



Power BI

Power BI Streaming DataFlow and Azure IoT Central

Marco Pozzan

Marco Parenzan

Marco Parenzan

- Senior Solution Architect @ beanTech
- 1nn0va Community Lead (Pordenone)
- Microsoft Azure MVP



- LinkedIn: <https://www.linkedin.com/in/marcoparenzan/>
- Slideshare: <https://www.slideshare.net/marco.parenzan>
- GitHub: <https://github.com/marcoparenzan>



Marco Pozzan

- Dal 2002 faccio consulenza sulla progettazione di data warehouse relazionali e sulla progettazione di cubi Tabular/Multidimensional con strumenti Microsoft.
- Dal 2007 Community Lead di 1nn0va (<https://www.facebook.com/1nn0va>)
- dal 2014 MVP per SQL Server
- Dal 2017 mi occupo di modern data warehouse con prodotti Azure: Synapse, Azure Data Factory, Stream Analytics, Data Lake e affini. Sono Docente all'Università di Pordenone nel corso Architetture Big Data e DWH: Tecniche di modellazione del dato
- Dal 2020 Founder e CTO presso Cargo BI: start-up di insurance data analytics
- MCP, MCSA, MCSE, MCT SQL Server
- Relatore in diverse conferenze sul tema.
 - info@marcopozzan.it
 - [@marcopozzan.it](https://www.instagram.com/marcopozzan)
 - www.marcopozzan.it



Microsoft
CERTIFIED
Trainer

Microsoft
CERTIFIED
Professional

Microsoft
CERTIFIED
Solutions Associate
SQL Server 2012/2014

Agenda

- Overview
- Presentazione di Azure IoT Central
- Setup - data export - demo
- Power BI Streaming Data Flow
- Setup Data Flow
- Anomaly detection
- Invocare Anomaly detection da power query - demo

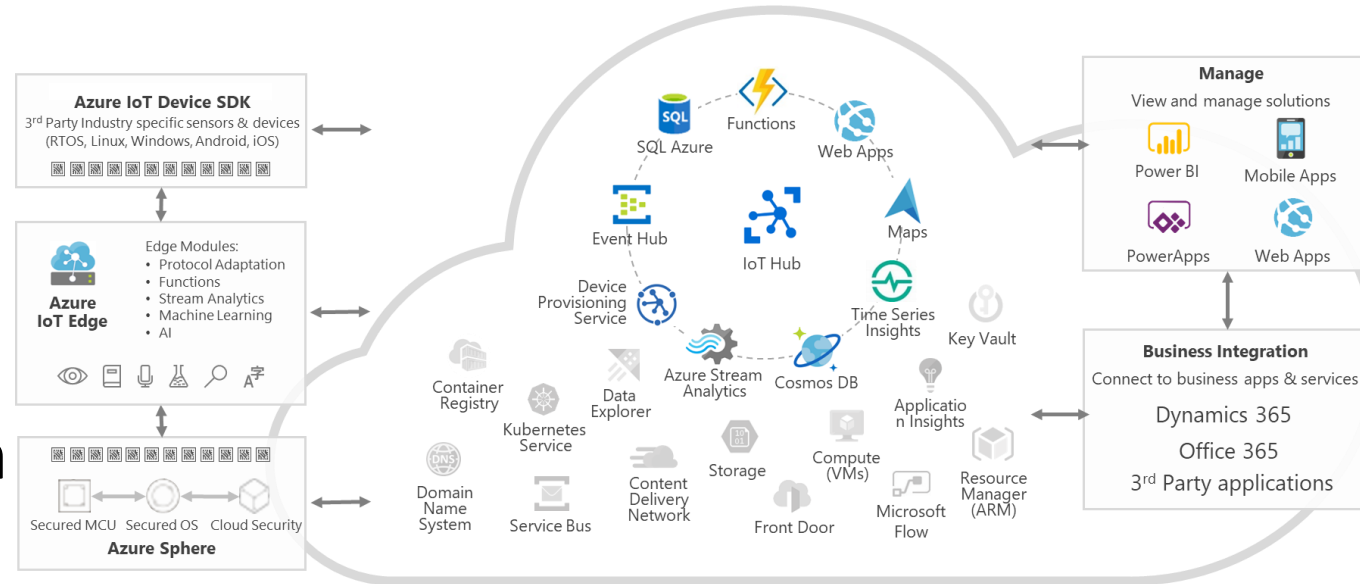
Overview



Power BI

Facts

- Everyone «needs» IoT
- Everyone receive requests for IoT
- IoT can become complex
- It is difficult to organize a team for that
- Biggest issues: cost, skills, manageability



Lot of IoT

Smart Buildings



- Space Management & Optimization
- Connected spaces
- Energy management & building operations
- Occupant experience and productivity

Retail



- Inventory Management
- Store Analytics
- Digital Distribution Center
- Connected logistics

Smart City



- Smart Lighting
- Smart water
- Smart parking
- Smart Waste
- Air quality monitoring

Healthcare



- Inventory Management for Medical Supplies
- Remote patient monitoring
- Cold Chain supply tracking
- Smart hospital building

Agriculture

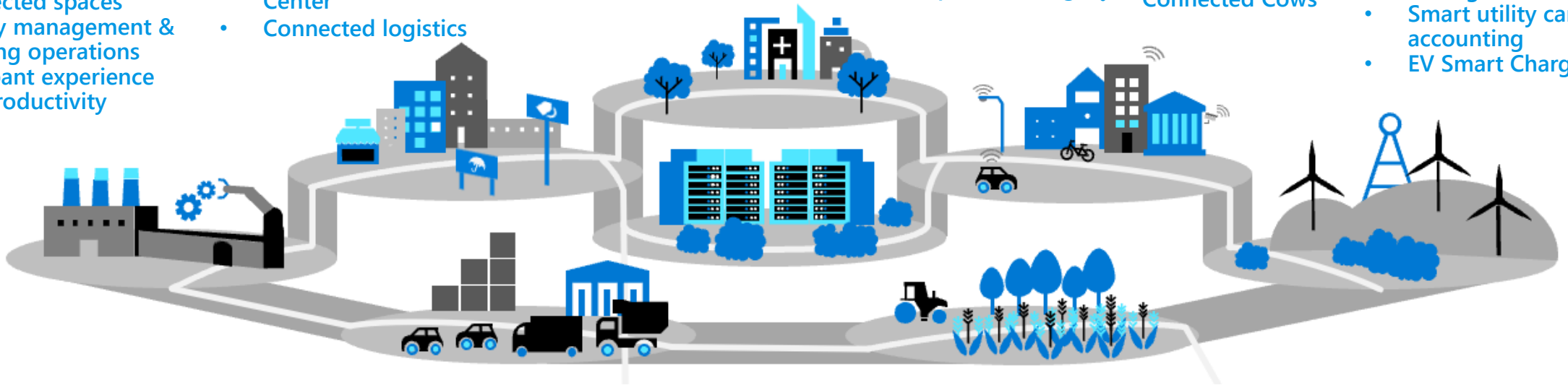


- Smart Equipment
- Precision Farming
- Connected Cows

Energy



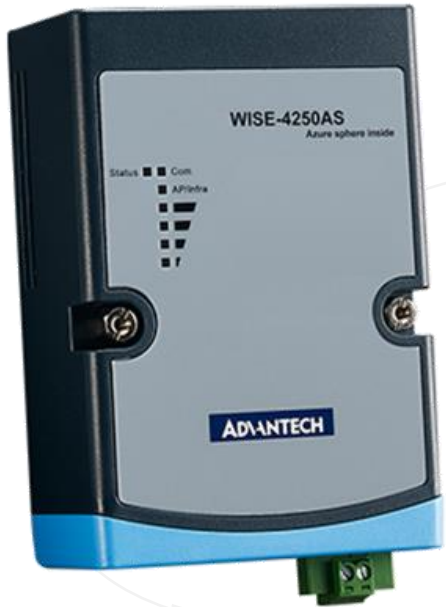
- Carbon Emission Reduction
- Grid assets management
- Smart utility carbon accounting
- EV Smart Charging



Which IoT?

- Let's consider a couple of IoT Scenarios mainly
 - Smart*
 - Industrial IoT
- Everything is (almost) useful in each, but let's consider Smart*
- Characteristics:
 - Short pipeline (Sensor → Low/Nothing edge intelligence → Gateway → Cloud)

Data Logger, Protocol Adapters, SCADA, ...



WISE-4250

https://devicemodels.azure.com/dtmi/advantech/wise_4250as_s231-1.expanded.json

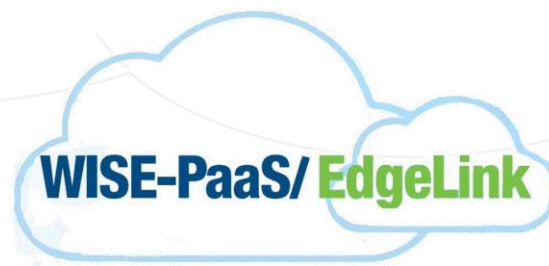


UNO-2372

https://github.com/Azure/azure-iot-device-ecosystem/blob/master/get_started/embedded-linux-ecu-1251-c.md



WISE-710



Needs

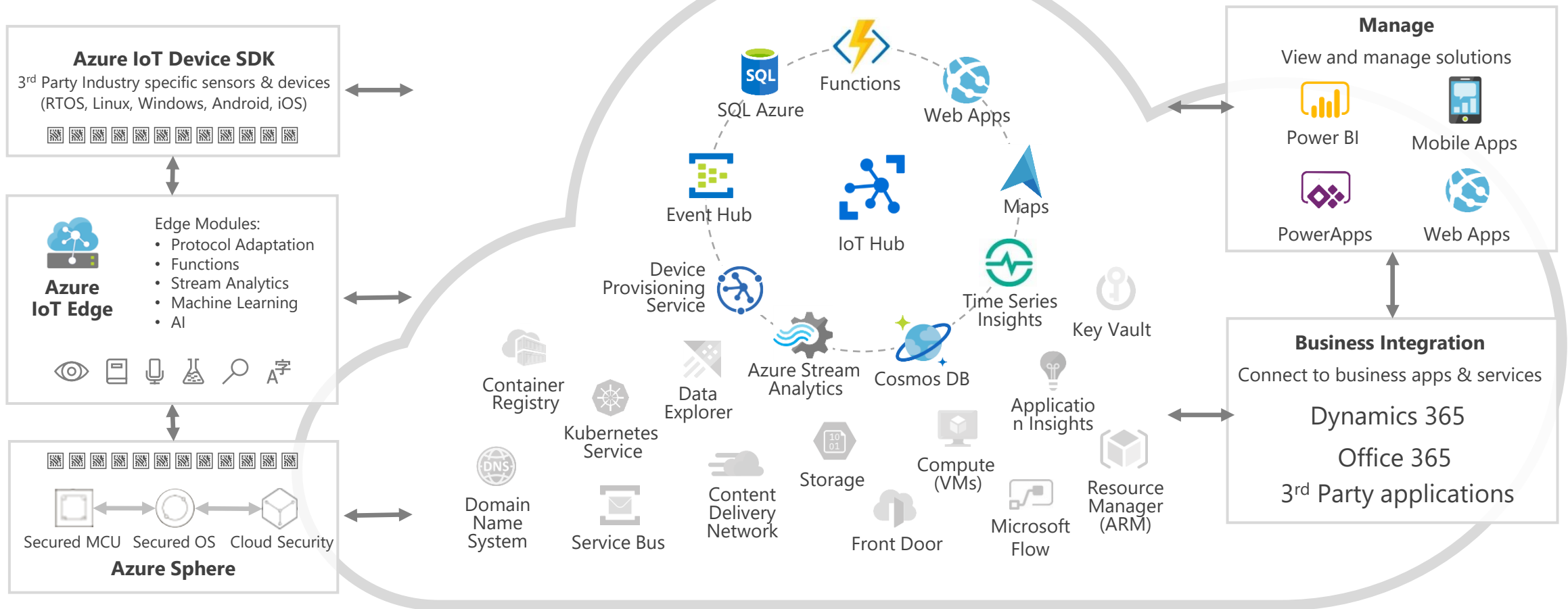
- Software as a Service approach
- Power BI set the line
- Any option for IoT in this space?

Azure IoT Central



Power BI

Your options for building IoT solutions



Your options for building IoT solutions



Pricing Model

	Standard Tier 0	Standard Tier 1	Standard Tier 2
Price per device	€0,07 al mese	€0,34 al mese	€0,59 al mese
Monthly device message allocation*	400 messages	5,000 messages	30,000 messages
Included free quantity per IoT Central application	2 free devices (800 included messages)	2 free devices (10,000 included messages)	2 free devices (60,000 included messages)
Use case	For devices sending a few messages per hour	For devices sending a few messages per hour	For devices sending a message every few minutes
Overage pricing per 1K messages ¹	€0.060 per 1K messages	€0.013 per 1K messages	€0.013 per 1K messages
*Total message allocation is shared across all devices in an IoT Central application			

Common Architectures

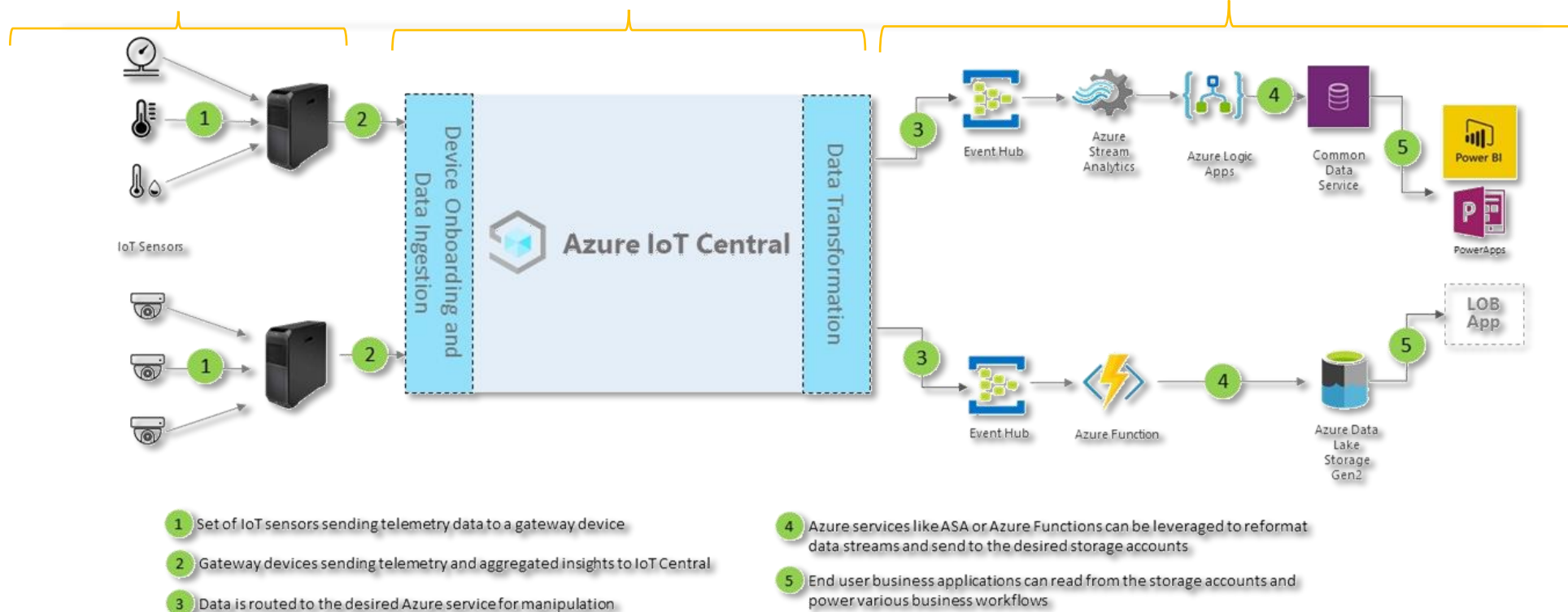
Telemetry from devices or sensors is sent to a gateway device or directly to IoT Central

Data is sent, aggregated and analyzed in IoT Central

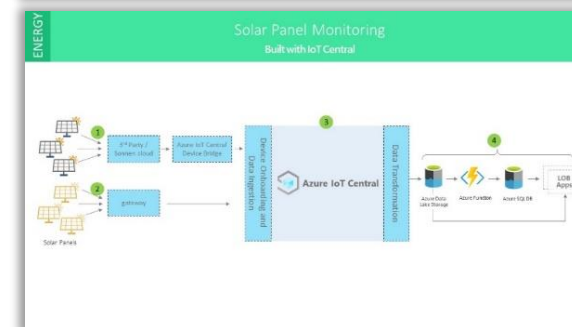
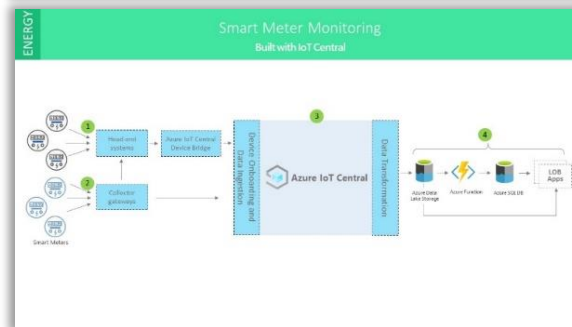
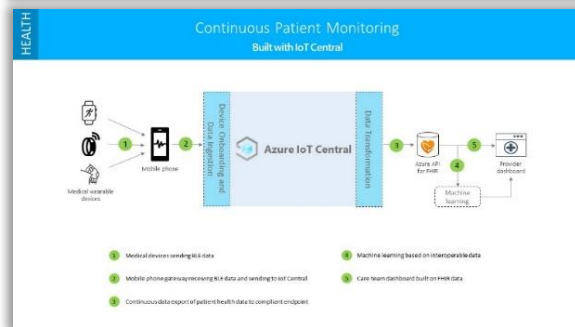
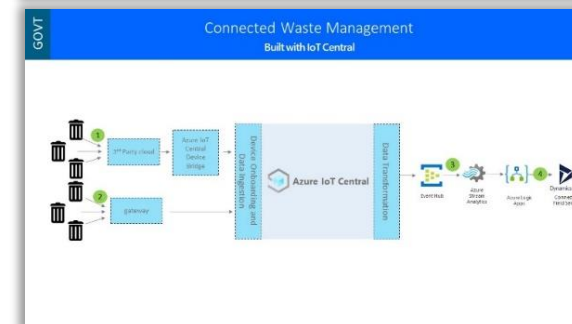
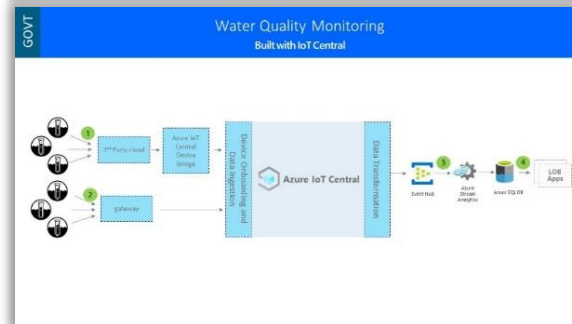
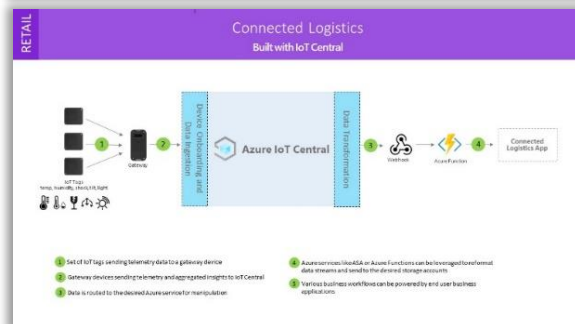
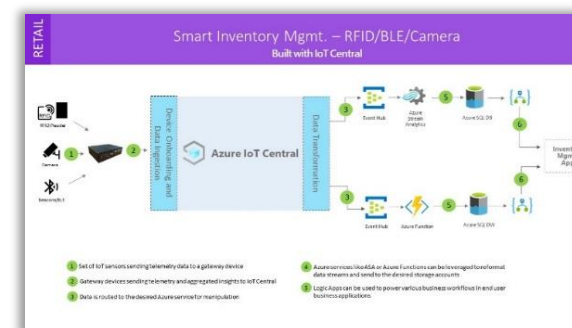
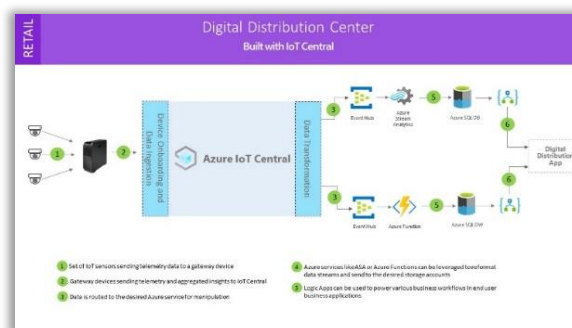
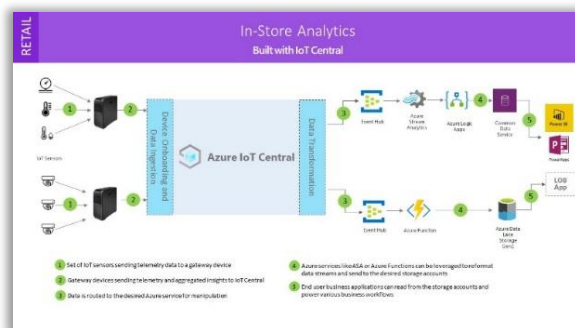
Data is routed to other Azure services for manipulation – storage, reformatting, integration to business workflows and applications, etc.

Azure services like ASA or Azure Functions can be used to reformat data streams and send them to storage accounts

Logic apps can be used to power business applications



“Build with” patterns (from Industry Application Templates)



Why Azure IoT Central is interesting and not a toy

- It is fully compliant with Azure IoT Platform (Azure IoT Hub, Azure IoT Device Provisioning Services, Azure IoT Device)
- It has an interesting pricing model (per Device)
- It forces you in terms of modeling (Device Twin and Plug and Play)
- It gives you a fast start...
- ...because it is a very interesting scenario based tool...
- ...with standard that gives you a growth path!
- It is extensible with other Azure services
- It shows how an IoT Solution should be

Azure IoT Central Values

- SaaS approach
- Scenario-based approach
- Pricing per Device (with message tier)
- IoT PnP, Device Provisioning Service
- IoT Hub full compatibility
- Dashboards
- Internal Data Consumption → Single Service
- REST API
- Routing to external paths

What is IoT Plug and Play?

- IoT Plug and Play enables solution builders to integrate smart devices with their solutions without any manual configuration. At the core of IoT Plug and Play, is a device model that a device uses to advertise its capabilities to an IoT Plug and Play-enabled application.
- To make IoT Plug and Play work with Azure Digital Twins, you define models and interfaces using the Digital Twins Definition Language (DTDL).
- IoT Plug and Play and the DTDL are open to the community, and Microsoft welcomes collaboration with customers, partners, and the industry. Both are based on open W3C standards such as JSON-LD and RDF, which enables easier adoption across services and tooling.
- There's no extra cost for using IoT Plug and Play and DTDL. Standard rates for Azure IoT Hub and other Azure services remain the same.

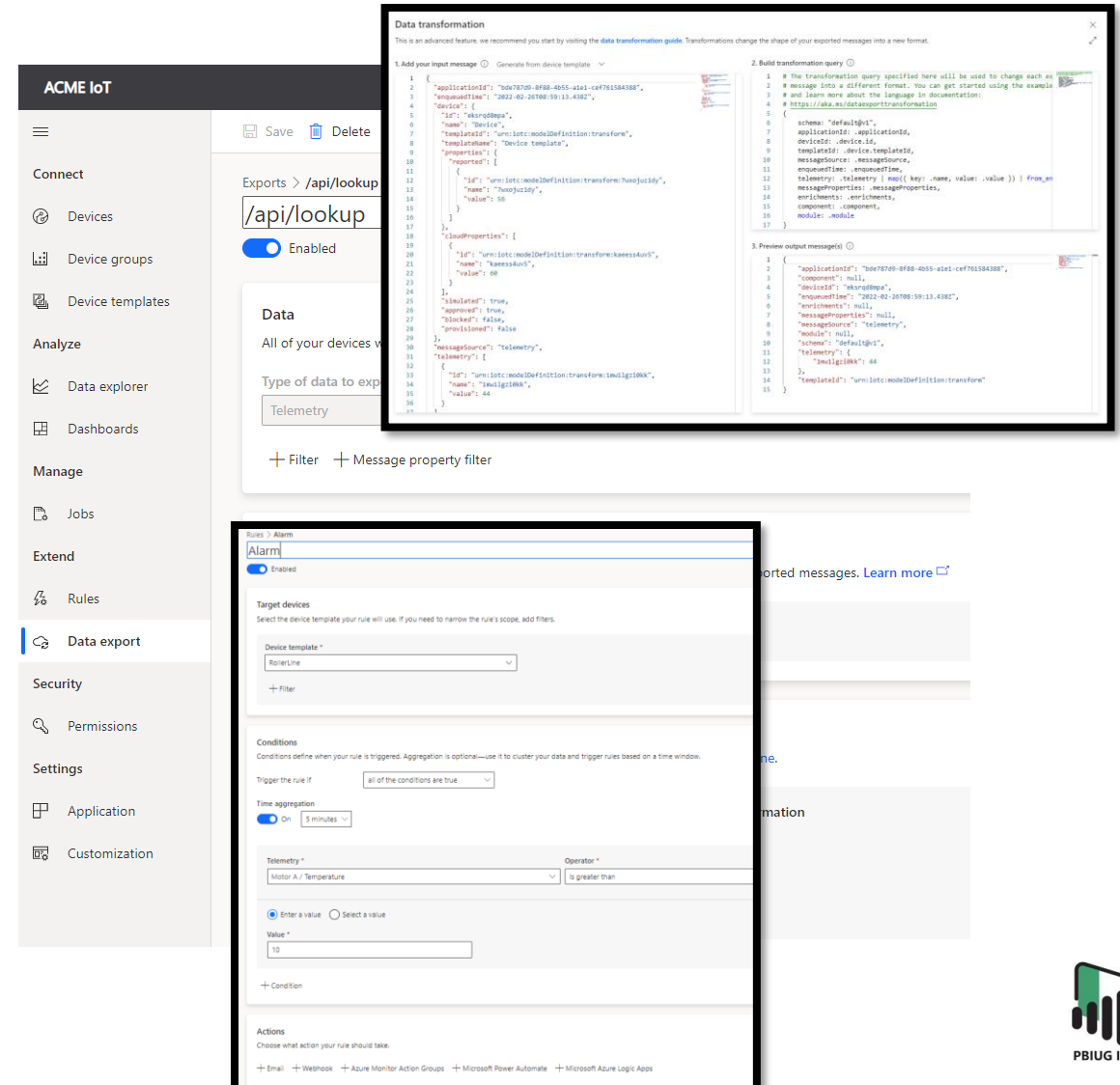
<https://devicecatalog.azure.com/devices/3467200c-3d58-409d-ab9a-b1c603bec01b>

Azure IoT Central REST API

- You can directly consume data from IoT Central without external routes
- All data are accessible from REST Api
- Management of Access tokens from the UI

The integration model of IoT Central

- Built upon the IoT Hub infrastructure
- Advantage from internal data consumption by IoT Central
 - Already include some services you always add
- New from IoT Hub
 - Device Provisioning Service
 - WebHooks
 - Rules (Stream Analytics)
 - REST API to query Central data
 - REST API to manage IoT Central
 - Data Transformation besides enrichment in data routing



Setup Azure IoT Central



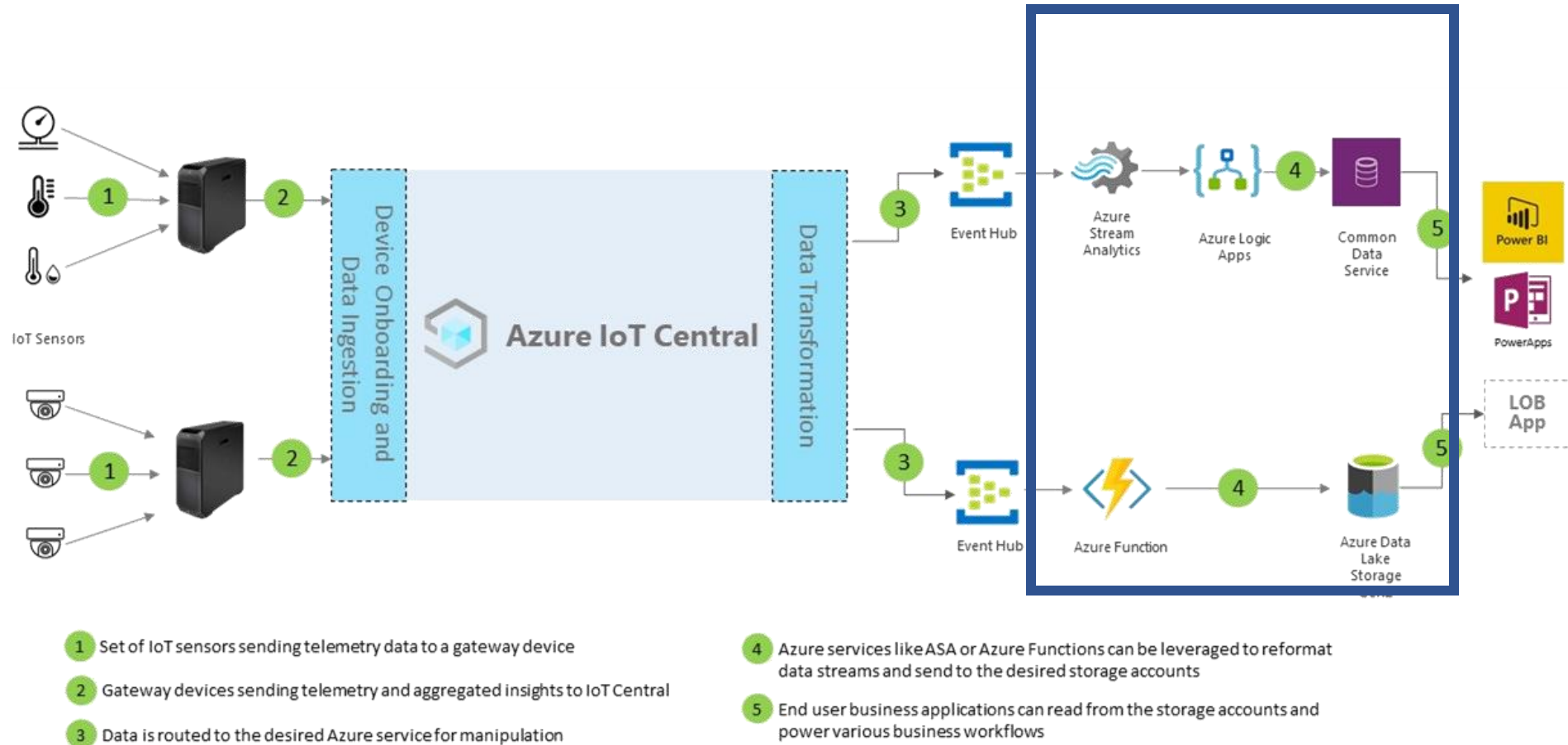
Power BI

4. Power BI Streaming Data Flow



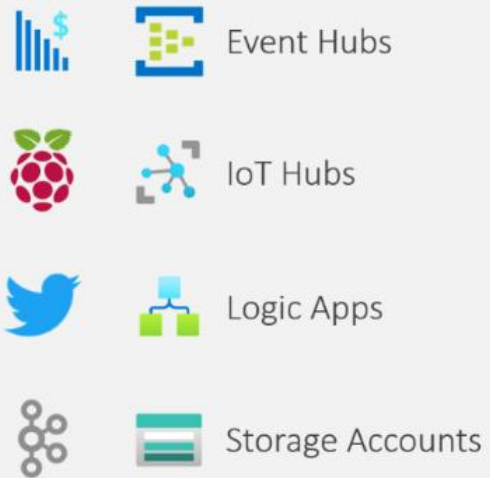
Power BI

Che cosa sono gli streaming dataflow

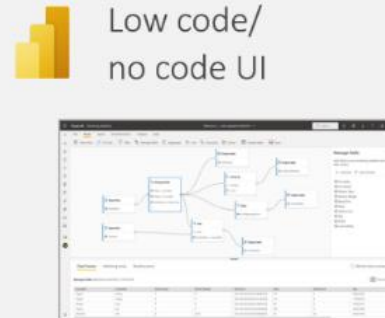


Che cosa sono gli streaming dataflow

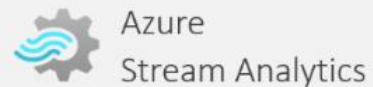
Streaming data sources



Streaming dataflows



Streaming engine



Power BI

Real-time reports



Real-time reports



Che cosa sono gli streaming dataflow

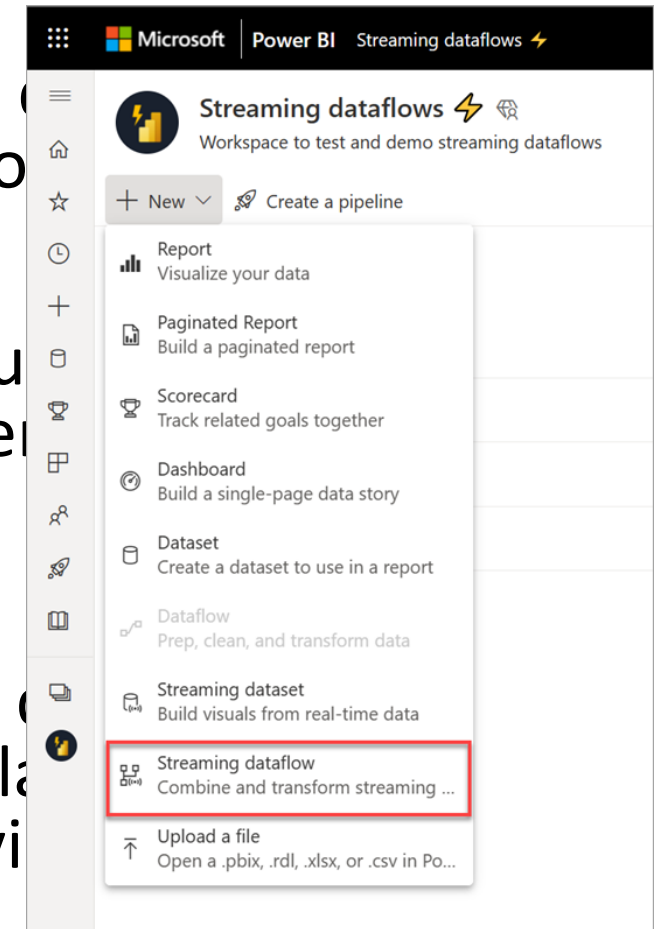
- Il servizio consente esperienze drag-and-drop senza codice.



- streaming dataflow in Power BI consente alle organizzazioni di:
 - Prendere decisioni quasi in tempo reale. Le organizzazioni possono essere più agili e intraprendere azioni significative sulla base delle informazioni più aggiornate.
 - Democratizza i dati in streaming. Le organizzazioni possono rendere i dati più accessibili e più facili da interpretare con una soluzione senza codice riducendo le risorse IT.
 - Accelera il time to insight utilizzando una soluzione di analisi dello streaming end-to-end integrata con l'archiviazione dei dati di BI.

Creazione di uno streaming dataflow

- Puoi aggiungere e modificare tabelle nel streaming direttamente dall'area di lavoro in cui è stato creato
- A causa della natura dei dati in streaming, c'è un flusso continuo di dati in entrata. L'aggiornamento è costante o infinito a meno che non venga interrotto
- Puoi avere un solo tipo di flusso di dati per area di lavoro. Se non disponi già di un flusso di dati regolare nell'area di lavoro, non potrai creare un flusso di dati in streaming (e viceversa)

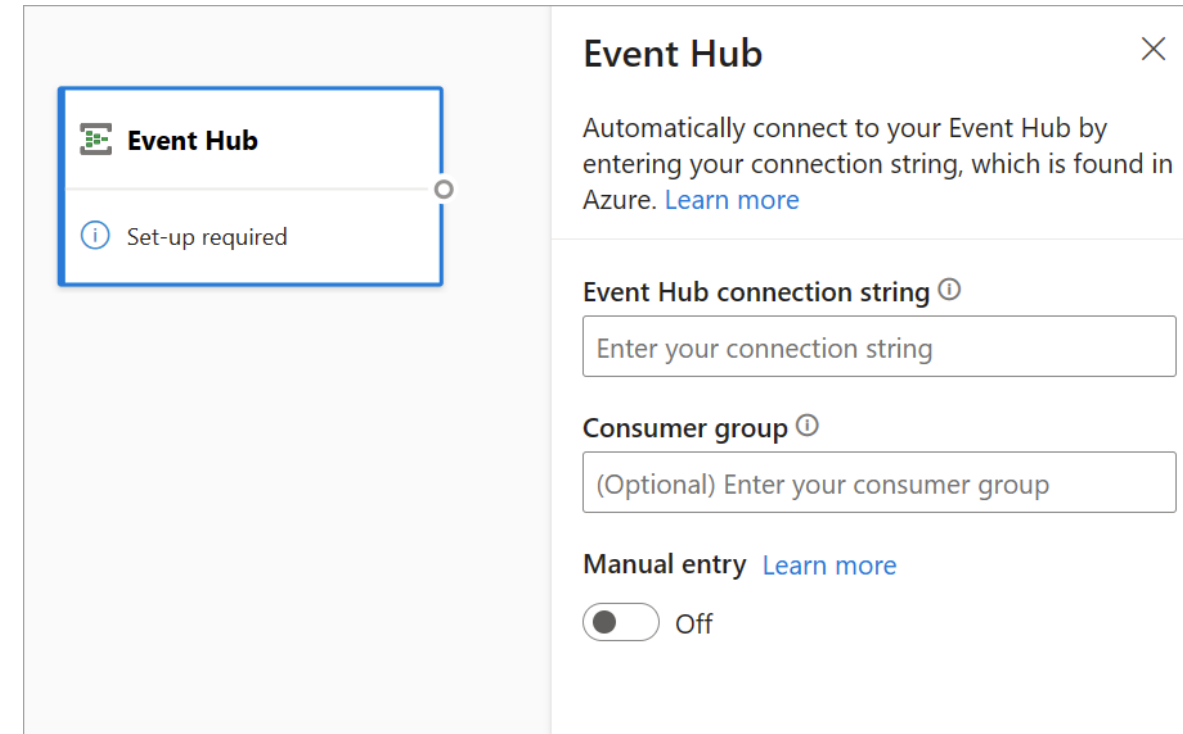


Componenti dello streaming dataflow

- Input
- Trasformazioni
- Output

Input: Event Hub

- Event Hub è una piattaforma di streaming per big data e un servizio di importazione di eventi.
- Può ricevere ed elaborare milioni di eventi al secondo.
- I dati inviati a un hub eventi possono essere trasformati e archiviati utilizzando qualsiasi provider di analisi in tempo reale o adattatori di batch/archiviazione.
- Viene visualizzata una scheda nella vista del diagramma, incluso un riquadro laterale per la sua configurazione.



Event Hub

Automatically connect to your Event Hub by entering your connection string, which is found in Azure. [Learn more](#)

Event Hub connection string ⓘ

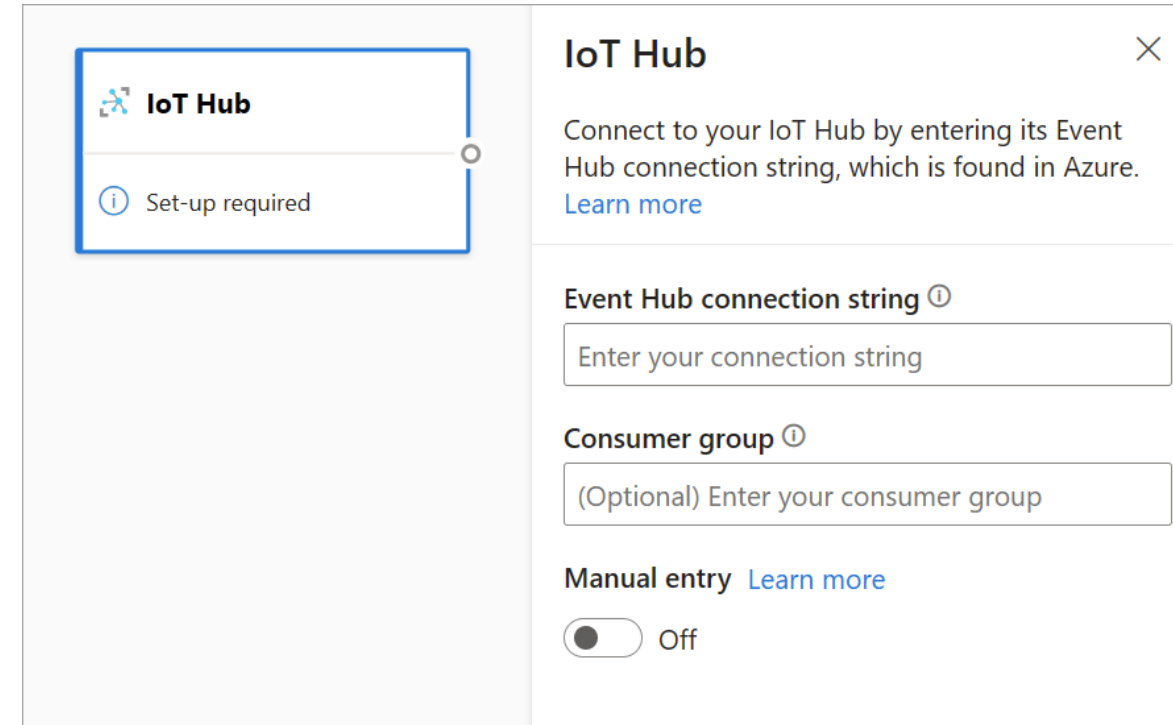
Consumer group ⓘ

Manual entry [Learn more](#)

☐ Off

Input: IoT Hub

- L'hub IoT è un servizio gestito ospitato nel cloud.
- Funge da hub centrale di messaggi per le comunicazioni in entrambe le direzioni tra un'applicazione IoT e i suoi dispositivi collegati.
- Puoi connettere milioni di dispositivi e le relative soluzioni back-end in modo affidabile e sicuro. Quasi tutti i dispositivi possono essere collegati a un hub IoT.



The screenshot shows the 'IoT Hub' configuration page in the Azure portal. On the left, a summary card displays the IoT Hub icon and a 'Set-up required' status. The main panel on the right is titled 'IoT Hub' and contains instructions to connect by entering the Event Hub connection string. It includes input fields for the 'Event Hub connection string' and an optional 'Consumer group'. At the bottom, there is a 'Manual entry' section with a toggle switch currently set to 'Off' and a 'Learn more' link.

IoT Hub ✕

Connect to your IoT Hub by entering its Event Hub connection string, which is found in Azure. [Learn more](#)

Event Hub connection string ⓘ

Enter your connection string

Consumer group ⓘ

(Optional) Enter your consumer group

Manual entry [Learn more](#)

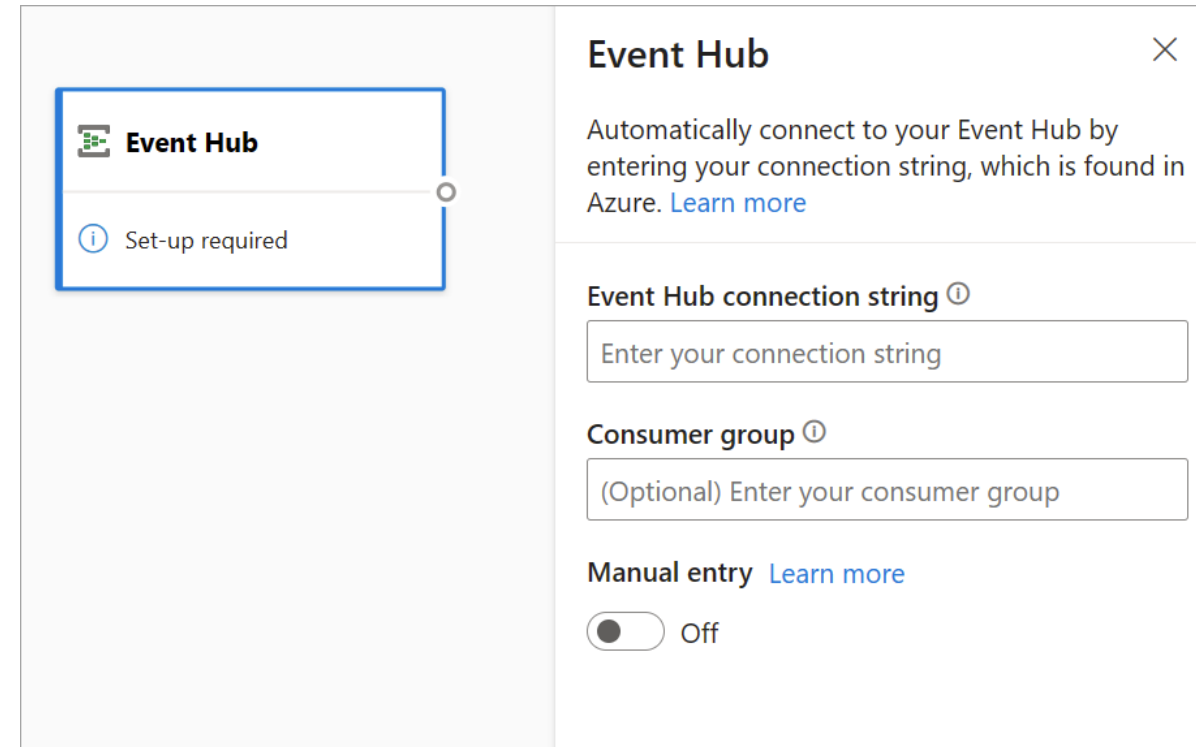
☐ Off

Considerazioni su Event Hub e IoT Hub

- Quando usi i dati di flusso da Hub eventi o Hub IoT, hai accesso ai seguenti campi temporali dei metadati nel flusso di dati di streaming:
- `EventProcessedUtcTime` : la data e l'ora in cui l'evento è stato elaborato.
- `EventEnqueuedUtcTime` : la data e l'ora di ricezione dell'evento.
- Nessuno di questi campi apparirà nell'anteprima di input. Devi aggiungerli manualmente.

Input: Blob Storage

- L'archiviazione BLOB è ottimizzata per l'archiviazione di enormi quantità di dati non strutturati.
- Possiamo usare i BLOB di Azure come input di streaming/riferimento.
- I BLOB in streaming vengono generalmente controllati ogni secondo per streaming. Mentre i BLOB di reference (≤ 50 MB) viene caricato solo all'inizio dell'aggiornamento



Event Hub ✕

Automatically connect to your Event Hub by entering your connection string, which is found in Azure. [Learn more](#)

Event Hub connection string ⓘ

Enter your connection string

Consumer group ⓘ

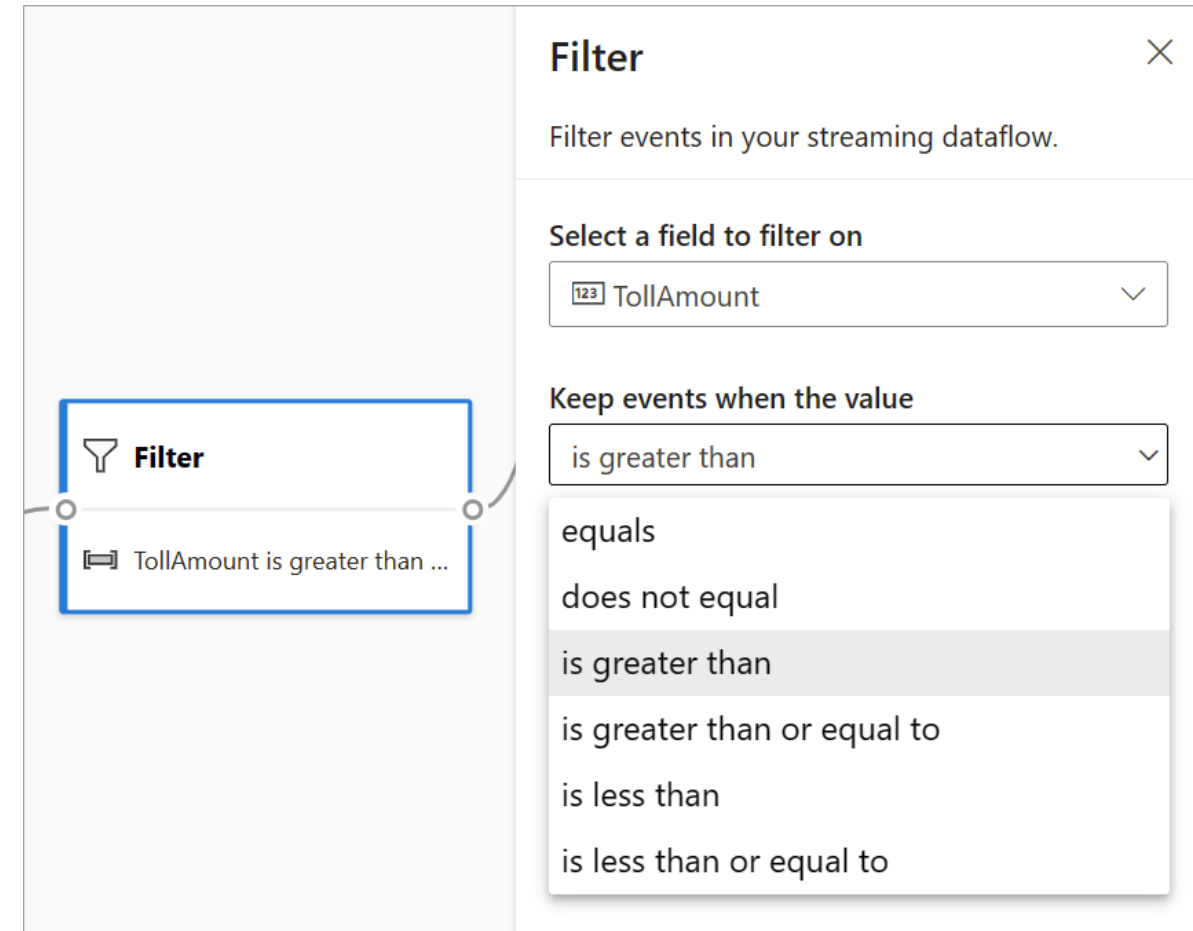
(Optional) Enter your consumer group

Manual entry [Learn more](#)

☐ Off

Trasformazione: Filtro

- Utilizzare la trasformazione Filtro per filtrare gli eventi in base al valore di un campo nell'input.
- A seconda del tipo di dati (numero o testo), la trasformazione manterrà i valori che corrispondono alla condizione selezionata.

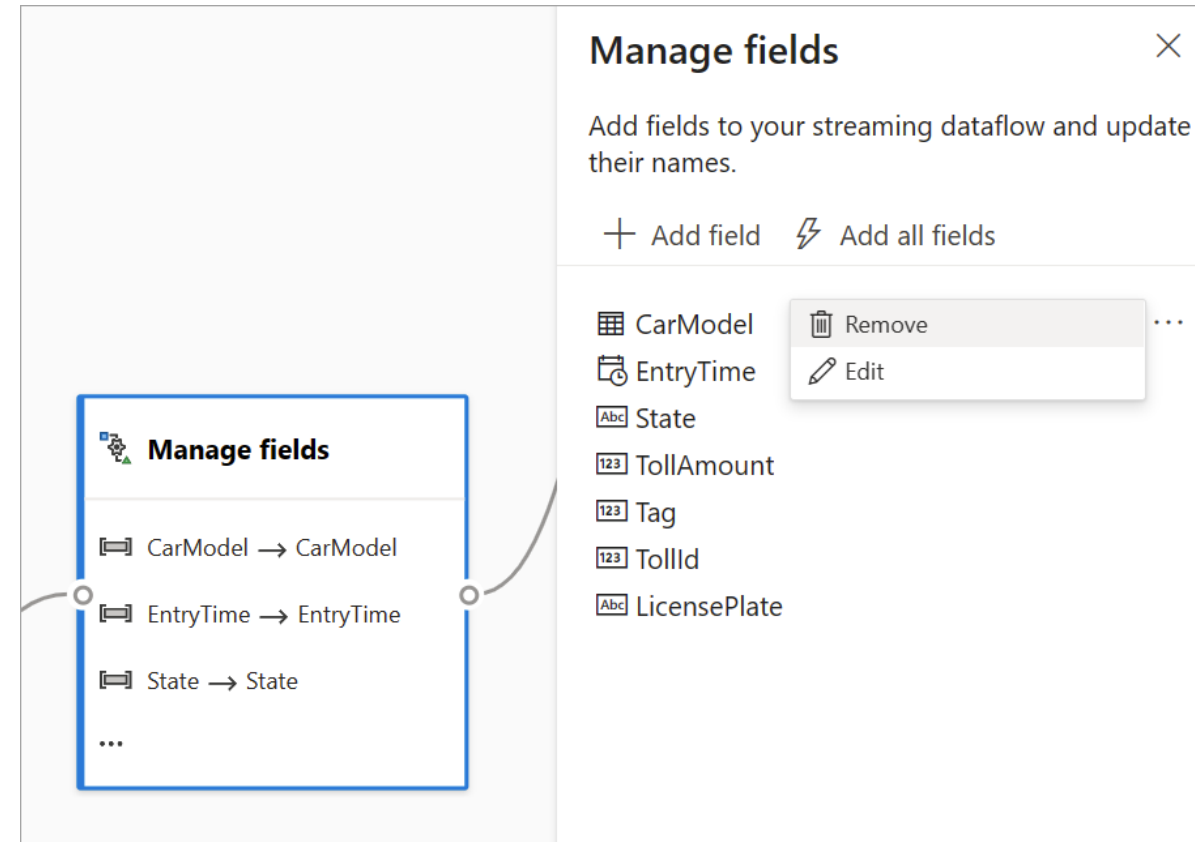


The image shows the configuration window for the 'Filter' transformation in a data processing tool. On the left, a visual representation of the 'Filter' node is shown with a blue border and a funnel icon. Below the icon, the text 'TollAmount is greater than ...' is visible. On the right, the configuration panel is titled 'Filter' and includes a close button (X). The description states: 'Filter events in your streaming dataflow.' The configuration options are as follows:

- Select a field to filter on:** A dropdown menu with 'TollAmount' selected.
- Keep events when the value:** A dropdown menu with 'is greater than' selected. A list of options is shown below the dropdown:
 - equals
 - does not equal
 - is greater than (highlighted)
 - is greater than or equal to
 - is less than
 - is less than or equal to

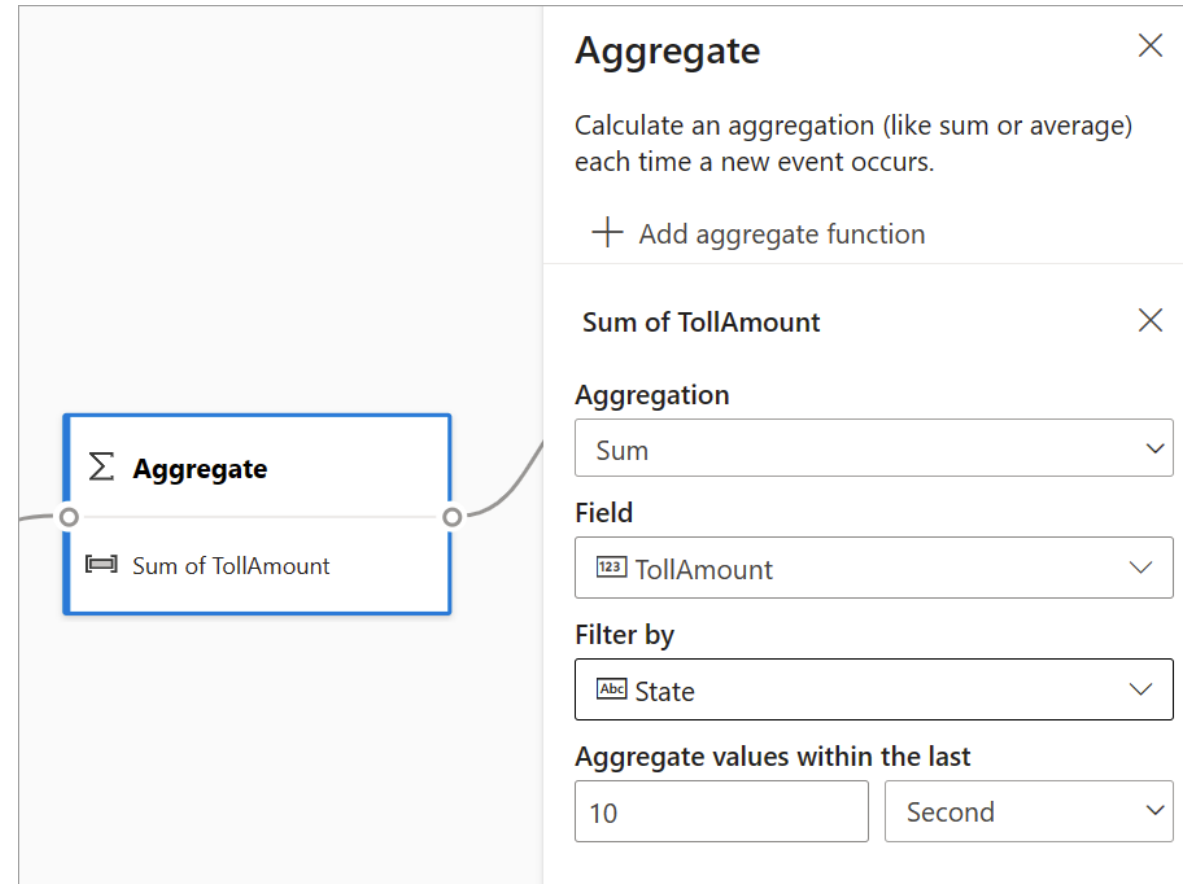
Trasformazione: Gestisci i campi

- La trasformazione Gestisci campi consente di aggiungere, rimuovere o rinominare i campi provenienti da un input o da un'altra trasformazione.
- Le impostazioni nel riquadro laterale ti danno la possibilità di aggiungerne uno nuovo selezionando Aggiungi campo o aggiungendo tutti i campi contemporaneamente.



Trasformazione: Aggregato

- È possibile utilizzare la trasformazione Aggrega per calcolare un'aggregazione (Sum , Minimum , Maximum o Average) ogni volta che si verifica un nuovo evento in un periodo di tempo.
- Puoi avere una o più aggregazioni nella stessa trasformazione.
- Possiamo aggiungere il periodo di tempo durante il quale verrà calcolata l'aggregazione e filtrare per altra dimensione



Aggregate

Calculate an aggregation (like sum or average) each time a new event occurs.

+ Add aggregate function

Sum of TollAmount

Aggregation

Sum

Field

TollAmount

Filter by

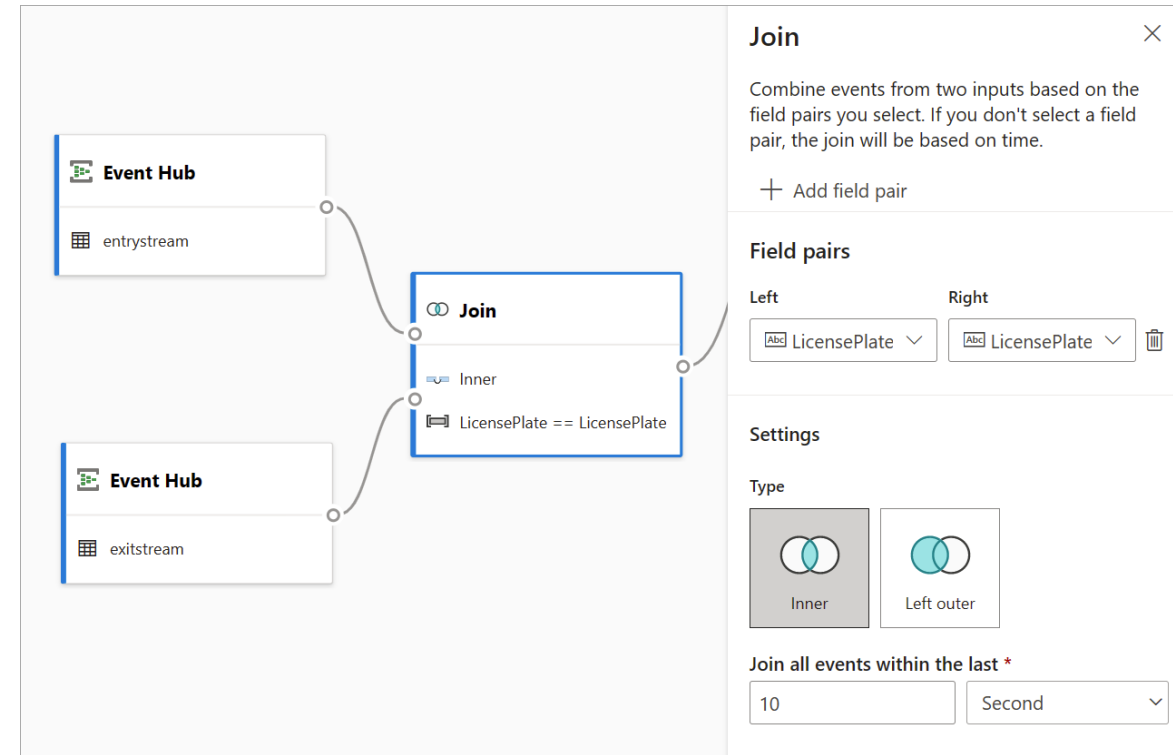
State

Aggregate values within the last

10 Second

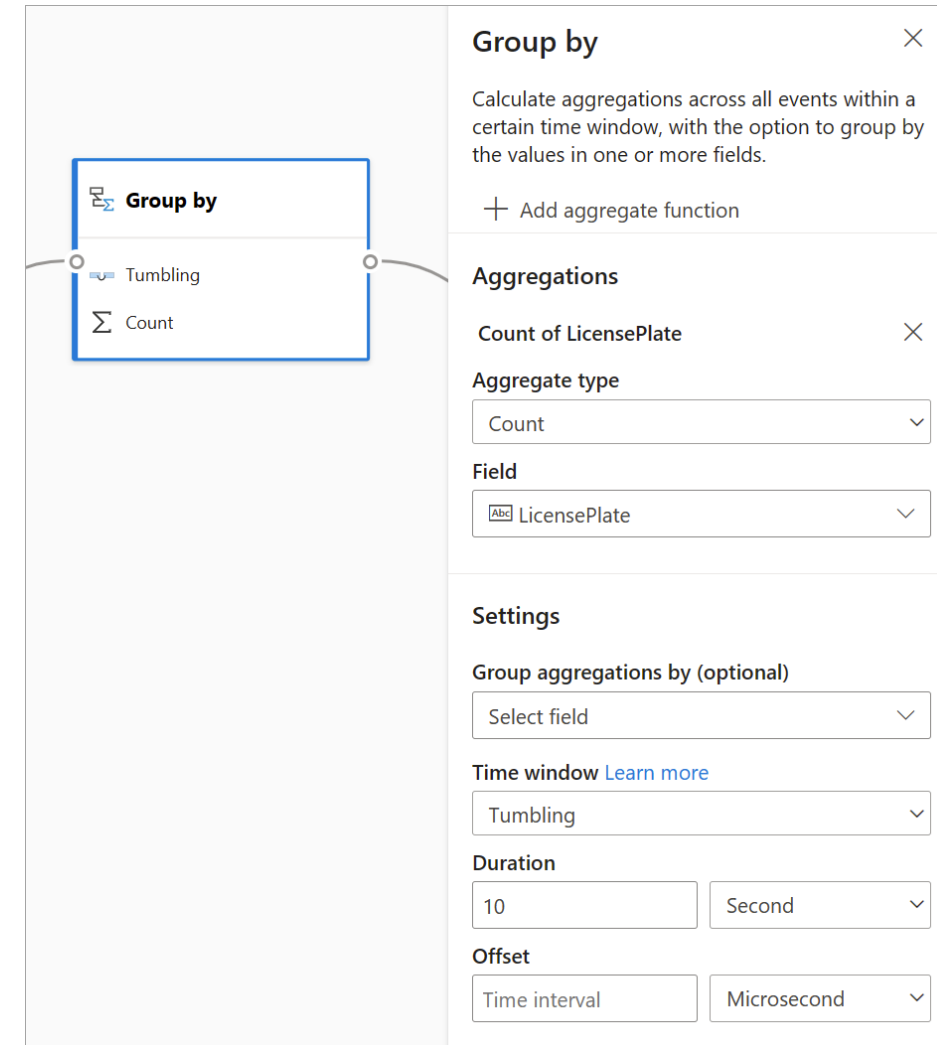
Trasformazione: Aggiungi (Join)

- Usare il Join per combinare eventi da due input in base alle coppie di campi selezionate. Come per i join regolari, hai diverse opzioni per la logica di join:
 - Inner join
 - Left join
- Infine, seleziona in quale periodo di tempo desideri che venga calcolato il join. In questo esempio, il join esamina gli ultimi 10 secondi.
- Per impostazione predefinita, sono inclusi tutti i campi di entrambe le tabelle.



Trasformazione: Raggruppa per

- Utilizzare la trasformazione Raggruppa per per calcolare le aggregazioni di tutti gli eventi arrivati entro un determinato intervallo di tempo.
- Include anche opzioni di finestre temporali più complesse. Puoi aggiungere più di un'aggregazione per trasformazione.
- Le aggregazioni disponibili in questa trasformazione sono: Media , Conteggio , Massimo , Minimo , Percentile (continuo e discreto), Deviazione standard , Somma e Varianza .



The screenshot shows the configuration interface for the 'Group by' transformation. On the left, a visual representation of the transformation is shown with a box labeled 'Group by' containing a 'Tumbling' window icon and a 'Count' aggregation icon. On the right, the configuration panel is divided into three sections: 'Group by', 'Aggregations', and 'Settings'.

Group by

Calculate aggregations across all events within a certain time window, with the option to group by the values in one or more fields.

+ Add aggregate function

Aggregations

Count of LicensePlate

Aggregate type: Count

Field: LicensePlate

Settings

Group aggregations by (optional): Select field

Time window [Learn more](#): Tumbling

Duration: 10 Second

Offset: Time interval Microsecond

Trasformazione: Raggruppa per

- Il tipo di finestra tumbling (cascata) è il tipo più comune temporale. Le caratteristiche chiave delle finestre a tumbling si ripetono, hanno la stessa durata (duration) e non si sovrappongono. Un evento non può appartenere a più di una finestra tumbling.

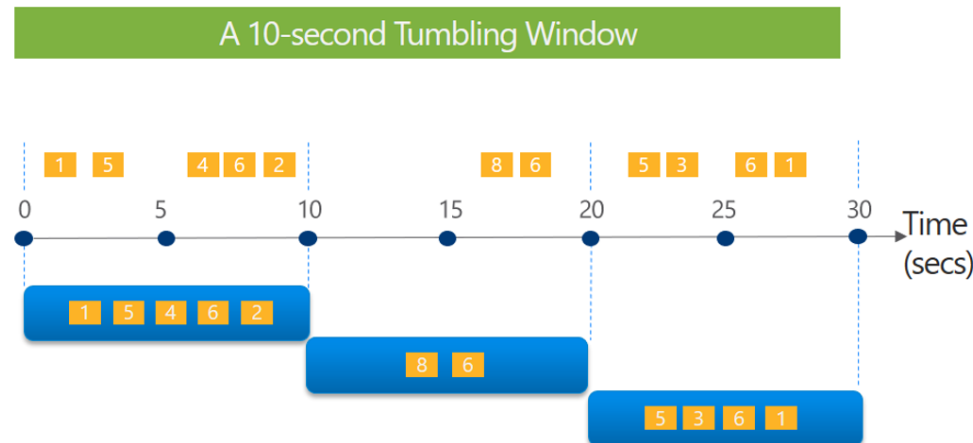
Time window [Learn more](#)

Time window: Tumbling

Duration: 5 Second

Offset: Time interval Microsecond

Tell me the count of Tweets per time zone every 10 seconds



Trasformazione: Raggruppa per

- Le finestre hopping (salto) servono per saltare in avanti un periodo fisso. Puoi pensarli come finestre tumbling sovrapporsi ed essere emesse più spesso delle dimensioni della finestra. Gli eventi possono appartenere a più di una finestra hopping.

Time window [Learn more](#)

Time window ▼

Hopping ▼

Hop size

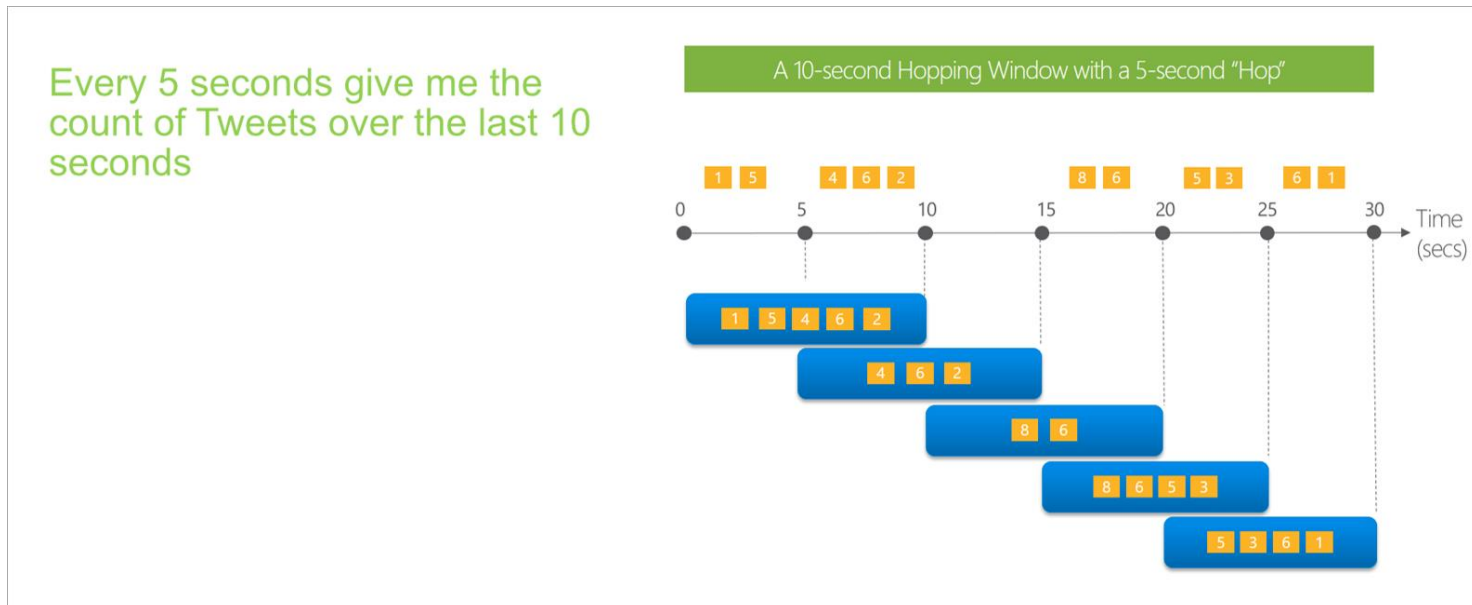
5 ▼ Second ▼

Duration

10 ▼ Second ▼

Offset

0 ▼ Second ▼



Trasformazione: Raggruppa per

- Le finestre sliding (scorrevole), a differenza delle finestre hopping, calcolano l'aggregazione solo per i momenti in cui il contenuto della finestra cambia effettivamente. Quando entra o esce dalla finestra, viene calcolata l'aggregazione. Quindi, ogni finestra ha almeno un evento.

Time window [Learn more](#)

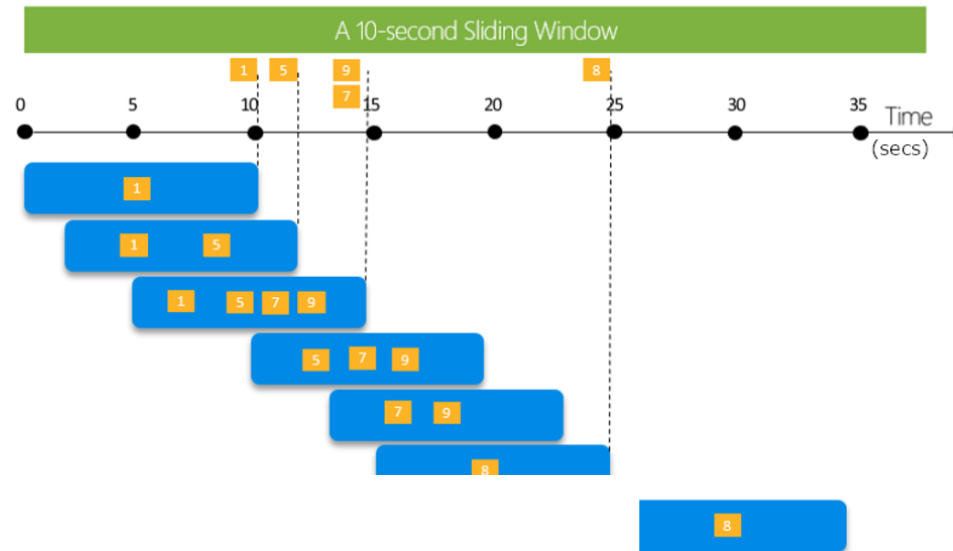
Sliding

Duration

10

Second

Give me the count of Tweets for all topics which are Tweeted more than 10 times in the last 10 seconds



Trasformazione: Raggruppa per

- Le finestre di session (sessione) sono il tipo più complesso. Raggruppano eventi che arrivano in orari specifici e periodi di tempo in cui non ci sono dati. Per questo è necessario fornire:
 - Un timeout: quanto aspettare se non ci sono nuovi dati.
 - Una durata massima: il tempo più lungo in cui verrà calcolato se i dati continuano ad arrivare.

Time window [Learn more](#)

Session

Max duration

10 Minute

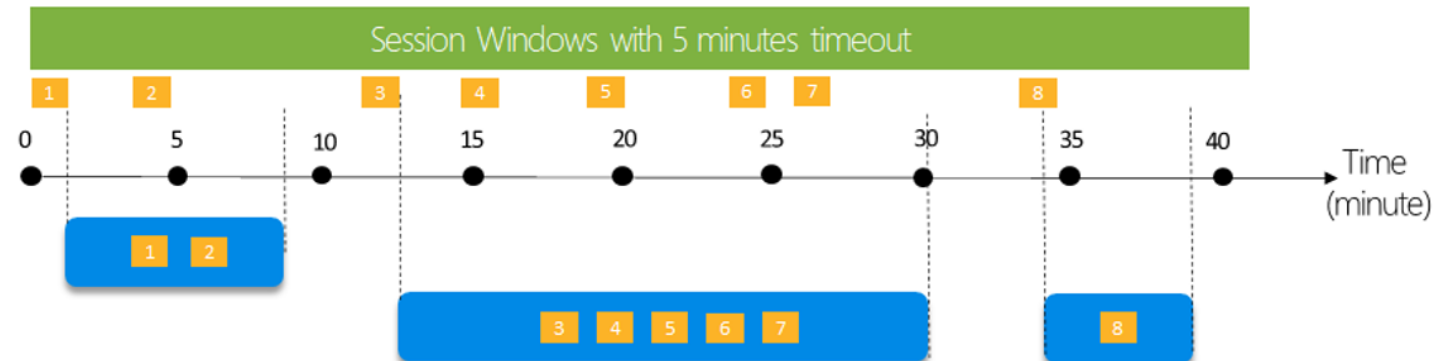
Timeout

5 Minute

Partition by [Learn more](#)

None

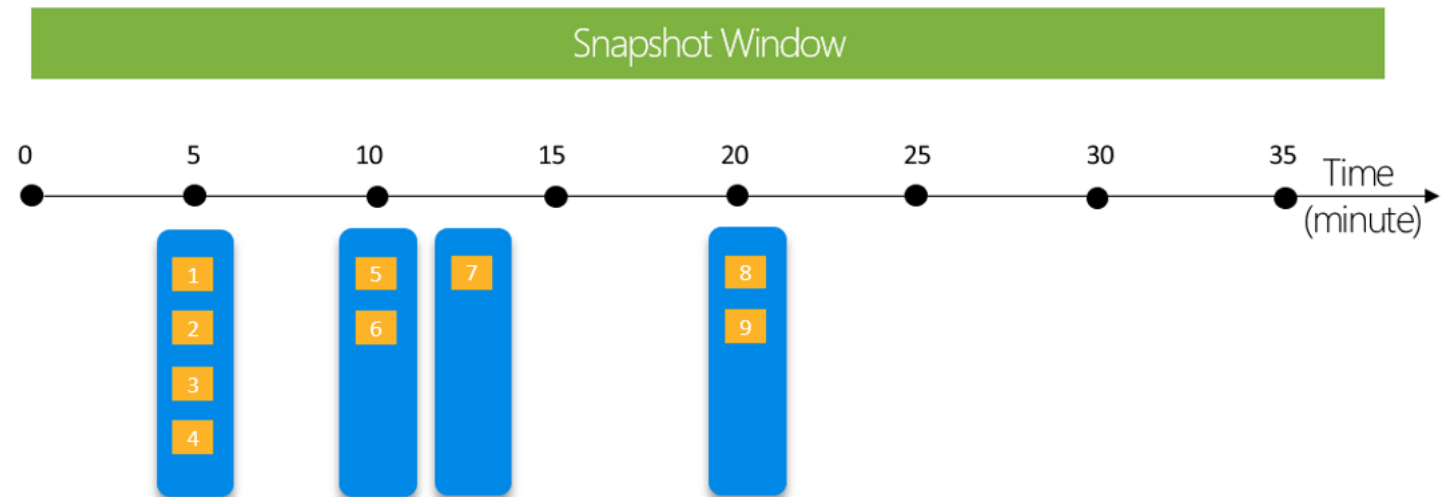
- Tell me the count of Tweets that occur within 5 minutes to each other



Trasformazione: Raggruppa per

- Le finestre di snapshot (istantanea) raggruppano gli eventi che hanno lo stesso timestamp. A differenza di altre finestre, uno snapshot non richiede alcun parametro perché utilizza il tempo del sistema.

- Give me the count of tweets with the same topic type that occur at exactly the same time



Output: Table

- Dopo che sei pronto con input e trasformazioni, è il momento di definire uno o più output. A partire da luglio 2021, i flussi di dati in streaming supportano un solo tipo di output: una tabella Power BI.
- Questo output sarà una tabella del flusso di dati (ovvero un'entità) che puoi usare per creare report in Power BI Desktop. Devi unire i nodi del passaggio precedente con l'output che stai creando per farlo funzionare.








Eseguire un flusso

- Per avviare il flusso di dati in streaming, prima salva il flusso di dati e vai all'area di lavoro in cui lo hai creato. Passa il mouse sopra il streaming dataflow e seleziona il pulsante di riproduzione che appare. Un messaggio pop-up informa che il flusso di dati in

+ New   Create a pipeline

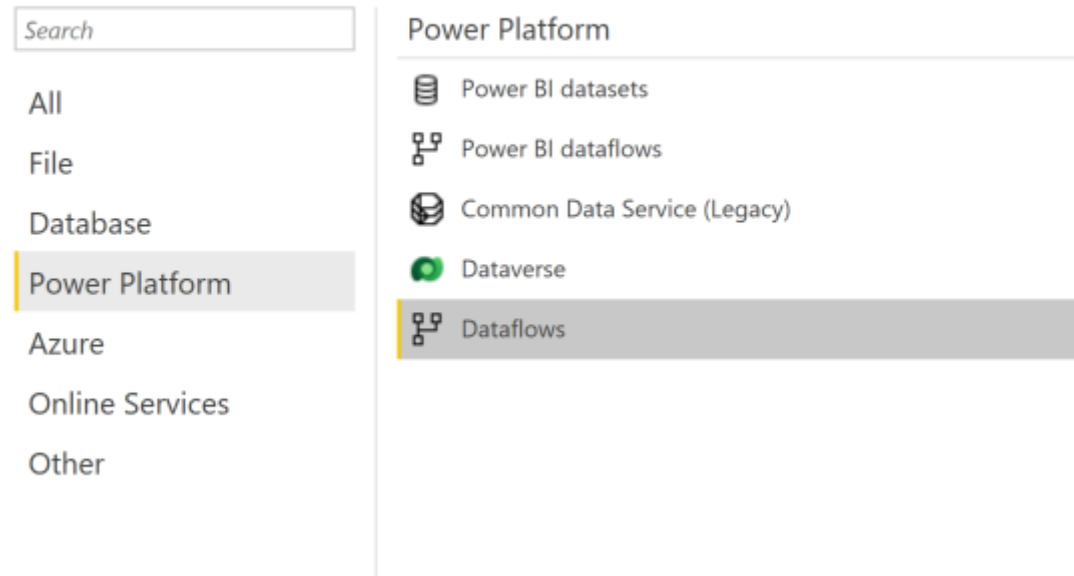
All  Content Datasets + dataflows

	Name		Type	Owner	Refreshed	Next refresh
	StreamingDataflow	  	Streaming dataflow	Miguel Martinez	—	N/A

Connettersi a un streaming dataflow

- Con la versione di luglio 2021 di Power BI Desktop, è disponibile un nuovo connettore denominato Dataflow di dati da usare. Come parte di questo nuovo connettore, per lo streaming dei flussi di dati, vedrai due tabelle che corrispondono all'archiviazione dei dati descritta in precedenza.

Get Data

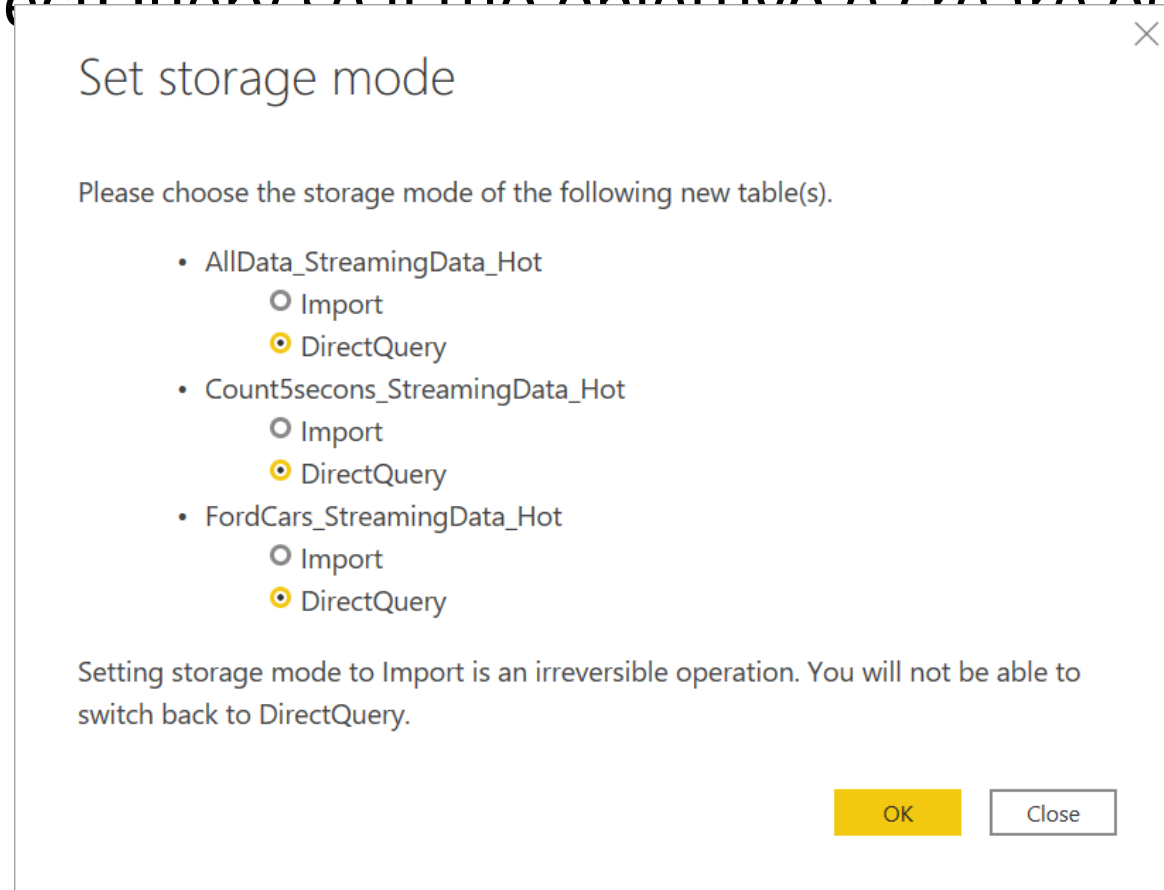


Connettersi a un streaming dataflow

- I Streaming Dataflow salvano i dati nelle due posizioni seguenti. L'uso di queste fonti dipende dal tipo di analisi che stai cercando di fare:
- Hot (analisi in tempo reale) : quando i dati arrivano in Power BI dai Streaming Dataflow, i dati vengono archiviati in una posizione attiva a cui puoi accedere con elementi visivi in tempo reale. La quantità di dati salvata in questo archivio dipende dal valore definito per Durata conservazione nelle impostazioni Streaming Dataflow. Il valore predefinito (e minimo) è 24 ore.
- Cold (analisi storica) : qualsiasi periodo di tempo che non rientra nel periodo definito per Durata di conservazione viene salvato in conservazione a freddo (BLOB) in Power BI per essere utilizzato se necessario. (solo con il connettore dei flussi di dati (Beta) di Power Platform)

Connettersi a un streaming dataflow

- Quando ti viene chiesto di scegliere una modalità di archiviazione, seleziona DirectQuery se il tuo obiettivo è creare oggetti visivi in tempo reale.



Limitazioni

- Per creare ed eseguire streaming dataflow è necessario un abbonamento a Power BI Premium (capacità o PPU).
- È consentito un solo tipo di flusso di dati per area di lavoro (streaming o regolare).
- Non è possibile collegare dataflow regolari e streaming dataflow.
- Capacità inferiori ad A3 non consentono l'uso di streaming dataflow.
- Se i dataflow o il motore di calcolo avanzato non sono abilitati in un tenant, non è possibile creare o eseguire streaming dataflow.
- Le aree di lavoro connesse a un account di archiviazione non sono supportate.
- Ciascun streaming dataflow può fornire fino a 1 megabyte al secondo di throughput.

Licenze

- Il numero di streaming dataflow consentiti per tenant dipende dalla licenza utilizzata:
- Per le capacità Premium, utilizzare la formula seguente per calcolare il numero massimo di streaming dataflow consentiti in una capacità:
 - Numero massimo di flussi di dati in streaming per capacità = vCore nella capacità x 5
 - Ad esempio, P1 ha 8 vCore: $8 * 5 = 40$ flussi di dati in streaming.
- Per Premium per utente, è consentito un flusso di dati in streaming per utente. Se un altro utente desidera utilizzare un flusso di dati in streaming in un'area di lavoro PPU, avrà bisogno anche di una licenza PPU.

Setup Data Flow



Power BI

Anomaly Detection



Power BI

Anomaly Detection

- Anomaly detection is the process of identifying unexpected items or events in data sets, which differ from the norm.
- And anomaly detection is often applied on unlabeled data which is known as unsupervised anomaly detection.

Anomaly Detection in Time Series

- In time series data, an anomaly or outlier can be termed as a data point which is not following the common collective trend or seasonal or cyclic pattern of the entire data and is significantly distinct from rest of the data. By significant, most data scientists mean statistical significance, which in order words, signify that the statistical properties of the data point is not in alignment with the rest of the series.
- Anomaly detection has two basic assumptions:
- Anomalies only occur very rarely in the data.
- Their features differ from the normal instances significantly.

Many Opportunities in Azure

- Programmatic:
 - ML.NET (in .NET space)
 - All Python space
- Data Platform
 - Spark and DataBricks
 - Azure Data Explorer
- All these solutions are in the PaaS space
- Any option in SaaS space?

Azure Cognitive Services

- Cognitive Services brings AI within reach of every developer—without requiring machine-learning expertise. All it takes is an API call to embed the ability to see, hear, speak, search, understand, and accelerate decision-making into your apps. Enable developers of all skill levels to easily add AI capabilities to their apps.
- Five areas:
 - **Anomaly Detector**
Identify potential problems early on.
 - **Decision**
 - **Content Moderator**
Detect potentially offensive or unwanted content.
 - **Language**
 - **Metrics Advisor PREVIEW**
Monitor metrics and diagnose issues.
 - **Speech**
 - **Vision**
 - **Personalizer**
Create rich, personalized experiences for every user.
 - **Web search**

Anomaly Detector

- Through an API, Anomaly Detector ingests time-series data of all types and selects the best-fitting detection model for your data to ensure high accuracy. Customize the service to detect any level of anomaly and deploy it where you need it most -- from the cloud to the intelligent edge with containers. Azure is the only major cloud provider that offers anomaly detection as an AI service.

It seems too much simple...☹️

Anomaly Detector Features

- Online sample
 - <https://algoevaluation.azurewebsites.net/>
- Notebooks
- .NET
- REST API
- Univariate and Multivariate analysis
- Univariate analysis
 - LastPoint
 - Batch
 - Trend
- Multivariate analysis

Invoke Anomaly detection API



Power BI

Invocare Anomaly detection da power query - demo



Conclusioni



Power BI

Power BI Streaming DataFlow and Azure IoT Central

- Nome e cognome
- email