



Introduction to Data Stream Processing



Meet the Presenters



Stewart Harper



Don Murray

Agenda

- Market trends
- Overview of data streams
- Overview of stream processing
- Cloud computing and stream processing

Market Trends



It's time to adopt Stream Data Integration

- The global streaming analytics market size is expected to grow from USD 12.5 billion in 2020 to **USD 38.6 billion by 2025**.
- More than half of major new business systems will incorporate **real-time data** and **continuous intelligence** to improve decisions.
- More and more organizations will use **streaming data** to support their data integration use cases.

A photograph showing four people working together at a long wooden table in a modern office or co-working space. They are looking at laptops and discussing their work. The office has large windows and a bright, open-plan layout.

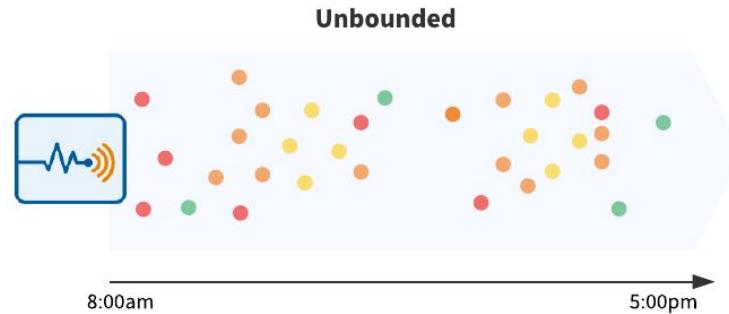
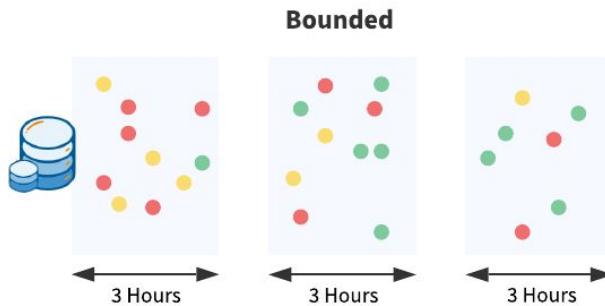
Why are more businesses leveraging stream processing?

- Explosive growth of Sensors and hardware that collect data.
- Technology to capture, store, and process data continues to improve.
- Actionable information delivered faster to decision makers is a competitive advantage.

Overview of Data Streams

Let's start with the data

- **Bounded data** is finite and has a discrete beginning and end. It is associated with batch processing.
- **Unbounded data** is infinite, having no discrete beginning or end. It is associated with stream processing.

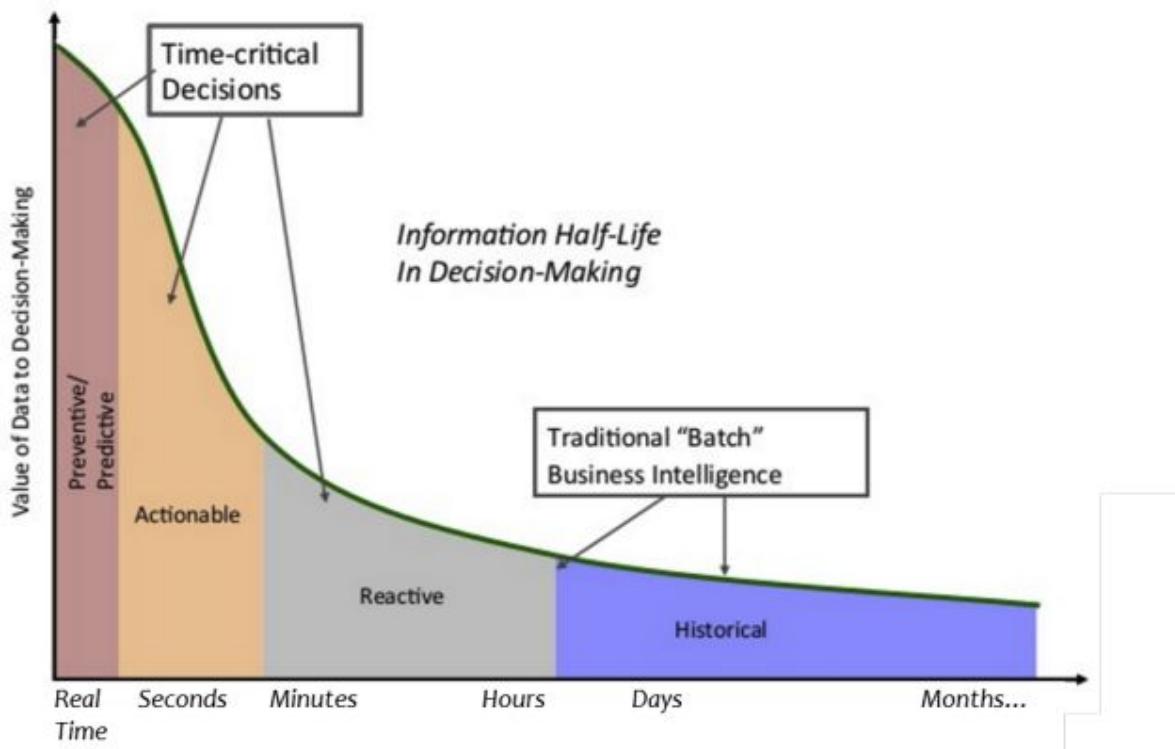


Characteristics of Data Streams

- Data records are **small in size**.
- Data volumes can be **extremely high**.
- Data distribution can be **inconsistent** with quiet and busy periods.
- Data can **arrive out of sequence** compared to when the event happened.



Diminishing Value of Data



Source: Amazon Web Services



What are the sources of Data Streams?

Common Event Streams

Business Applications

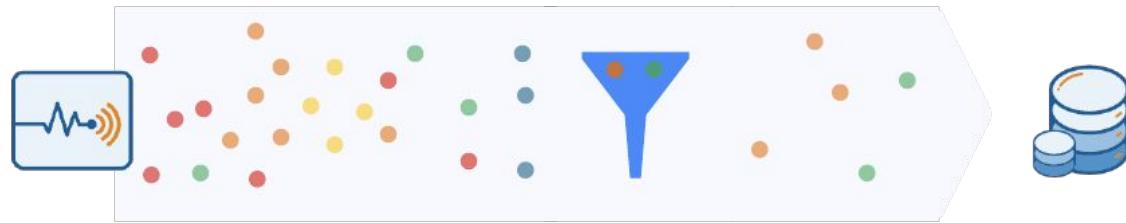
- Customer orders
- Airline reservations
- Insurance claims
- Bank transactions
- Telco call detail records

Digital Information

- Clickstreams
- Social computing
- Customer call logs
- News, weather feeds
- IT, network logs
- Market data
- Email

Internet of Things

- Radio-frequency Identification
- Telemetry SCADA
- Geolocation
- Machine logs





Industries benefiting from streaming data:

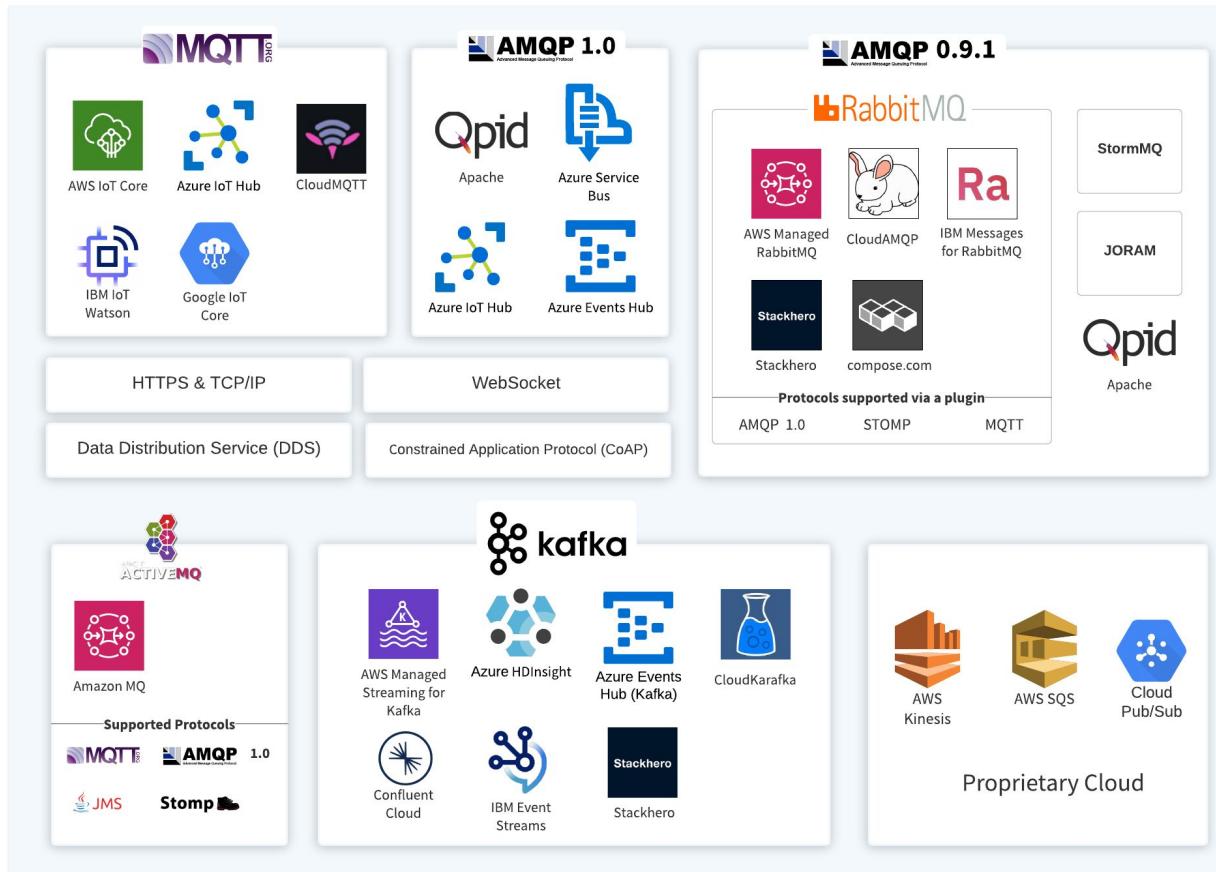
- Utilities
- Energy
- Telecommunications
- Transportation
- Commercial
- Government
- Healthcare
- ...



Business Improvements from Steaming Data:

- Operational efficiency
- Asset management
- Customer experience
- Situational awareness
- Medical care

Data Streaming Technologies



Poll Question:
**Is your organization leveraging or
planning to leverage streaming data?**

Overview of Stream Processing

What is Stream Processing?

Stream processing is a term that groups together the collection, integration, and analysis of unbounded data.

Stream Processing delivers insights to organizations on a continuous basis.

High and Low Volume Message Streams

There are three ways organizations work with unbounded data.

1. **Batch Processing of Stored Data** - Unbounded data is stored and processed at specified intervals.
2. **Event Processing** - Each event in the unbounded stream is handled separately with connections between events being stored in persistent storage.
3. **Stream Processing** - Continuous processing of high-volume data streams. Data is processed in memory before storing.

Near Real-time

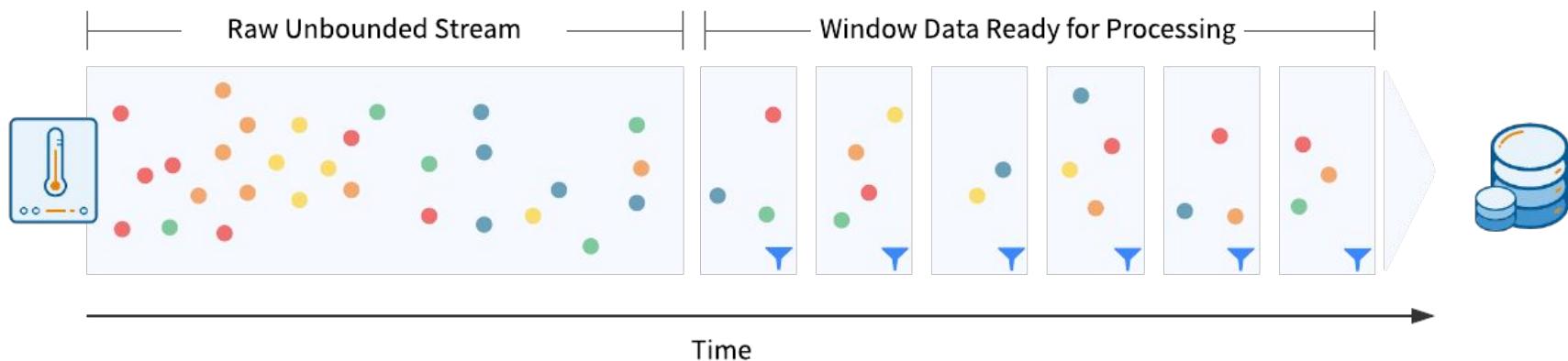
Real-Time Low Velocity

Real-Time Any Velocity

Batch Processing vs Streaming Processing

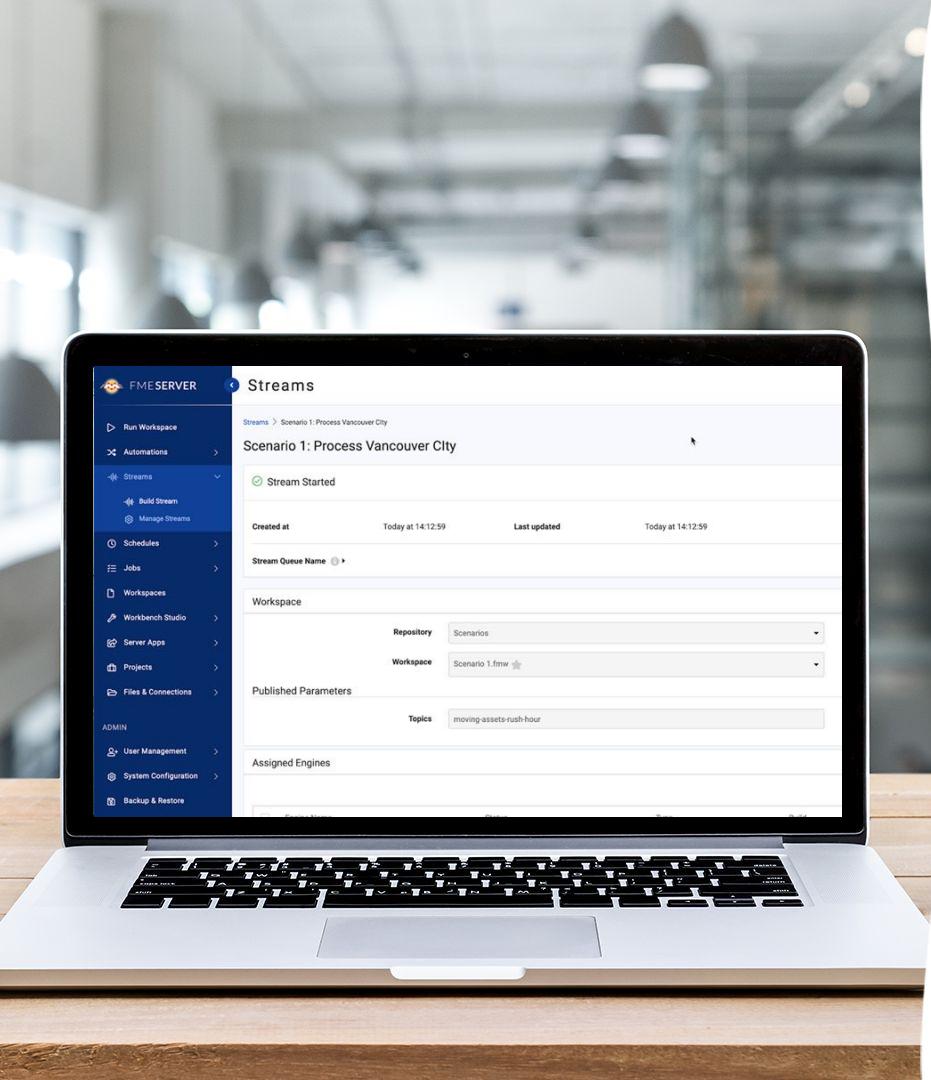
	Batch Processing	Stream Processing
Frequency	Infrequent jobs that produce results once the job has finished running.	Continuously running jobs that produce constant results.
Performance	High latency (minutes to hours).	Low latency (seconds to milliseconds).
Data Sources	Databases, APIs, and static files.	Message queues, event streams, and transactions.
Analysis Type	Complex analysis is possible.	Simpler analysis, including aggregation, filtering, enrichment, proximity analysis, and event detection.
Processing	Process after storing the data.	Process and then maybe store the data.

Breaking Up Data Streams for Processing



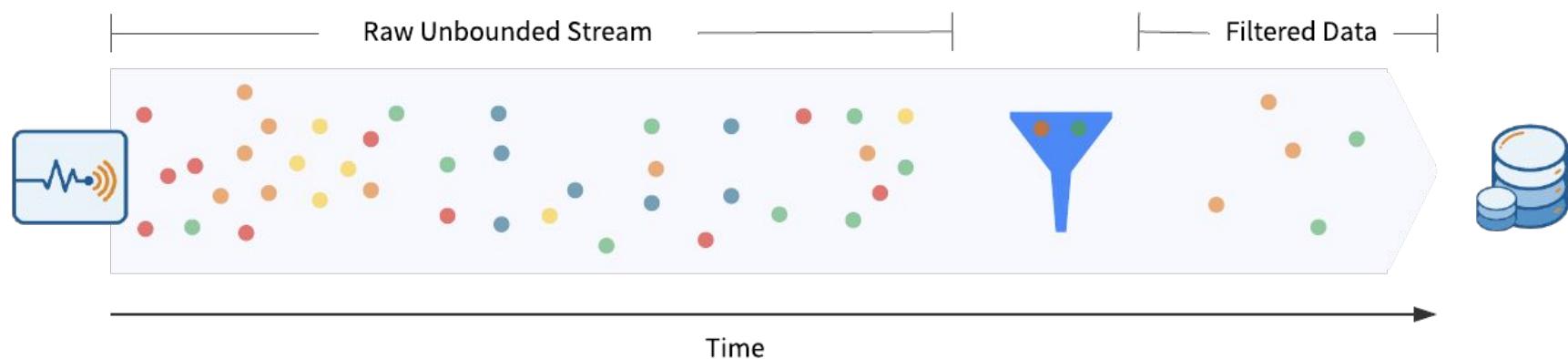
Overview of Core Stream Processing Workflows

- Filtering
- Enriching
- Aggregating
- Event detection



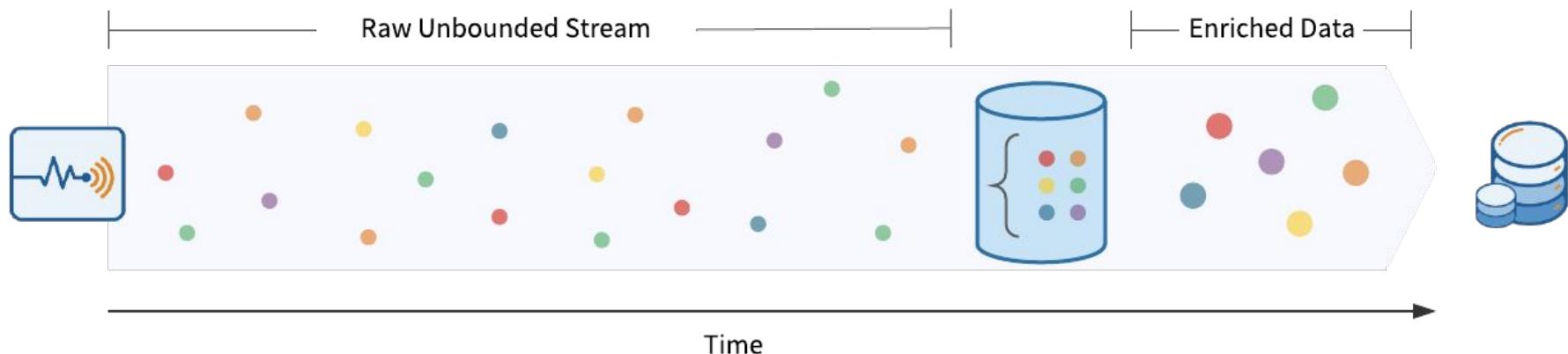
Filtering

Reduce data volumes in memory before committing data to disk.



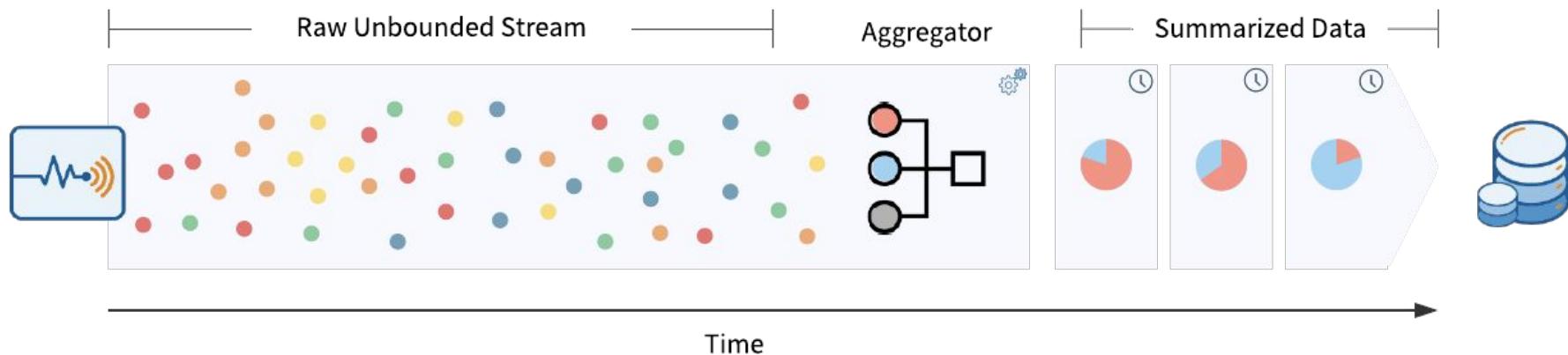
Enriching

Join the unbounded data to other datasets (databases, APIs) before committing data to disk.



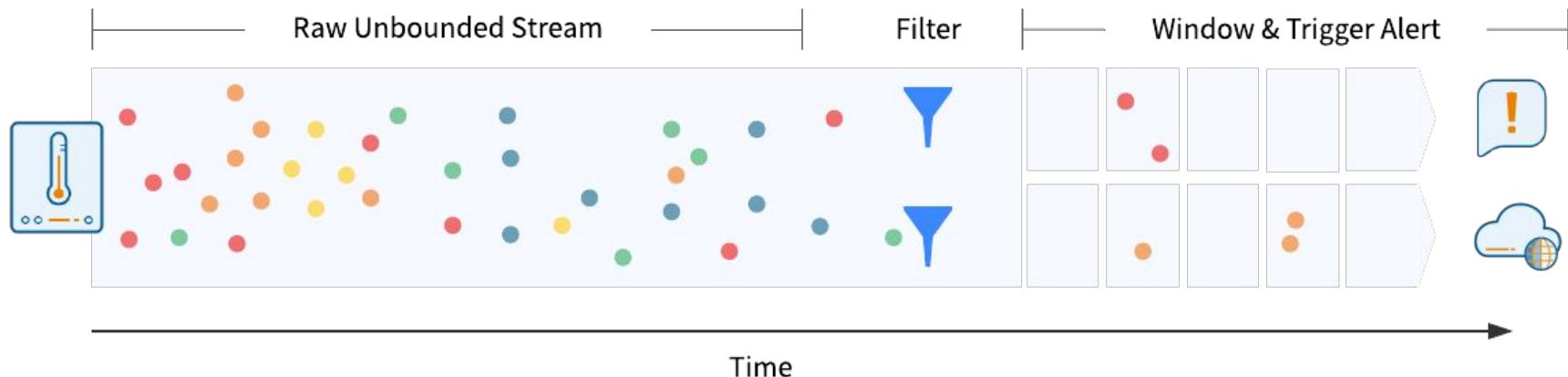
Aggregating

Summarize the unbounded data by calculating time-windowed aggregations before committing data to disk.



Event Detection

Detect patterns in memory and then trigger an event when certain criteria are met.

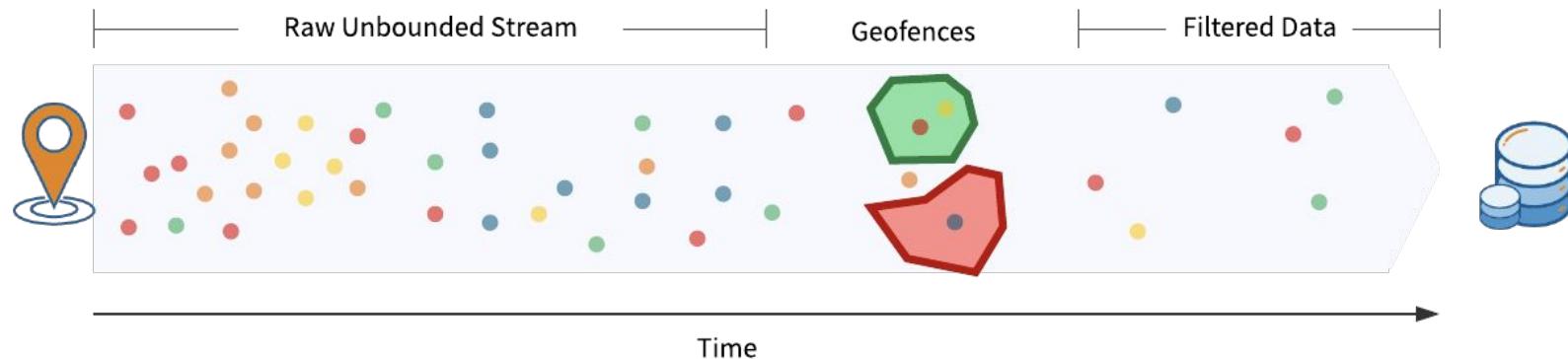


Geospatial Data and Analysis in Stream Processing

*Leveraging real-time geospatial or location data expands
the insights delivered to your organization.*

Geospatial Processing

- Filtering with Geofences
- Proximity Analysis
- Snapping Data to a network
- Calculate Distance



The Role of Cloud Computing in Stream Processing



Cloud Services Related to Stream Processing

- Data Warehouses
- Data Lakes
- Stream Processing
- IoT Device Connection
- Machine Learning
- Scalable Compute



Summary

Stream Processing delivers **real-time insights to organizations.**

There is no better time to get started than now!

- Sensors and data collection are exploding.
- Technologies continue to improve.
- All industries can benefit.
- Customer satisfaction will skyrocket.

See you on April 8th for part 2.

Empowering Real-Time Decision Making with Data Streaming

- Common use cases
- Customer Stories
- How FME supports streaming data
- Demos

Register: safe.com/webinars

FME World Fair 2021

Come one, come all from May 4 - 14! Discover tips & tricks, the newest updates, and innovative ways to use FME.

Register: safe.com/world-fair





Thank You!

Read our Data Streaming blog
at safe.com/blog

Connect with us for more FME

