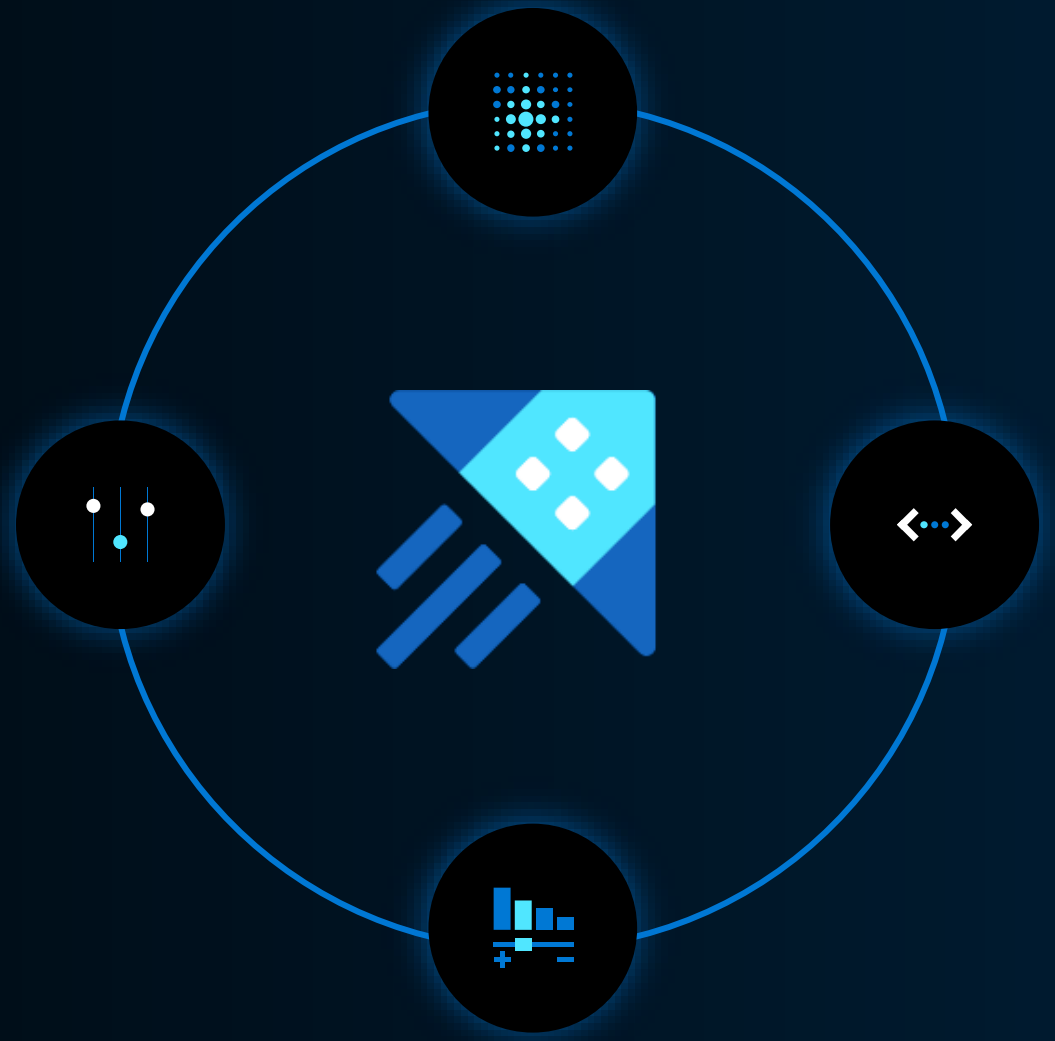
Microsoft

TSI to ADX Migration



Agenda

Lesson 1: Why TSI to ADX migration

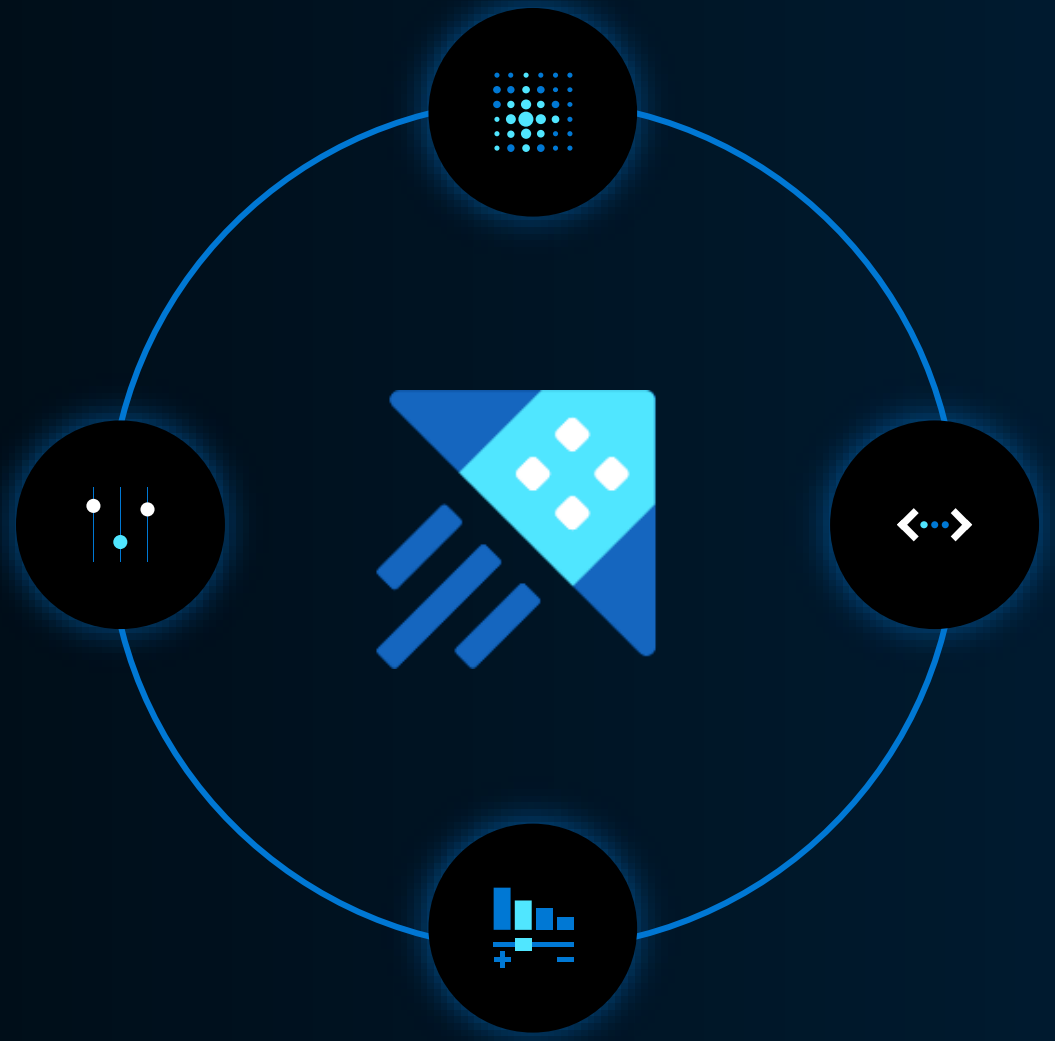
Lesson 2: Understanding ADX

Lesson 3: Migrating from TSI to ADX

Lesson 4: Finalizing the migration

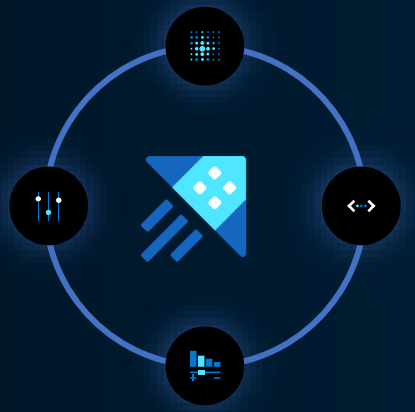
Why Migrate

Lesson 1



Objectives

After completing this Learning, you will be able to understand:



1. Announcement
2. Path, options & timelines
3. What is possible and what is not
4. Gather TSI info

Time Series Insights (TSI) Announcement

- No longer supported after **March 2025**.
- Published docs (https://aka.ms/tsi2adx) regarding deprecation and migration.

Note:

- If you are unable to migrate to Time Series Insights to Azure Data Explorer by **31 March 2025**, your Time Series Insights resources **will be automatically deleted.**
- You'll be able to access Gen2 data in your storage account.
- However, you'll only be able to perform management operations (such as updating storage account settings, getting storage account properties/keys, and deleting storage accounts) through Azure Resource Manager.
- For Gen1 data, if you have a support plan, please create a support ticket to retrieve your Gen1 data. We will keep your Gen1 data until **30 April 2025**.

Feature Comparison

Feature	TSI	ADX
Ingestion	Hubs, limit 1 MB/s	Many data connection methods, SDKs, APIs, No limits (scalable), 200MBs per second per node benchmarked on 16-cores.
Storage & Retention	Warm – multitenant ADX	Cold – Azure Blob storage distributed columnstore, with Hot - Highly optimized SSD (locally on compute nodes).
Formats	JSON	JSON, CSV, Avro, Parquet, ORC, TXT and more.
Querying	TSQ	KQL, SQL
Visualization	TSI Explorer, PBI	PBI, ADX Dashboards, Grafana, Kibana, plus more via ODBC/JDBC
ML	n/a	Built-in plugins, R, Python to train and score data. Native forecasting, anomaly detection, and clustering.
PBI Connector	Public preview	Optimized native PBI connector (GA), supports Direct Query, Import mode, parameters and filters.
Export	Parquet in Blob	Supports continuous export to Azure storage, and External tables to read exported data
HA/DR	Depends on config	HA SLA of 99.9%, AZ supported, built on durable Azure Blob storage.
Security	Private link for incoming, but open for storage and hubs	VNet injection, Private Link, Encryption at rest with customer managed keys
RBAC & RLS	Limited, no RLS	Granular, RLS and DM supported

Migration Paths

TSI Gen1

1. Create ADX Cluster
2. Setup parallel ingestion from hubs to ADX
3. Continue ingesting data until retention or fixed period is met
4. Start using ADX Cluster
5. Delete TSI Env.

Detailed FAQ: How to migrate TSI Gen1 to ADX

TSI Gen2

1. Create ADX Cluster
2. Redirect data ingestion to ADX
3. Import TSI cold data using lightingest
4. Start using DX Cluster

Detailed FAQ: How to migrate TSI Gen2 to ADX

Demo 1*

- Understand the TSI Env - TSI Explorer (azure.com)

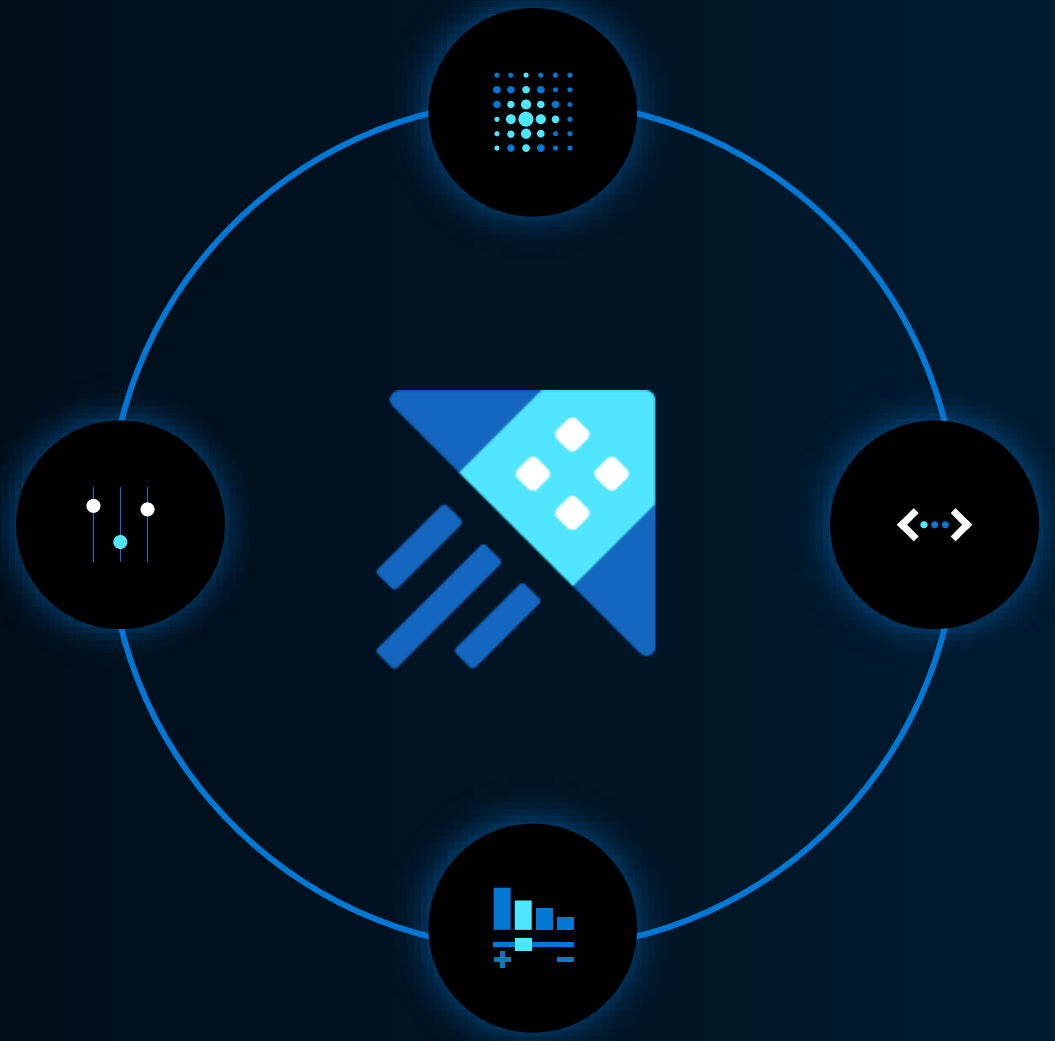
- Gather details
- Blob folder, lateness in data) - ie. Data arriving later than expected

- Give access

* Estimate 10 minutes

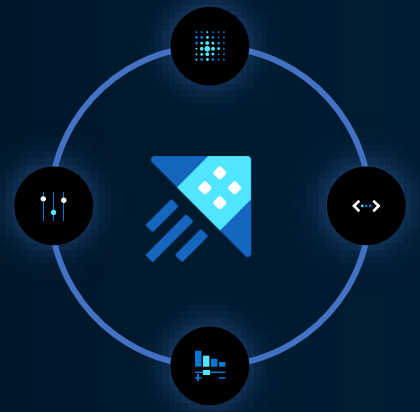
ADX Overview

Lesson 2



Objectives

After completing this Learning, you will be able to understand:



1. Common use-cases
2. What ADX is
3. Architecture
4. Batching vs streaming ingestion
6. Web UI capabilities
7. Extensibility
8. How ADX is Enterprise Ready

ADX is Best Suited

Now also in Synapse

for time stamped data: logs, time series, telemetry

for free text, structured and semi-structured data

Near real-time data analytics at scale

for ad-hoc analytics

ADX Key Differences

Experience	TSI	ADX
Service kind	SaaS - simple	PaaS - deep control and scalability
Models	Hierarchies	Folders, Tables, Data, UDF, Views
User-Interface	TSI Explorer UI	ADX Web-UI, Kusto Explorer, SDKs, Embedding
AI	n/a	Built-in ML plugins, Kqlmagic

Common use cases across verticals

Retail

Web Analytics, IoT Analytics

Financial

Audit Logs

Oil/Gas & Energy

Industrial IoT, Historian, Time Series, Grid analytics

Security

Security Analytics, Threat Detection, SIEM

Healthcare

IoT device Analytics

Advertising

Personalized offers, campaign management

Media Entertainment

Content Delivery Network analytics, Viewer experience analytics, Live Event Analytics

Automotive

Manufacturing, Connected cars, Fleet management

Central Observability (Logs, Traces, Metrics) and Time Series

What is Azure Data Explorer?

- Fully managed
- High-performance
- Big data analytics platform
- Analyze high volumes of data in near real time
- End-to-end solution for data ingestion
- Query, visualization, and management
- Useful for log analytics, time series, IoT, and general-purpose exploratory

What makes ADX unique?

- Data velocity, variety, and volume
- User-friendly query language
- Advanced analytics
- Easy-to-use wizard
- Versatile data visualization
- Automatic ingest, process and export

Major components of ADX

An ADX cluster does all the work to ingest, process, and query your data. The clusters are auto-scalable according to your needs. It stores the data on Azure Storage and caches some of this data on the cluster compute nodes to achieve optimal query performance.

What is an ADX cluster?

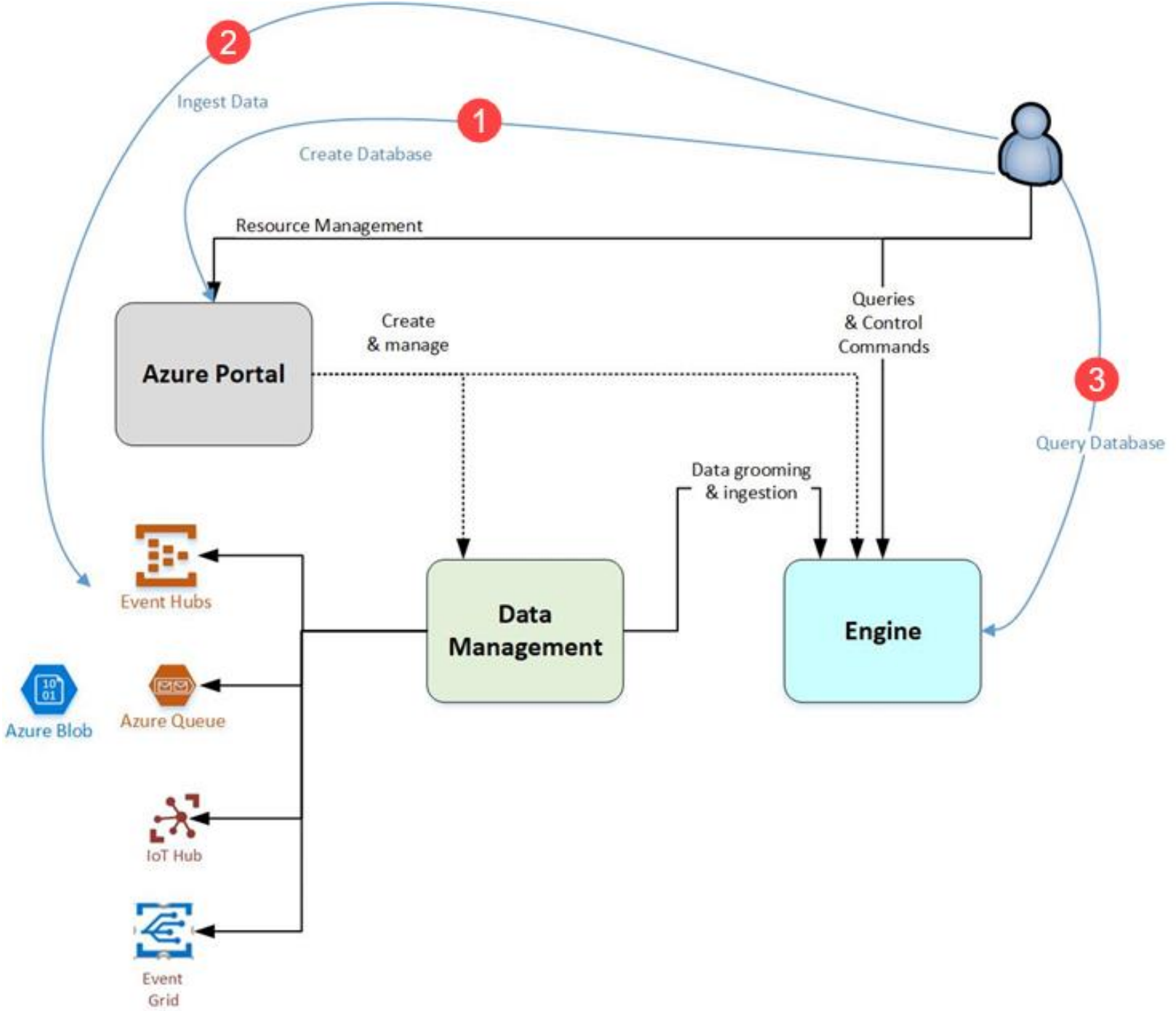
- Each ADX cluster can hold up to **10,000** databases and each database upto **10,000** tables.
- The data in each table is stored in data shards also called "**extents**".
- All data is **automatically indexed** and partitioned based on the ingestion time.
- There are no primary foreign key constraints or any other constraints, such as uniqueness.

The logical structure of a database is similar to many other relational databases. An ADX database can contain:

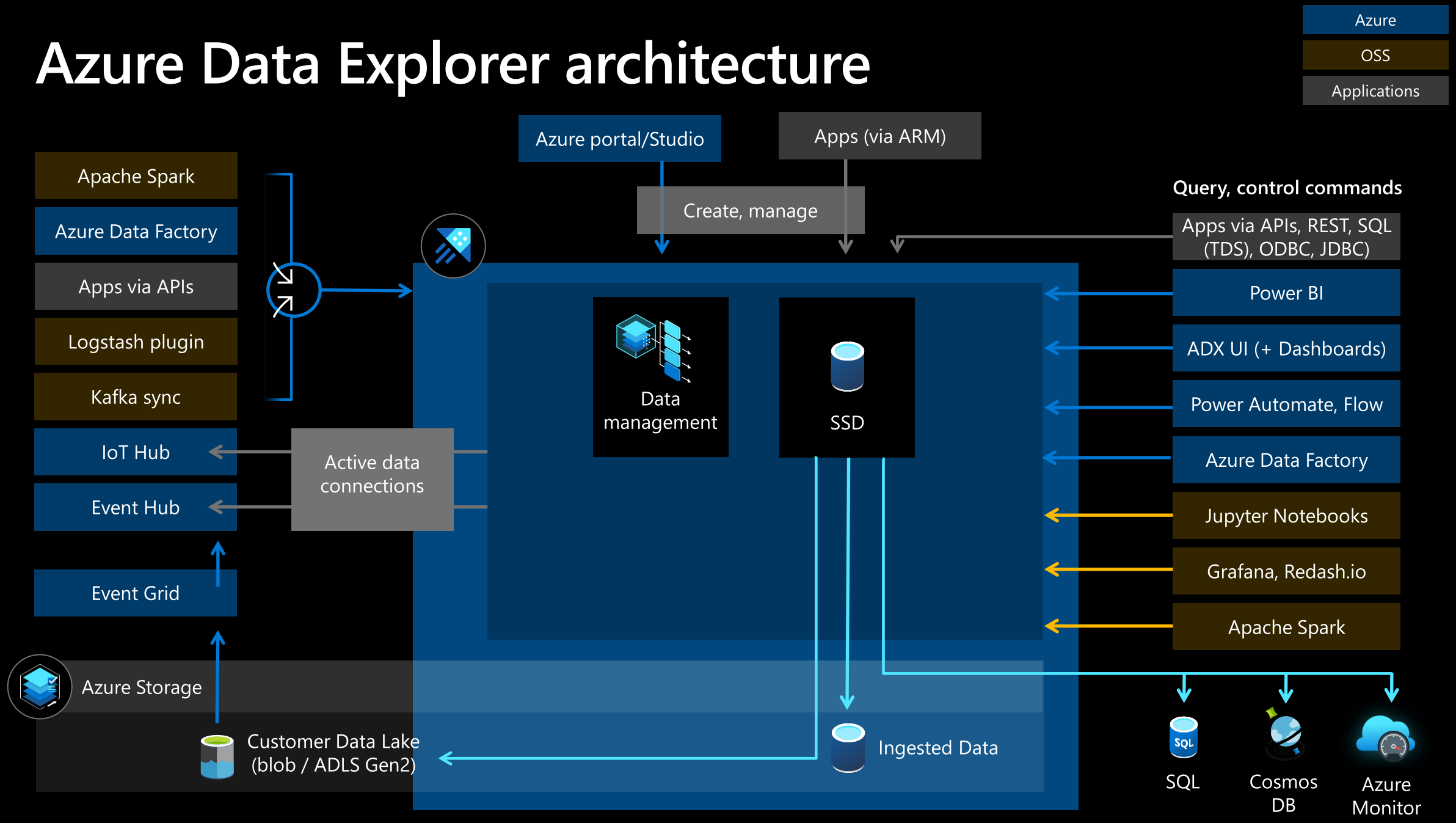
- **Tables**: Made up of a set of columns. Each column has one of nine different data types.
- **External tables:** Tables whose underlying storage is in other locations such as Azure Data Lake.

Working with ADX

1. Create Database
2. Ingest Data
3. Query Database
4. Visualize results



Azure Data Explorer architecture



Batching vs streaming ingestion

Batching

- Optimized for high ingestion throughput
- Preferred method and most performant
- Data is batched according to properties
- Set ingestion batching policy on databases or tables
- Default max batching value is 5 minutes, 1000 items or total of 1 GB
- 4 GB data size limit for a batch ingestion command

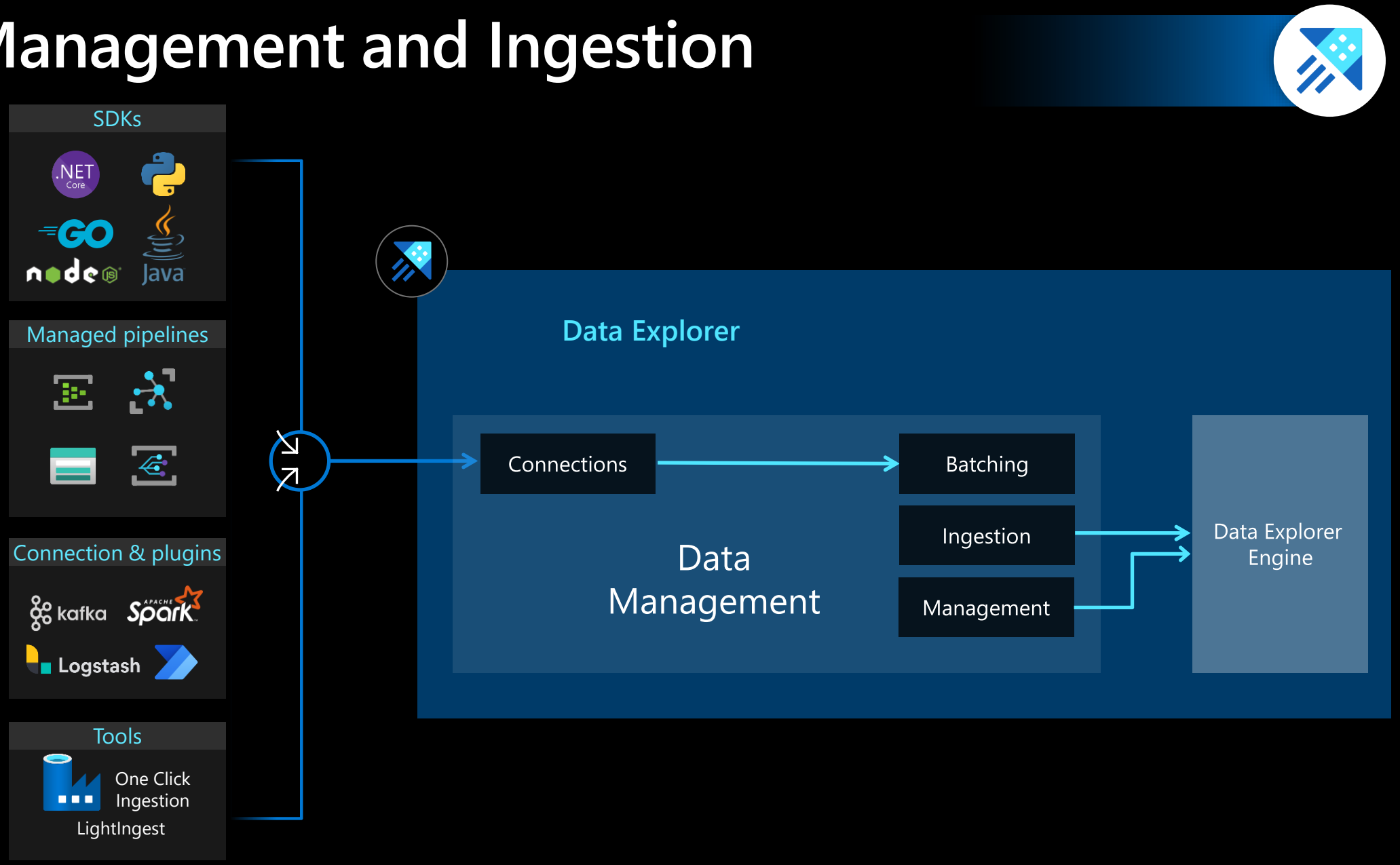
Streaming

- Ongoing data ingestion from a streaming source
- Near real-time latency for small sets of data per table
- Initially ingested to row store
- Then moved to column store extents
- Steaming can be done using ADX client library or supported pipelines

NOTE: The recommendation is to ingest files between 100 MB and 1 GB.

See more: comparing-ingestion-methods-and-tools

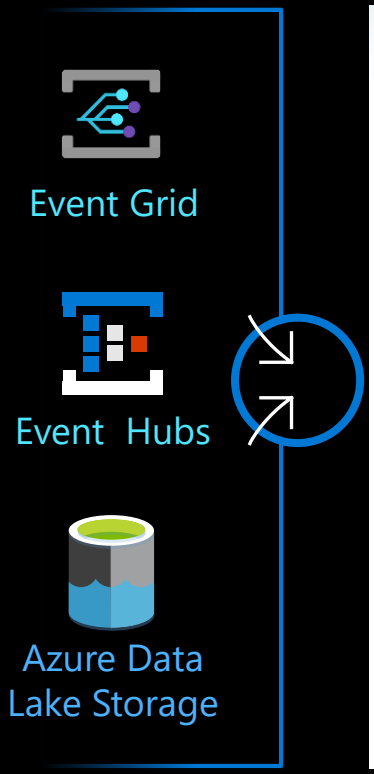
Data Management and Ingestion



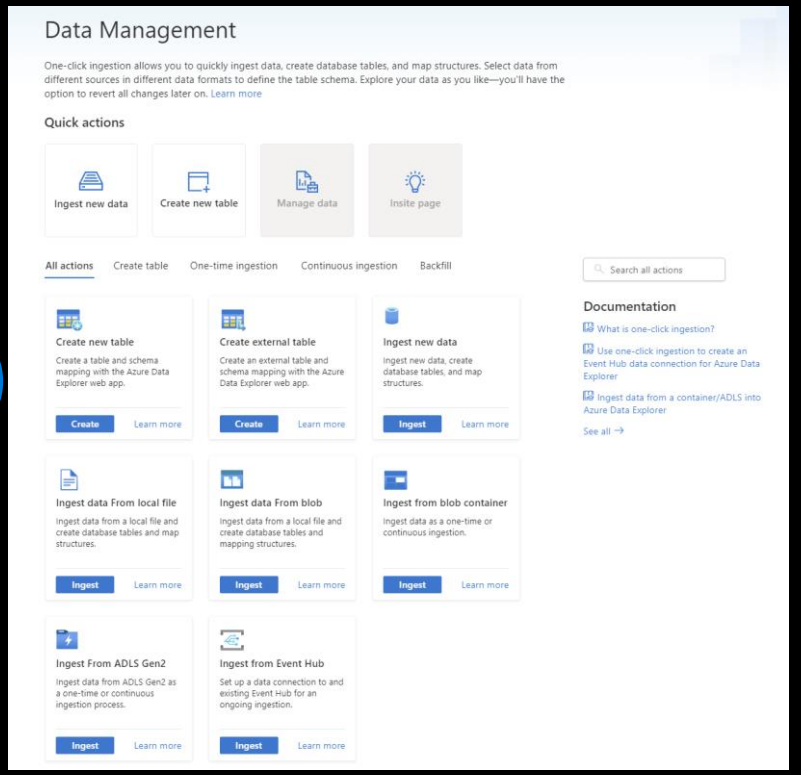
ADX Web UI One-Click (wizard)

Friction free on-boarding of your Azure data sources to Data Explorer

1. Get data from Azure Data Lake, Blob, Event Hub, or local file
2. Infer the schema automatically from source
3. Create table and mapping automatically from source
4. Generate custom code to start ADX project with one of the supported SDKs
5. Manage data-policies (e.g. retention, batching policies, streaming)
6. Get insight on your data management
7. Ingest one-time data or create connection for continuous data ingestion
   - Continuous (EventHub/EventGrid), one time or backfill (LightIngest)



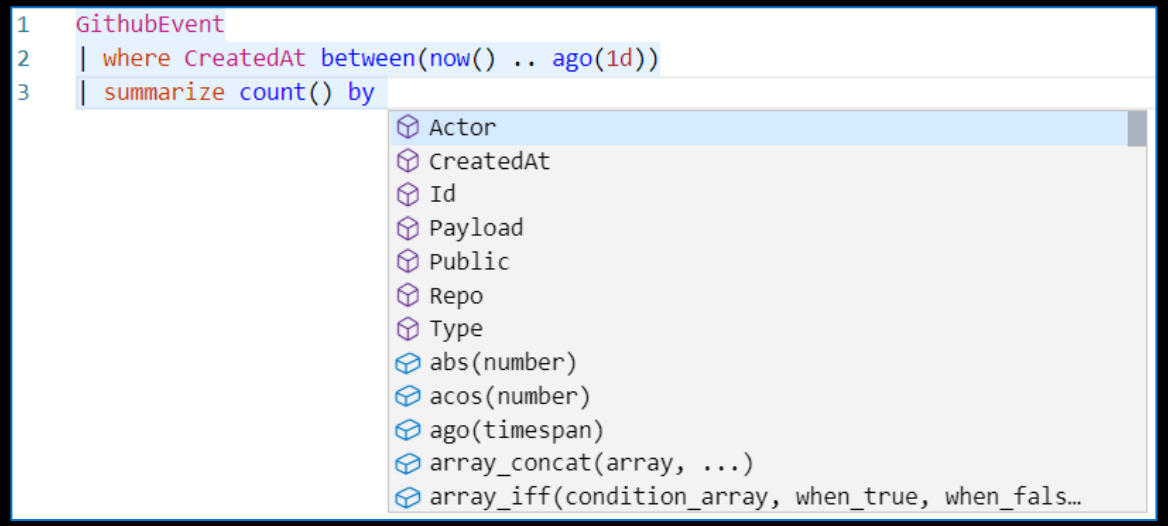
Data Explorer



Intuitive querying

Simple and powerful

- Rich rational query language (filter, aggregate, join, calculated columns, and more)
- Built-in full-text search, time series, user analytics, geospatial, and machine learning operators
- Out-of-the box visualization
- Easy-to-use syntax + Microsoft IntelliSense
- Highly recognizable hierarchical schema entities



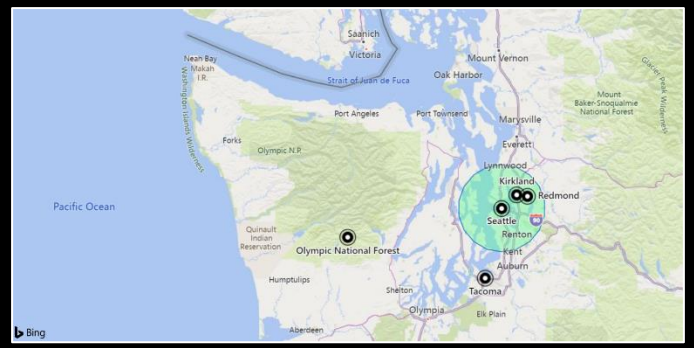
Extensible

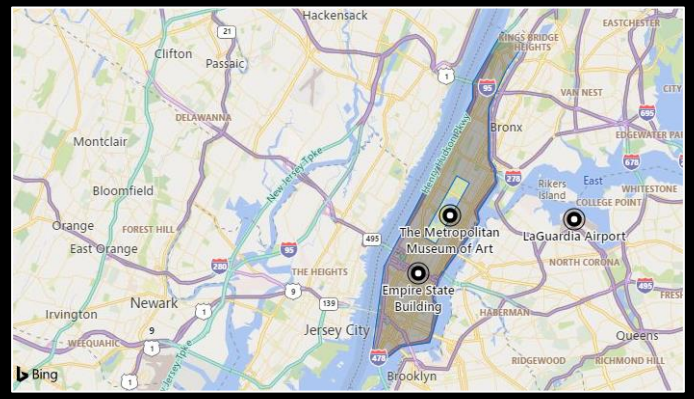
- In-line Python and R
- T-SQL

Geospatial query

1. **Geohash support**
   - Transformation from coordinates to geohashes and back
   - Use-case: Summarization by geographical buckets, store locations based on a single column
2. **Distance**
   - Calculate the distance between two points
3. **Contains**
   - Check whether a point is in a given circle
   - Next: support for lines & polygons (i.e. check whether a point/line/polygon is in a polygon)
4. **Next: Intersection**
   - Support for lines & polygons

Geospatial coordinates are interpreted as represented per the WGS-84 reference system.





Advanced Analytics – Built-in & Extendible

Out of the box

- Auto Clustering for Diagnosis and RCA
- Anomaly Detection
- Regression
- Forecasting
- Time Series Analysis library

Distributed Custom Code Execution

- Distributed Python and R execution
- Custom code is embedded in the KQL query

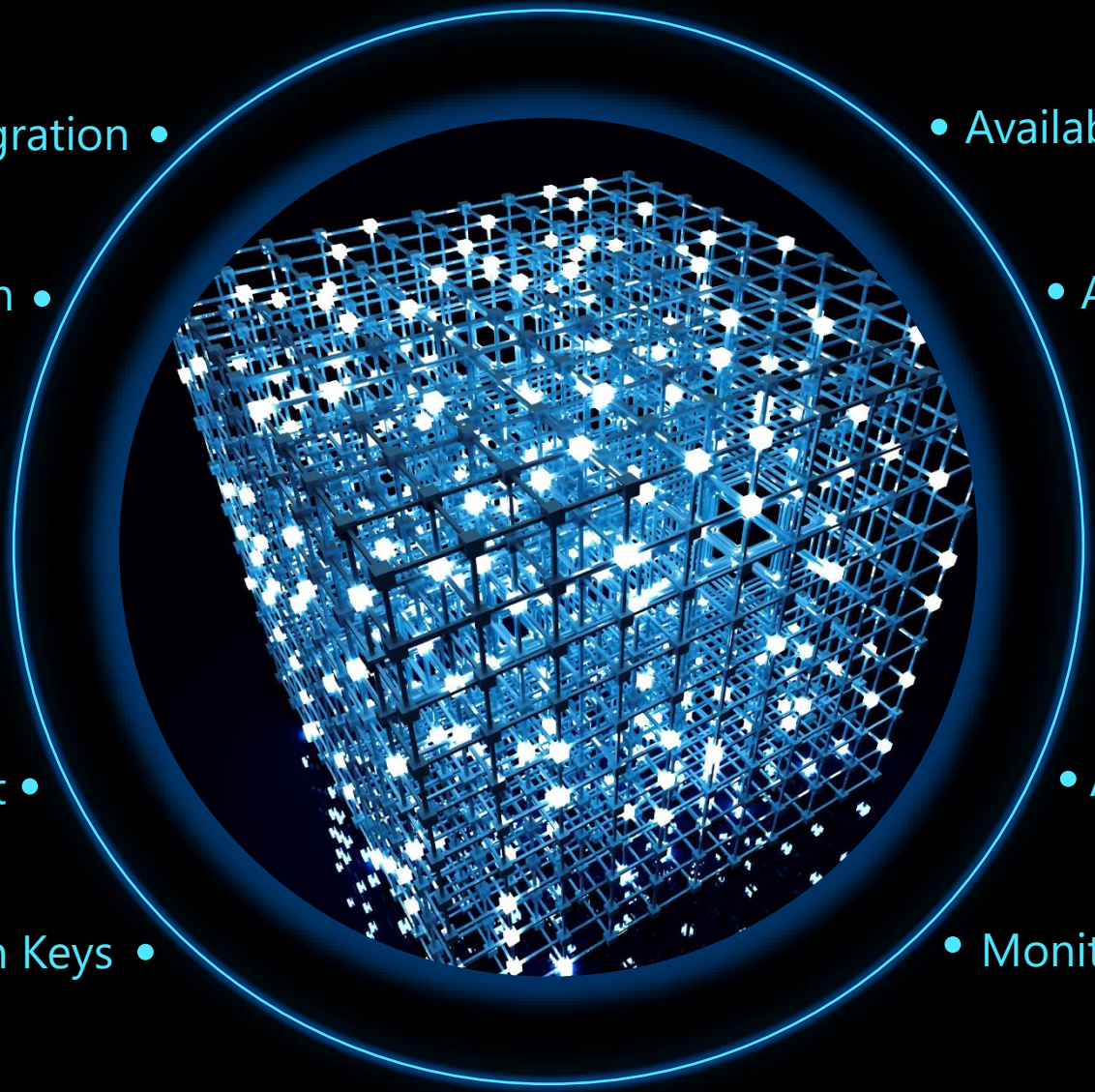
Spark Integration

- Native Spark connector for heavy duty model training
- Operationalize model into ADX for scoring

Tools

- Jupyter Integration with KQL Magic
- Python, Java SDKs

Enterprise Ready – Mission Critical



- Azure Active Directory Integration
- Role based authorization
- Network Perimeter Security
- Encryption with customer managed keys/"BYOK"
- Build Azure Policy Support
- Bring Your Own Keys
- Availability Zones
- Auto Scale-Up/In
- Globally available
- CI/CD Integration
- Automated provisioning
- Monitoring

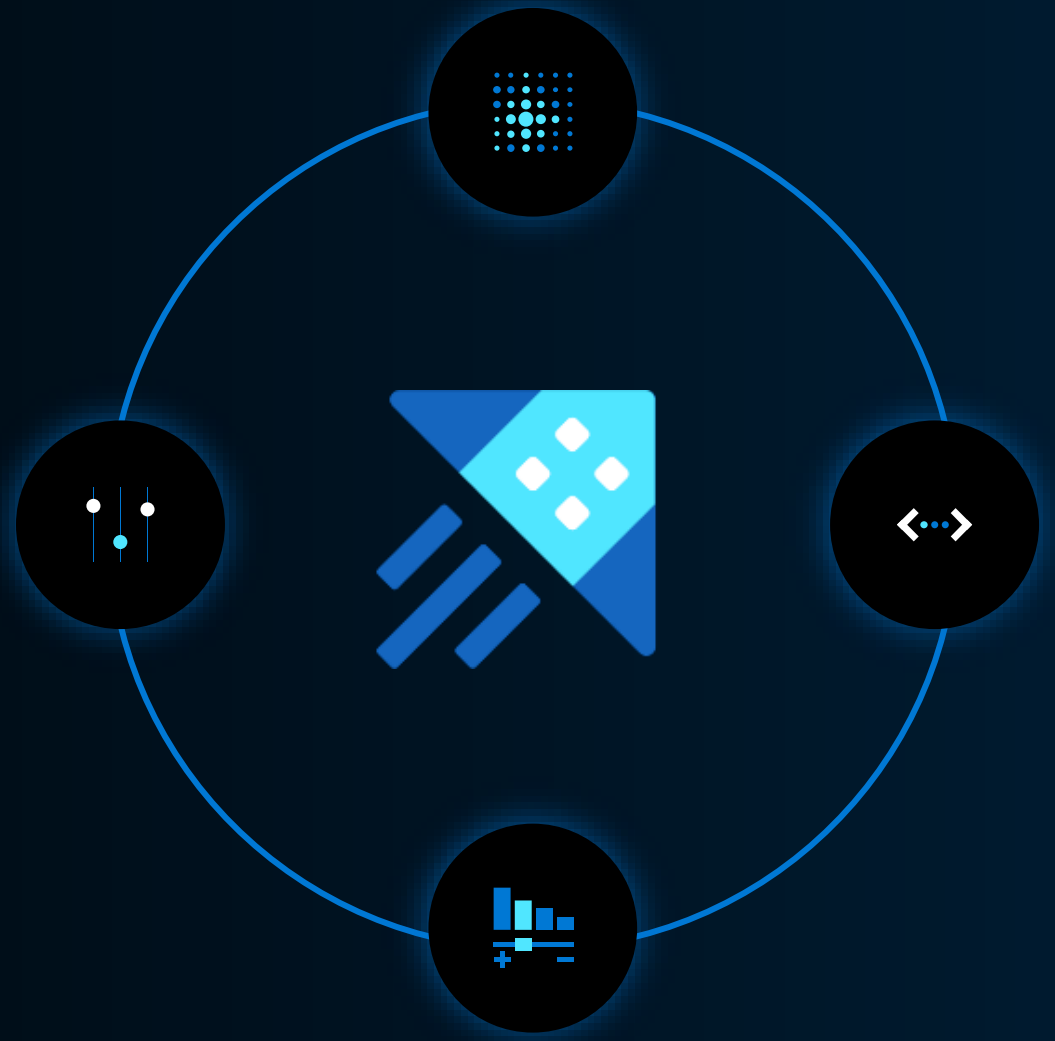
Demo 2*

- Thermostat data
- JSON
- Render
- Fill gaps
- Forecast
- Anomalies
- Dashboards
- Materialized Views
- External Tables

* Estimate 20 minutes

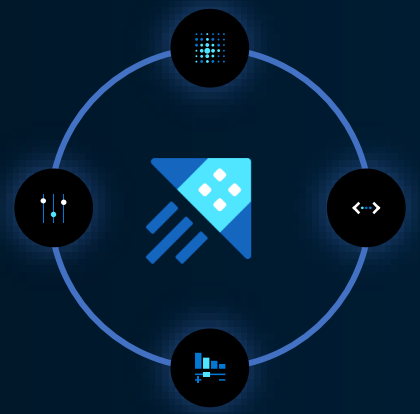
Migration

Lesson 3



Objectives

After completing this Learning, you will be able to understand:



1. Walkthrough key steps
2. Explain automation options
3. Importing historical

TSI Gen1 – Migration to ADX Key Steps

1. Create ADX Cluster
2. Set up parallel ingestion from hubs to ADX Cluster
3. Continue ingesting data for the period of fixed retention
4. Start using ADX Cluster
5. Delete TSI environment

Considerations

- If telemetry data is required to be exported, use TSI Query API to download the events in batches and serialize in required format.
- TSI Explorer or Reference Data API can be used to download reference data set and upload it into ADX as another table.
- Then create materialized views in ADX can be used to join reference data with telemetry data.

Post-migration

1. Translate Time Series Insights Queries to KQL

TSI Gen2 – Migration to ADX Key Steps

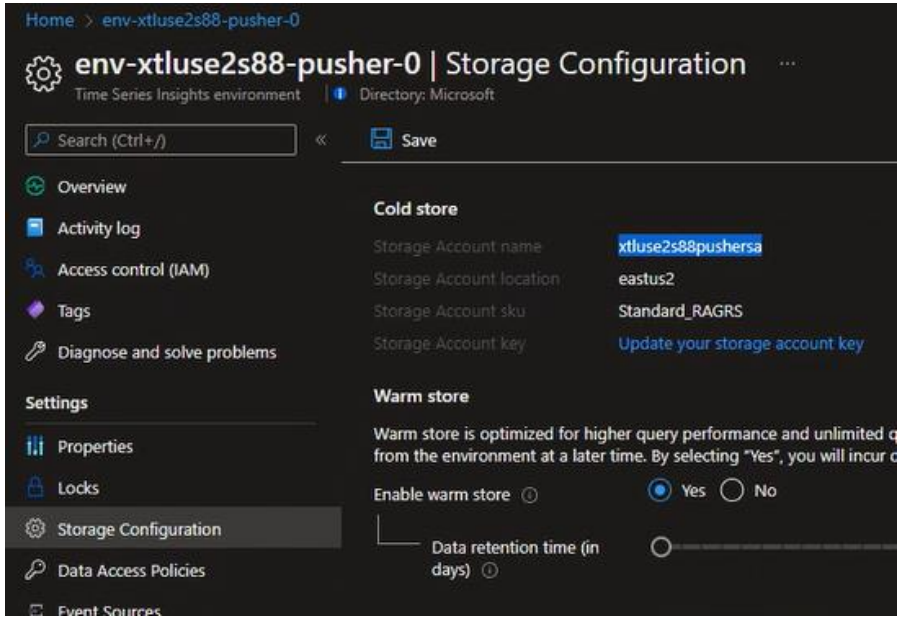
1. Get migration telemetry size, path and ingestion behavior
2. Migrate telemetry to ADX
3. Redirect TSI Live Data

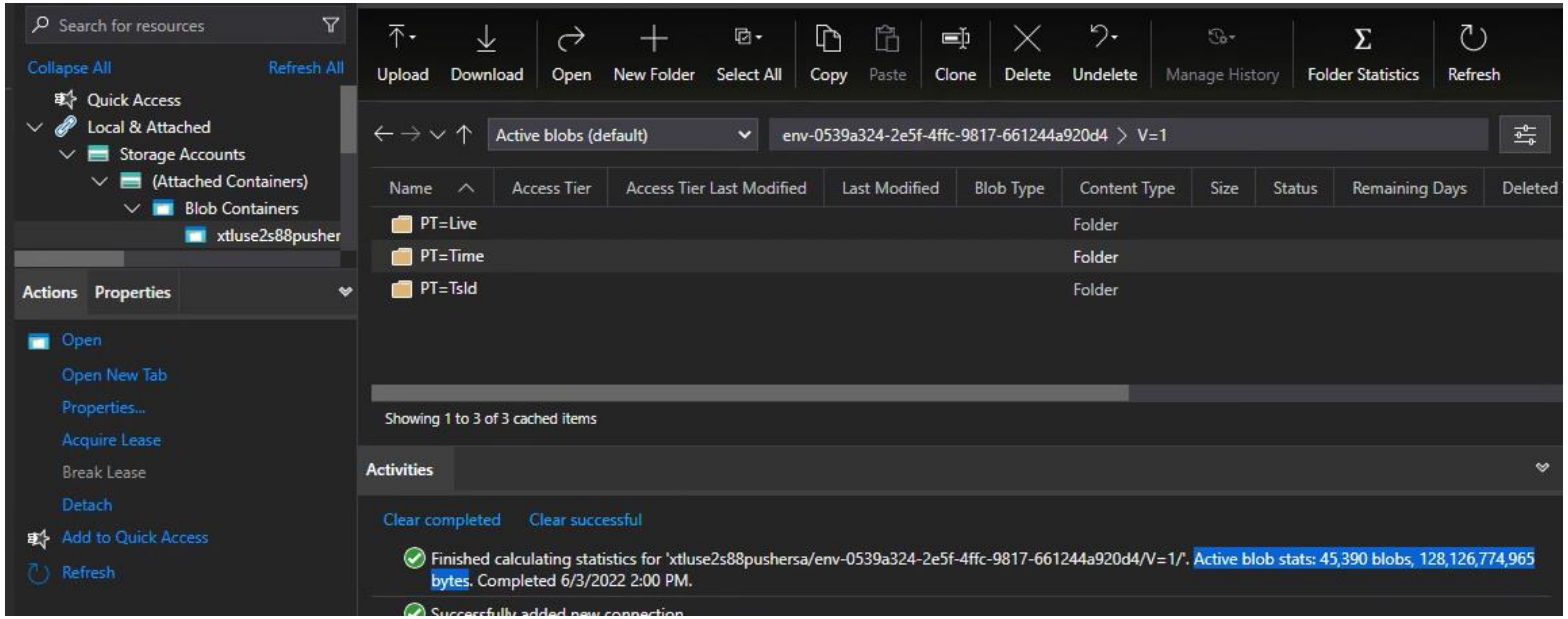
Post-migration

1. Migrating Time Series Model (TSM) to ADX
2. Translate Time Series Queries (TSQ) to KQL
3. OPTIONAL: Migration from TSI Power BI Connector to ADX Power BI Connector

Step 1 - Migrating Telemetry Prerequisites

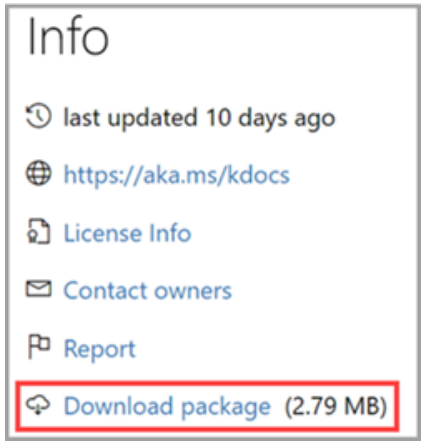
1. Sum of data size and volume to migrate
2. Where data is stored
3. If there is a lot of late arriving data in comparison to ingestion point
4. If TSI is Gen 2, then use `PT=Time` folder to retrieve historical data.
   - Azure portal > TSI environment > Settings > Storage Configuration.
   - Inside you will find a folder per environment, year & month.

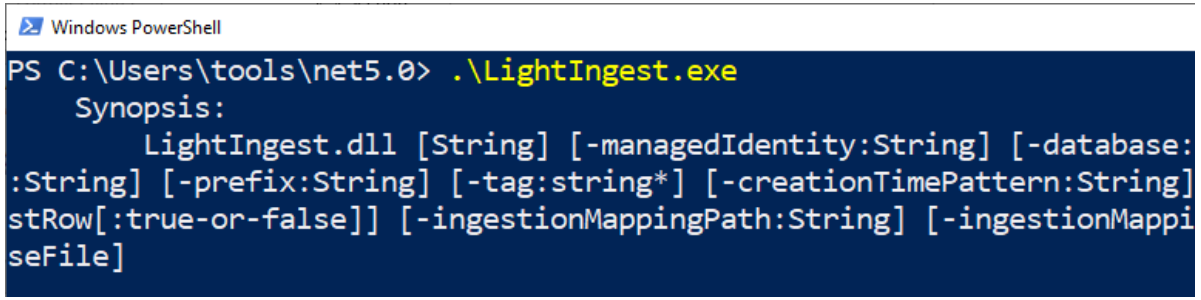




Con’t

5. Download LightIngest





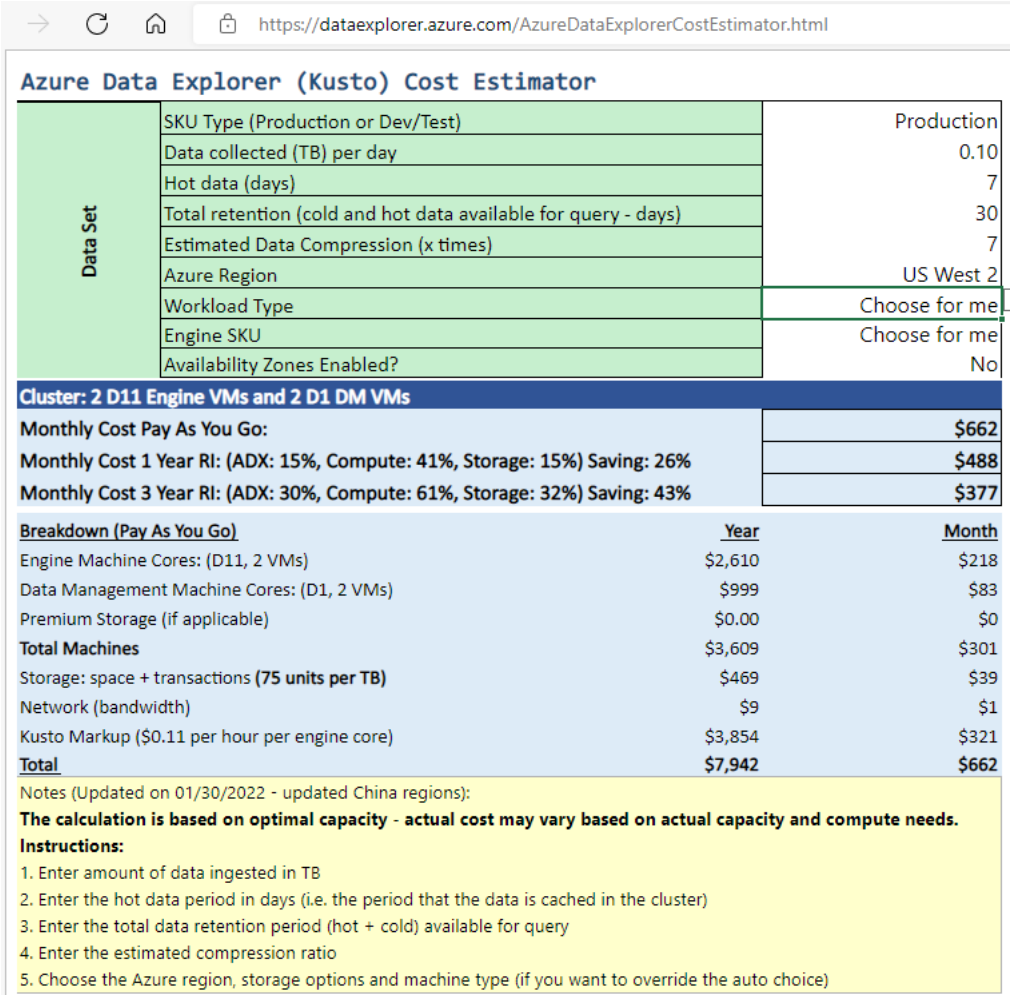
6. Review if there is significant late arriving data. (very unlikely)
7. Review blob files in `PT=Time` subfolder of the TSI storage account(s).
8. Compare when blob is created to min event using the filename format:
   `<**BlobCreation**Timestamp>_<**MinEvent**Timestamp>_<**MaxEvent**Timestamp>_*.parquet`

9. If BlobCreation is far from MinEvent for a lot of files, then there’s a lot of late arriving data.
10. If true, then use ADX data partitioning policy before we ingest the data.

Step 2 - Prep ADX Cluster

Create the ADX Cluster

1. Size your cluster for your data
2. Estimate cost
3. Consider growth
4. Consider more nodes during migration
5. Enable Diagnostics
6. Set data partitioning (if needed)
    1. If non-default ADX partitioning is needed, then don't use -CreationTimePattern flag in LightIngest.exe tool.



https://dataexplorer.azure.com/AzureDataExplorerCostEstimator.html

Azure Data Explorer (Kusto) Cost Estimator

Data Set	SKU Type (Production or Dev/Test)	Production
	Data collected (TB) per day	0.10
	Hot data (days)	7
	Total retention (cold and hot data available for query - days)	30
	Estimated Data Compression (x times)	7
	Azure Region	US West 2
	Workload Type	Choose for me
	Engine SKU	Choose for me
	Availability Zones Enabled?	No

Cluster: 2 D11 Engine VMs and 2 D1 DM VMs	
Monthly Cost Pay As You Go:	**$662**
Monthly Cost 1 Year RI: (ADX: 15%, Compute: 41%, Storage: 15%) Saving: 26%	**$488**
Monthly Cost 3 Year RI: (ADX: 30%, Compute: 61%, Storage: 32%) Saving: 43%	**$377**

Breakdown (Pay As You Go)	**Year**	**Month**
Engine Machine Cores: (D11, 2 VMs)	$2,610	$218
Data Management Machine Cores: (D1, 2 VMs)	$999	$83
Premium Storage (if applicable)	$0.00	$0
Total Machines	$3,609	$301
Storage: space + transactions **(75 units per TB)**	$469	$39
Network (bandwidth)	$9	$1
Kusto Markup ($0.11 per hour per engine core)	$3,854	$321
Total	**$7,942**	**$662**

Notes (Updated on 01/30/2022 - updated China regions):

The calculation is based on optimal capacity - actual cost may vary based on actual capacity and compute needs.

Instructions:

1. Enter amount of data ingested in TB
2. Enter the hot data period in days (i.e. the period that the data is cached in the cluster)
3. Enter the total data retention period (hot + cold) available for query
4. Enter the estimated compression ratio
5. Choose the Azure region, storage options and machine type (if you want to override the auto choice)

Step 3 - Migrate Telemetry to ADX

1. Prep for Data Ingestion
   - Go to https://dataexplorer.azure.com.
   - Copy the LightIngest command and store it somewhere so you can use it in the next step.

2. Execute Data Ingestion.
   - **Option 1:** Ingest All Data. For **smaller** environments.
   - **Option 2:** Ingest Data by Year or Month. For **larger** environments.

3. Monitor Ingestion.
   - IMPORTANT: use the -dontWait flag in LightIngest command.
   - Use the **Insights** tab via Azure Portal for the ADX cluster to monitor ingestion progress.
   - Ingestion is complete once metrics go to 0 for source table.

Step 4 - Redirect TSI Live Data

Data will be flowing into TSI via EventHub or IoTHub.

1. Ingest live data by creating a data connection (managed pipeline) to the ADX database table
   - Create another consumer group on EventHub and IoTHub
   - Setting up the ADX Data Connection using the new consumer group.

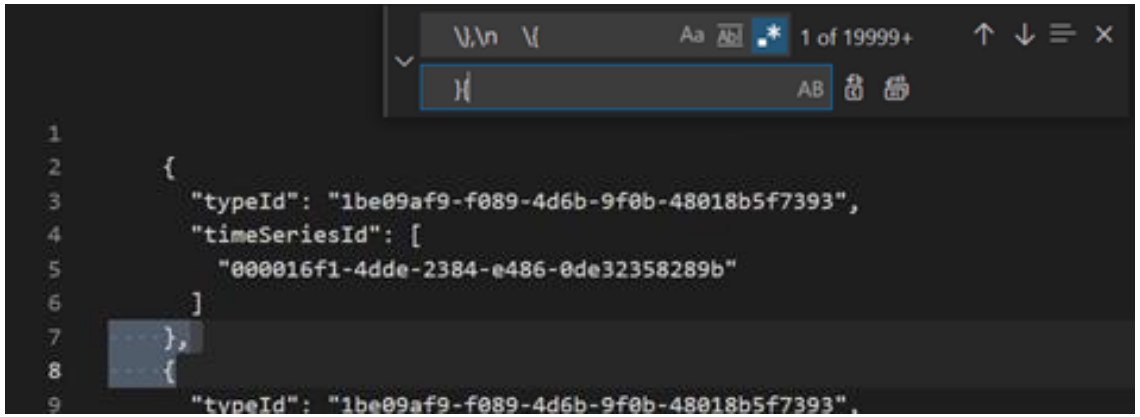
- **Option 1: Historical** data & **Hot** data in **same table**.
  - **Benefit**: Single table simplifies the process and queries.
  - **Considerations**:
    - Almost certainly leads to duplicate data, then handle duplicate data in ADX.
    - Less flexible than option 2.

- **Option 2: Historical** data & **Hot** data in **2 separate tables**.
  1. Use one-click ingestion to ingest data from EventHub into Azure Data Explorer.
  2. Create a KQL function to unify the data using union.

Step 5 - Migrating TSI Model (TSM) to ADX

1. Download models in JSON format via TSI Explorer UX or TSM Batch API.

2. Edit it using VSCode or another editor.
   - Search and replace as regex `\},\n \{` with `}{`.



3. Ingest JSON file to separate ADX table via ADX One-Click UI.

Demo 3*

- Setup ADX
- Setup table + mapping
- Setup LightIngest
- Setup ADF
- Setup ADF pipelines, etc

* Estimate 20 minutes

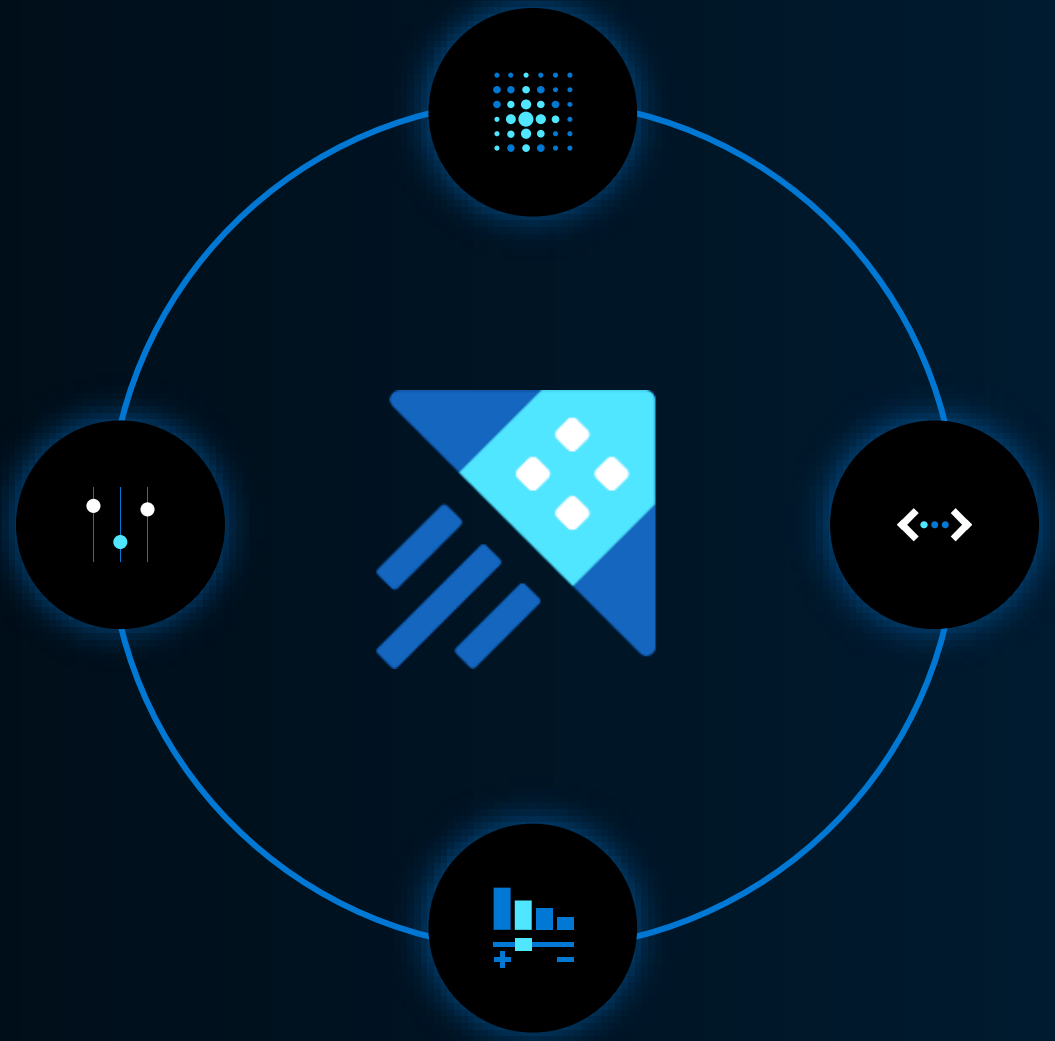
Demo 4*

- Execute LightIngest
- .show commands to monitor
- ADX Insights (Ingestion)
- Trigger ADF pipelines, etc
- Monitor ADF pipelines

* Estimate 20 minutes

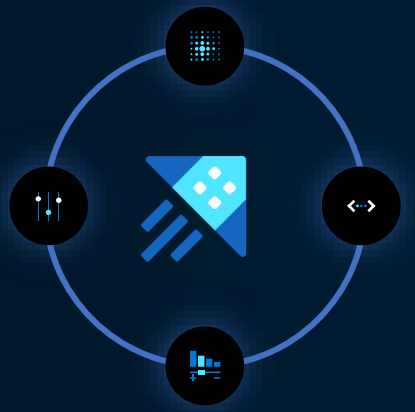
Finalization

Lesson 4



Objectives

After completing this Learning, you will be able to understand:



1. Finalize changes that you have made from TSI to ADX
2. Resource cleanup
3. Final testing of ADX to ensure that all your data is migrated

Beyond data migration and visual recreation

- If you were using TSI API, you will need to change your application to use ADX API
- ADX provides a private endpoint and disables public access feature so you will need to adjust consumer applications accordingly or use gateways to ensure that the consumers can query data from ADX
- It is advisable to keep the TSI environment **for 1 to 2 weeks** after all the data has been migrated to ensure that you can still retrieve any missing data

Resource cleanup

Delete TSI environment after you have tested and used ADX sufficiently and are confident that most of your use cases are being fulfilled by ADX

ADF Pipelines

- If you used ADF pipelines to migrate data from TSI to ADX, you can cleanup the pipelines after the TSI environment has been deleted

LightIngest

- If you used a separate VM to run LightIngest, you can cleanup this VM

PBI to TSI

- If you created a new PBI report to connect to ADX, then you can archive the old PBI report that connected to TSI

ADX usage monitoring and optimization

- Use **Insights blade** in ADX to monitor the usage pattern of ADX
- Use the **Advisor** recommendations and cluster boundedness recommendations to adjust the SKU, size and node count of the ADX cluster to ensure that your costs stay in control
- Implement stop-start script to control ADX costs for usage outside business hours
  - This can have implication on how data is ingested into ADX. So, discuss this with your CSA/CE before deciding to implement this.

Post Data Migration

- Verify all the data was migrated
- Check data in Log Analytics for Successful Ingestion and verify it matches with data on blob for historical data. For example, this will provide the number of blobs ingested:

```
SucceededIngestion
| where Table == 'TSITelemetry'
| summarize dcount(IngestionSourcePath)
```

- Run some basic KQL queries to explore the data in ADX
- Start building visualization based on decision on which option to utilize
- Configure RBAC on the ADX database to provide needed permissions

Demo 5* - Visualization Options

- Power BI: Connected Devices - Power BI
- ADX Dashboard: ADT Integration Demo (azure.com)
- Grafana: Patient Monitoring with ADT - Grafana (azgrafana.io)
- TSI JavaScript Controls: Time Series Insights JavaScript SDK Examples (tsiclientsample.azurewebsites.net)
- Seeq: Microsoft Azure Data Explorer (ADX) Connector - Seeq Knowledge Base - Confluence (atlassian.net)

* Estimate 20 minutes

Summary

Lesson 1: Why TSI to ADX migration

Lesson 2: Understanding ADX

Lesson 3: Migrating from TSI to ADX

Lesson 4: Finalizing the migration

Resources

- Migration docs: aka.ms/tsi2adx
- Keep up with news: aka.ms/adx.blog
- Learn more: aka.ms/adx.iot
- Free online courses: Azure Data Explorer (pluralsight.com)
- Technical Whitepaper: azure.microsoft.com/resources/azure-data-explorer
- IoT Reference Architecture: docs.microsoft.com/azure/architecture/solution-ideas/articles/iot-azure-data-explorer

Microsoft

Thank You

aka.ms/adx.tsi.eval