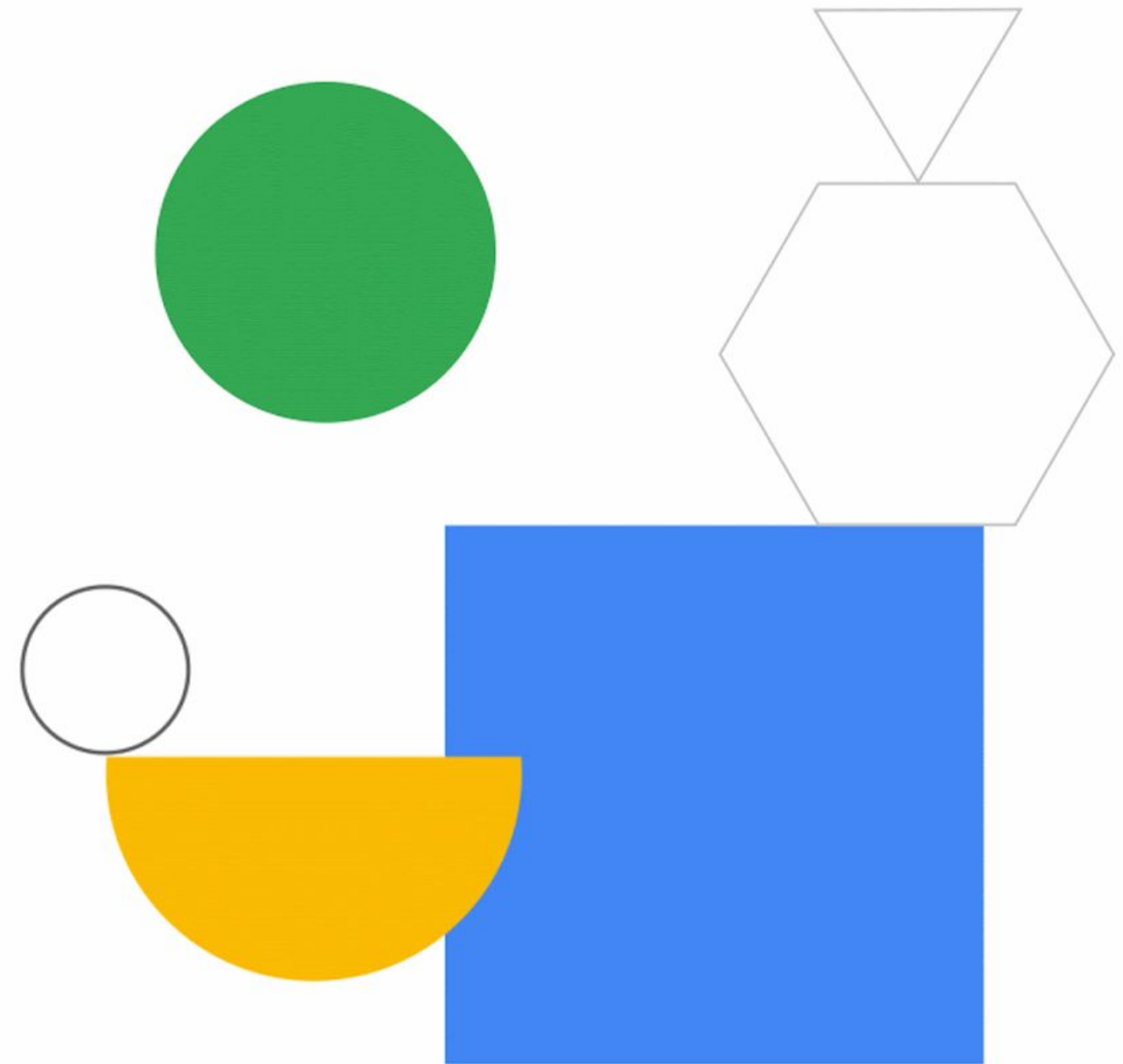


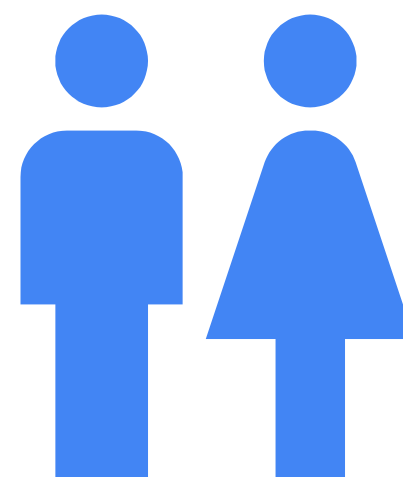
Building Resilient Streaming Analytics Systems on Google Cloud



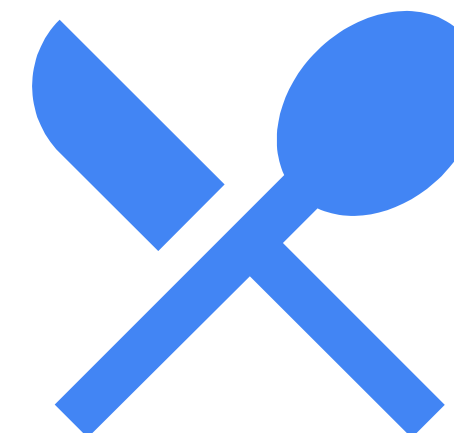
Facilities



Parking

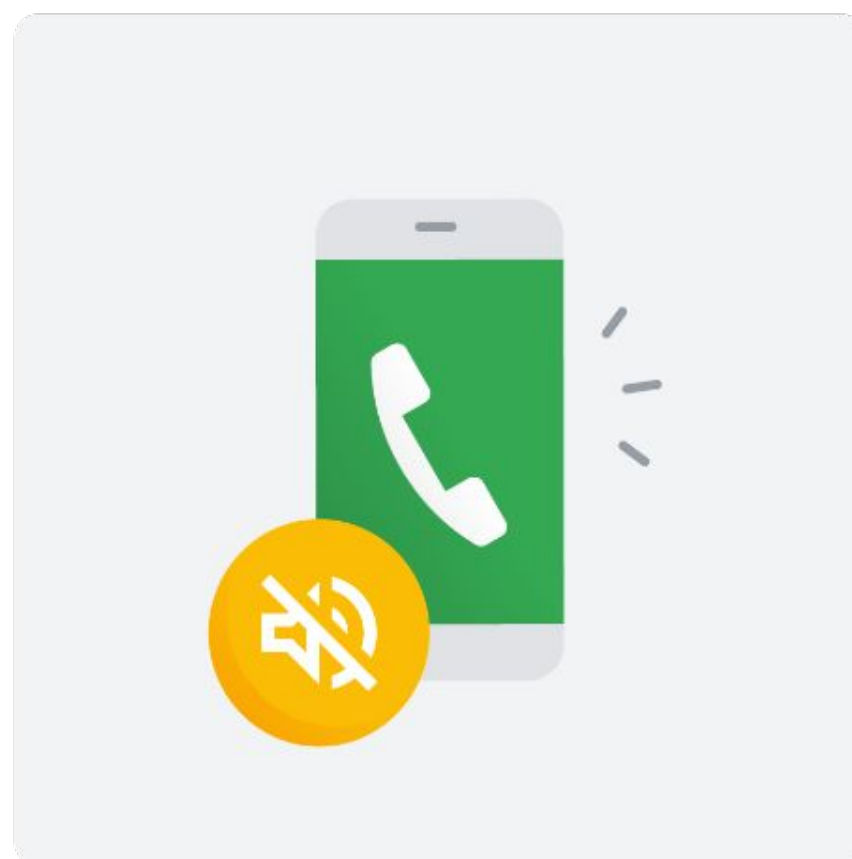


Facilities

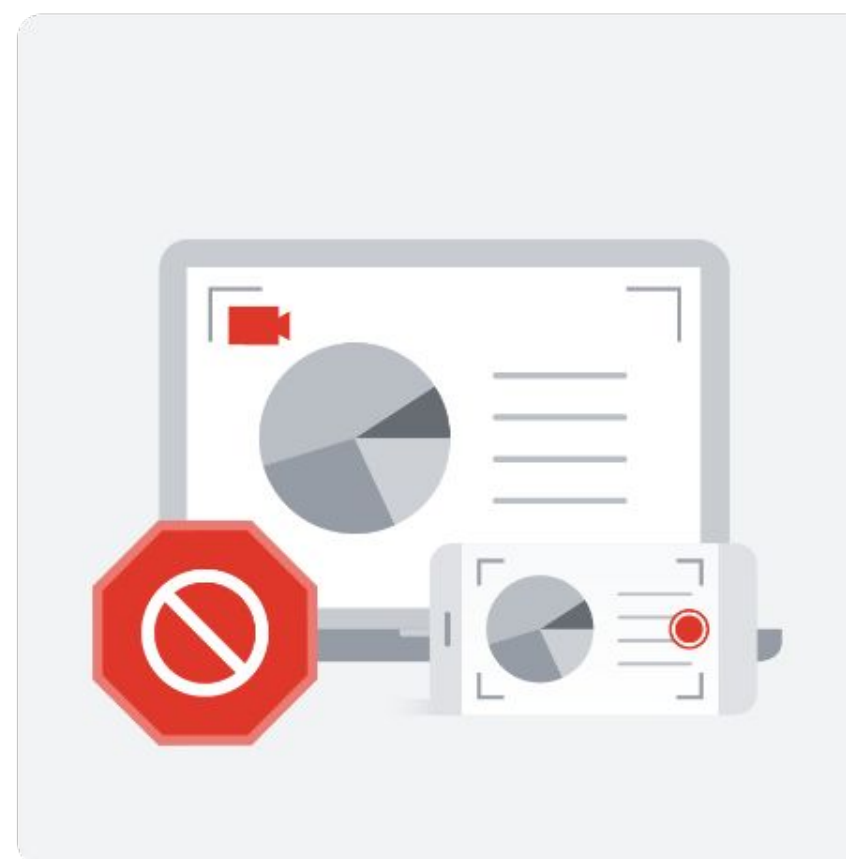


Food

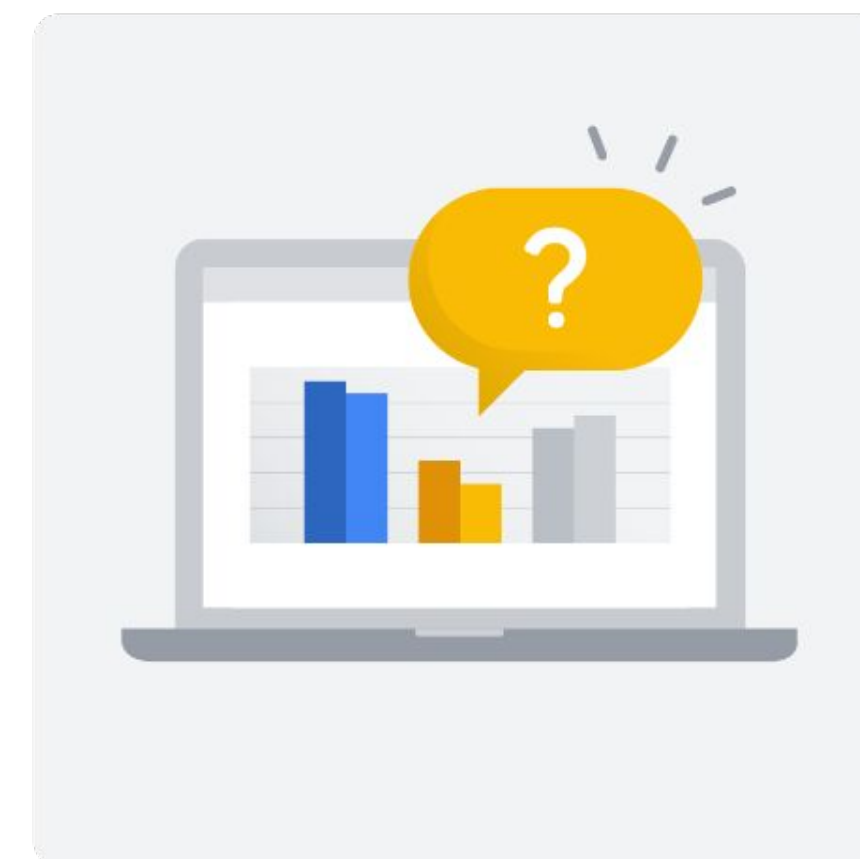
Etiquette



No calls



No recording



Ask questions

Data Engineering on Google Cloud learning path



Modernizing Data Lakes and Data Warehouses with Google Cloud



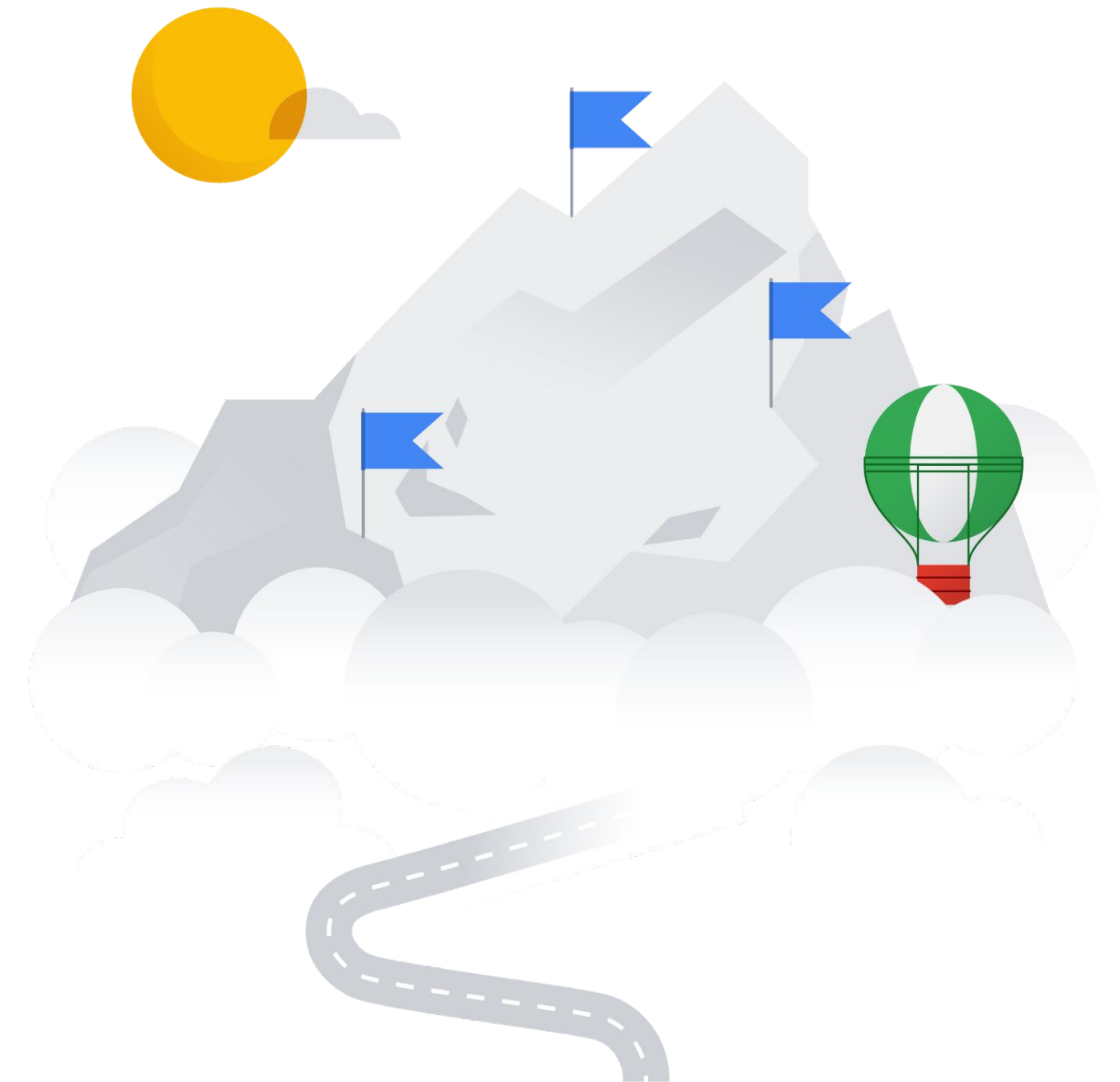
Building Batch Data Pipelines on Google Cloud



Building Resilient Streaming Analytics Systems on Google Cloud




Smart Analytics, Machine Learning and AI on Google Cloud





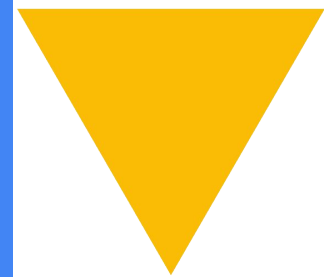
Day 3 agenda



- 01 Introduction to Processing Streaming Data
- 02 Serverless Messaging with Pub/Sub
- 03 Dataflow Streaming Features
- 04 High-Throughput BigQuery and Bigtable Streaming Features
- 05 Advanced BigQuery Functionality and Performance

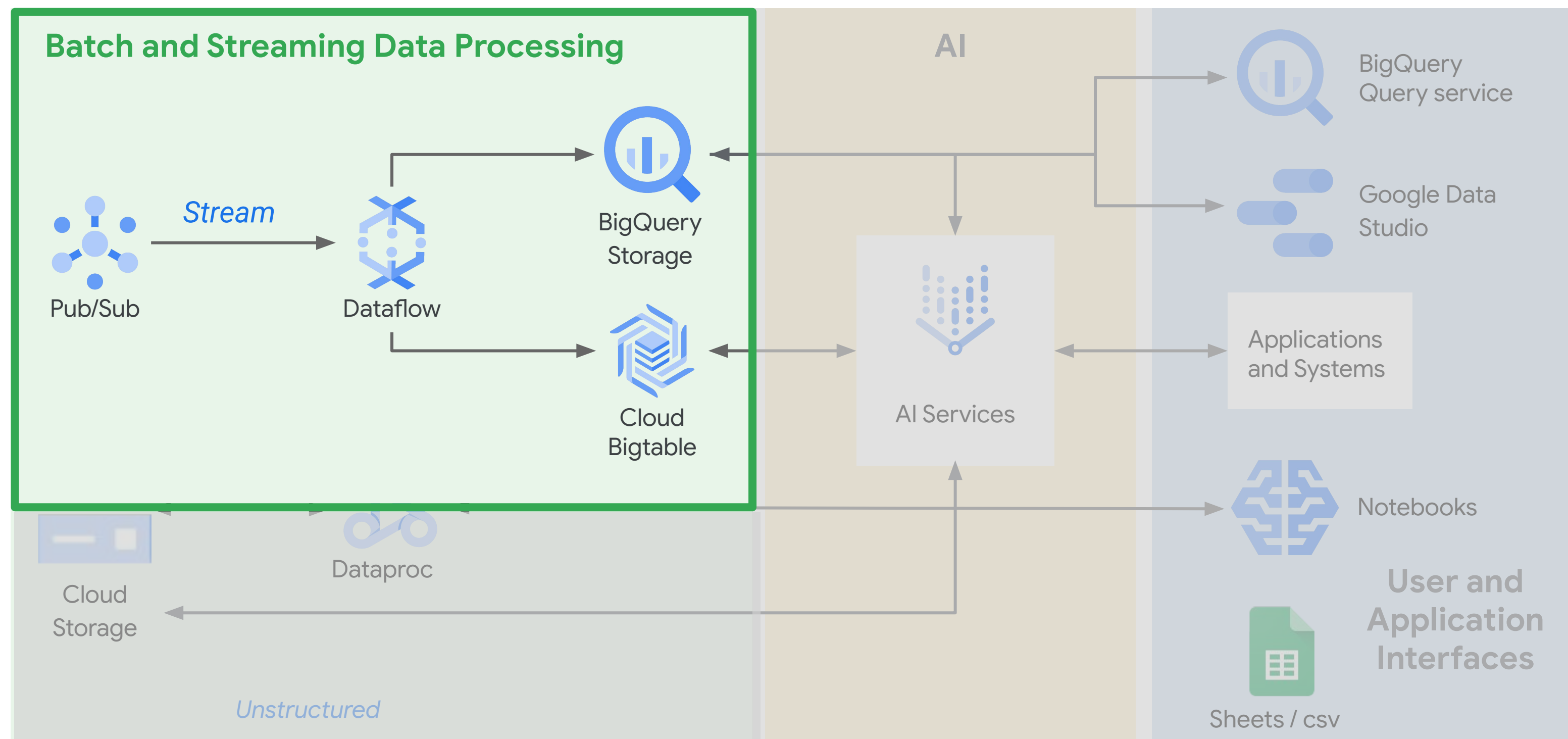


01

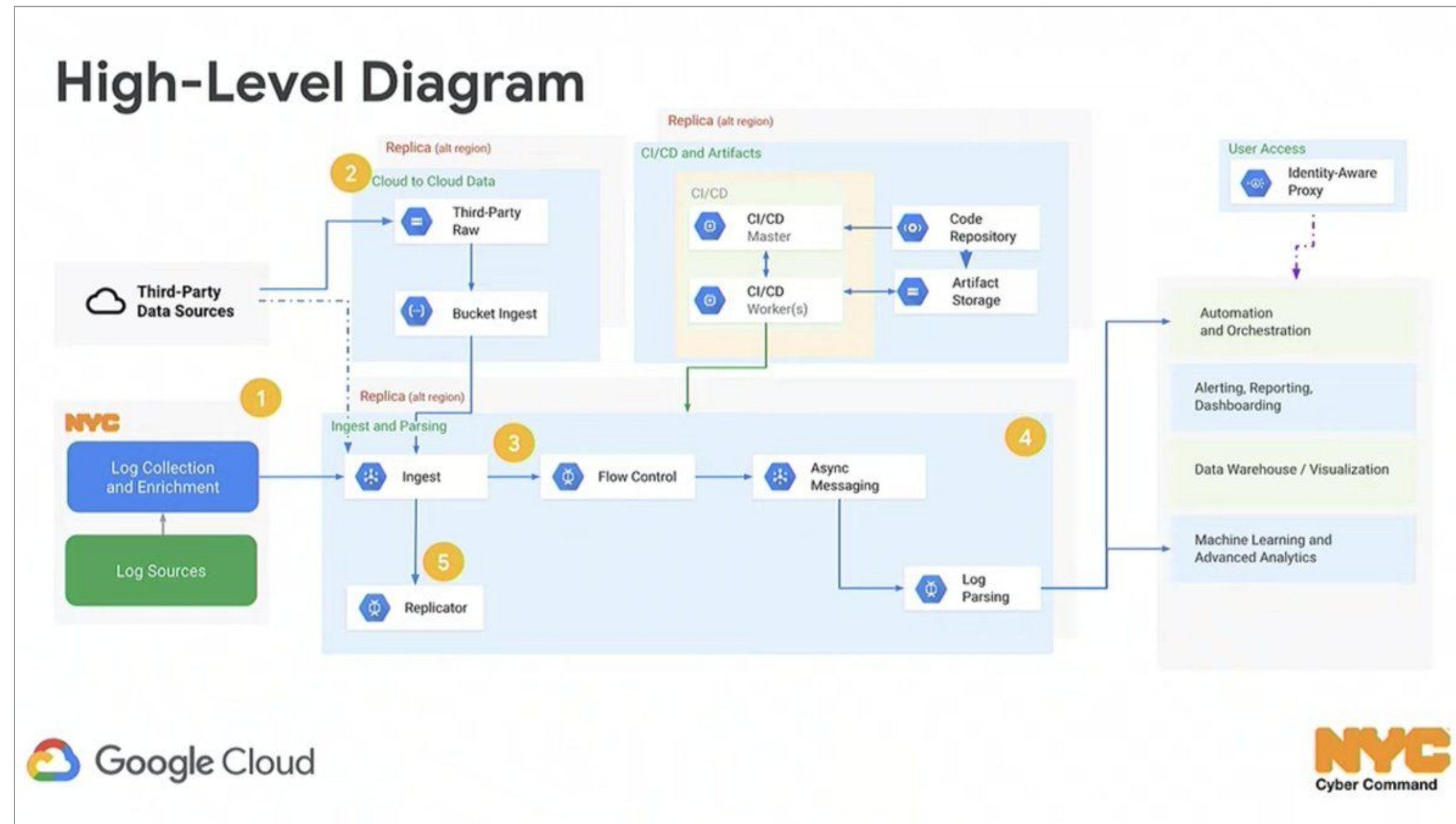


Introduction to Processing Streaming Data

Streaming data processing



Many enterprises want to enable their analysts to be able to make decisions in real-time; NYC3 did it



“

Real time is king, and that's the only data valuable to us

Noam Dorogoyer
New York City Cyber Command

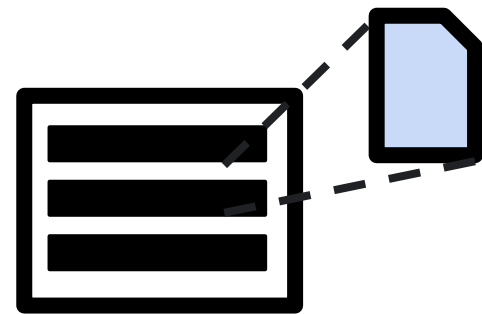
Article in GCN:

<https://gcn.com/cloud-infrastructure/2019/08/nyc-s-real-time-cyber-defense-platform/297548/>

Talk at NEXT 2019 (High-Level Diagram at 19:48):

<https://www.youtube.com/watch?v=x4yQY8yhVJY>

Streaming is data processing for unbounded data sets



Bounded Data (Batch)

- Finite data set
- Usually complete
- Time of elements is usually disregarded
- Typically at rest
- Held in durable storage

Data Stream



Unbounded Data (Stream)

- Infinite data set
- Never complete
- Time of elements is usually significant
- Typically in motion
- Held in temporary storage

Stream analytics has many applications

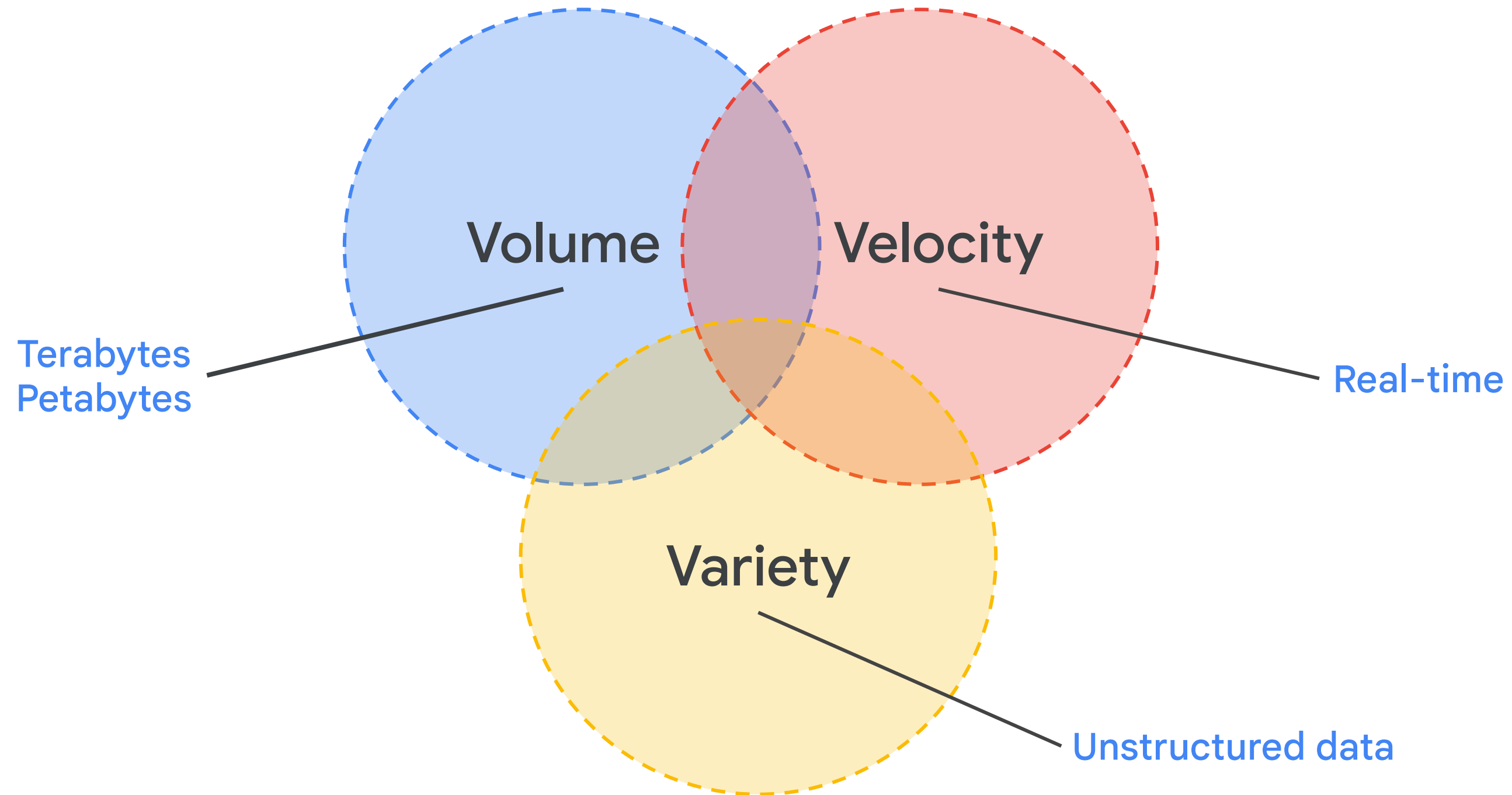
Data integration (10 sec - 10 min)

- Data warehouses become real-time
- Take load off source databases with change data capture (CDC)
- Microservices require databases and caches

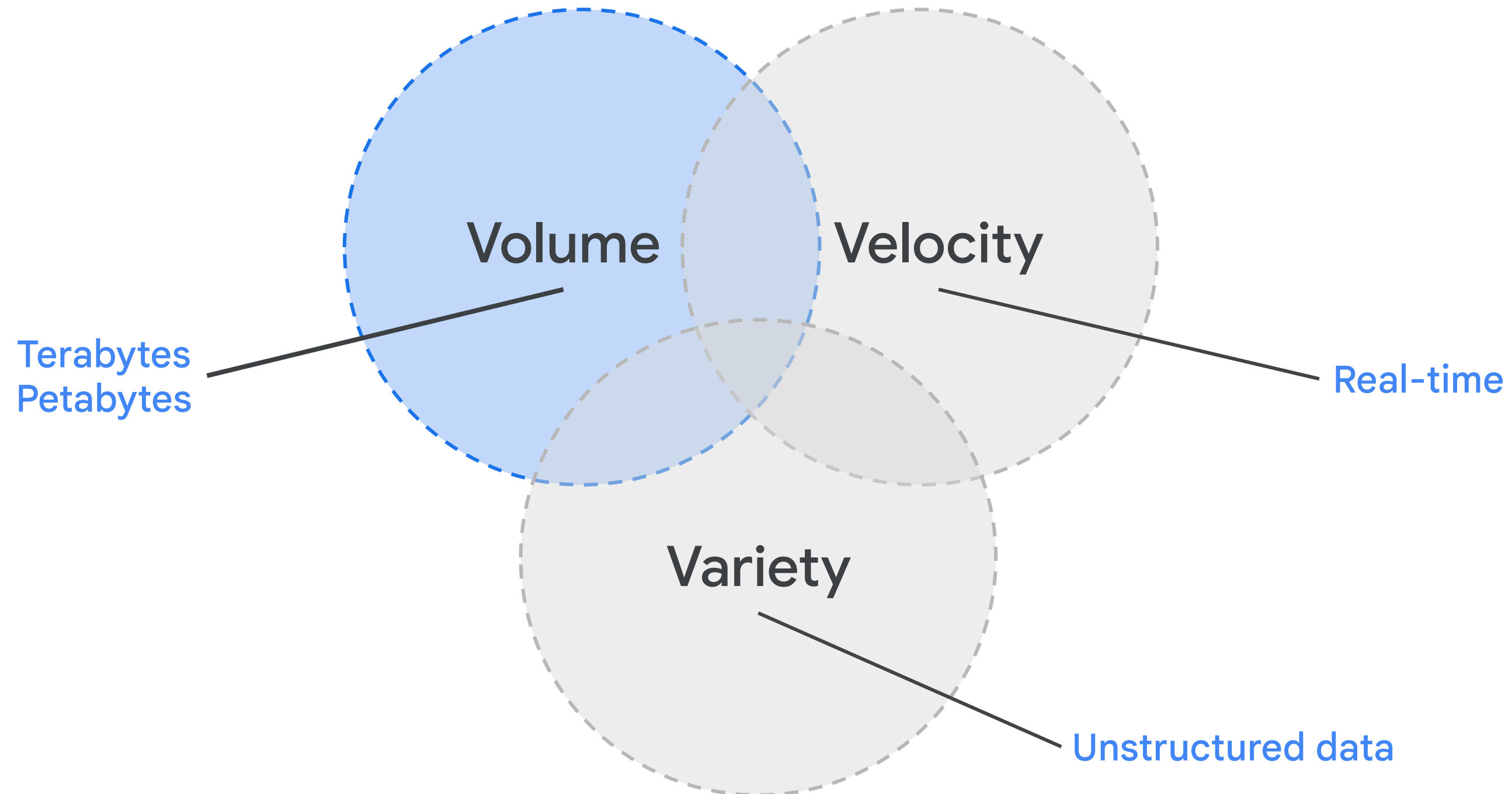
Online decisions (100 ms - 10 sec)

- Real-time recommendations
- Fraud detection
- Gaming events
- Finance back office apps

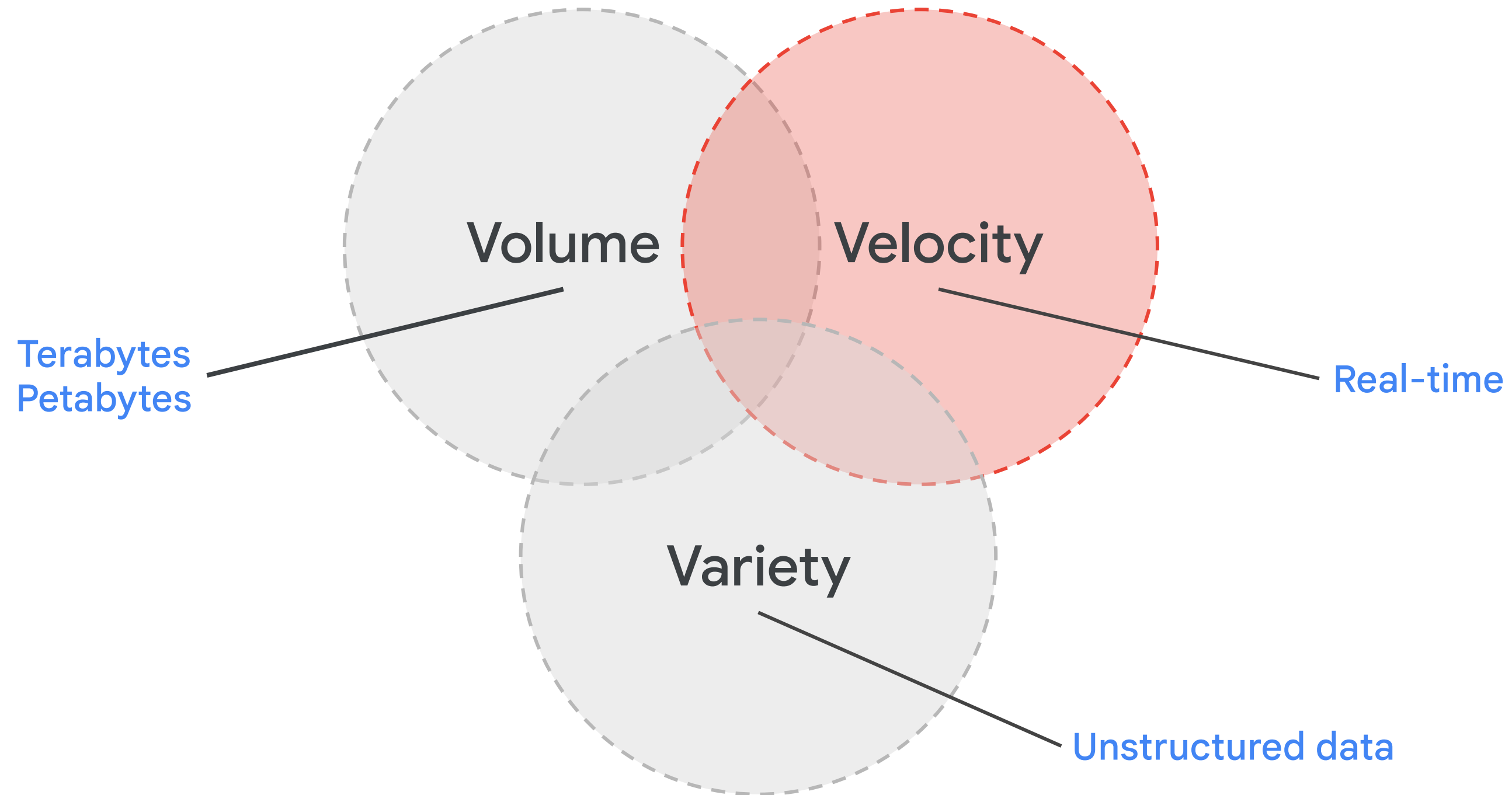
How to handle data volume, velocity, and variety?



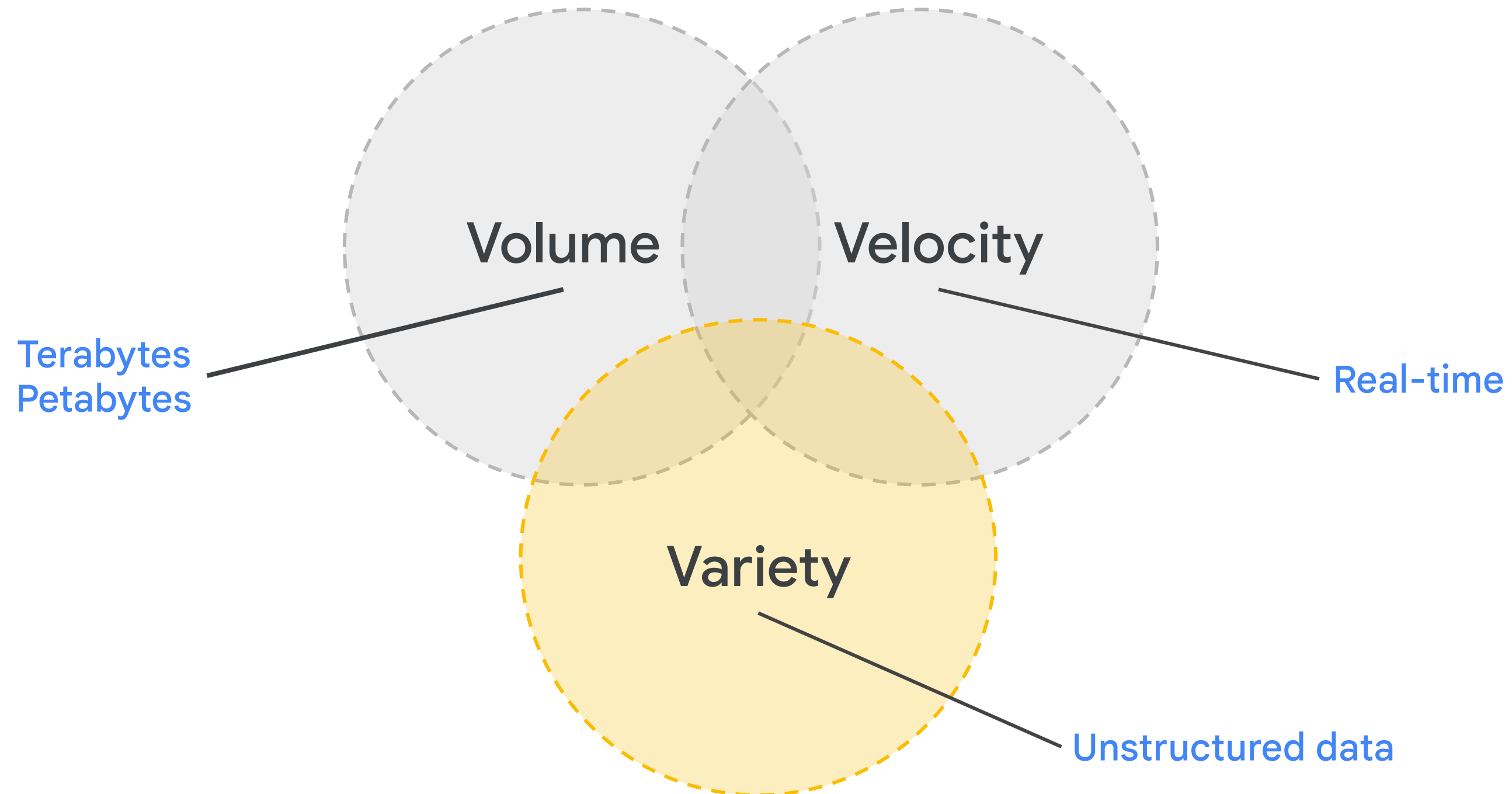
How to handle data volume, velocity, and variety?



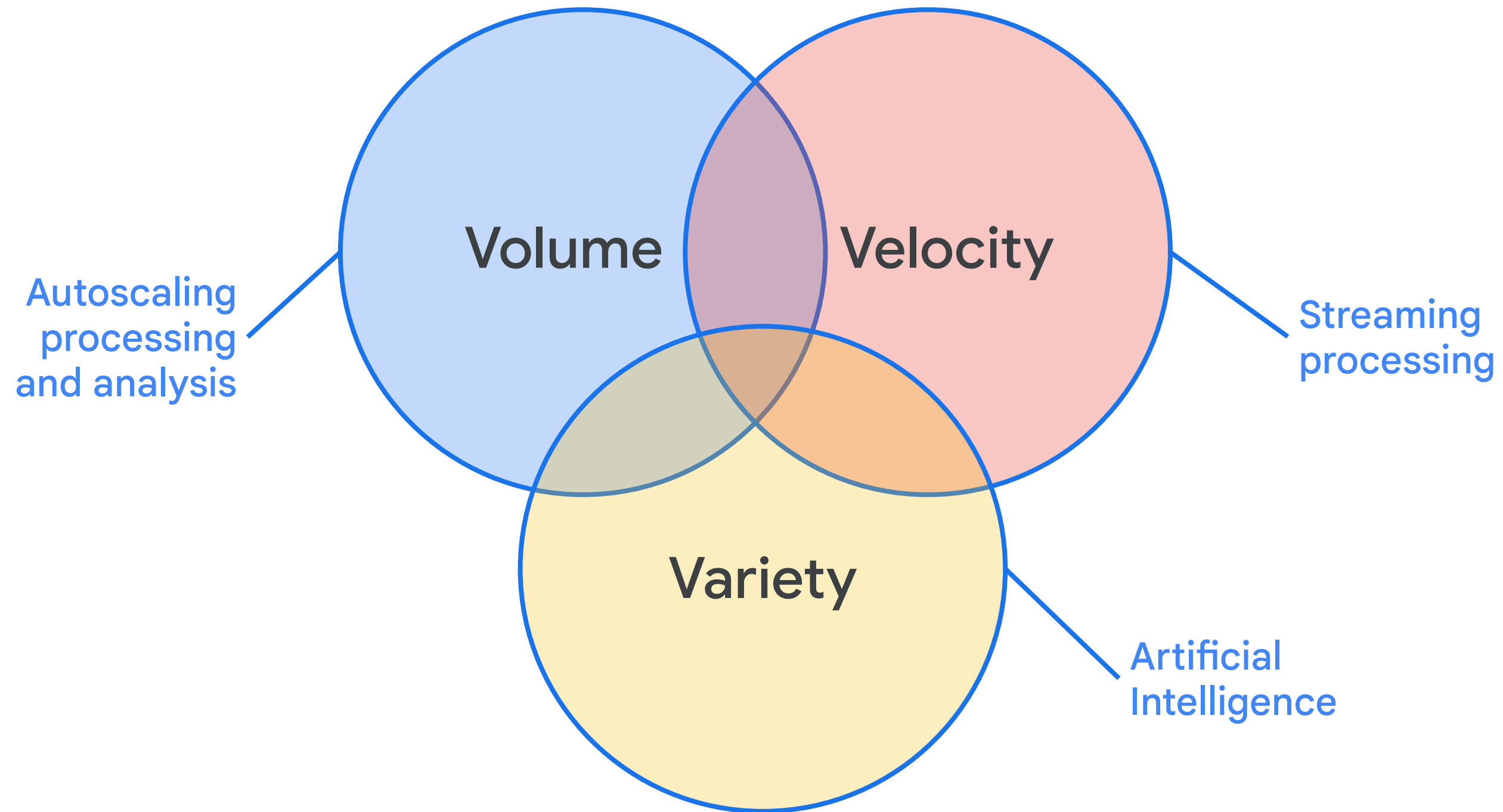
How to handle data volume, velocity, and variety?



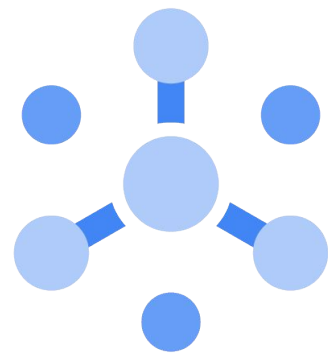
How to handle data volume, velocity, and variety?



Autoscaling, Machine Learning, and Streaming



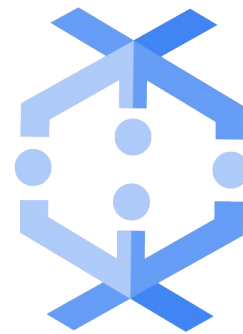
Google Cloud products help you address key challenges in stream data processing and analytics



Pub/Sub

1

Changing and variable
volumes of data



Dataflow

2

Process data without
undue delays



BigQuery

3

Need ad-hoc analysis
and immediate insights

Stream analytics includes some common steps

