

# Data Ingestion vs. Data Migration

Harvesting, Storing, and Retrieving Data  
Dr. Boyce

# Data Ingestion

# Goals of Data Ingestion

Accelerate

- Accelerate Data Ingestion

Understand

- Understand Your Data

Govern

- Govern Your Data

Remove

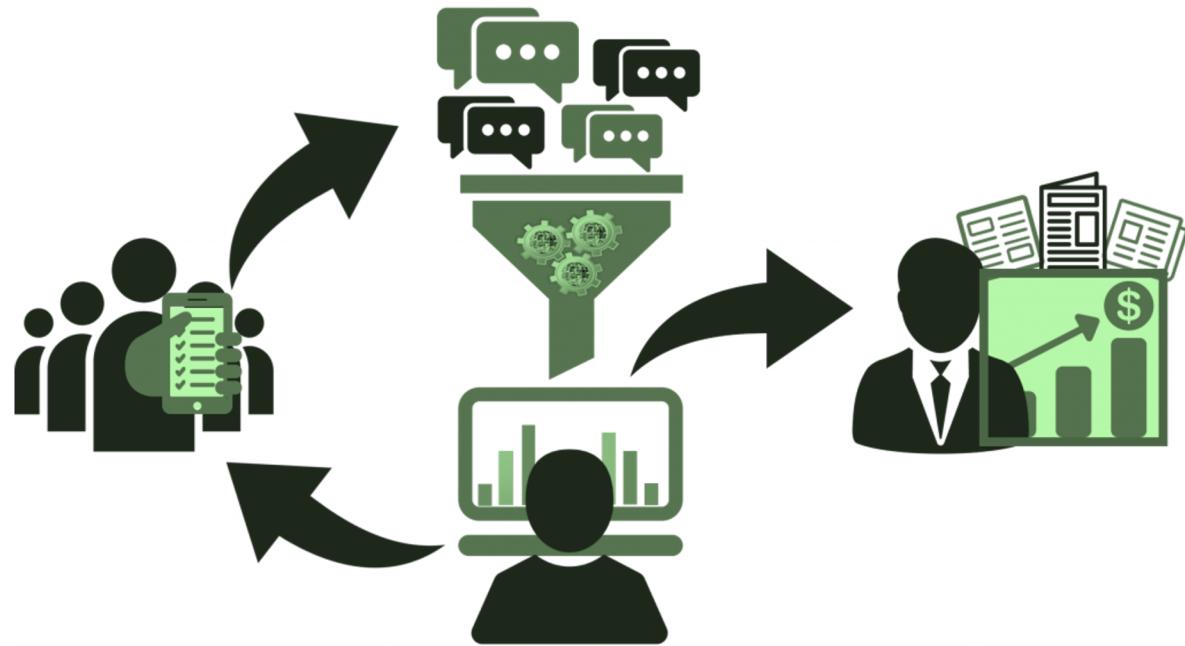
- Remove Silos

Deliver

- Deliver Data to a Wide Audience

# Parameters of Data Ingestion

---



- Data Velocity
- Data Size
- Data Frequency
- Data Format

# Challenges of Data Ingestion

Sluggish Processes

Increased Complexity

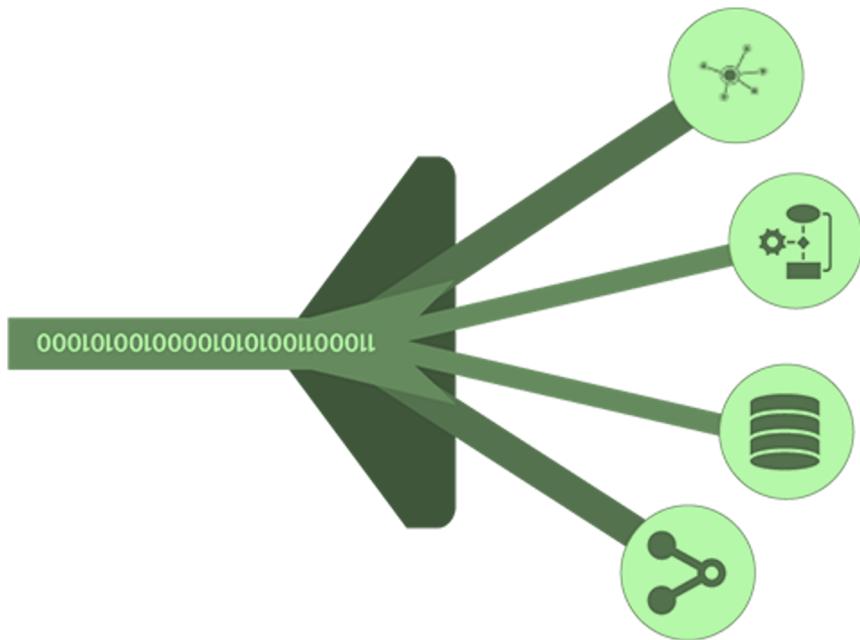
Cost Factor

Risk of Data Security

Unreliability

# Best Practices of Data Ingestion

---



---

Network Bandwidth

---

Anticipate Difficulties

---

Automate the Process

---

Support for Unreliable Network

---

Heterogenous Technologies & Systems

---

Choose Right Data Format

---

Business Decision

---

High Accuracy

---

Connections

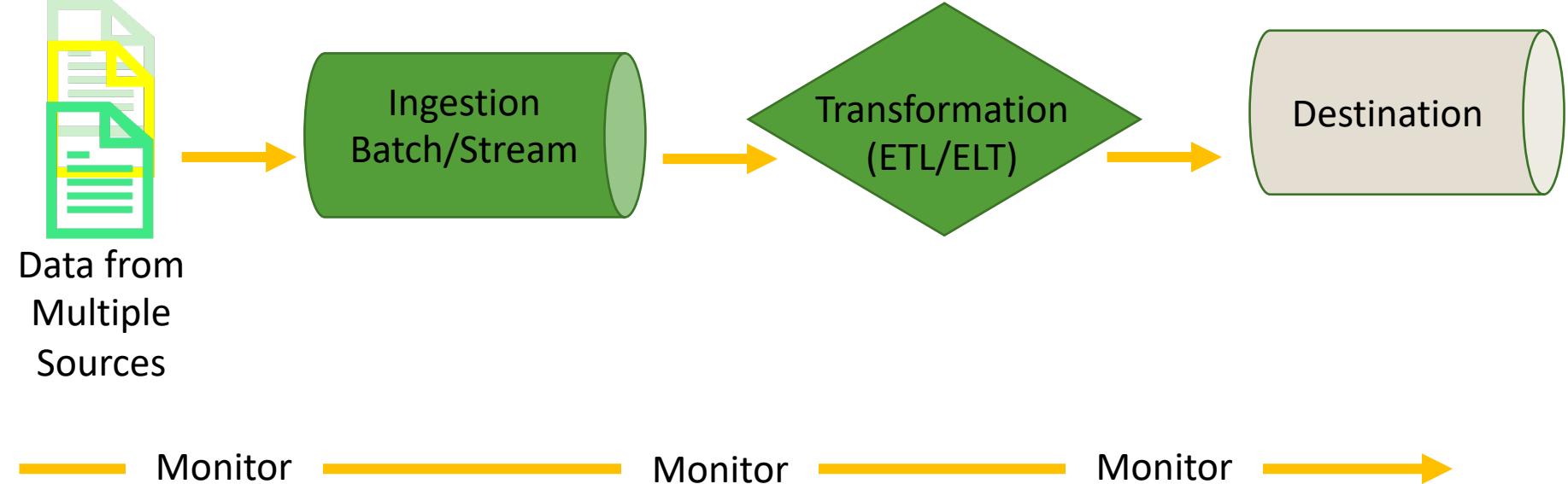
---

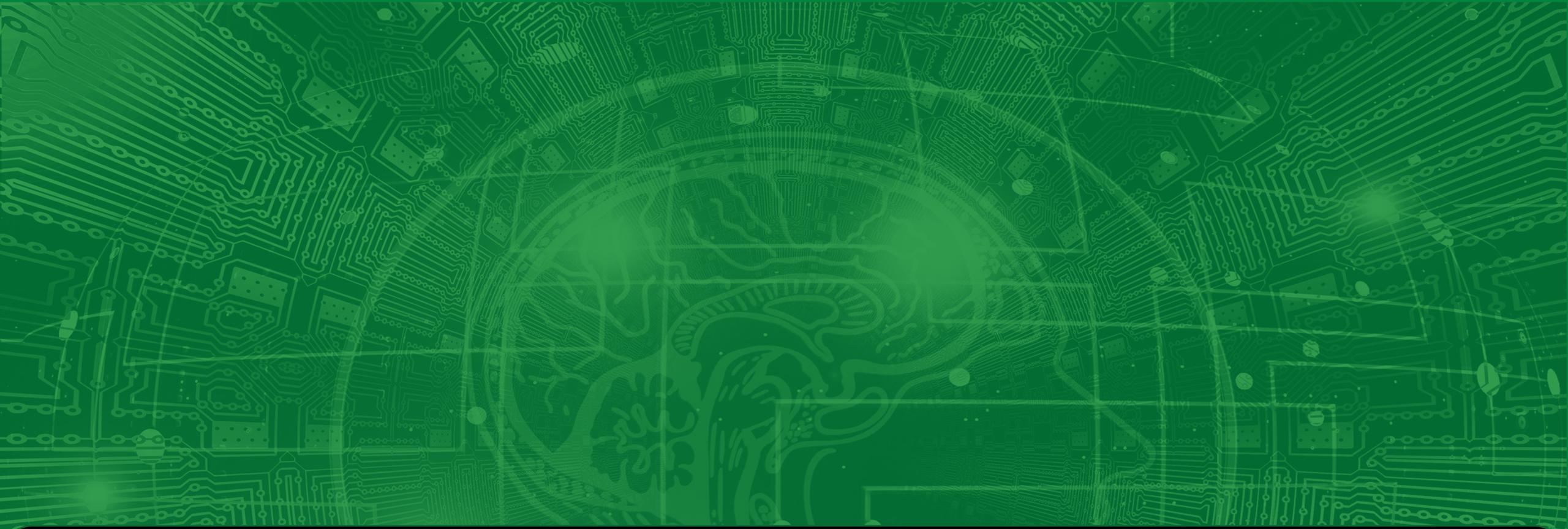
Latency

---

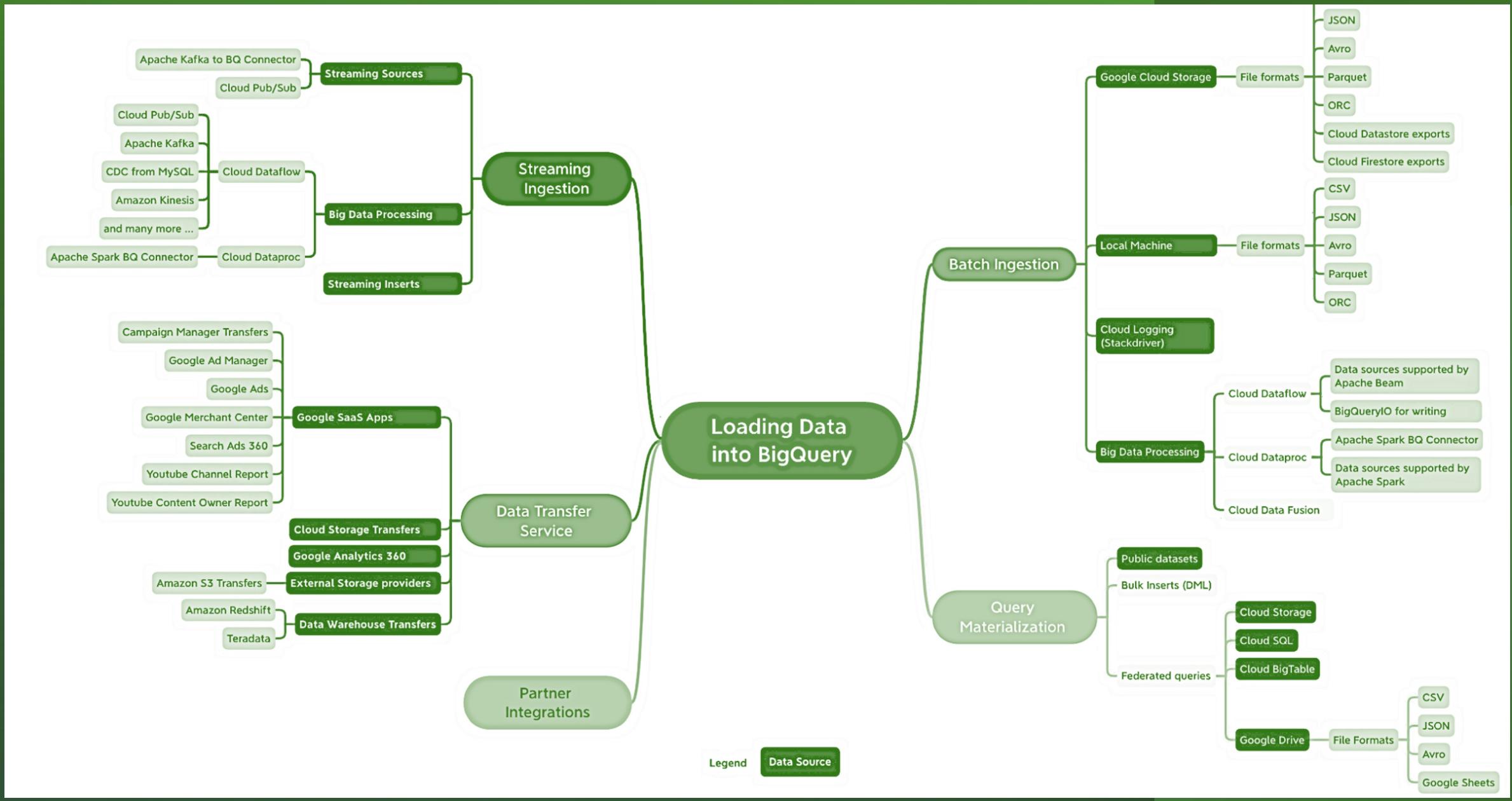
Maintain Scalability

# Typical Stages of Building a Data Pipeline

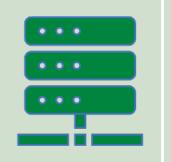




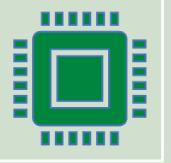
# Data Ingestion with GCP



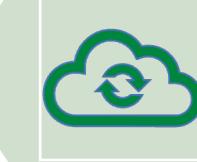
# Data Ingestion and Collection



**Using APIs on the data provider—**  
pull data from APIs at scale using  
Compute Engine instances (virtual  
machines) or Kubernetes



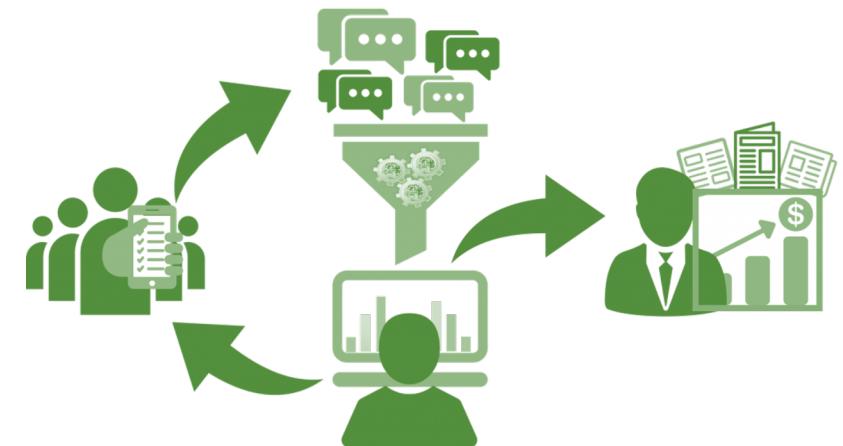
**Large volume of data on-premises—**  
most suitable for Google transfer  
appliance or GCP Online Transfer,  
depending on volume



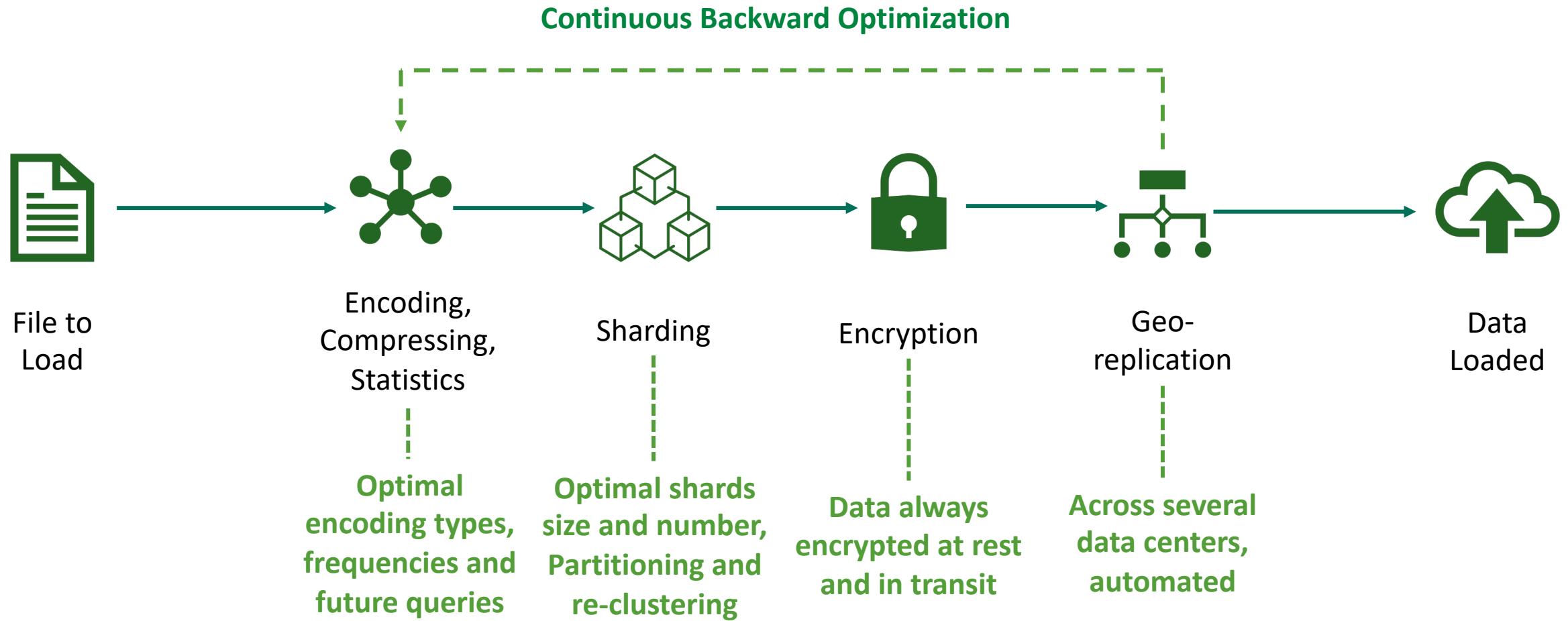
**Real time streaming—**  
best with Cloud Pub/Sub



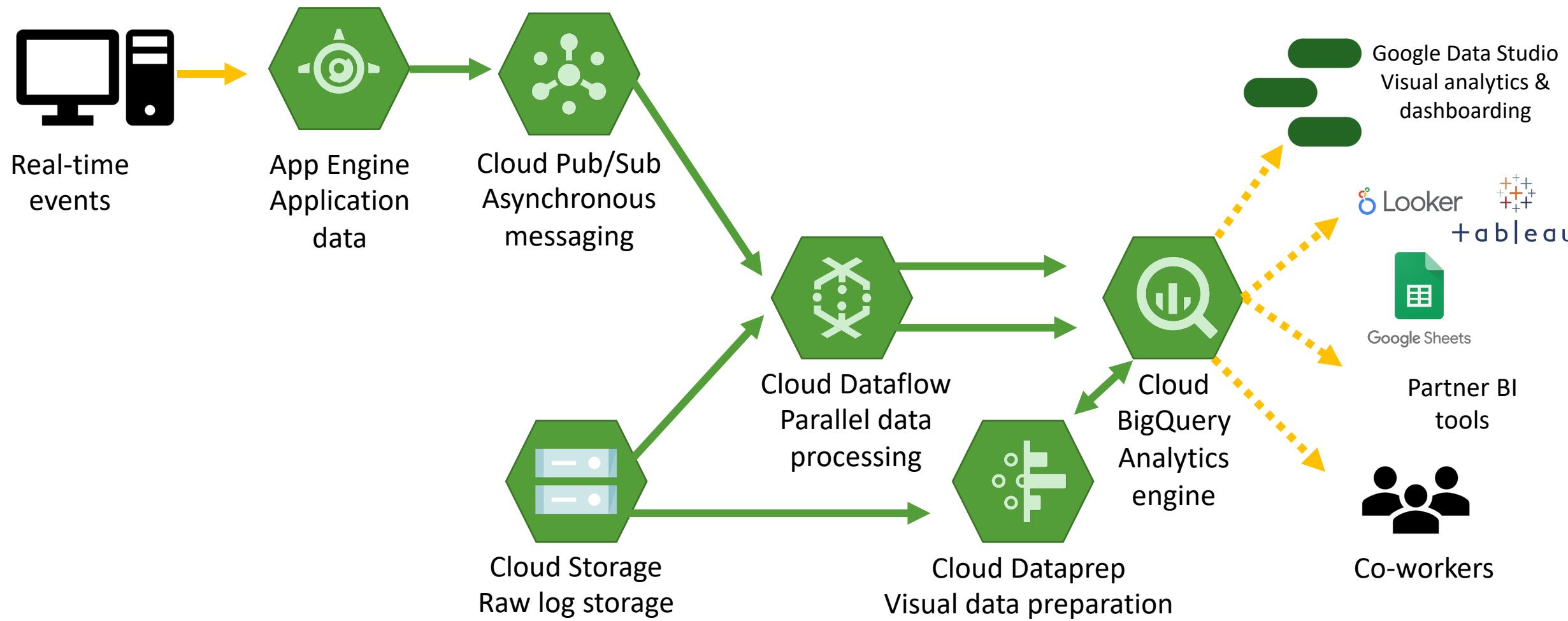
**Large volume of data on  
other cloud providers**



# Batch Ingestion: Behind the Scenes



# Streaming Ingestion





# In Your Homework, you will

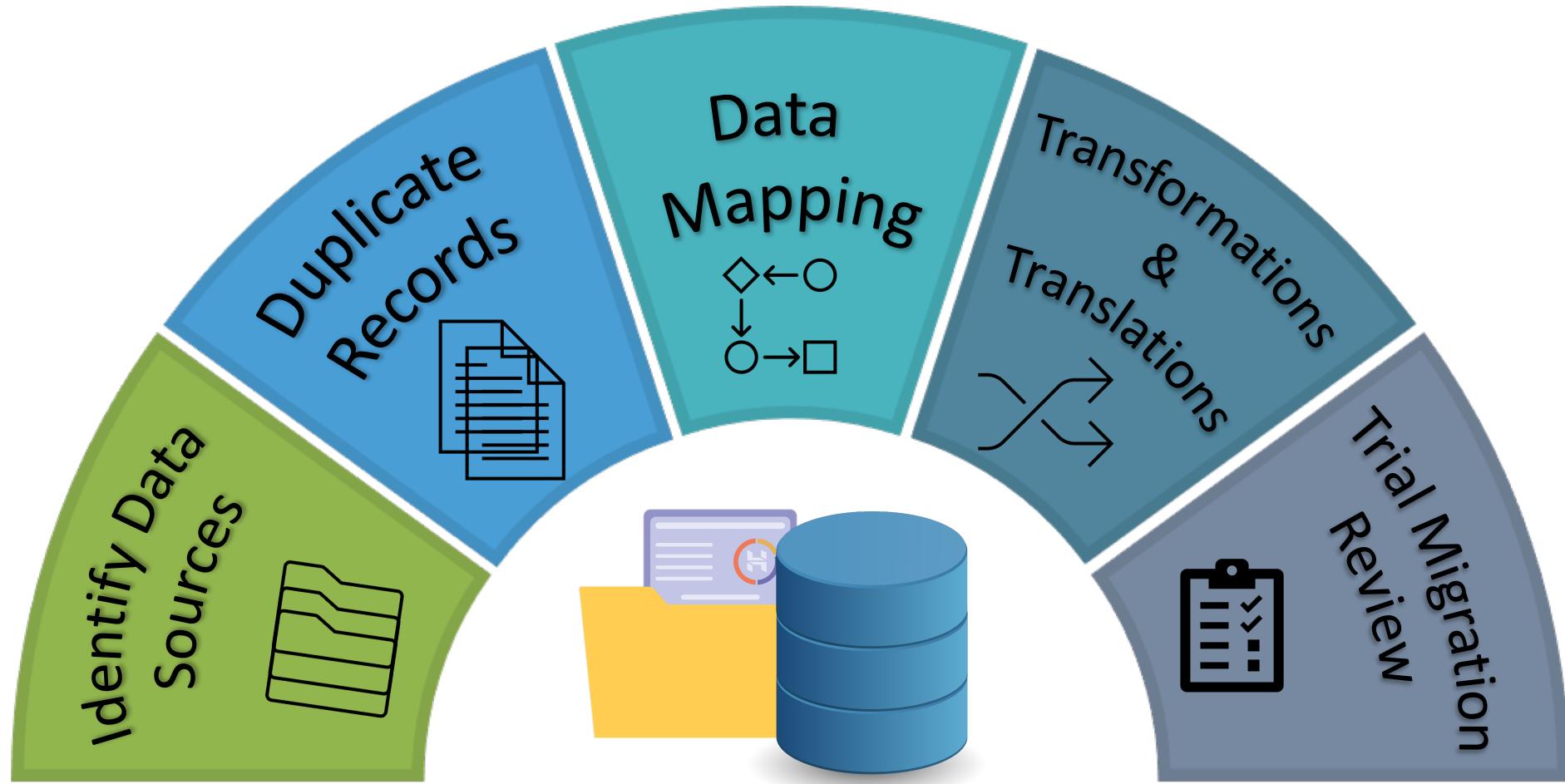
---

Collect Twitter Data

Ingest the collected data into

- GCP Storage bucket
- BigQuery
- DataPrep

# Data Migration



# Data Migration

# Goals of Data Migration

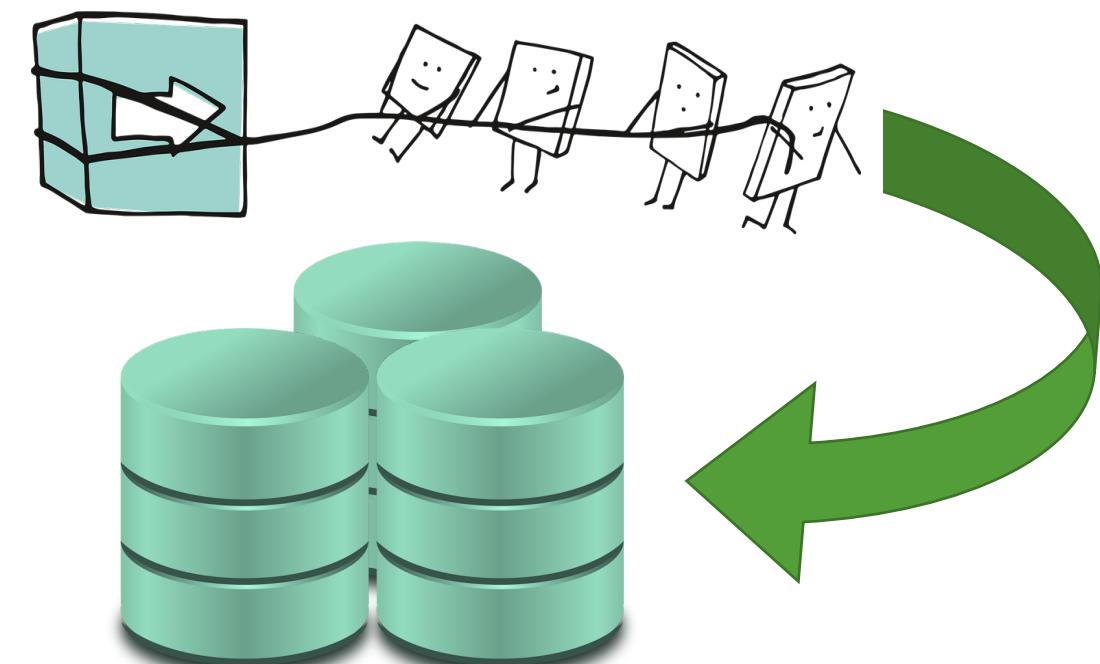
High Data  
Fidelity  
During  
Migration

History and  
Traceability  
for all  
Migrated  
Data

No System  
Downtime  
During  
Migration

# Parameters of Data Migration

---



- Data Integrity
- Business Impact
- Cost
- User Experience and Impact
- Potential Downtime
- Data Assessment
- Data Quality

# Challenges of Data Migration

Identifying Disparate Databases

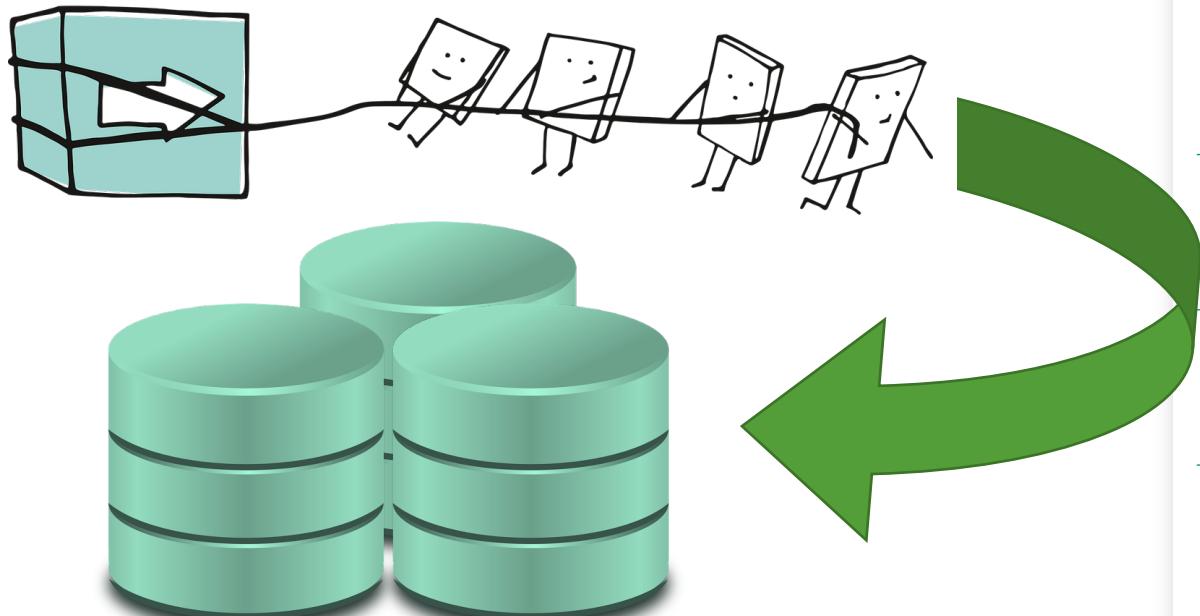
Devising a Migration Strategy

Data Cleansing and Coding

Risk of Data Security

# Best Practices of Data Migration

---



---

Explore and Assess the source

---

Backup the Data

---

Define and Design the Migration

---

Adhere to the Plan

---

Consider Security

---

Test, Test, Test

---

Go Live

---

Audit