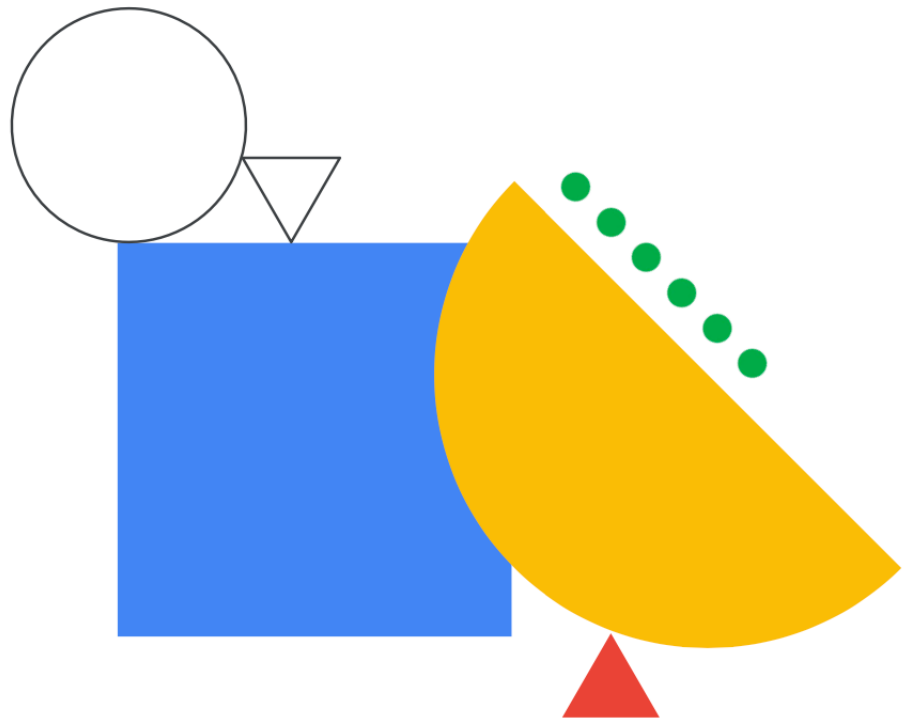
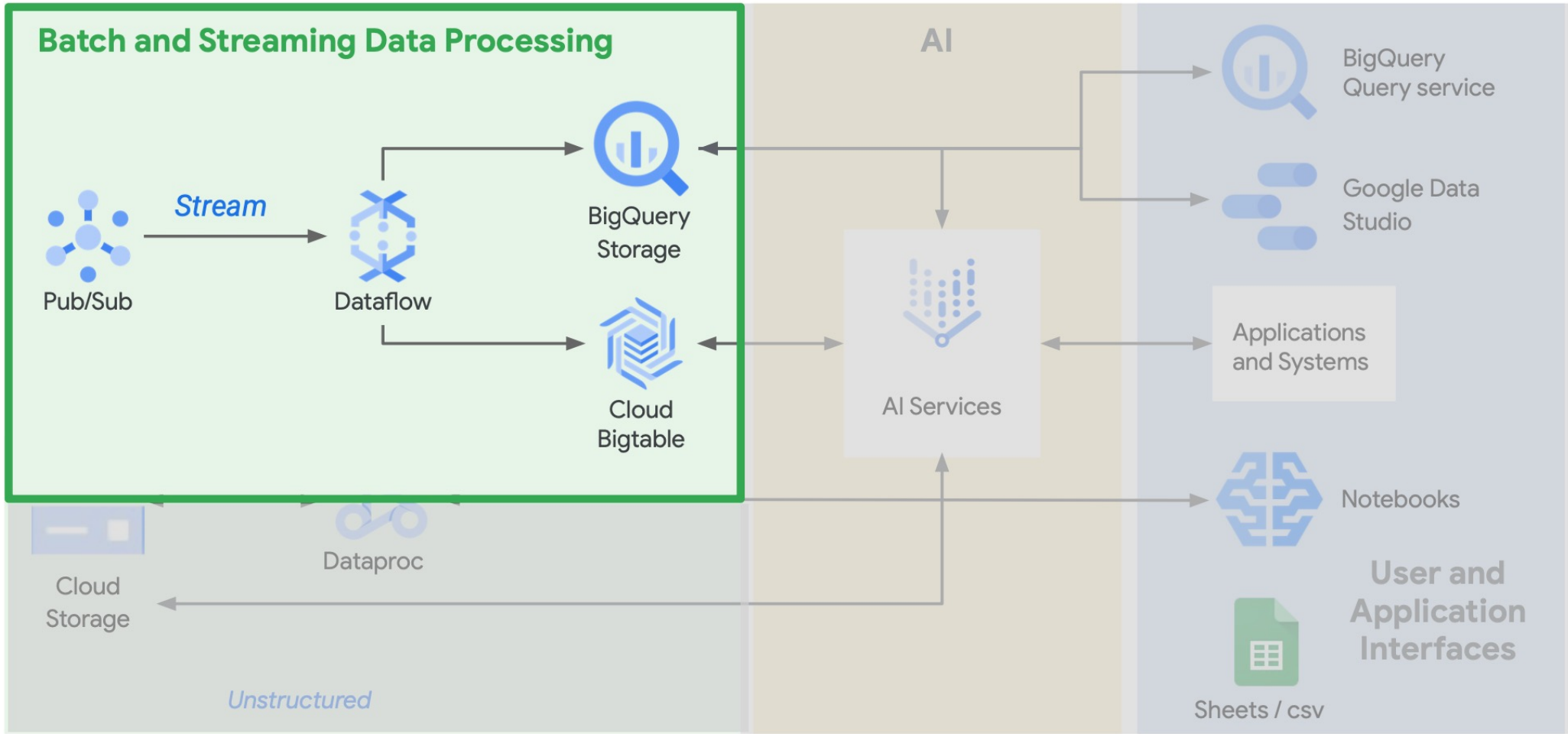
Google Cloud

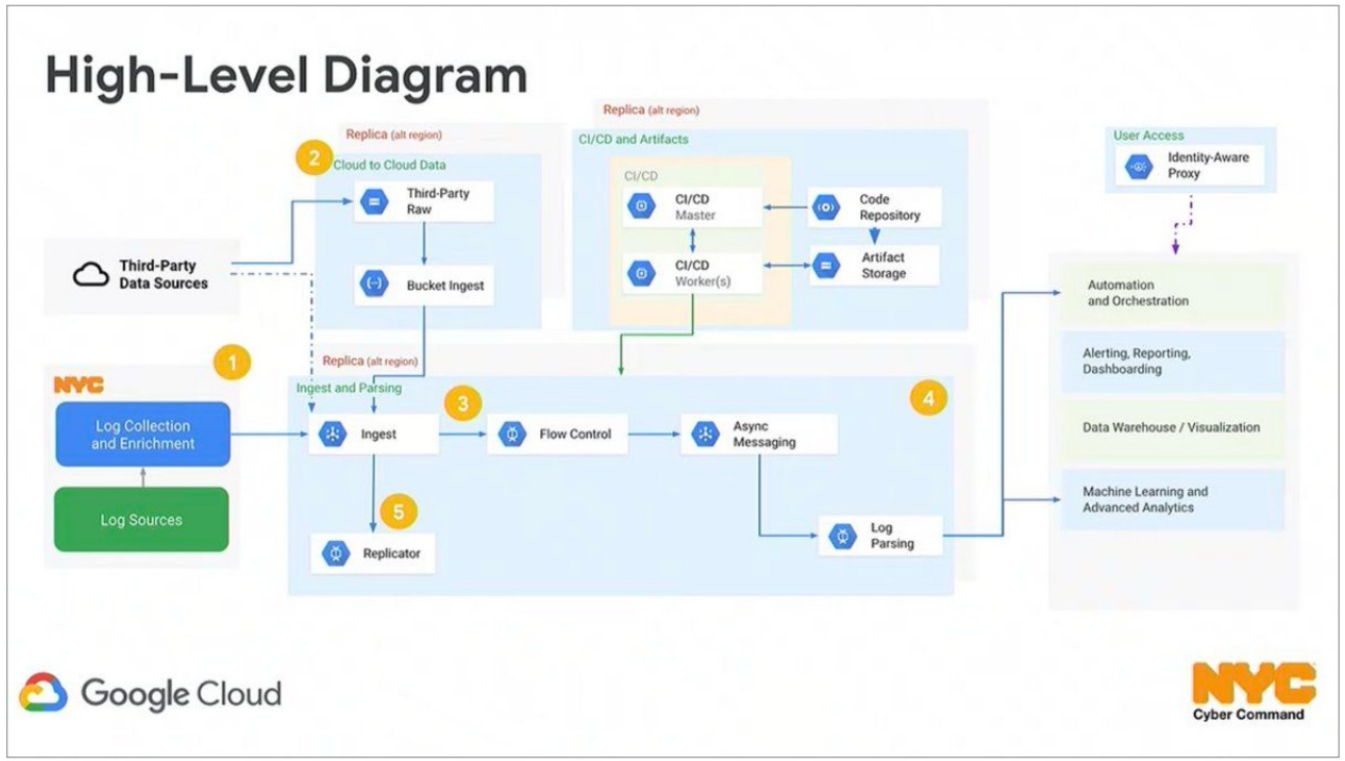
Introduction to Processing Streaming Data



Streaming data processing



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

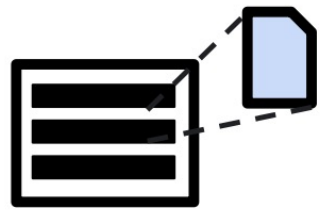


> Real time is king, and that's the only data valuable to us
>
> **Noam Dorogoyer**
> New York City Cyber Command

Article in GCN: https://gcn.com/articles/2019/08/01/nyc-cloud-cyber-pipeline.aspx

Talk at NEXT 2019: 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

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

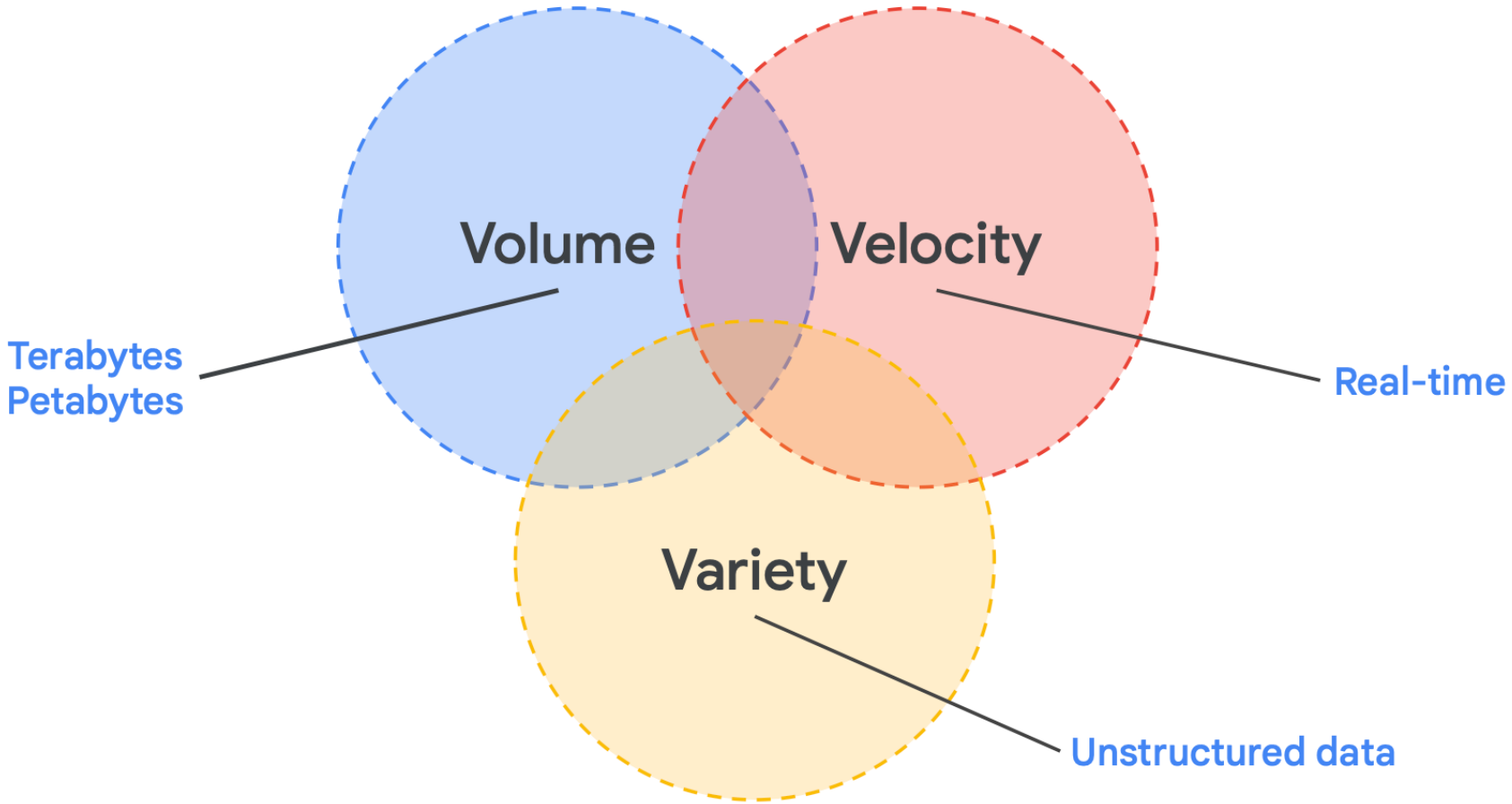
Data integration (10 sec - 10 min)

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

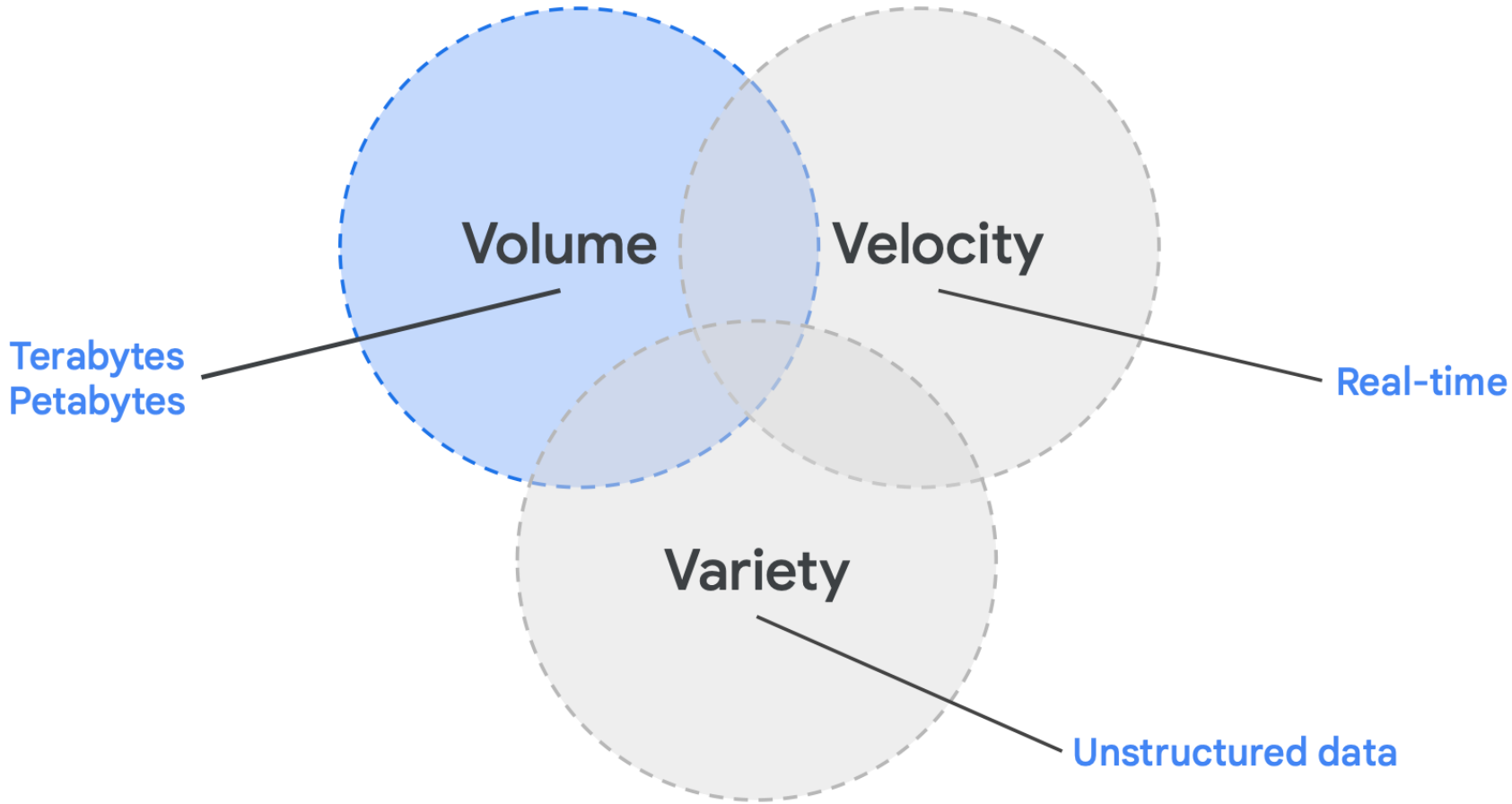
Online decisions (100 ms - 10 sec)

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

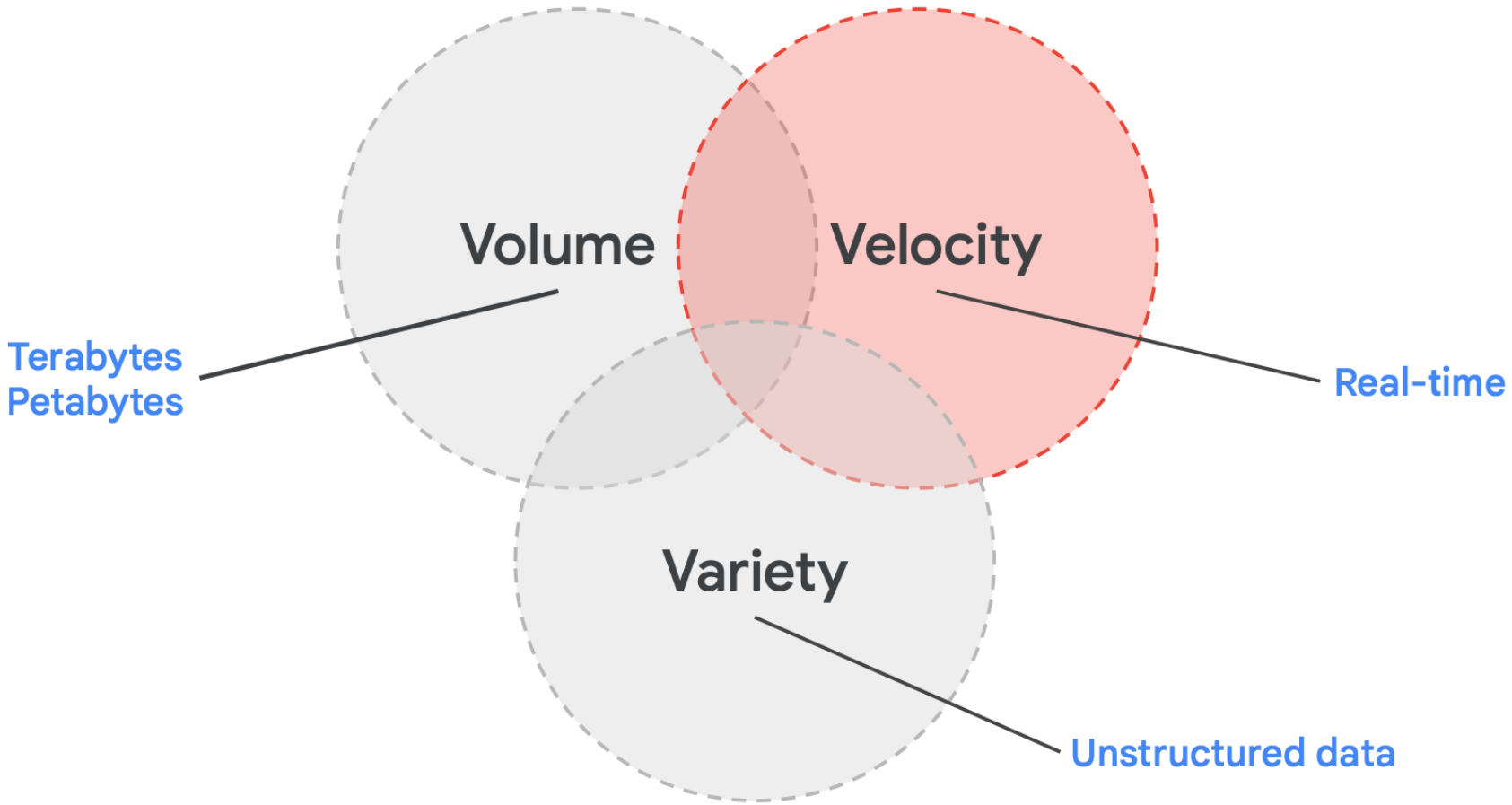
How to handle data volume, velocity, and variety?



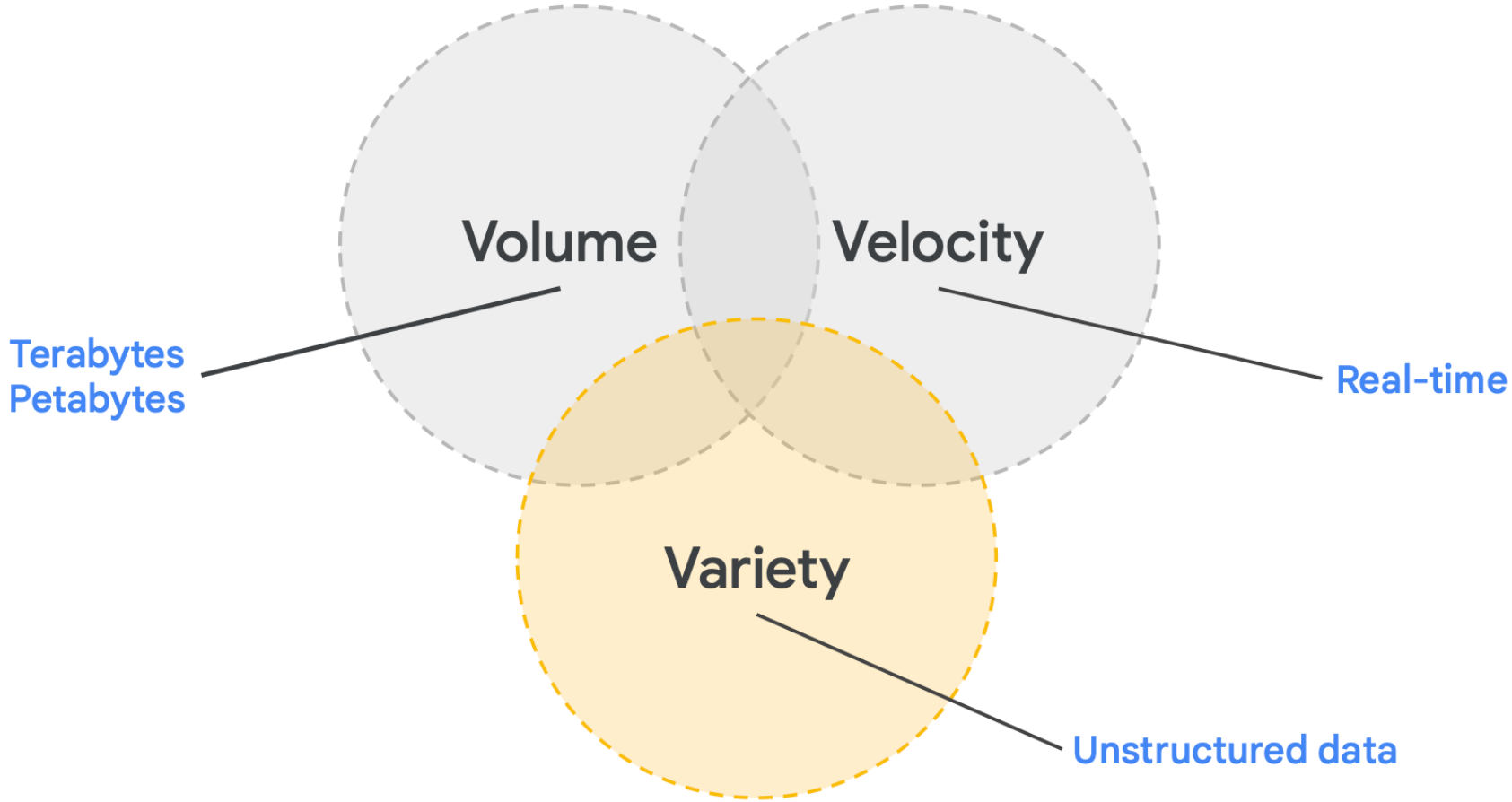
How to handle data volume, velocity, and variety?



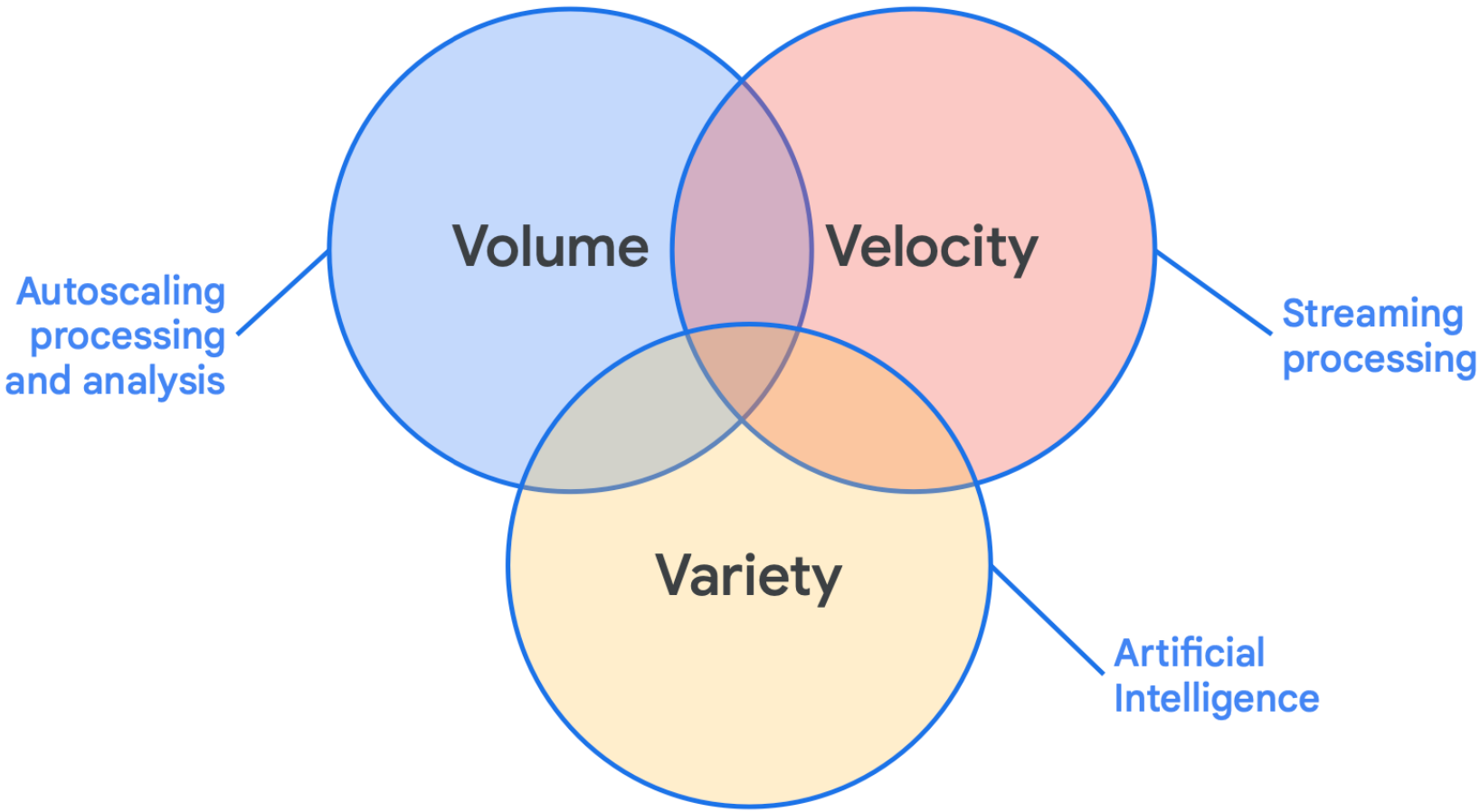
How to handle data volume, velocity, and variety?



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