



#GlobalAzureTorino



# AI supermercato con Azure Data Explorer

Nicola Paro



# About Me

## Nicola Paro

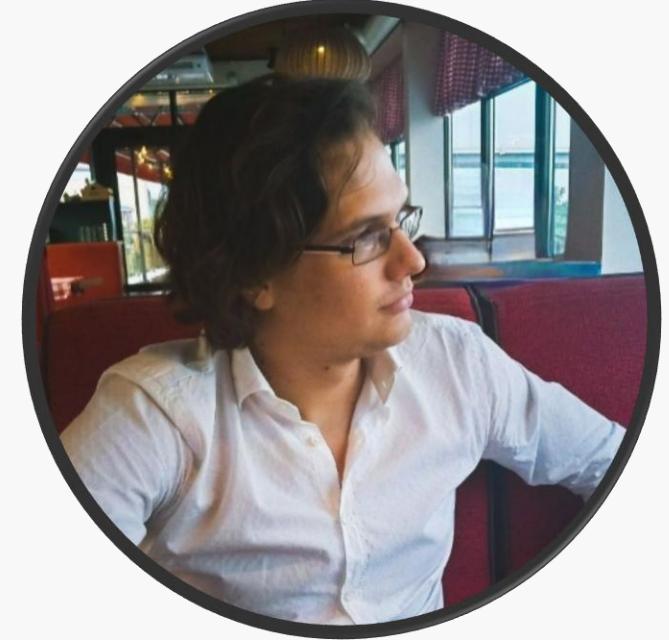
Cloud Solutions Engineer | Technical Team Leader  
@ beanTech



[linkedin.com/in/nicolaparo](https://linkedin.com/in/nicolaparo)



[github.com/nicolaparo](https://github.com/nicolaparo)



The background of the slide features a complex, abstract digital design. It consists of a dark blue gradient background with numerous thin, light-colored lines forming a grid-like network. Overlaid on this are several large, glowing, semi-transparent circular nodes of varying sizes, primarily in shades of white, yellow, and green. Some of these nodes contain small, illegible text or symbols. In the upper right corner, there is a solid yellow rectangular box containing the text '#GlobalAzureTorino' in black.

#GlobalAzureTorino

# A Real Life Story



#GlobalAzure

# A Real Life Story



Creerò degli scraper per leggere i prezzi dei prodotti giorno per giorno.

Scraper 1

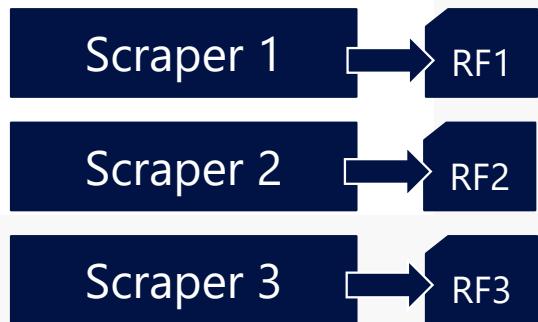
Scraper 2

Scraper 3

# A Real Life Story



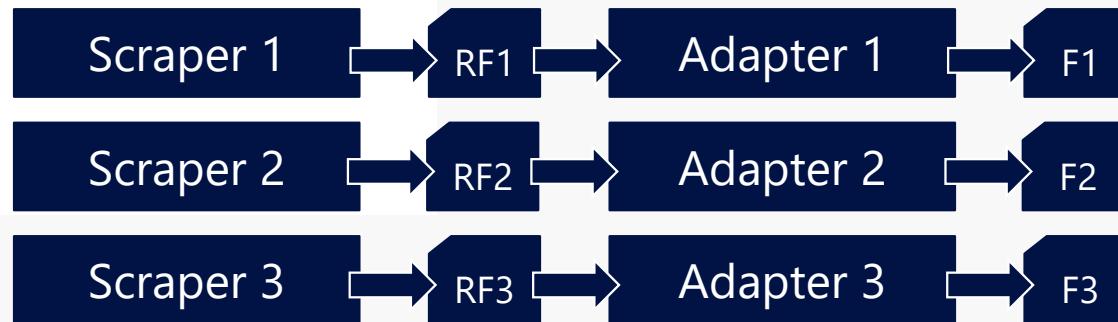
Salverò i dati estratti in alcuni RawFiles in formato TXT



# A Real Life Story



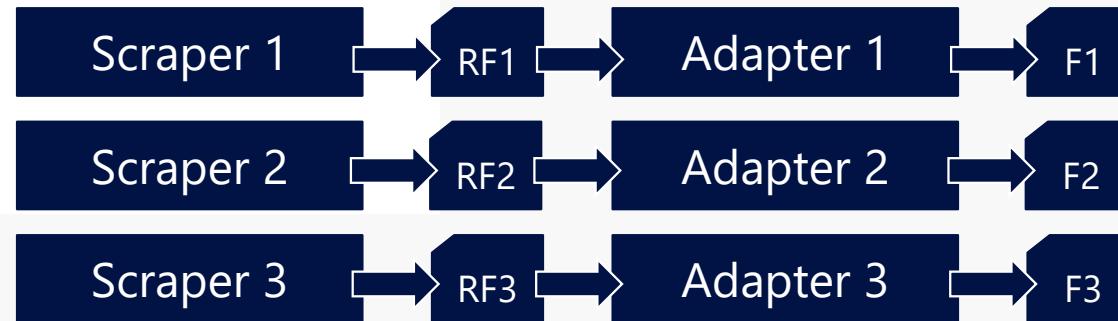
I miei adapters estrarranno i dati rilevanti e li persisteranno in un formato comune



# A Real Life Story



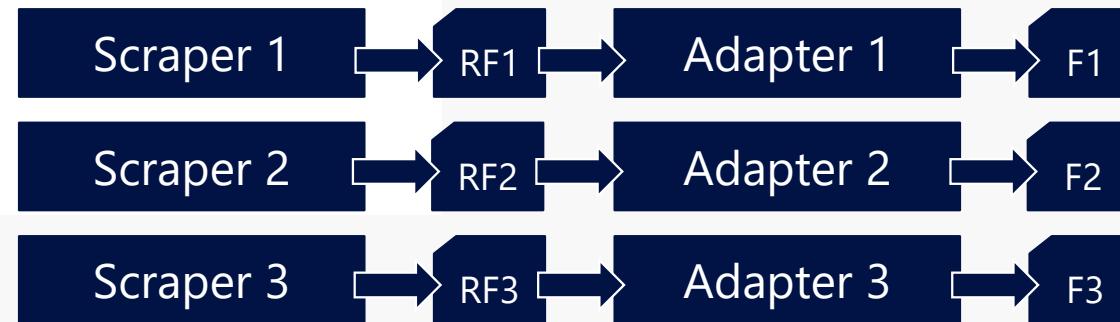
Sono troppi dati e non riesco a leggerli.



# A Real Life Story



Sono troppi dati e non riesco a leggerli.



# A Real Life Story



Caricherò i files in uno Storage Account e li analizzerò con Azure Data Explorer



# A Real Life Story

## Prima di fare Web Scraping

- **Controlla prima i termini di licenza** del sito web
- Non fare scraping di dati protetti da **copyright**
- Non fare scraping di contenuti protetti da login
  - Non fare scraping dopo aver fatto una login
- Analizza solamente informazioni disponibili pubblicamente
- Controlla il **/robots.txt** per essere avere la certezza di poter analizzare un determinato path
- Controlla la **/sitemap.xml** per un elenco agevolato delle pagine da analizzare



#GlobalAzureTorino

# Azure Data Explorer



#GlobalAzure

# Azure Data Explorer

Azure Data Explorer è un servizio di analisi dati veloce e totalmente gestito per l'analisi in tempo reale di grandi volumi di dati, provenienti da applicazioni, siti web, dispositivi IoT ecc...

# Azure Data Explorer

## Main Features

Fast Data Ingestion

Interactive Data Exploration

Real Time Analytics on Streaming data

Scalability

Integration with other Azure Services

# Azure Data Explorer

When does it makes sense to use ADX?

Sliding  
Window of  
data

Read A Lot!

Insert /  
Append A Lot!

Few Delete

NEVER  
Update

# Azure Data Explorer

Pricing for Business Purposes

**ADX Cost + VMs Cluster Cost + Storage Cost**

Dev/Test  
**Free** (No SLA)

Standard  
\$0.11/core per  
hour

Varies on VM  
Size

\$0.126/hour  
for the  
smallest VM

Pay per use

More info on <https://azure.microsoft.com/en-us/pricing/details/data-explorer/>

# Azure Data Explorer

Pricing for Education Purposes with Limited Functionalities

**ADX Cost + VMs Cluster Cost + Storage Cost**

Dev/Test

Varies on VM

Free Cluster Available with Microsoft Account / AAD User Identity  
**No need for an Azure Subscription or Credit Card**

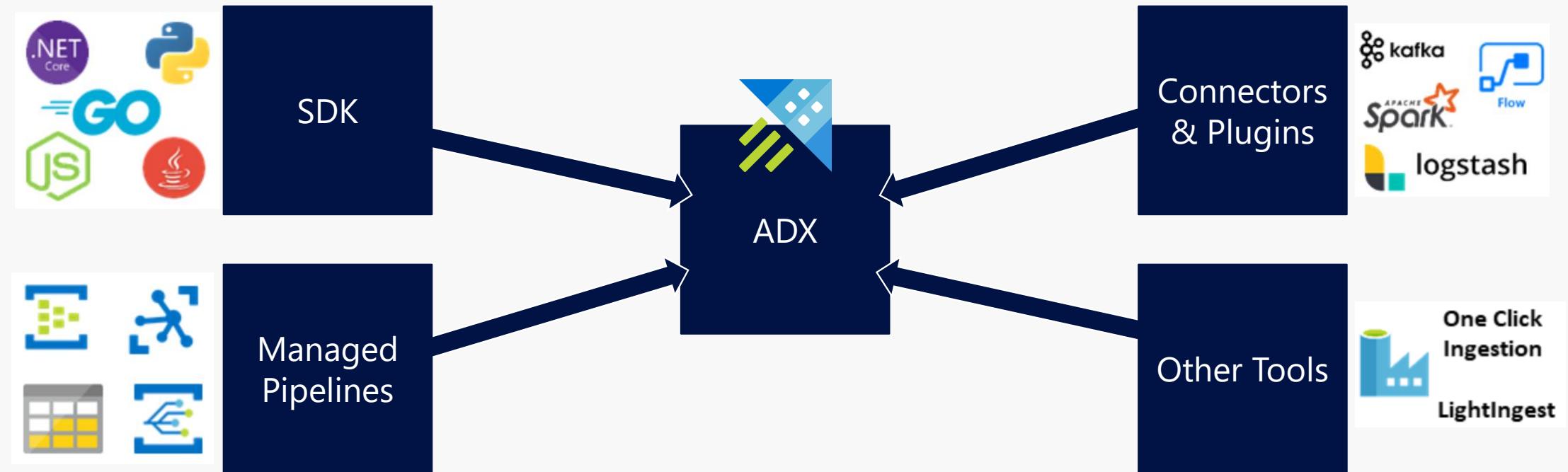
hour

smallest VM

More info on <https://azure.microsoft.com/en-us/pricing/details/data-explorer/>

# Azure Data Explorer

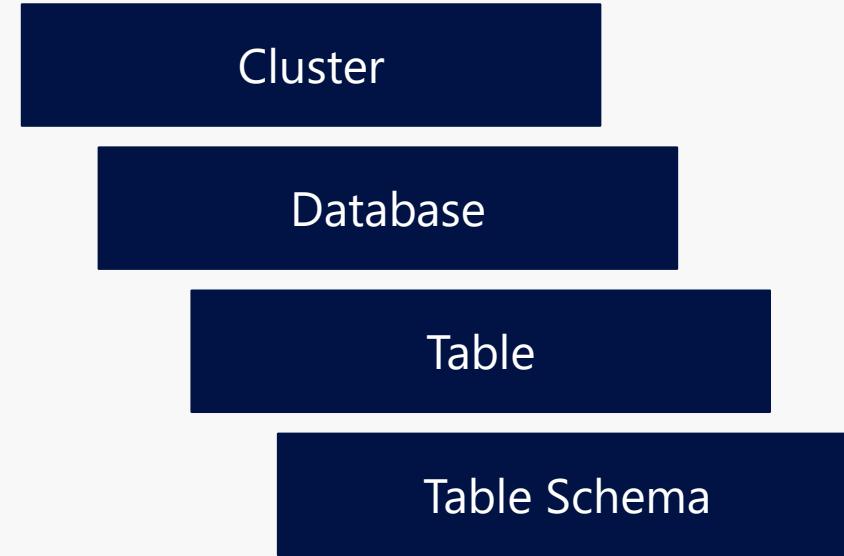
## Data Ingestion



# Azure Data Explorer

## Data Organization

Simile ai database  
relazionali (ex: SqlServer)



# Azure Data Explorer

Data Organization – Differenze rispetto ai RDMS

No Primary  
Key

No Unique  
Keys

No Foreign  
Keys

Columnstore  
Indexes

Data Sharding  
(Extents)

# Azure Data Explorer

# Behind the scenes

# Azure Data Explorer

Behind the scenes

I dati delle tabelle sono divisi in extents  
(aka shards, partizioni, ...)

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla
Bla	Bla	Bla

Bla	Bla	Bla
Bla	Bla	Bla

# Azure Data Explorer

Behind the scenes

I dati delle tabelle sono divisi in extents  
(aka shards, partizioni, ...)

Un extent è una mini-tabella che  
contiene dati e metadati.

Un extent **non può mai essere  
modificato**, ma può essere cancellato

I dati sono organizzati in colonne

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla
Bla	Bla	Bla

Bla	Bla	Bla
Bla	Bla	Bla

# Azure Data Explorer

Behind the scenes

I dati delle tabelle sono divisi in extents  
(aka shards, partizioni, ...)

Un extent è una mini-tabella che  
contiene dati e metadati.

Un extent **non può mai essere  
modificato**, ma può essere cancellato

I dati sono organizzati in colonne

Extent più piccoli possono essere uniti  
in extent più grandi

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla

# Azure Data Explorer

Behind the scenes

Gli extent sono creati durante le operazioni di inserimento

Un extent è unito ad altri

- Shard rebuild
- Shard merge

Un extent può essere cancellato con una retention-policy

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla

Bla	Bla	Bla

# Azure Data Explorer

## Behind the scenes

Gli shards sono distribuiti tra i nodi del cluster

## Node 1

## Node 2

# Azure Data Explorer

# Behind the scenes - A query to the cluster

# Logs

| where Timestamp > ago(1h)

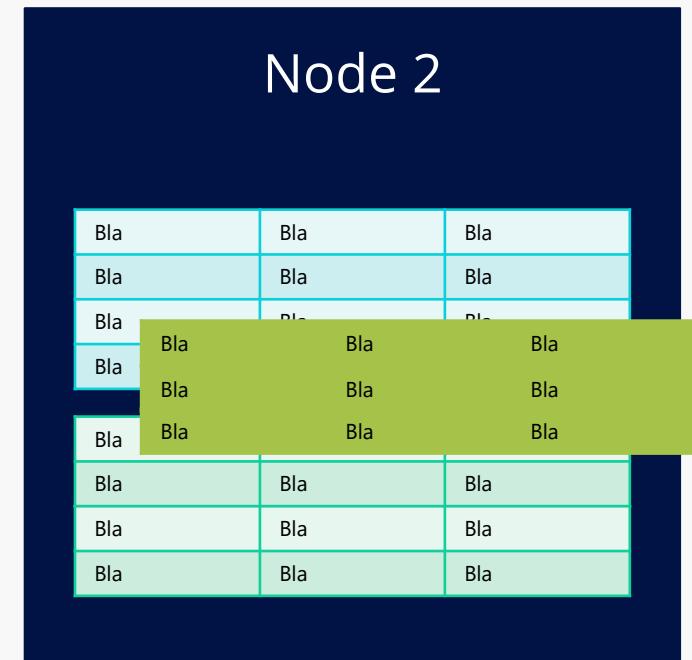
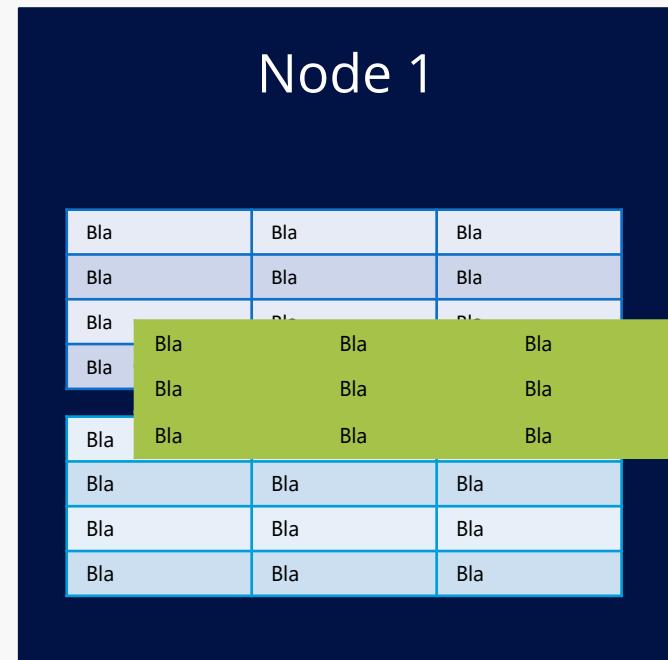
# Node 1

# Azure Data Explorer

# Behind the scenes - A query to the cluster

## Logs

| where Timestamp > ago(1h)

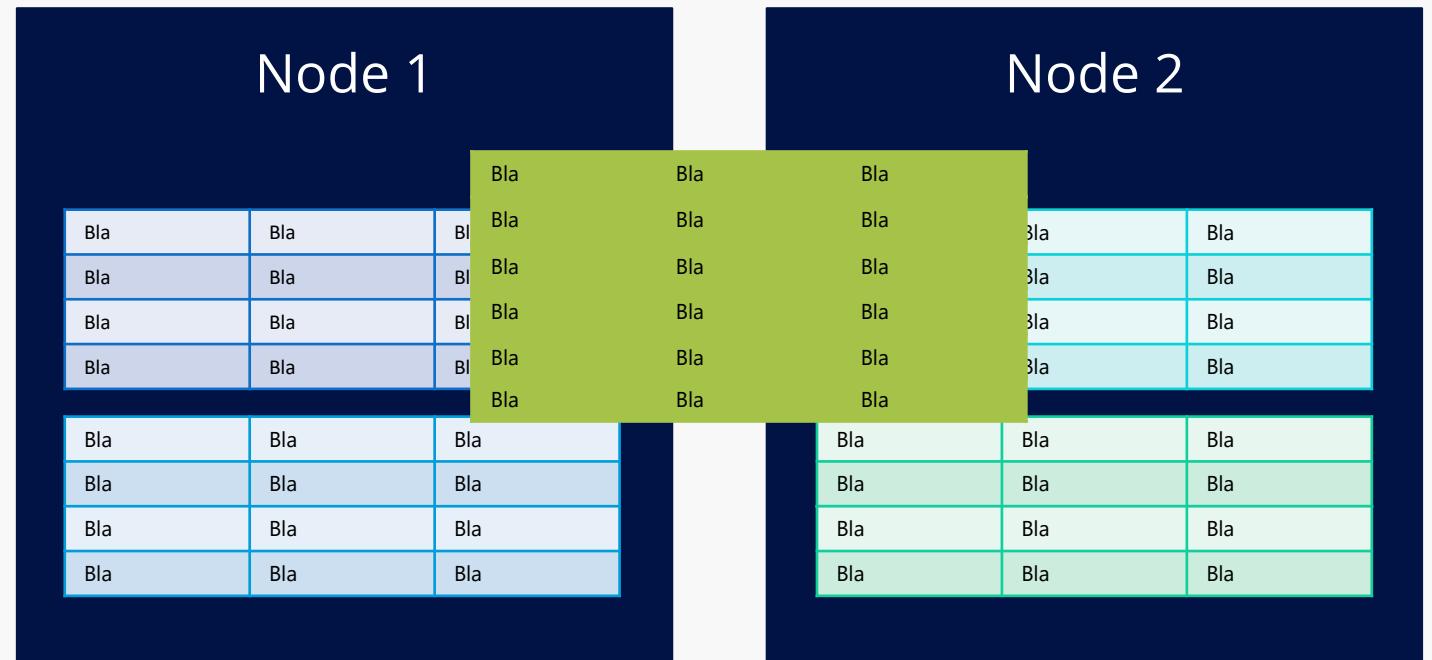


# Azure Data Explorer

Behind the scenes - A query to the cluster

Logs

```
| where Timestamp > ago(1h)
```



# Azure Data Explorer

## Kusto Query Language

Una query Kusto è una richiesta in sola lettura per il processamento dei dati e la produzione di risultati.

La richiesta è effettuata tramite testo, utilizzando un modello di data-flow che è semplice da leggere, scrivere ed automatizzabile.

Le query Kusto sono composte di una o più istruzioni.



# Azure Data Explorer

## Kusto Query Language

SQL	KQL
<b>SELECT</b>	<b>project, extend, project-away, project-keep, ...</b>
<b>WHERE</b>	<b>where, search, ...</b>
<b>JOIN, LEFT JOIN, ...</b>	<b>join kind=inner, kind=leftouter, ...</b>
<b>UNION</b>	<b>union</b>
<b>GROUP BY</b>	<b>summarize</b>
<b>ORDER BY</b>	<b>sort by, order by, top by</b>
<b>TOP, LIMIT</b>	<b>take</b>

More on <https://learn.microsoft.com/en-us/azure/data-explorer/kusto/query/>



#GlobalAzureTorino

# Demo

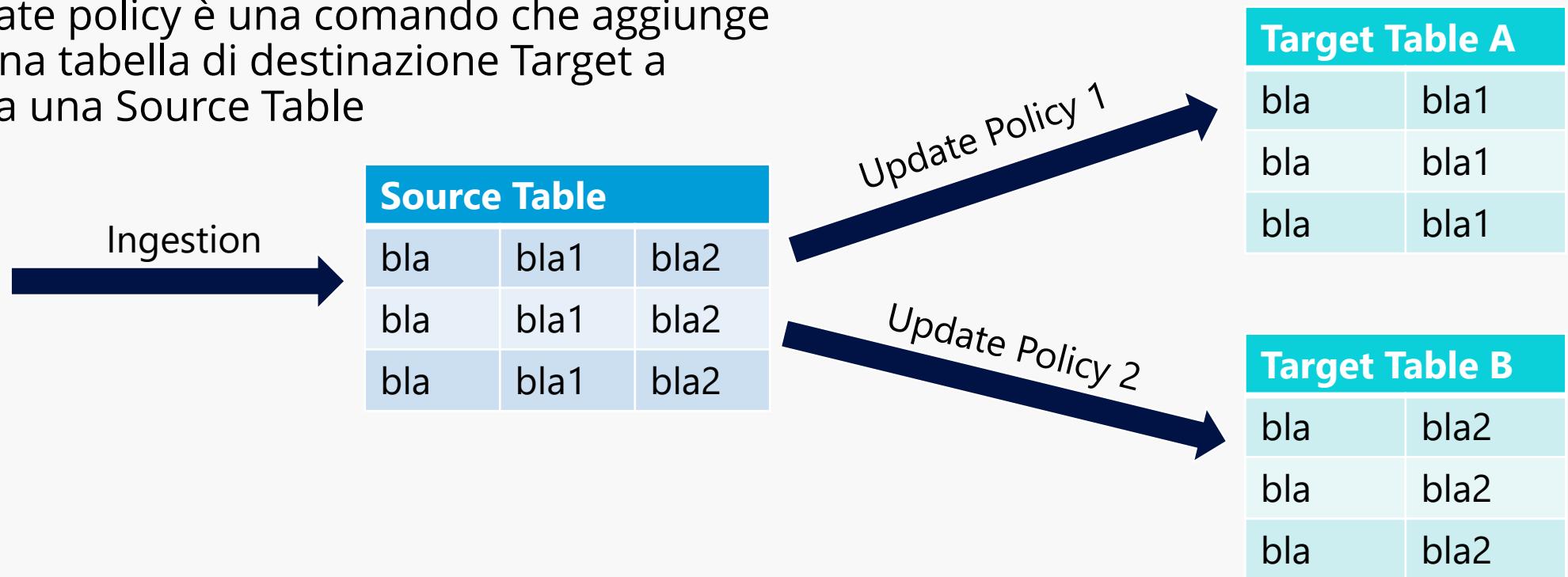


#GlobalAzure

# Azure Data Explorer

## Update Policies

Una update policy è un comando che aggiunge dati ad una tabella di destinazione Target a partire da una Source Table



# Azure Data Explorer

## Materialized View

Una Materialized View è una vista aggregata sui dati di una tabella ADX. I dati sono materializzati anche su disco.

Vantaggi nell'adozione delle Materialized View

Performance  
Improvement

Data  
Freshness

Cost  
Reduction

# Azure Data Explorer

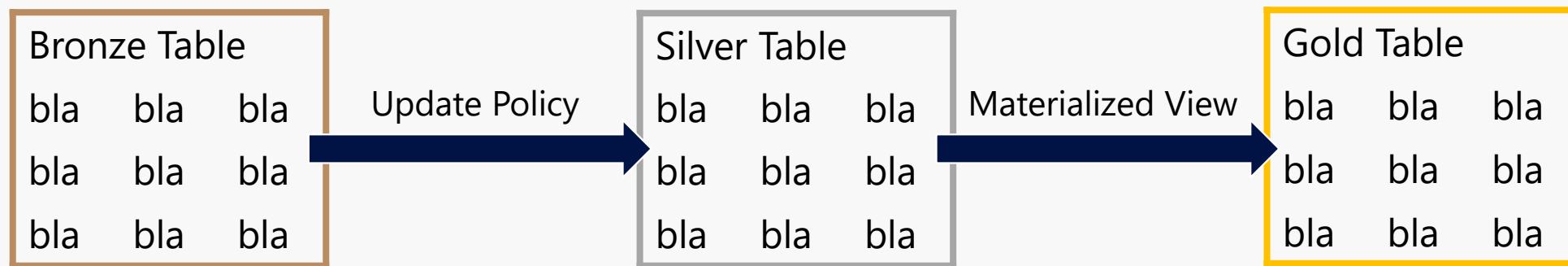
## Update Policy vs Materialized View

### Update Policy

- Data Transformation
- Data Enrichment

### Materialized View

- Data Aggregation

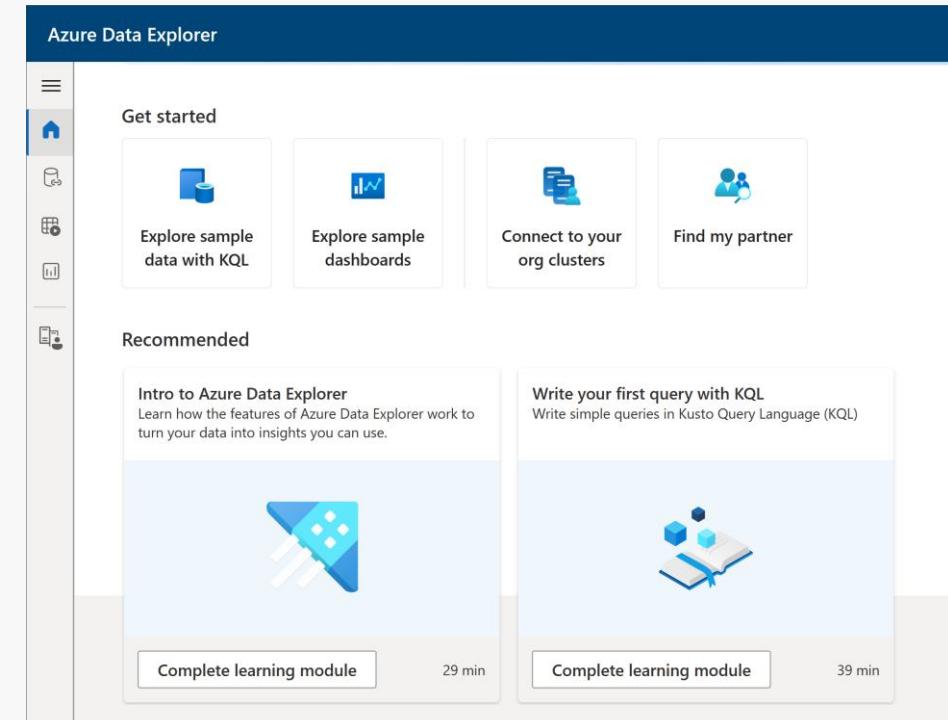


# Azure Data Explorer

Where to start?

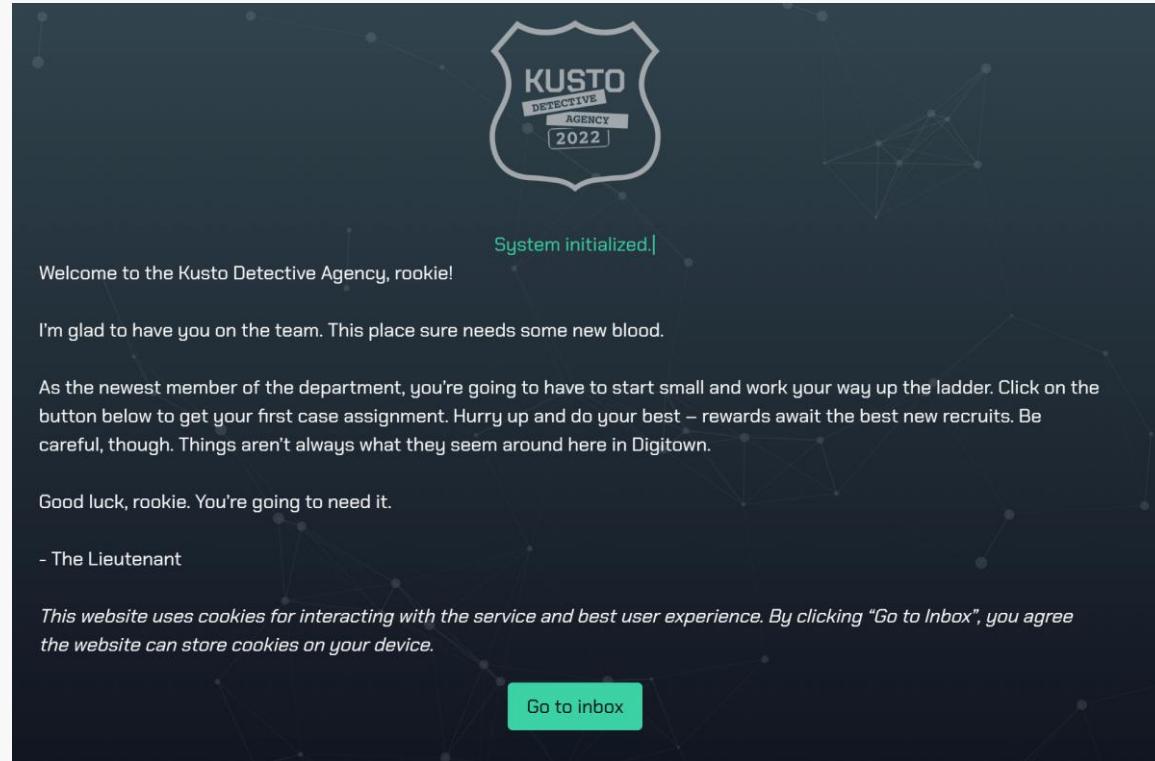
Azure Data Explorer supporta dei database **sample gratuiti** su cui è possibile effettuare delle interrogazioni per prendere confidenza con lo strumento

<https://dataexplorer.azure.com/home>



# Azure Data Explorer

Where to start?



**Kusto Detective Agency:** una gamification per imparare ad usare kusto! (e non solo)

<https://detective.kusto.io/>



# Azure Data Explorer

Where to start?

Find my partner: trova il tuo partner per consulenza, formazione e supporto in Azure Data Explorer

Welcome to Azure Data Explorer  
Discover valuable business insights and make data-driven decisions

Get started

- Explore sample data with KQL
- Explore sample dashboards
- Connect to your org clusters
- Find my partner

Recommended

Azure Data Explorer | All dashboards > Find my partner

Editing Add Parameters File Share

Partner : All Focus Geography : All Partner Type : All

As of 1 minu

Reset

Partner details

Partner	FocusGeography	PartnerType	Website	Contact
Uptake Technologies Inc.	["Global"]	["Global ISV"]	<a href="https://www.uptakefusion.com/">https://www.uptakefusion.com/</a>	leanna.chan@uptake.com
Alten	["Global", "Europe", "Neth..."]	["System Integrator", "Con..."]	<a href="https://www.alten.com/">https://www.alten.com/</a>	adx@alten.nl
beanTech srl	["Italy"]	["System Integrator", "Solu..."]	<a href="https://www.beantech.it/">https://www.beantech.it/</a>	MSFT-ADX@beantech.it
Codit	["Western Europe", "UK", ...]	["System Integrator"]	<a href="https://www.codit.eu">https://www.codit.eu</a>	francis.defauw@codit.eu
InCycle	["USA", "Canada"]	["System Integrator", "Bout..."]	<a href="https://www.incyclesoftware.com">https://www.incyclesoftware.com</a>	barry.paquet@incyclesoftware.com
MAQ Software	["Worldwide, with office..."]	["SI mostly with SaaS Offe..."]	<a href="https://maqsoftware.com">https://maqsoftware.com</a>	engage@maqsoftware.com
Data Mastery	["Australia"]	["System Integrator"]	<a href="https://www.datamastery.ai/">https://www.datamastery.ai/</a>	info@datamastery.ai
CyberProof – a UST company	["Global"]	["MDR", "Consulting", "Cust..."]	<a href="https://www.CyberProof.com">https://www.CyberProof.com</a>	info@cyberproof.com
Mesh Systems	["Global"]	["IoT Solution Provider", "I..."]	<a href="https://meshsystems.com">https://meshsystems.com</a>	<a href="https://meshsystems.com/contact">https://meshsystems.com/contact</a>
OPTION 4.0 AG	["European and UK Time..."]	["System Integrator"]	<a href="https://www.option40.com">https://www.option40.com</a>	operations@option40.com

<https://learn.microsoft.com/it-it/azure/data-explorer/find-my-partner>



#GlobalAzureTorino

# Q & A



#GlobalAzure



#GlobalAzureTorino

# Thank You!



#GlobalAzure

# Contacts

## Nicola Paro

Cloud Solutions Engineer | Technical Team Leader  
@ beanTech



[linkedin.com/in/nicolaparo](https://linkedin.com/in/nicolaparo)



[github.com/nicolaparo](https://github.com/nicolaparo)



