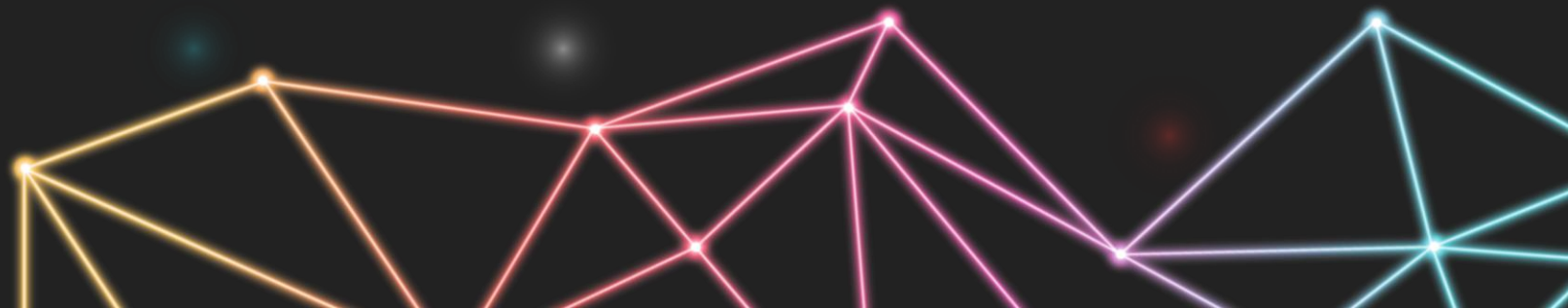
OVERNET.

2024 WPC

Time Series with Azure Data Explorer

Gianluca Sartori
Founder
Quantumdatis



Gianluca Sartori

He/Him

Founder

QUANTUMDATIS

- Independent SQL Server consultant
- Data platform MVP
- Works with SQL Server since 7.0
- DBA @ Scuderia Ferrari

@spaghettidba.bsky.social

spaghettidba.com

quantumdatis.com

Agenda

1. What is Time Series data?
2. Options on the market
3. Azure Data Explorer
4. Storing Time Series data
5. Retention Policies & Downsampling
6. Querying Time Series data with KQL

What is Time Series data?

Definitions

- What is **time series data**?

> *"Time series data (or time-stamped data) is a collection of observations for a single object (entity) at different time intervals."*

- What is a **time series database**?

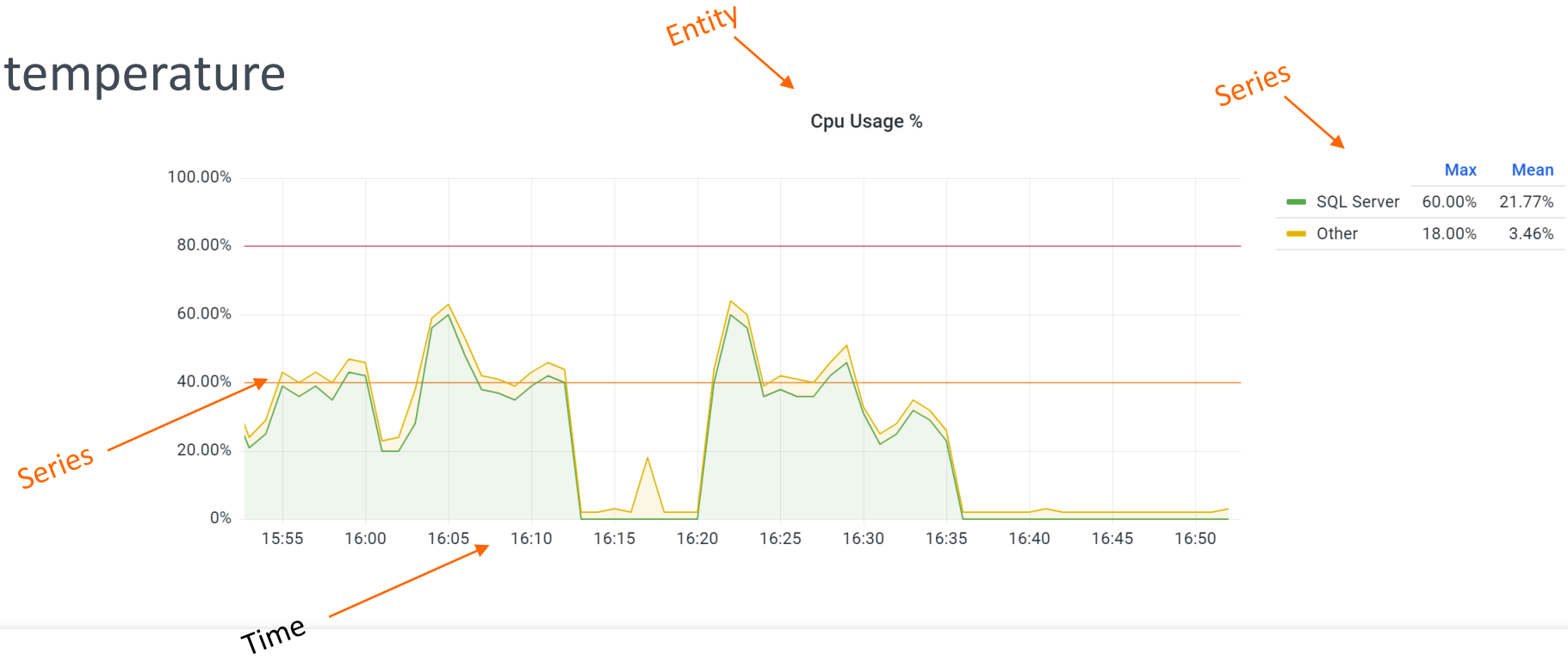
> *"A time series database (TSDB) is a database optimized for **time series data** and for measuring change over time."*

Time Series Data

Time series data is obtained by performing repeated measurements over time

Examples:

- Atmospheric temperature
- Stock prices
- CPU usage %
- Emails/sec
- Sensor data



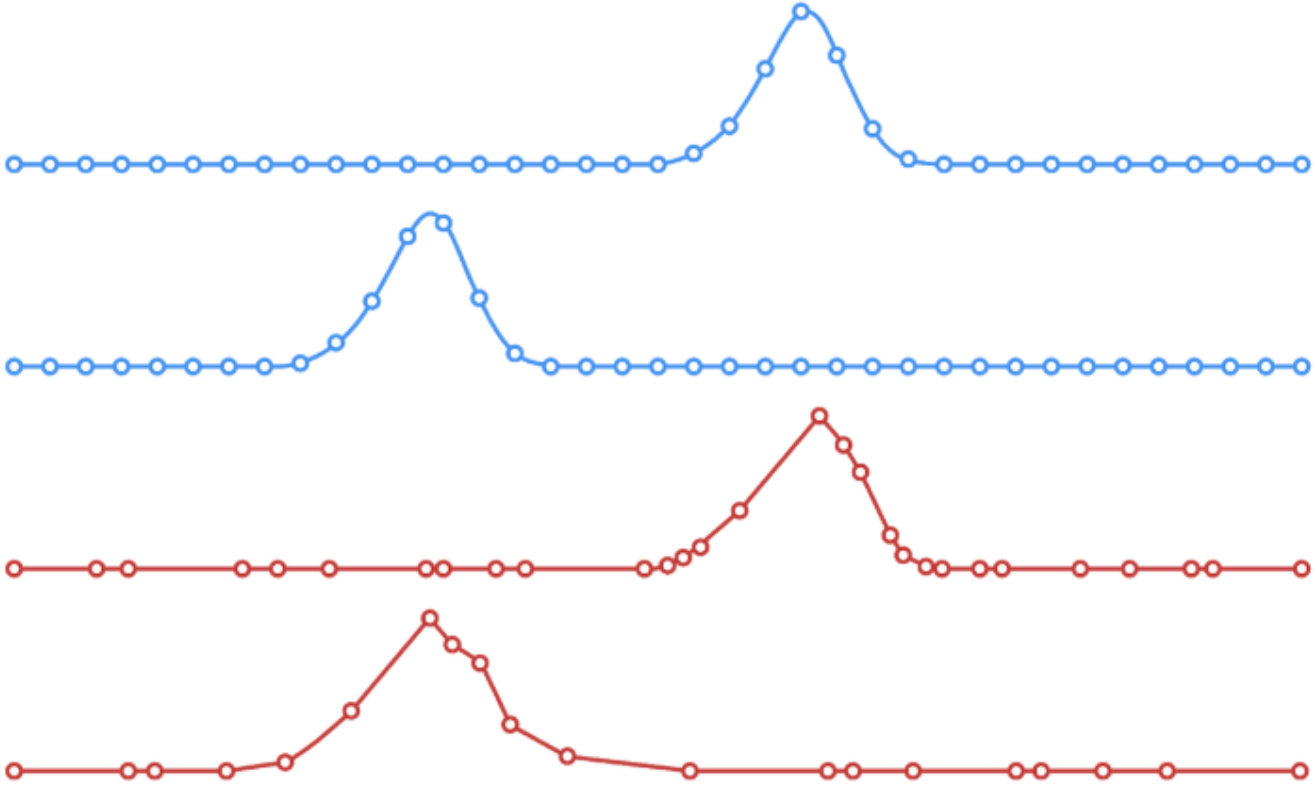
Time Series Data

Metrics (Regular)

Measurements gathered at regular time intervals

Events (Irregular)

Measurements gathered at irregular time intervals



Time Series Data

How is this different from relational data?

- Continuous Stream of data in time order
- Bulk uploads of large sets of data
- High volume data
- Data is append-only – no updates
- Delete large volumes of data when it goes out of scope
- Downsampling and aggregating high resolution data to save space

Why does it matter?

Aircraft Engine

10 terabytes every 30 min.
14 million hours of flight in 2018
2 engines
= 560 million TB

2020s

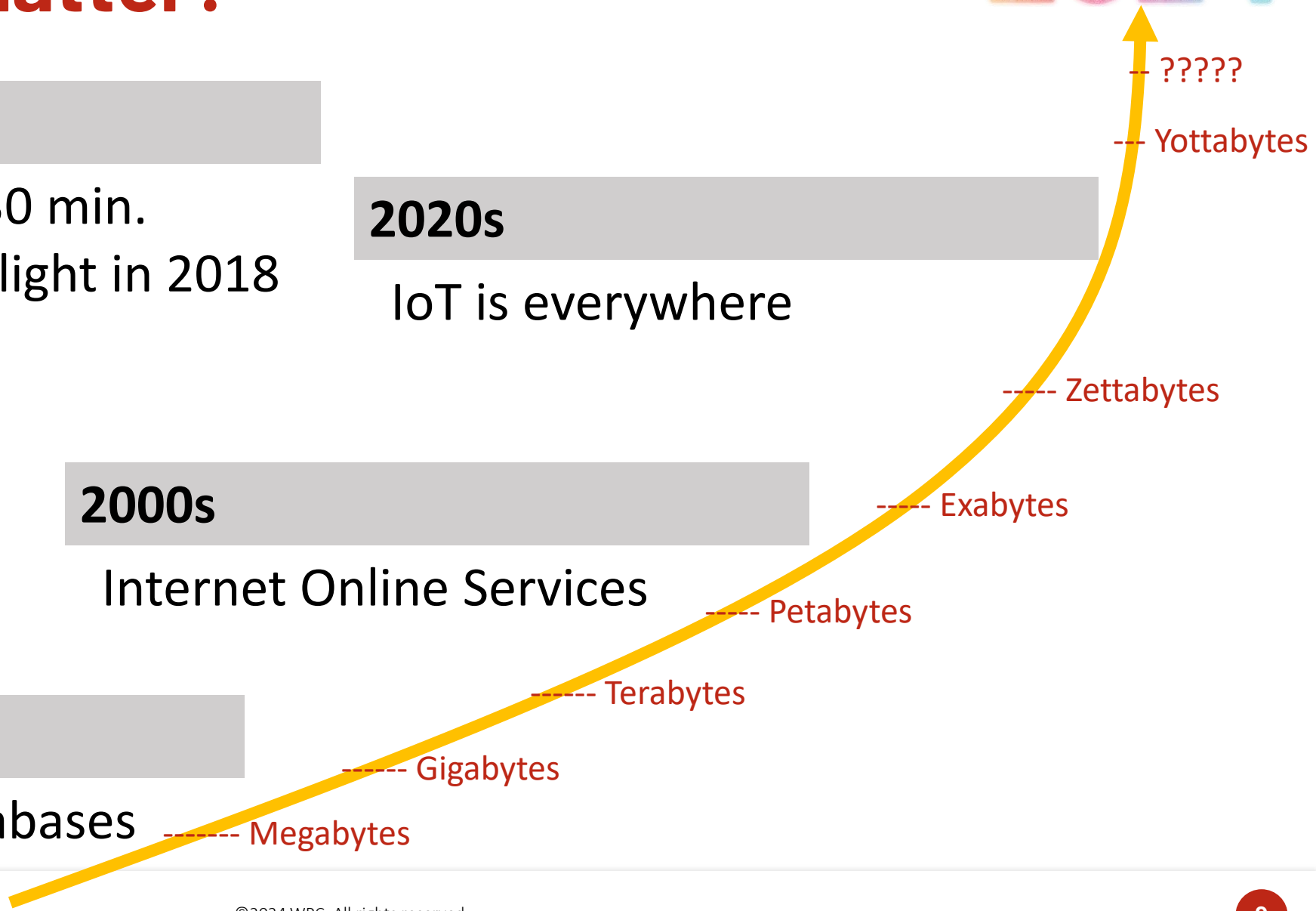
IoT is everywhere

2000s

Internet Online Services

1970s

Early Relational Databases



Time Series Database Options

RANK	DBMS	SCORE	
NOV 2024		NOV 2024	24 MONTH
1	**InfluxDB**	**21.47**	**-8.55**
2	KDB+	7.06	-1.72
3	Prometheus	6.92	+0.61
4	Graphite	4.91	-0.45
5	TimescaleDB	3.68	-0.88
6	QuestDB	2.91	+2.01
7	Apache Druid	2.70	-0.48
8	DolphinDB	2.60	+0.98
9	TDEngine	2.27	+1.08
?????	Azure Data Explorer	????	?????

WHY SO LOW?

Azure Data Explorer

- Fully managed, high performance, big data analytics platform
- Can store and analyze structured, semi-structured and unstructured data
- Uses a «relational» model (tables, columns and rows) with fixed strongly-typed schemas
- Tables are stored in databases
- Clusters contain databases
- Log analytics, IoT, time series data

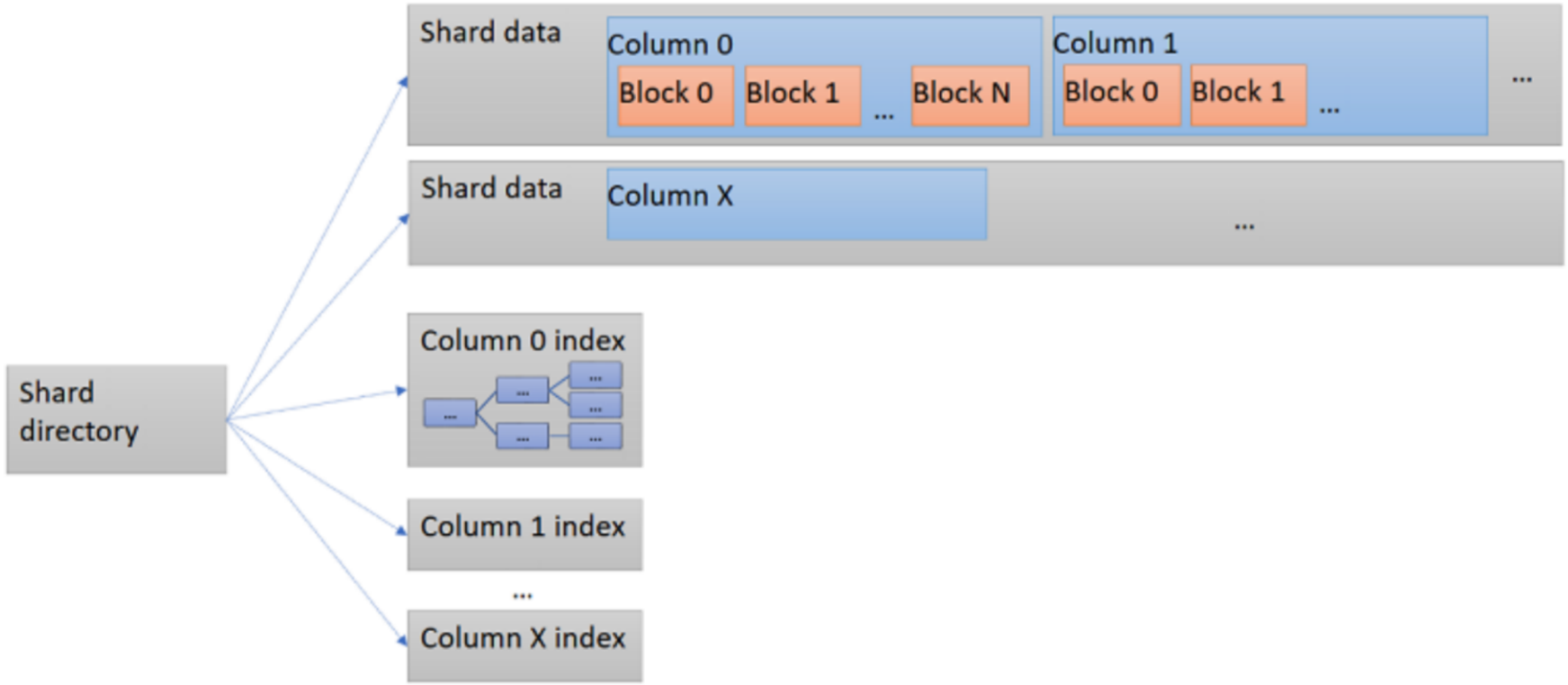
➔ not strictly a time-series database!

Azure Data Explorer Internals

- Separates storage and compute resources
- Persistent data resides in Azure Blob Storage
  - Data is stored in **extents** (shards)
  - Extents are spread across cluster nodes
  - Extents are cached in SSD and memory
  - Data is compressed with columnar compression
- Compute uses a cache for persistent storage
- Row store is used when ingesting data (streaming ingestion)

Shard Format

- Compressed column store with free text support and full-text inverted index



Partitioning Policies

```
{
    "ColumnName": "timestamp",
    "Kind": "UniformRange",
    "Properties": {
        "Reference": "2021-01-01T00:00:00",
        "RangeSize": "7.00:00:00",
        "OverrideCreationTime": false
    }
 }
```

Working with Azure Data Explorer

Working with ADX – Free cluster

- Azure Data Explorer is available **for free**!
- You can create a free cluster with your Microsoft account
- Be careful, it ends in a weird subscription…

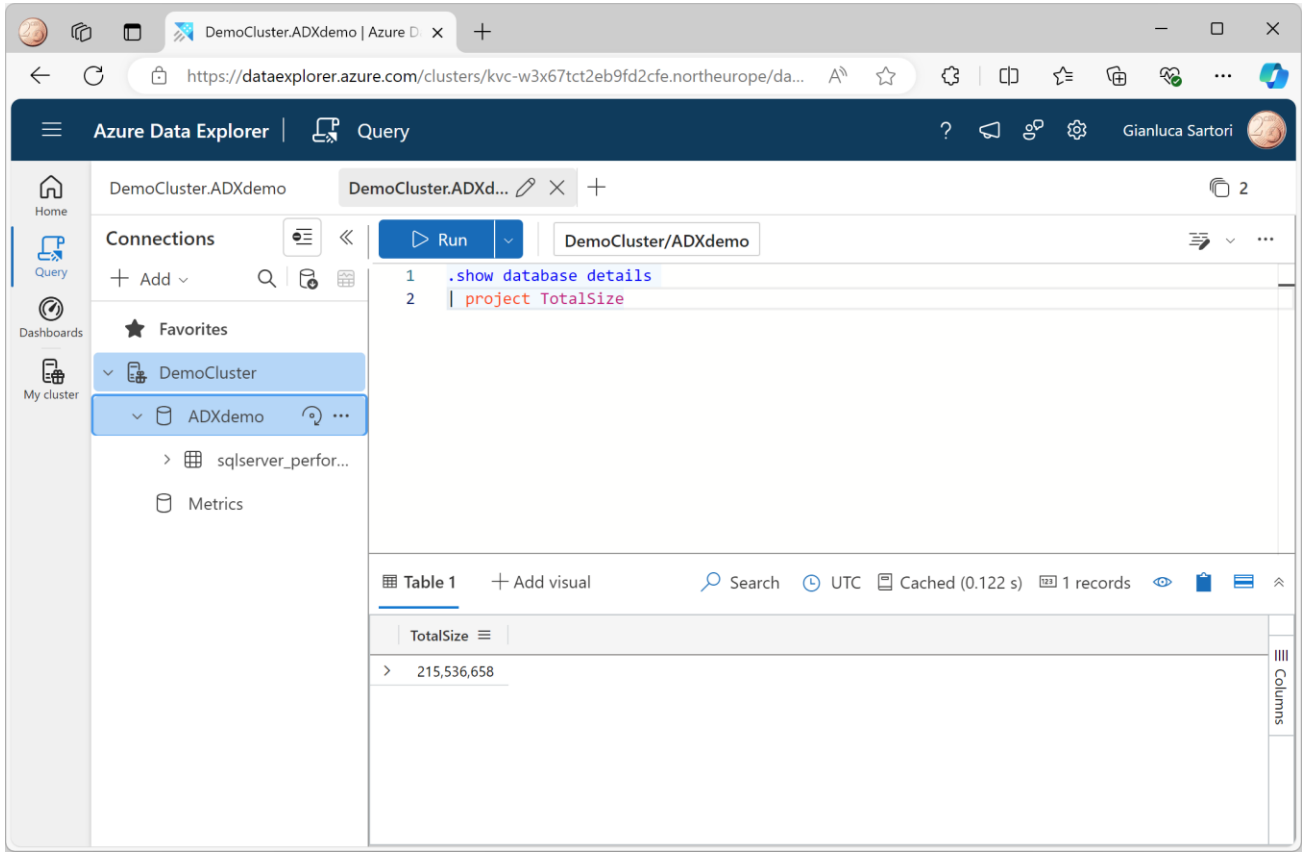
Storage (uncompressed)	100 GB
Databases	Up to 10
Tables per database	Up to 100
Columns per table	Up to 200
Materialized views per database	Up to 5

- Not all features are available. See the docs: https://learn.microsoft.com/en-us/azure/data-explorer/start-for-free
- May or may not renew after 1 year

Working with ADX – Query Tools

https://dataexplorer.azure.com

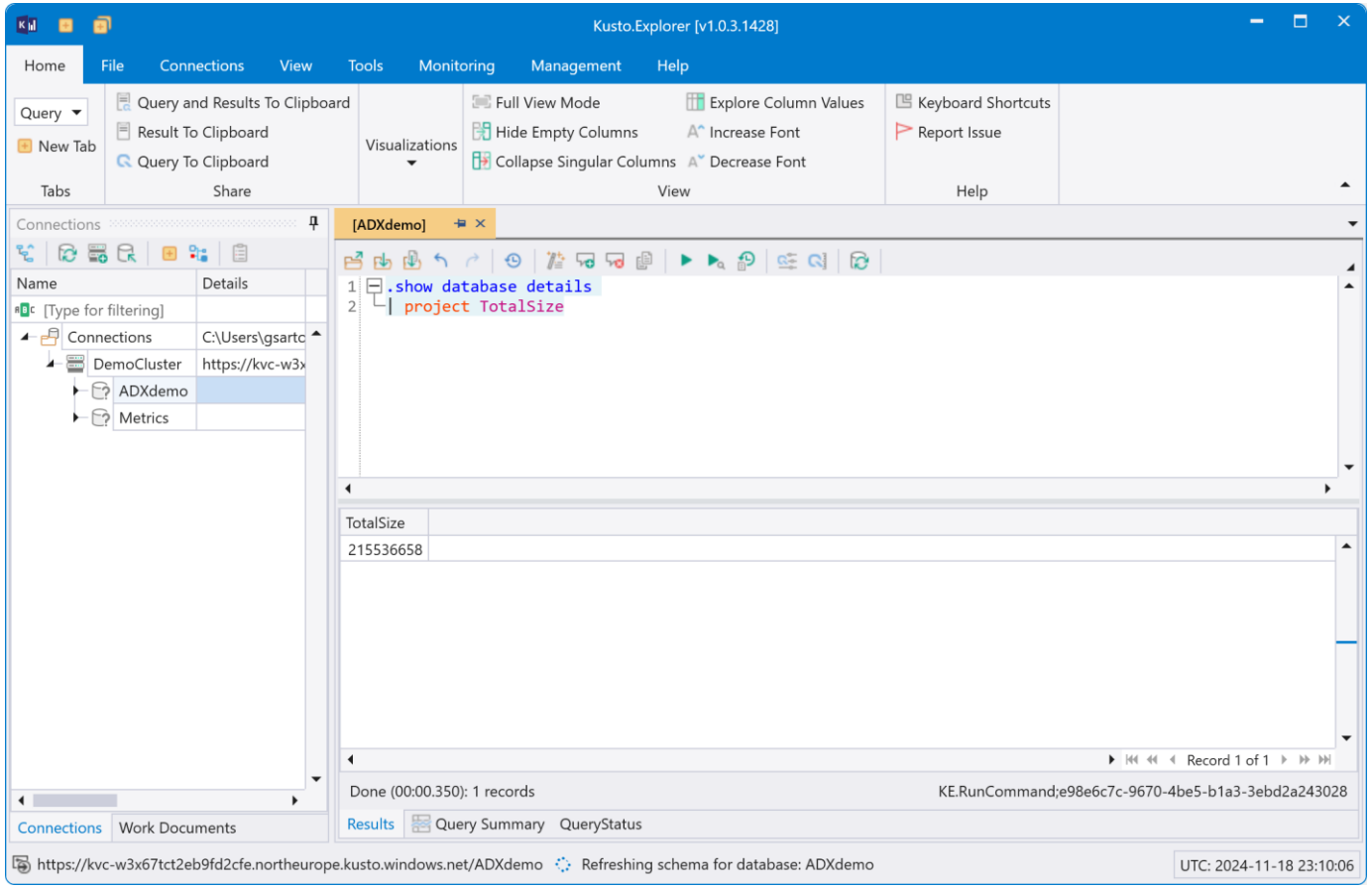
- Native portal for ADX
- Supports query commands in KQL
- Supports administrative commands
- Rich UI for creating clusters, databases, tables, ingesting data etc…



Working with ADX – Query Tools

Kusto Explorer

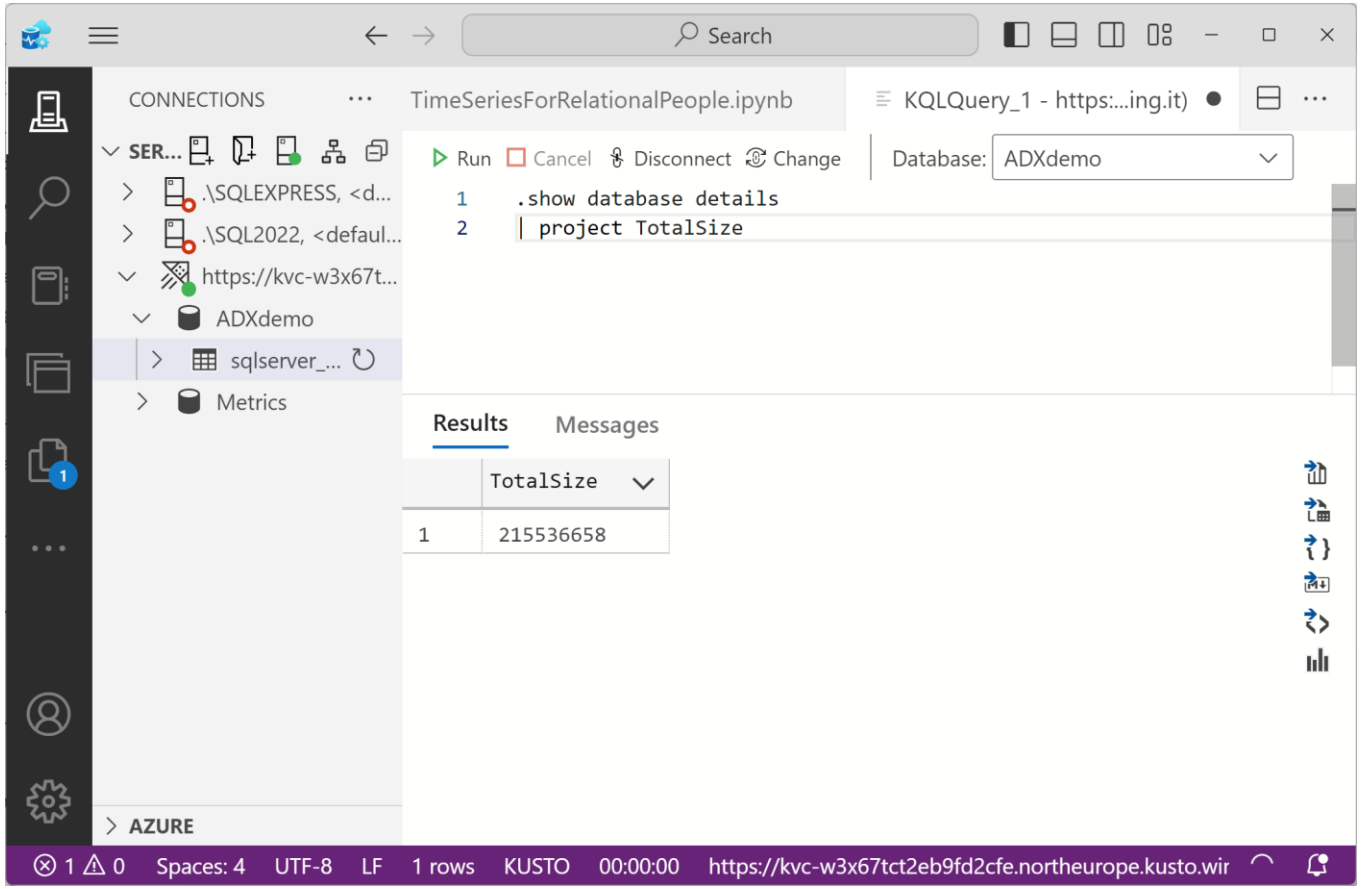
- Windows application for querying Kusto clusters
- Some extra features like copy to clipboard, export CSV and so on



Working with ADX – Query Tools

Azure Data Studio

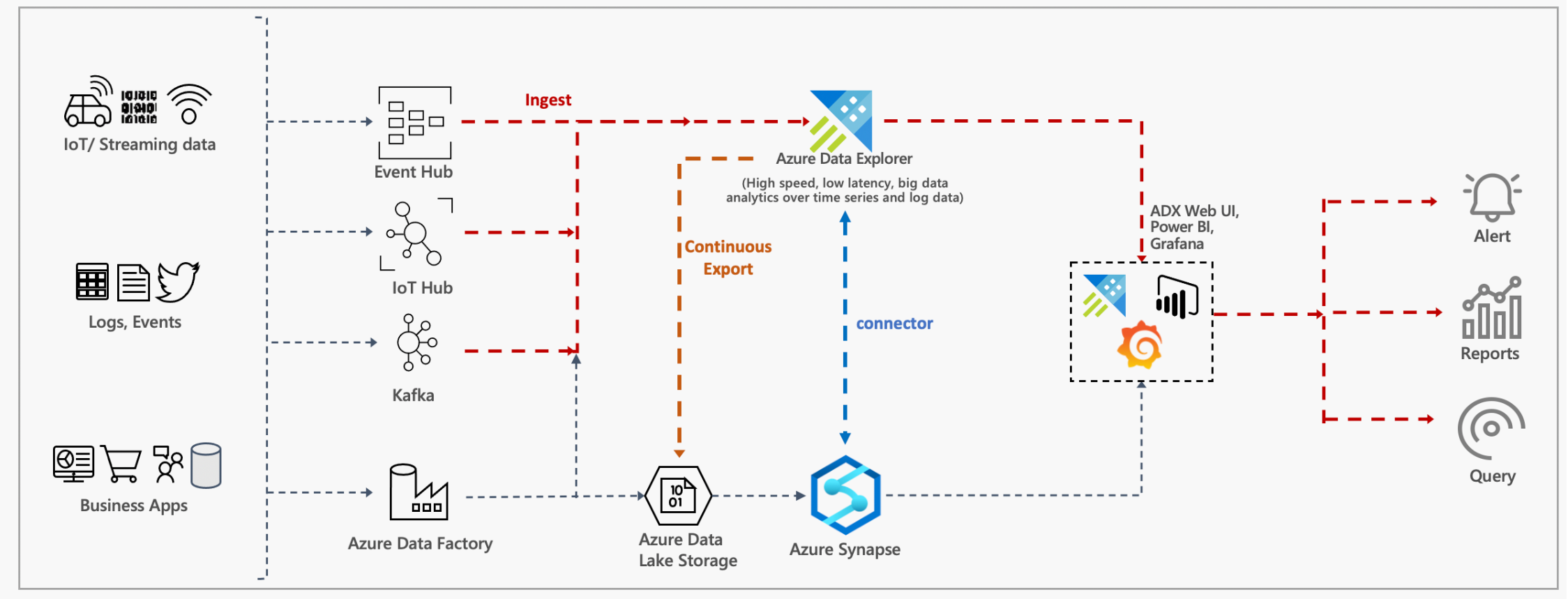
- Installs as an (official) extension
- Supports KQL queries to ADX
- Adds KQL query cells support in Notebooks



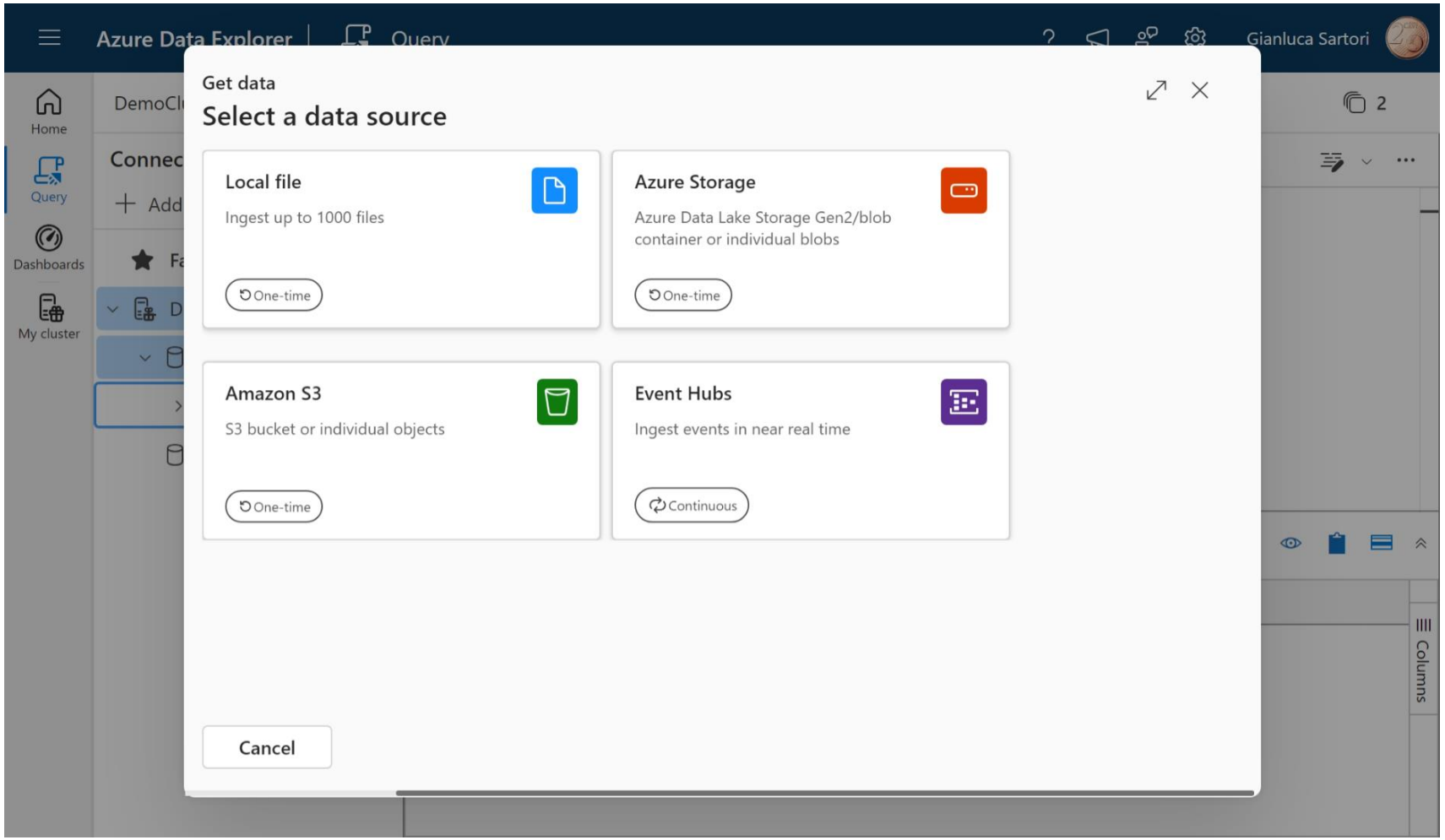
Working with ADX – Writing Data

- Writing data to ADX is called «ingestion»
- Two main ways to ingest data
  - **Batch Ingestion**
    - One-time ingestion from the web UI
    - Batch ingestion from Blob storage
  - **Streaming Ingestion**
    - From IoT Hub/Event Hub/Event Grid/ADF
    - From API client libraries

Working with ADX – Writing Data



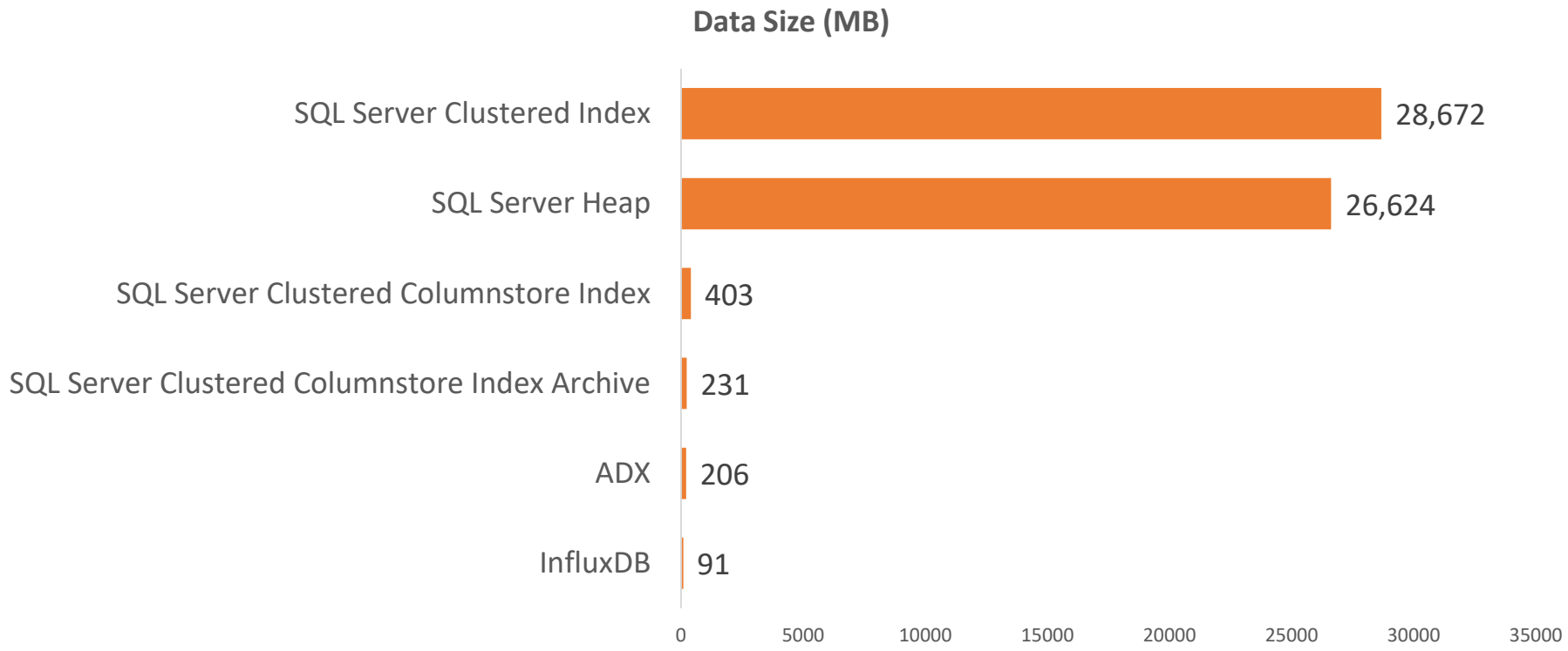
Working with ADX – One time ingestion



DEMO!

How efficient is the storage?

Data size for SQL Server performance counters, 2 instances,
15 days of data, 80 million rows



Working with ADX – UPDATE and DELETE?

- You can insert data with the **.append** command
- delete is supported with **.clear** / **.purge** / **.delete** depending on what you want to do
  - Columnar databases are terrible at deleting data
  - Time series data is append-only, so not really a problem
- **.update** is also supported, but it is a delete/insert instead (surprise)
- Changing data after ingestion has performance implications and is frowned upon

Querying data in Azure Data Explorer

Working with ADX: Querying Data

- ADX supports the Kusto Query Language (KQL)
- It's a brand new language, no similarities with SQL
- SQL is not supported but you can convert SQL to KQL with EXPLAIN
- KQL is simple and powerful
- … but it's not SQL!!!
- SQL support for KQL databases is coming in 2025!!!

```
1  Perf
2  | where TimeGenerated >= ago(10m)
3  | where CounterName == "% Free Space"
4  | project PerfComputer = Computer
5          , CounterName
6          , CounterValue
7          , PerfTime=TimeGenerated
8  | join ( InsightsMetrics
9          | where TimeGenerated >= ago(10m)
10         | project IMComputer = Computer
11                 , Namespace
12                 , Name
13                 , Val
14                 , IMTime=TimeGenerated
15         )
16     on $left.PerfComputer == $right.IMComputer
17
18
```

DEMO!

Working with ADX – Time-Series queries

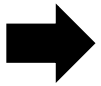
- Typical time-series query patterns:
  - Filter by time

```
where time > now(-7d)
        where time > ago(7d)
        where time between ago(7d) .. now()
```

Working with ADX – Time-Series queries

Group by time

time	lazy_writes_sec	
16-10-2022 10:00:00	398091	10:00
…	…	
16-10-2022 10:59:59	395644	
16-10-2022 11:00:00	432211	11:00
…	…	
16-10-2022 11:59:59	452325	
16-10-2022 12:00:00	456541	12:00
…	…	
16-10-2022 12:59:59	433505	
16-10-2022 13:00:00	456654	13:00
…	…	
16-10-2022 13:59:59	456788	
16-10-2022 14:00:00	417545	…
…	…	



time	avg_lazy_writes_sec
16-10-2022 10:00:00	398,091.00
16-10-2022 11:00:00	400,187.50
16-10-2022 12:00:00	406,095.00
16-10-2022 13:00:00	411,167.76
16-10-2022 14:00:00	417,963.40
16-10-2022 15:00:00	425,683.27
16-10-2022 16:00:00	433,001.40
16-10-2022 17:00:00	441,937.02
16-10-2022 18:00:00	449,937.73
16-10-2022 19:00:00	450,737.81
16-10-2022 20:00:00	453,467.40
16-10-2022 21:00:00	455,703.76
16-10-2022 22:00:00	457,270.19
16-10-2022 23:00:00	458,743.47
17-10-2022 00:00:00	461,533.64
…	…

Working with ADX – Time-Series queries

Fill missing data with defaults

time	value
16-10-2022 10:35:45	398091
16-10-2022 10:36:01	398091
16-10-2022 12:16:15	406095
16-10-2022 12:16:30	406095
16-10-2022 13:06:45	411167
16-10-2022 13:07:01	411167

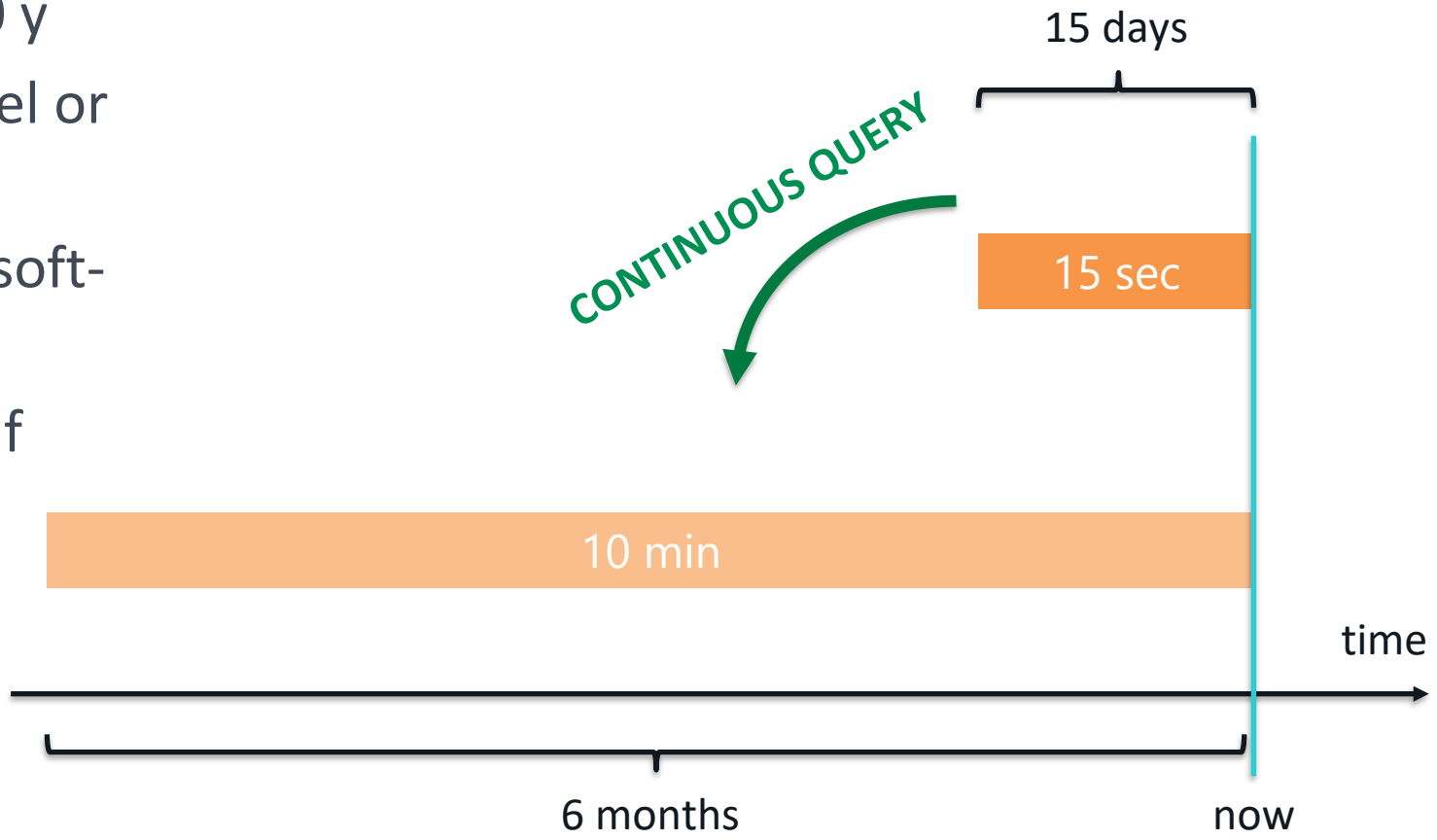
time	avg_lazy_writes_sec
16-10-2022 10:00:00	398,091.00
16-10-2022 11:00:00	NULL
16-10-2022 12:00:00	406,095.00
16-10-2022 13:00:00	411,167.76

DEMO!

Retention Policies and Downsampling

Retention Policies and downsampling

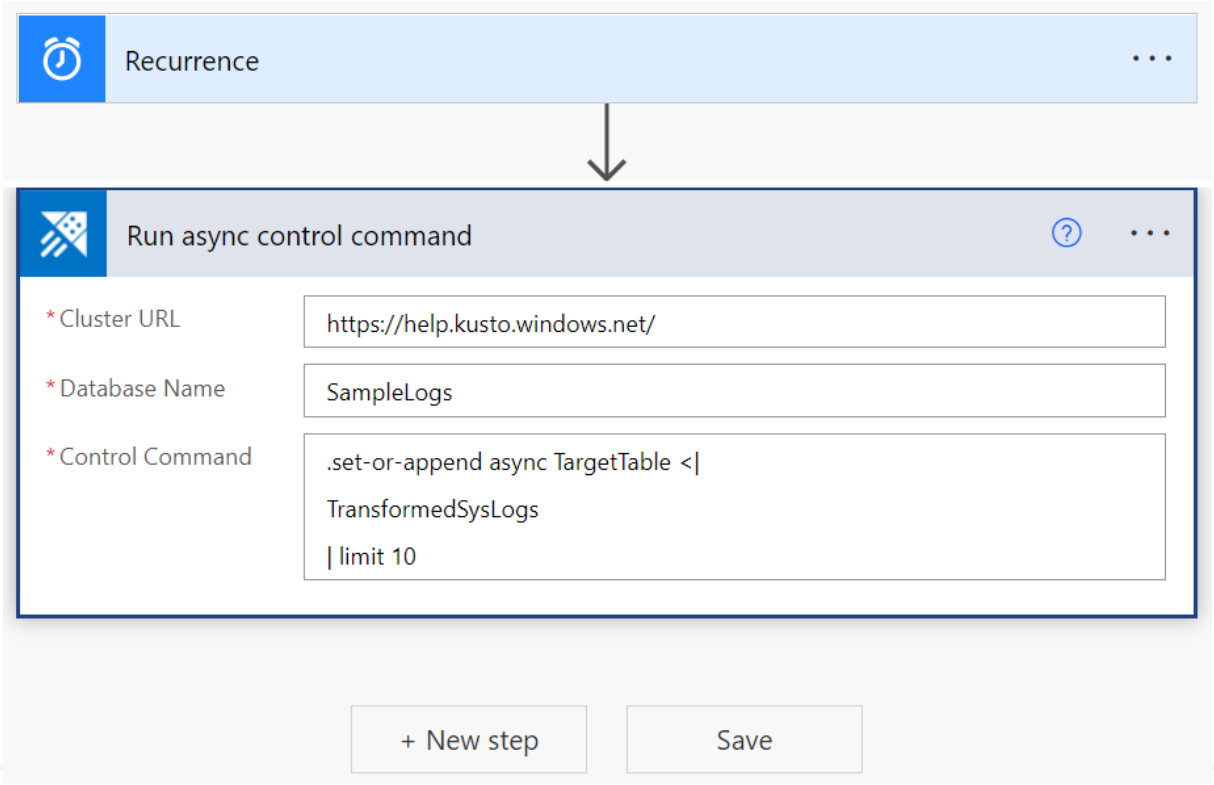
- Default retention policy = 1000 y
- Can be applied at database level or table level
- After retention period, data is soft-deleted
- Can be recovered for 14 days (if recoverability enabled)



Downsampling

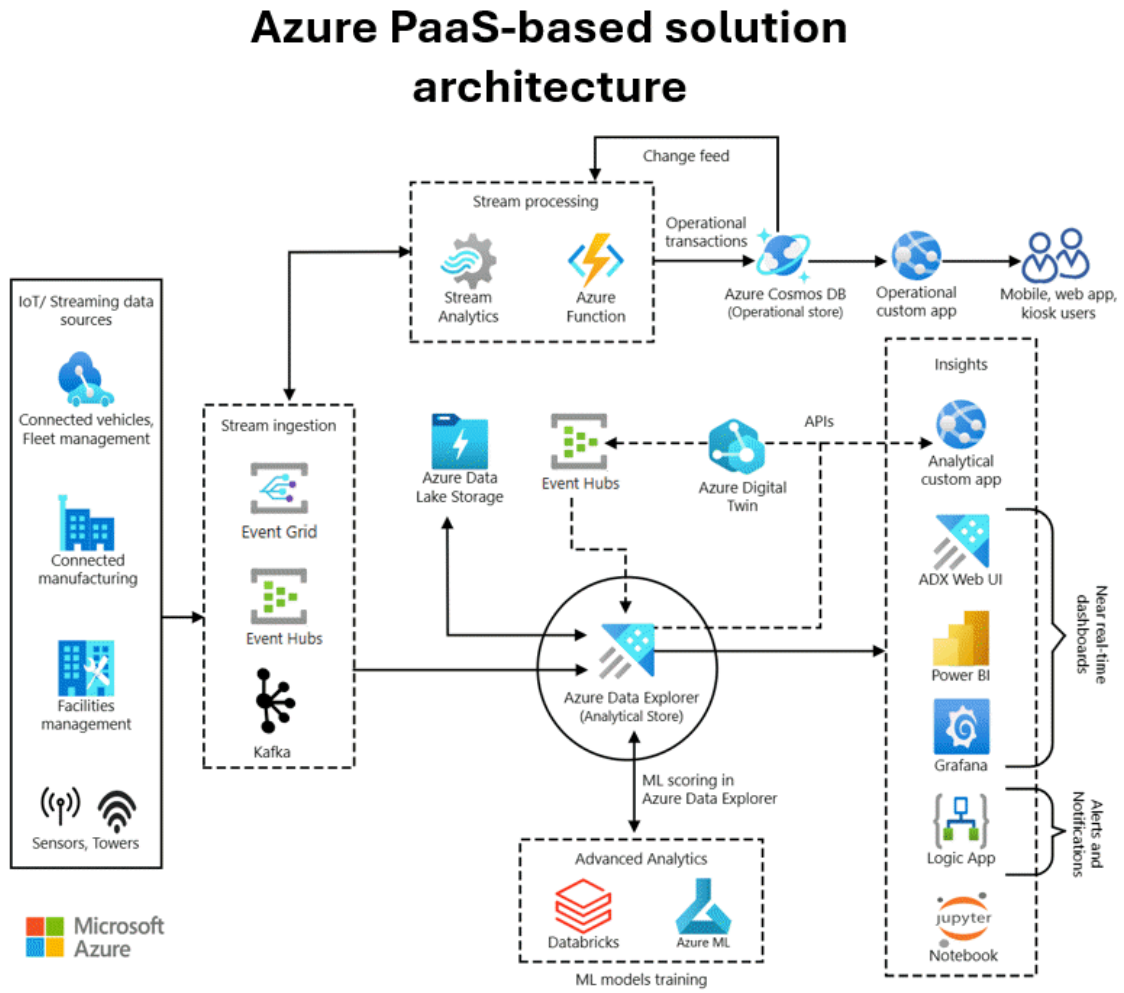
- Data can be summarized with the **make-series** operator and piped to a new table with a lower frequency and longer retention
- How do I automate it? Power Automate has a connector for ADX
- Pay extra attention to overlapping windows
- ADX is not a proper time-series database!

DEMO!

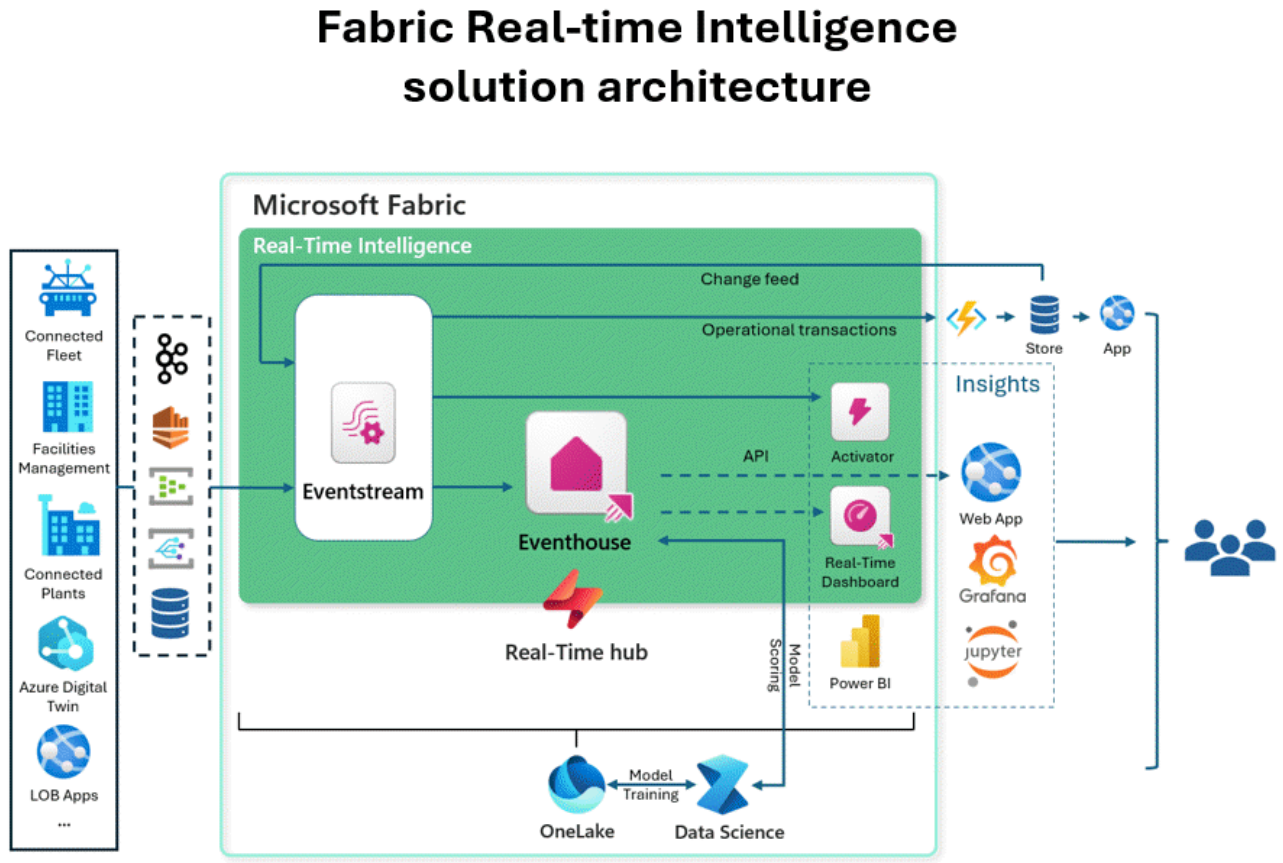


Real-Time Analytics

ADX – Real-Time Analytics



ADX – Real-Time Analytics



15:30

SALA NERA

15:30 → 50 min

Real Time Intelligence Experience in Microsoft Fabric

Nicola Paro Marco Parenzan

Data & Analytics IoT

Conclusion

Can Azure Data Studio do Time-Series?

- Not really a Time-Series database
- Not only a Time-Series database
- It is a data analytics platform with some Time-Series capabilities

✓ Efficient storage*
✓ Retention and downsampling*
✓ Time-Series Query capabilities*

- Can be a very capable platform for your application

2024 WPC

26, 27, 28 NOVEMBRE MILANO

Valuta la sessione

GRAZIE!