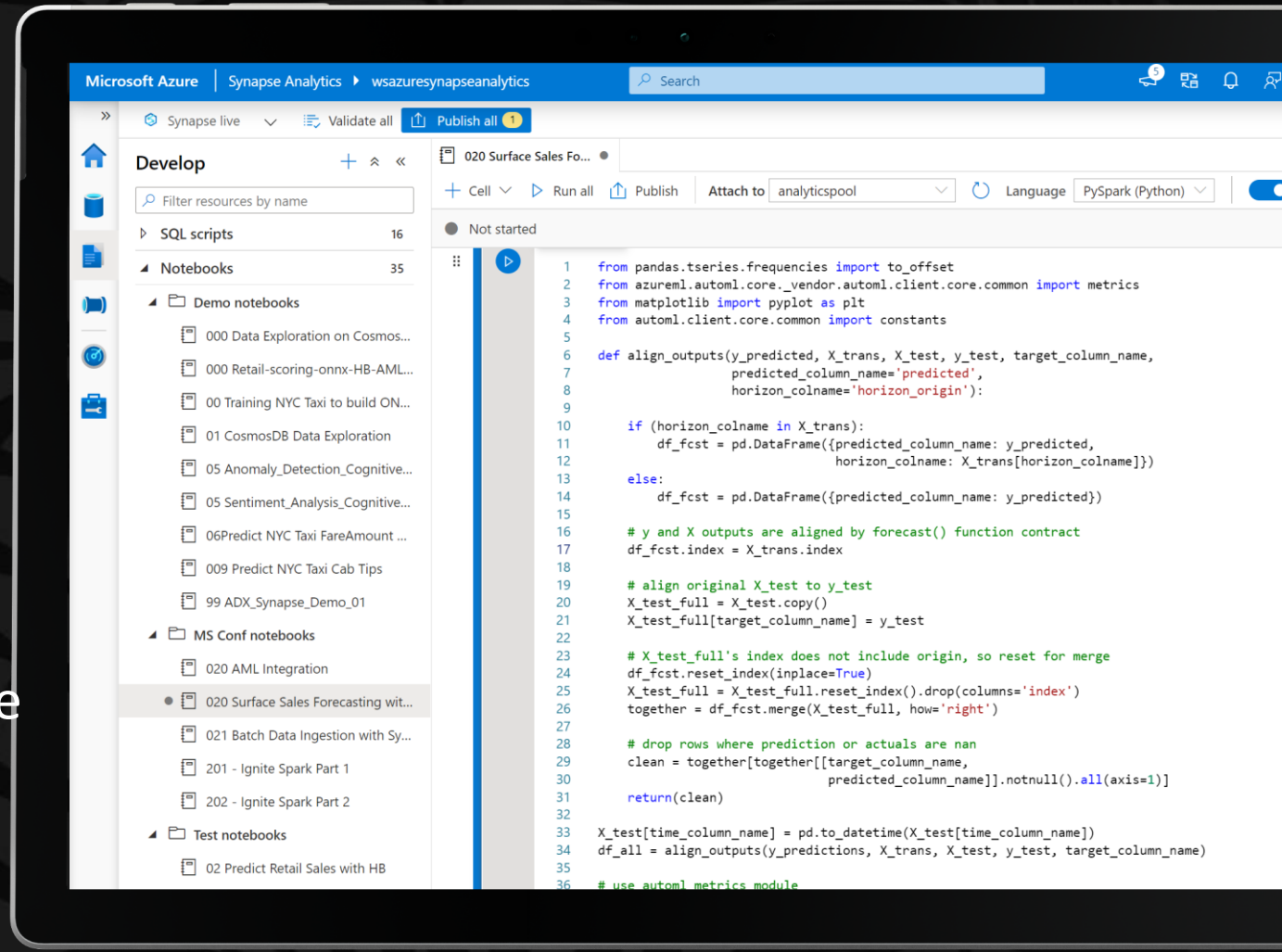
- Automatic code completion (Intellisense)
- Script collaboration within the Workspace
- Built-in visualizations
- Easily switch between clusters



Synapse Studio

Notebook IDE code authoring

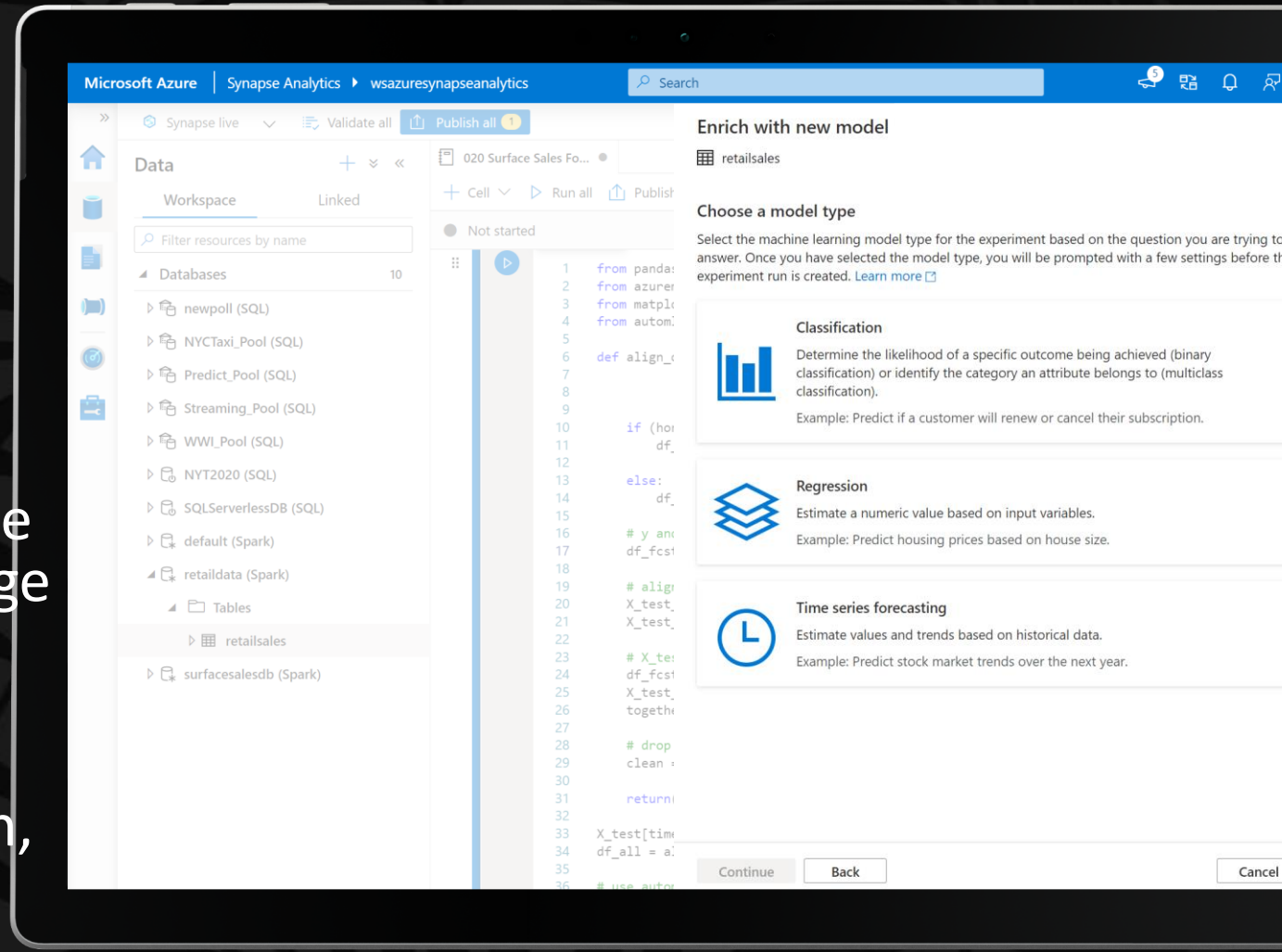
- PySpark, Scala, and C# languages supported
- Automatic code completion (Intellisense)
- Author multiple languages in a single notebook
- Analyze data from the data warehouse, data lake, and real-time operational data from one place



Synapse Studio

Automatic machine learning

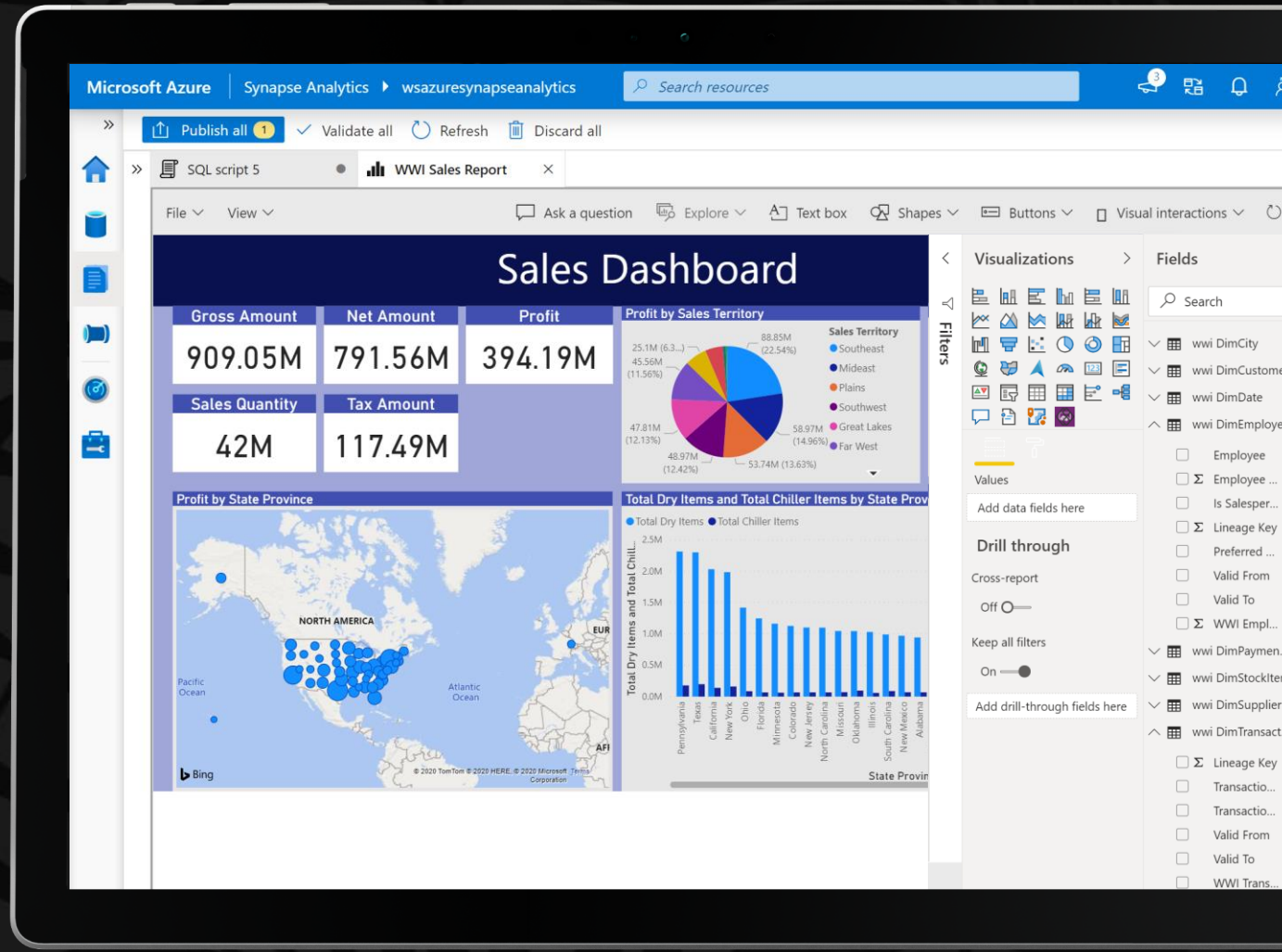
- No-code creation on Machine Learning models
- Democratize ML to everyone since no data science domain knowledge required
- Support for ensemble models
- Supports classification, regression, and time-series forecasting



Synapse Studio

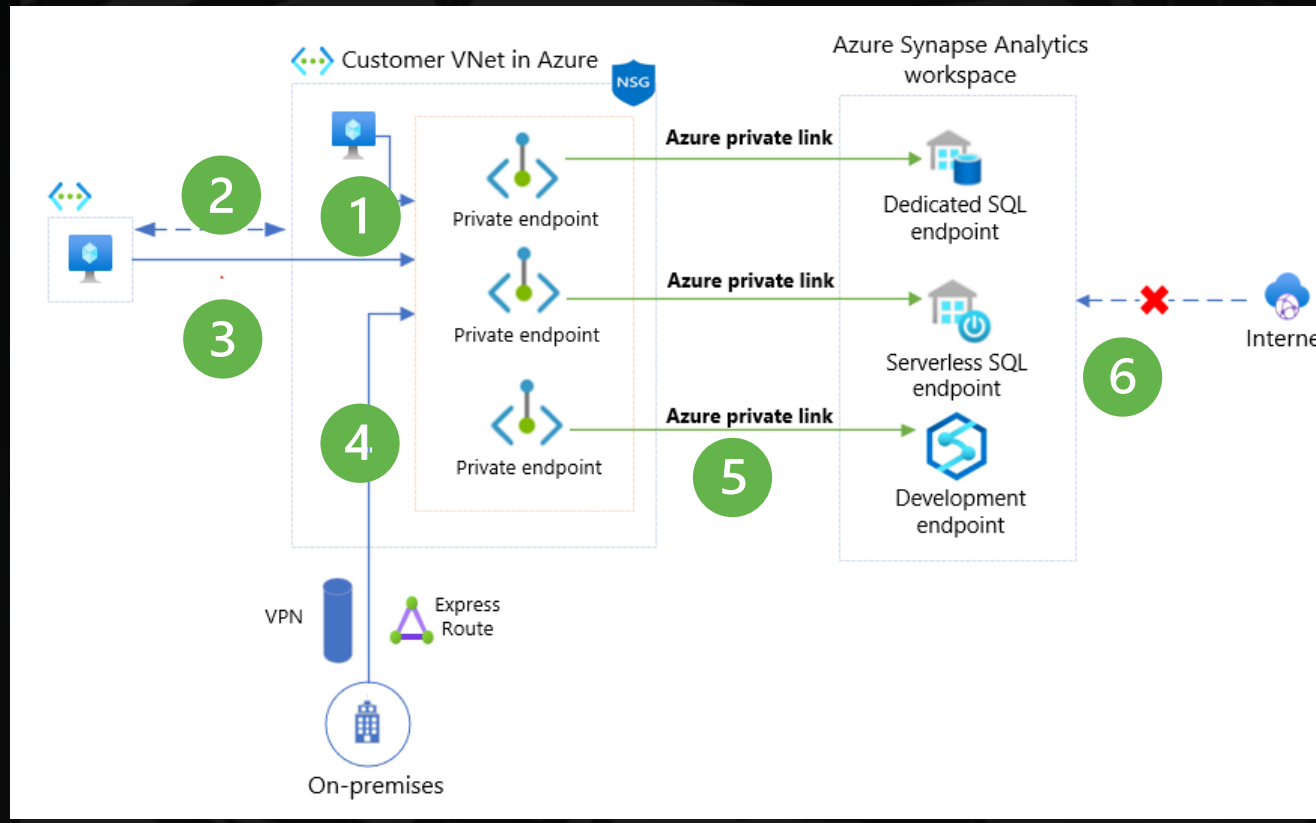
Power BI Integration

- Build dashboard in Synapse Studio



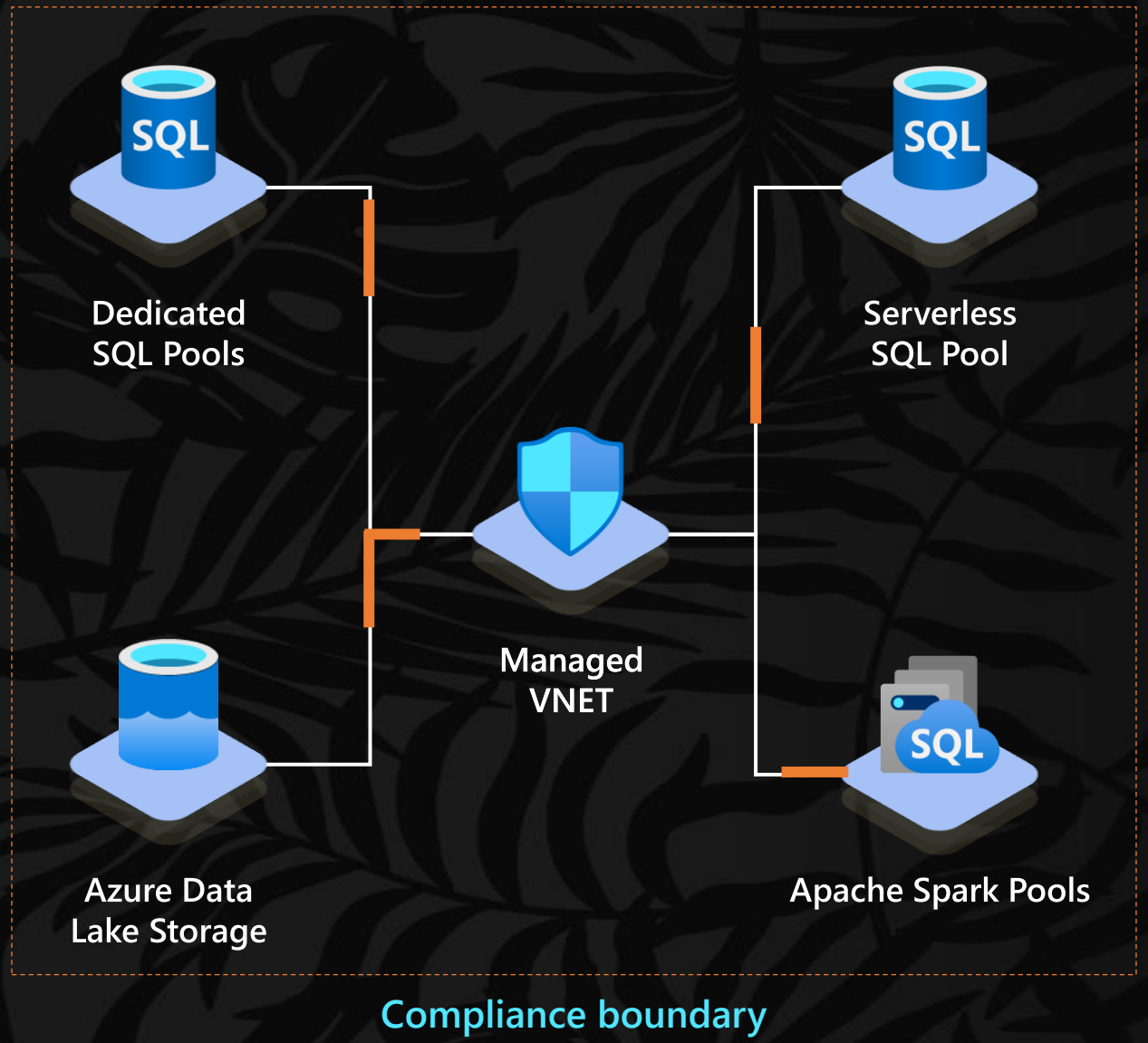
Synapse Network Security

- Workstations from within the customer VNet access the Azure Synapse private endpoints.
- Peering between customer VNet and another VNet.
- Workstation from peered VNet access the Azure Synapse private endpoints.
- On-premises network access the Azure Synapse private endpoints through VPN or ExpressRoute.
- Workspace endpoints are mapped into customer's VNet through private endpoints using Azure Private Link service.
- Public access is disabled on the Synapse workspace.



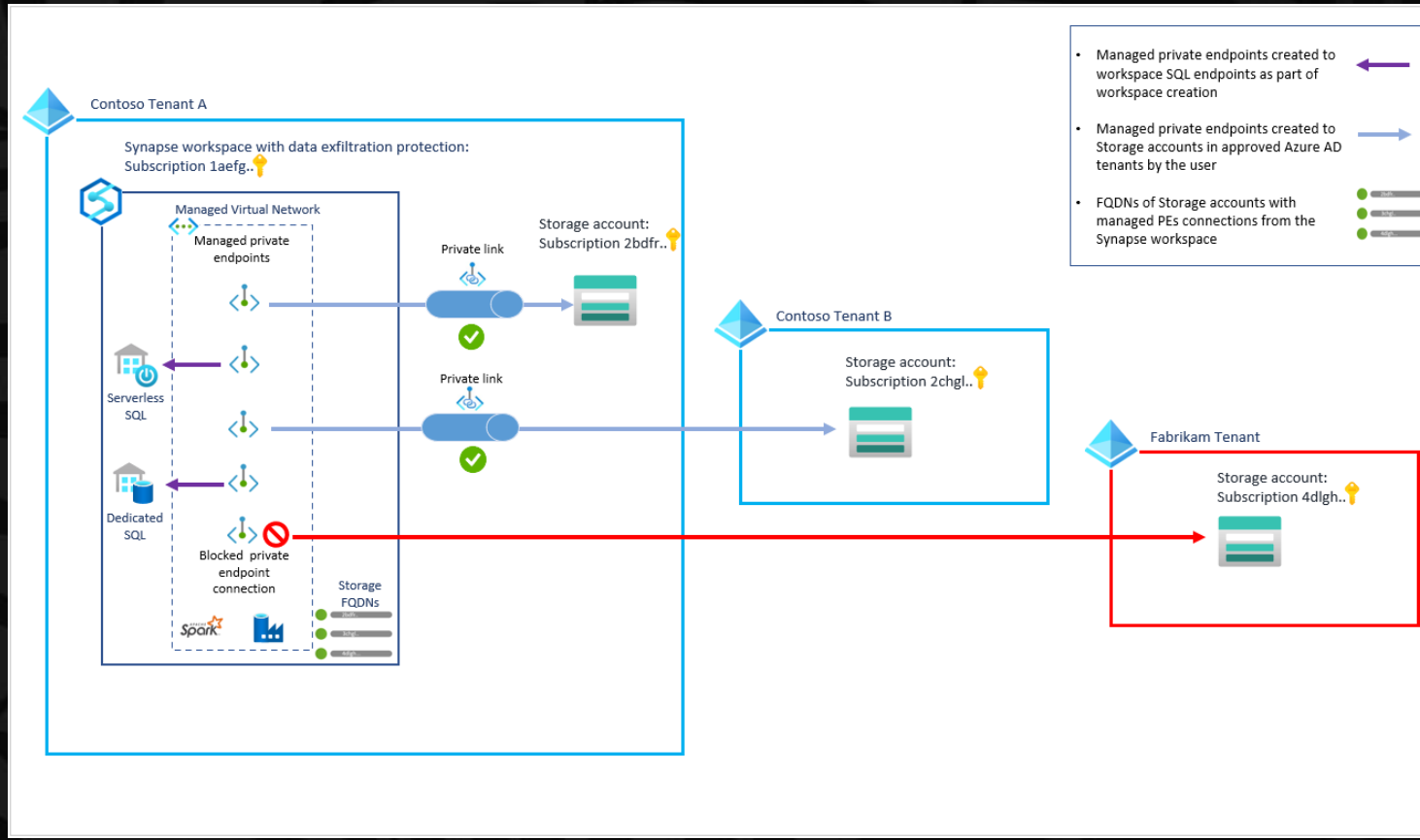
Managed Virtual Networks

- Eliminate network maintenance
- One-click enables automated management of virtual networks between cluster endpoints
- Synapse resources only ever interop with private endpoints
- No management of subnets or IP Ranges
- Prevents data exfiltration



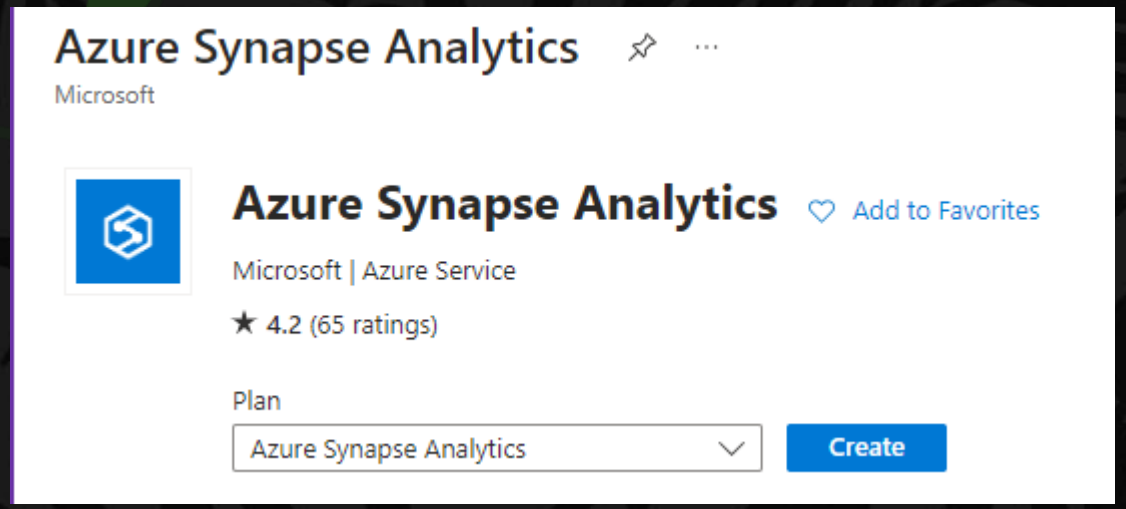
Data exfiltration protection

- Managed Virtual Network
- Exfiltrating sensitive data to locations outside



DEMO

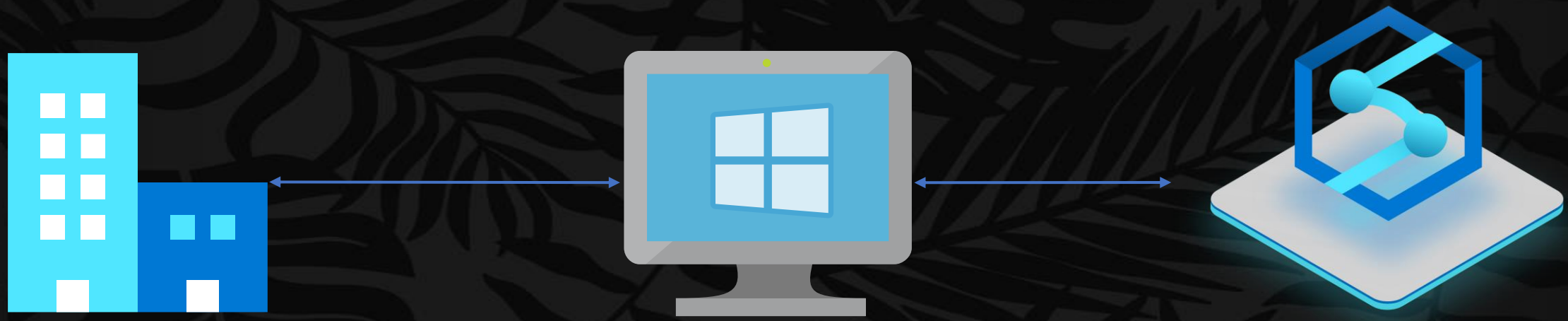
Creating a Synapse Analytics Environment



Governance

- Native integration with Azure Purview
- Automatically discover and classify data assets
- End-to-end data lineage

Connecting to On-Prem



Connecting Power BI /Fabric

Power BI Gateway

public access allowed/Denied

Public network access

Choose whether to permit public network access to your workspace. You can modify the firewall rules after you enable this setting. Learn more

Public network access to workspace endpoints

○ Enable ◉ Disable

You must use private endpoints to connect to your workspace when this setting is disabled. Selecting the **Disable** option will not apply any firewall rules that you may configure.

public access allowed/Denied

▷ Azure Private Link
Enabled for the entire organization

▷ Block Public Internet Access
Enabled for the entire organization

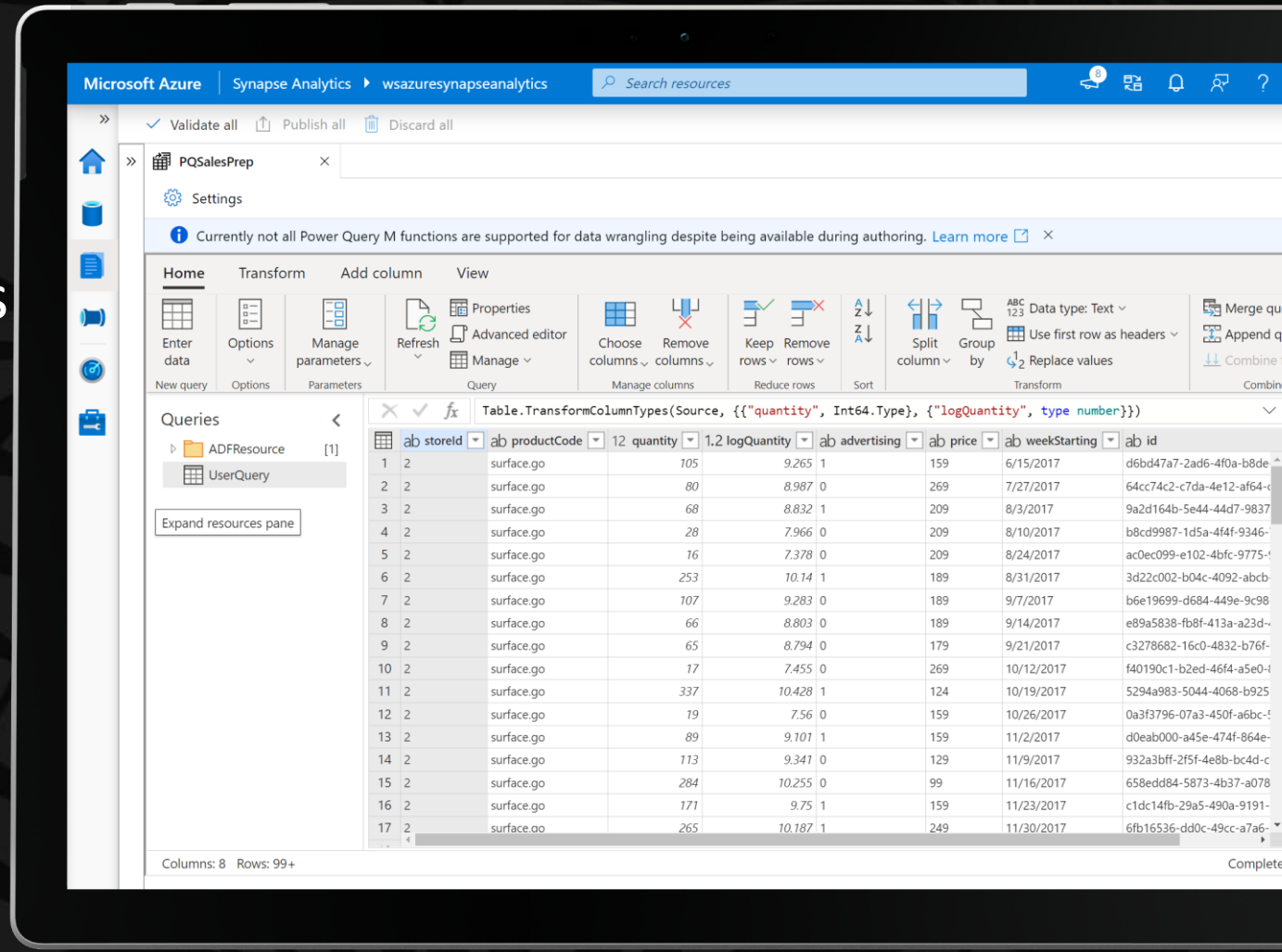
Data Integration

Synapse Pipelines

- Ingest Data from 120+ sources
- Cloud native ETL/ELT
- Secure connectivity to on-premise data sources, other clouds, and SaaS applications
- Code-first and low/no code design interfaces
- Schedule and Event based triggering

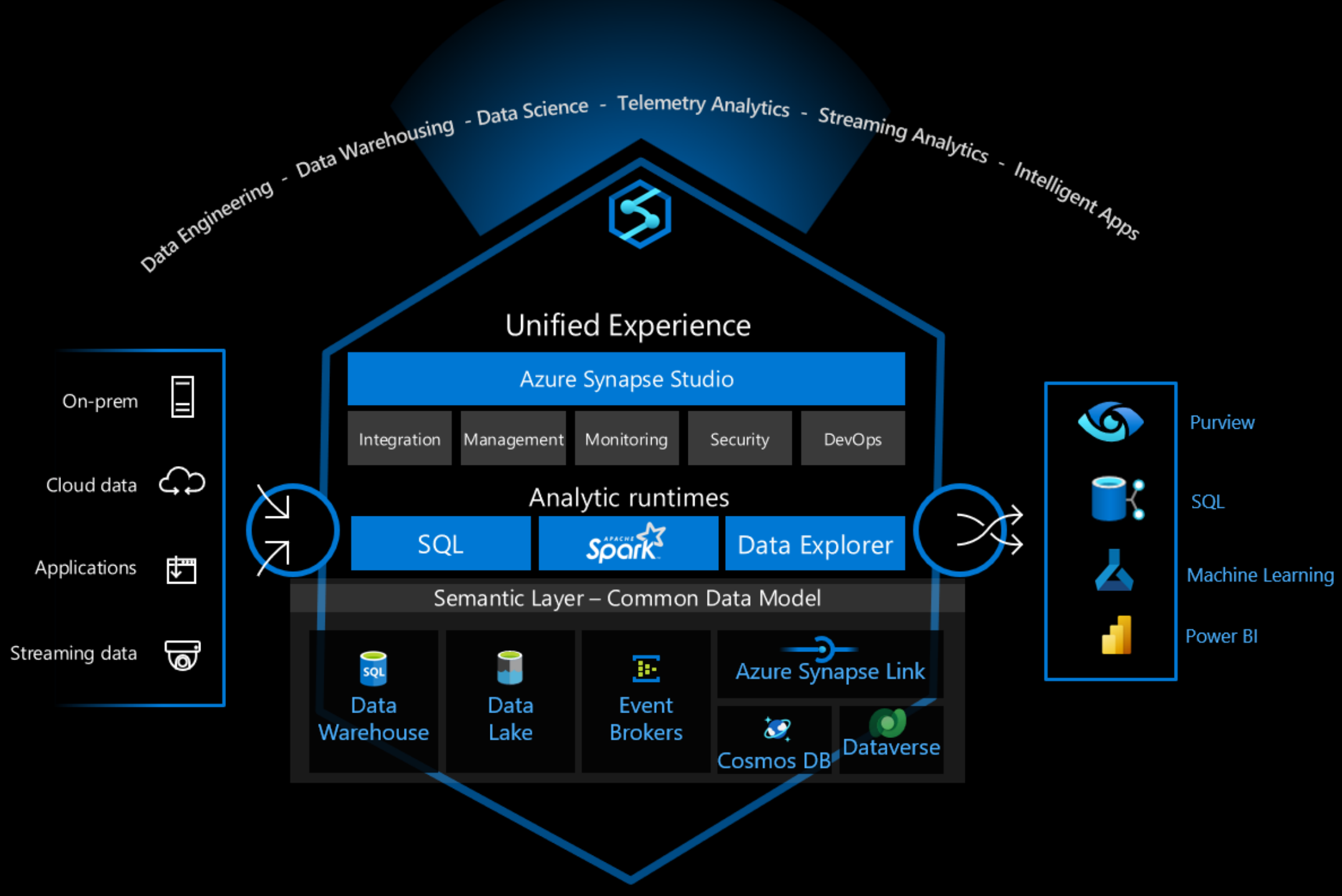
Synapse Pipelines

- No/low-code data transformation
- Excel-like interface is familiar to users
- Transform data to desired shape completely visually
- Operationalize into pipelines



DEMO

Querie Data


Data Engineering - Data Warehousing - Data Science - Telemetry Analytics - Streaming Analytics - Intelligent Apps
Unified Experience
Azure Synapse Studio
Integration
Management
Monitoring
Security
DevOps
Analytic runtimes
SQL
Spark
Data Explorer
Semantic Layer – Common Data Model
Data Warehouse
Data Lake
Event Brokers
Azure Synapse Link
Cosmos DB
Dataverse
On-prem
Cloud data
Applications
Streaming data
Purview
SQL
Machine Learning
Power BI

Final Words

- Azure Synapse Analytics is Unified Data Platform
- Blending Data Integration, Data Warehouse and Big Data Analytics
- SQL/Spark/Data Explorer analytical runtime’s
- Easily query on Data Lake files
- Power BI Integrates
- Native integration with Microsoft Purview

Thank You

Erwin de Kreuk
Principal Consultant – Lead Data & AI
InSpark

Microsoft® Most Valuable Professional

@erwindekreuk

linkedin.com/in/erwindekreuk

erwindekreuk.com

github.com/edkreuk

https://sessionize.com/erwin-de-kreuk/

Let's connect