Little Devices and Big Data

Applying Cloud Big Data to RAIN RFID Use Cases

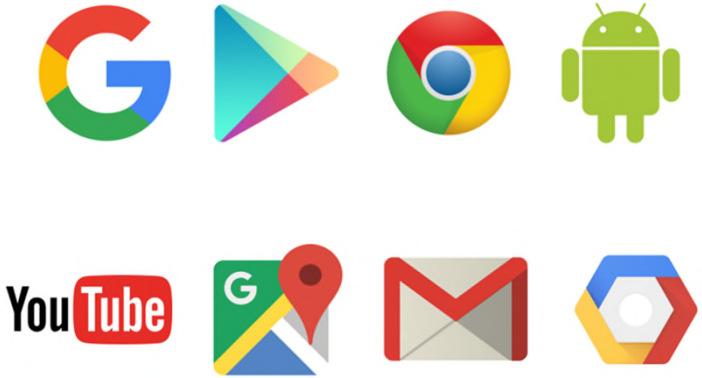
Stephen Johnson

Customer Engineer, Google Cloud

Google Cloud Platform

Outline

About “Big Data”



Big Data is a term used to describe the workloads and tools used to process data at an enterprise or internet scale.

It does not necessarily mean you have huge files or Petabytes of data, though it can handle that.

The Cloud allows customers access to enormous resources at utility prices.

About “Big Data” Strategies

This report said that enterprises lacked a “strategy” for big data, IoT, and machine learning.

Strategy does not imply specific solutions or vendors (though many vendors suggest this).

Cloud allows for low cost experimentation and iteration.

My goal is to present a methodology to help you successfully reach your business goals.

Outline

Always start with Use Cases

Smart Shelf (Inventory Location) : A supermarket wants to track inventory and misplaced items on a shelf.

Commercial Forestry: Tagging of saplings to lumber products help optimize supply chain.

Pharmaceutical Inventory (High Value Assets): Tracking medicines and controlled substances in hospitals to ensure proper disposition.

Attendee Tracking (People Movement): Following movement of individuals at events for marketing and security.

Data Characteristics (for design)

Immediacy - How soon do you need data?

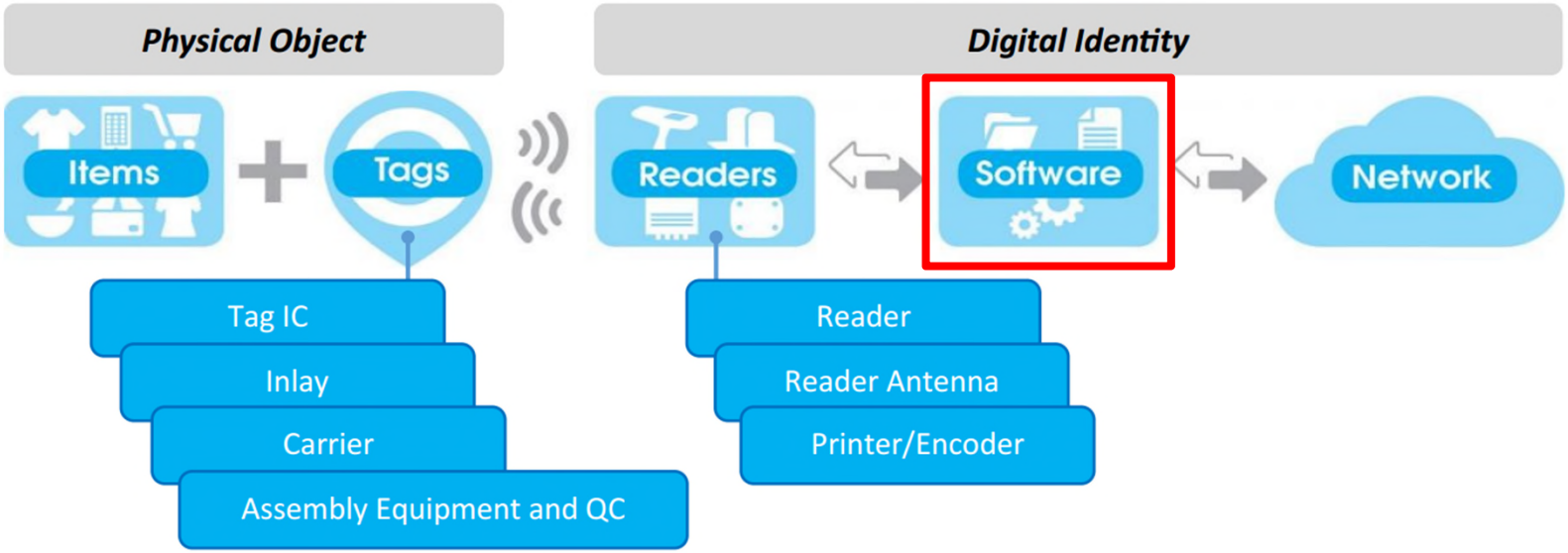
Volume - How much data is gathered?

Complexity - Do we need to supplement the data?

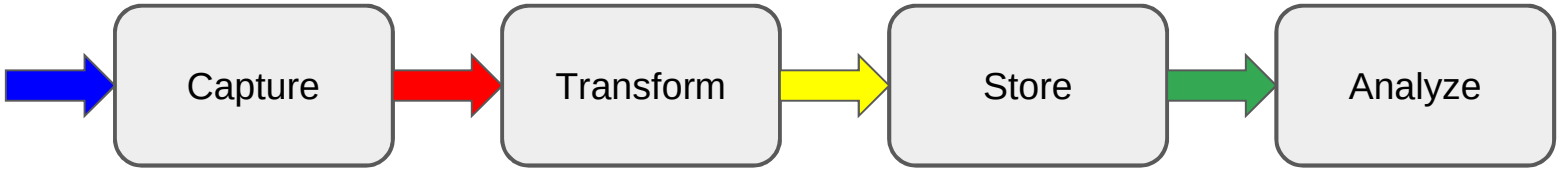
Security - What is the risk of “sharing” the data?

Outline

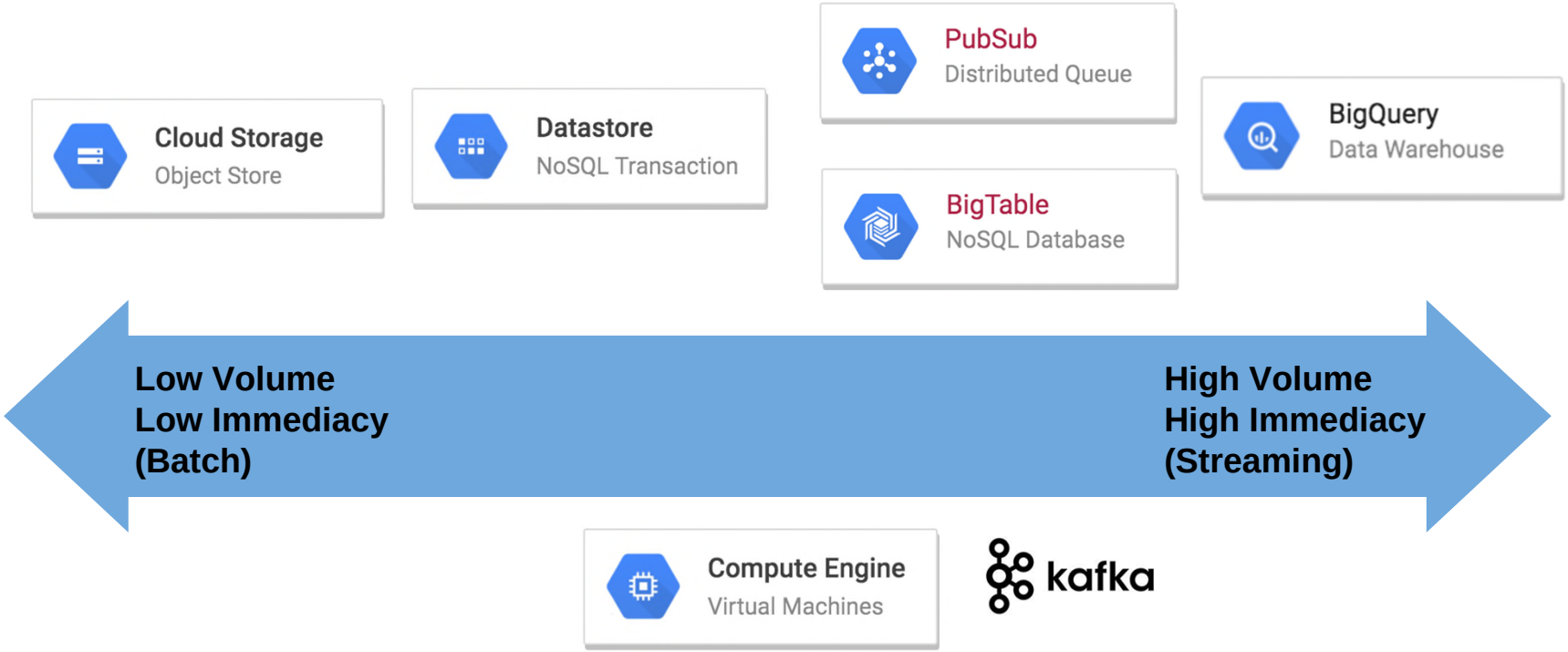
RAIN RFID Solution Profiles



Zoom in on “Network”



Data Capture



Data Transformation

DataProc
Managed Hadoop

Dataflow
Batch and Stream

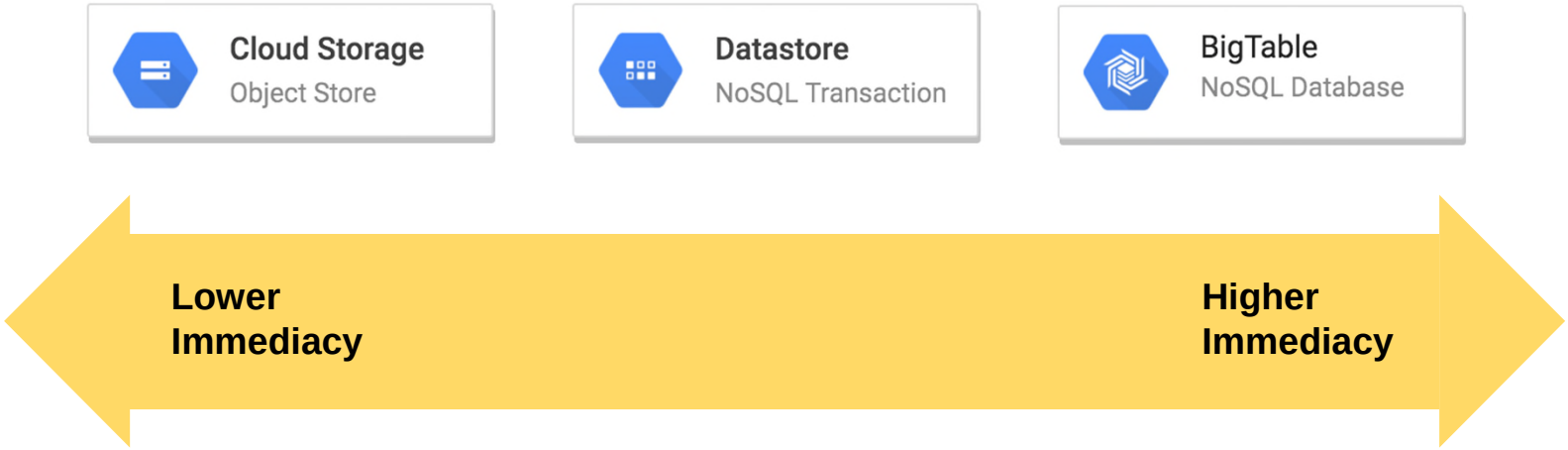
Low Complexity
Low Immediacy

High Complexity
High Immediacy

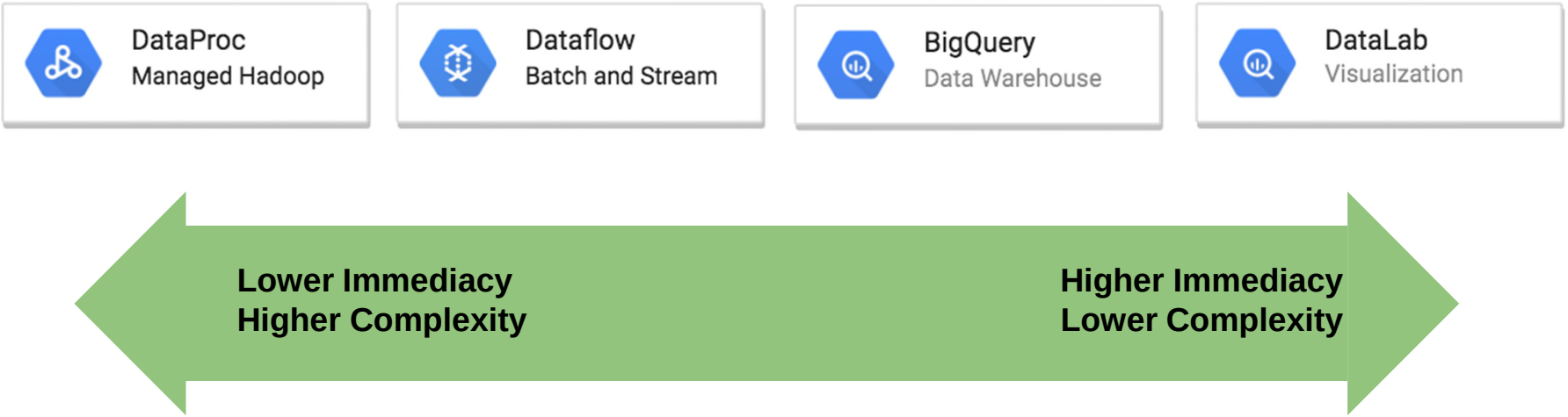
hadoop

beam

Data Storage



Data Analytics



Outline

Mapping to Use Cases

	Commercial Forest Tagging	Retail Smart Shelf	Attendee Tracking	Pharmaceutical Inventory
Immediacy	**Low**	**Medium**	**Low**	**High**
Volume	**Low**	**Medium**	**High**	**High**
Complexity	**Low**	**Low**	**High**	**High**
Security	**Low**	**Low**	**Medium**	**High**

Commercial Forestry

Customer is a small lumber company that wants to track trees from planting to harvesting. They are also interested in using custom RFID tags that can transmit state about humidity, temperature, and sunlight. They are sending data from one geographic location.

Immediacy: Data can be collected daily to every few hours.

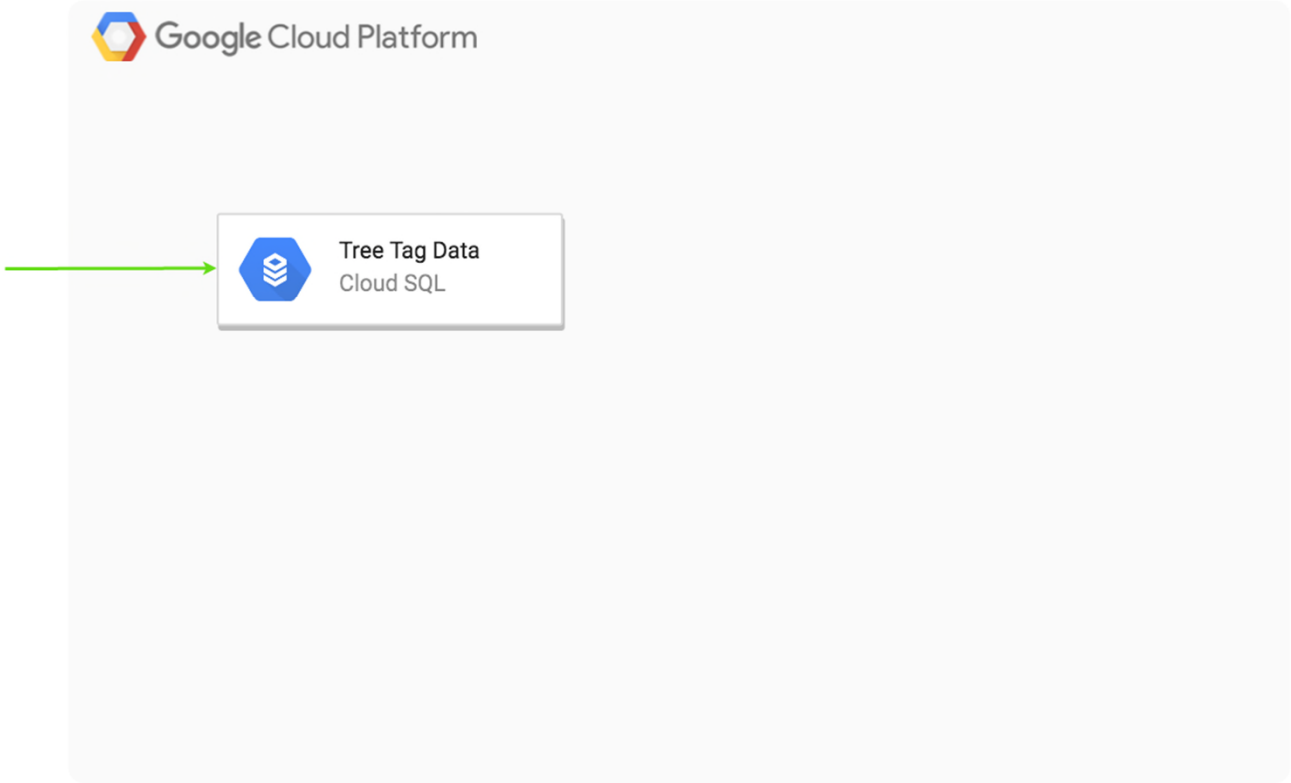
Volume: Forest contains 1000s of trees, volume depending on amount of data returned (location + environment sensor data).

Complexity: Reads can be logged directly as time-series data.

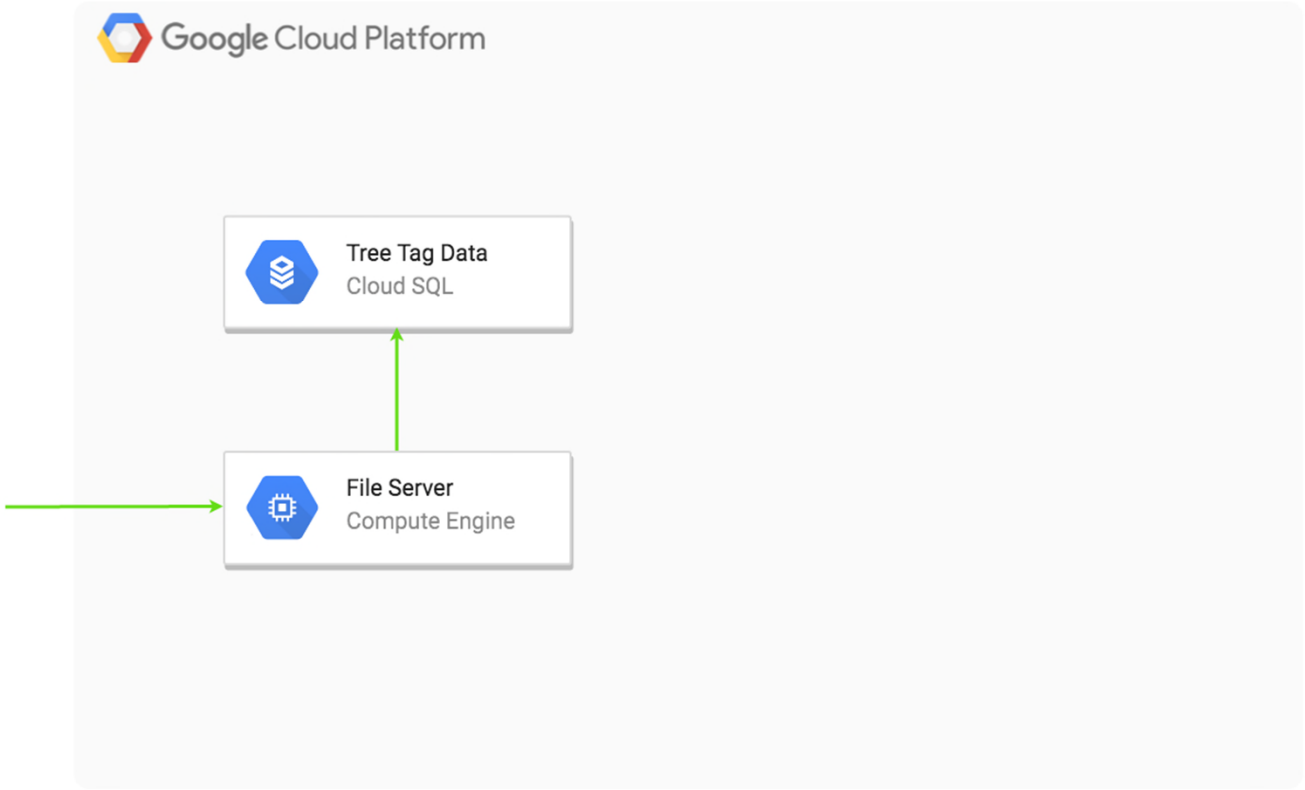
Security: Low risk of compromise. Data can be approximated without hacking

Courtesy TECTUS

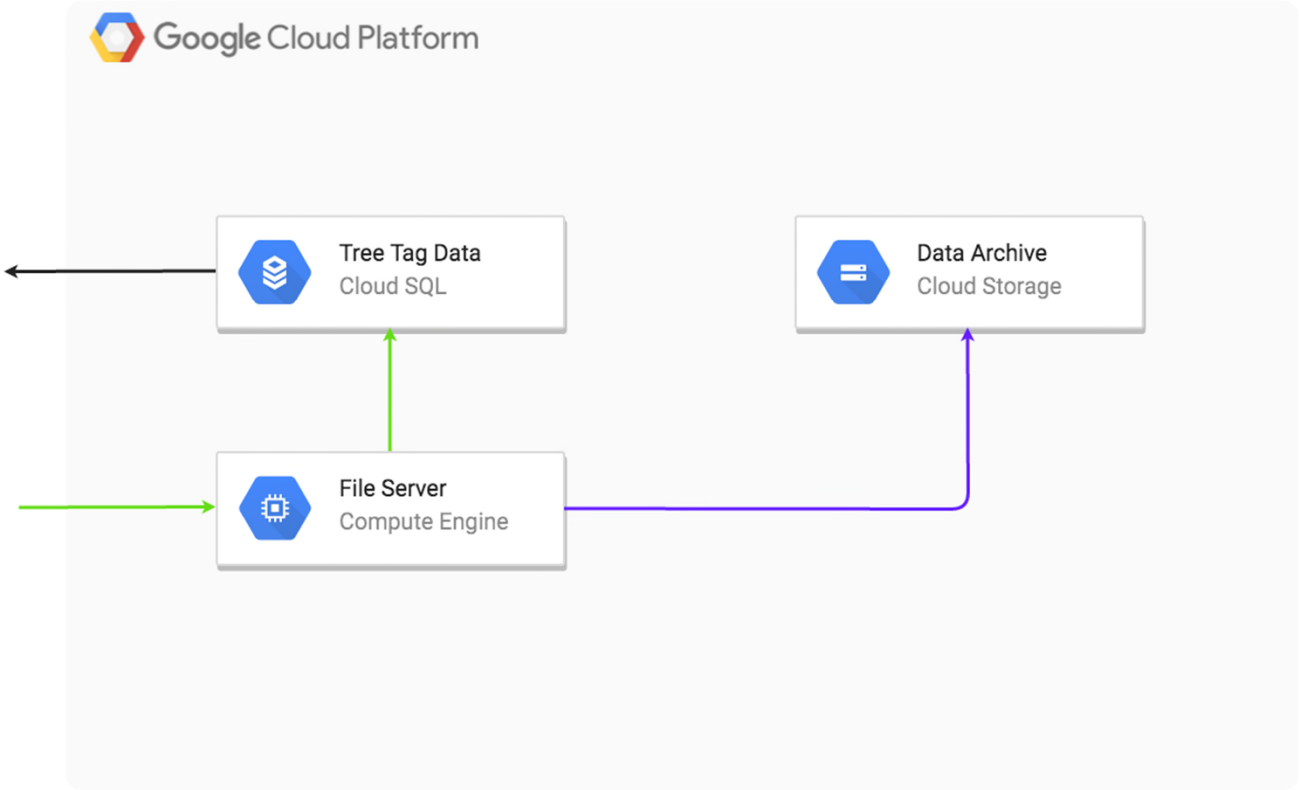
Commercial Forestry: Low Volume, Low Immediacy, Low Complexity, Low Security



Commercial Forestry: Low Volume, Low Immediacy, Low Complexity, Low Security



Commercial Forestry: Low Volume, Low Immediacy, Low Complexity, Low Security



Retail Smart Shelf

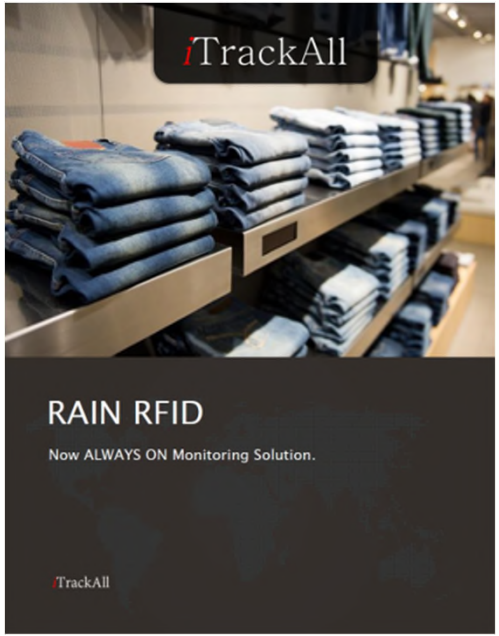
Customer is a retailer that has an RFID reader on a shelf containing various items for sale. The Customer is interested in tracking inventory, detecting movement, and locating misplaced items. Customer has 100 locations in a single region.

Immediacy: Data does not need to be acted on immediately, hourly reporting will be sufficient.

Volume: Each reader will scan 100s of items, detecting presence or movement.

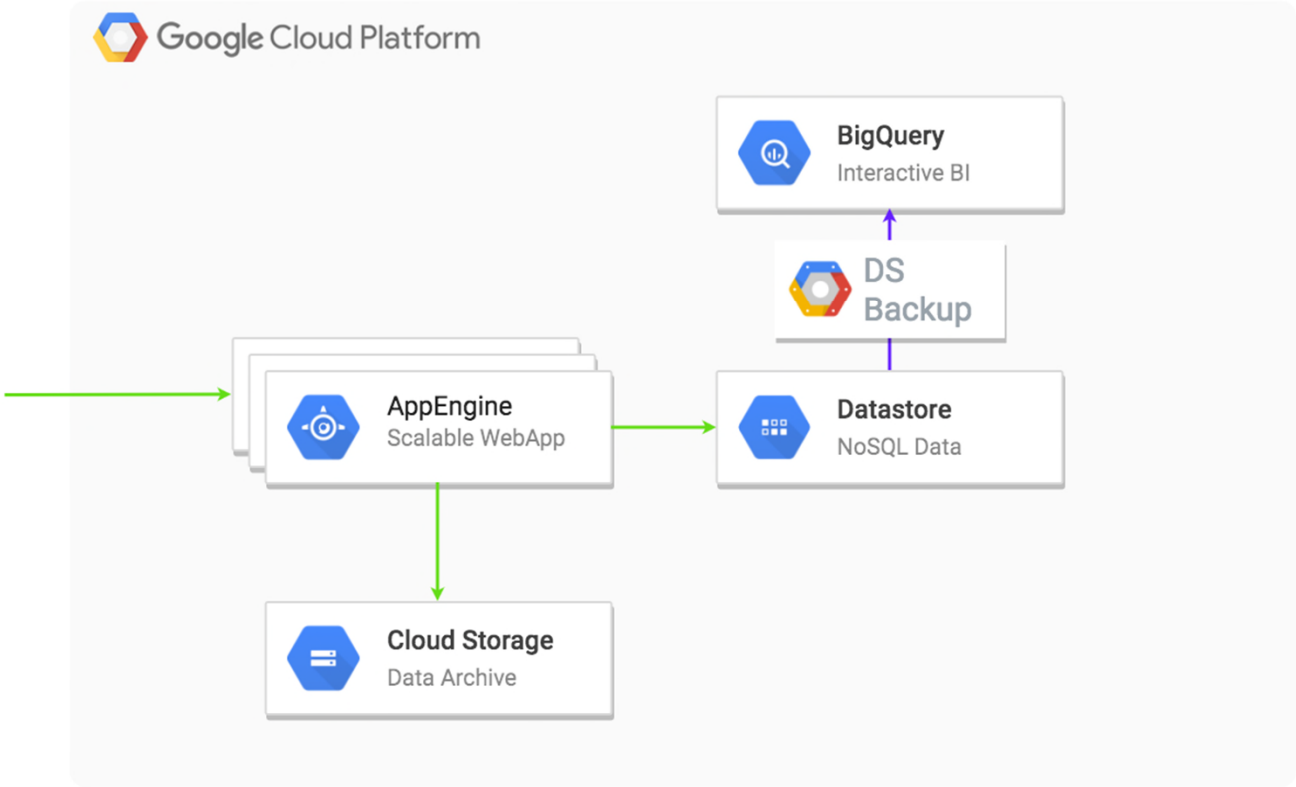
Complexity: Presence of item is primary metric, but movement can be reported to marketing database. Proper location can be determined from database as well.

Security: In our case, the presence of items (or identity of items) is low risk information. Security may be more of an issue if the RFID tags are used at checkout.



Courtesy iTrack All

Smart Shelf: Medium Volume, Medium Immediacy, Low Complexity, Low Security



Attendee Tracking

Customer is running a SaaS offering across Europe that provides event attendee tracking. Events typically have paid admission and have “red zones” requiring access permissions. They also track the movement of individuals within the event to measure the popularity of certain presentations and booths.

Immediacy: Gathering of data is done primarily for analytic purposes. No immediate action needs to be taken on reads.

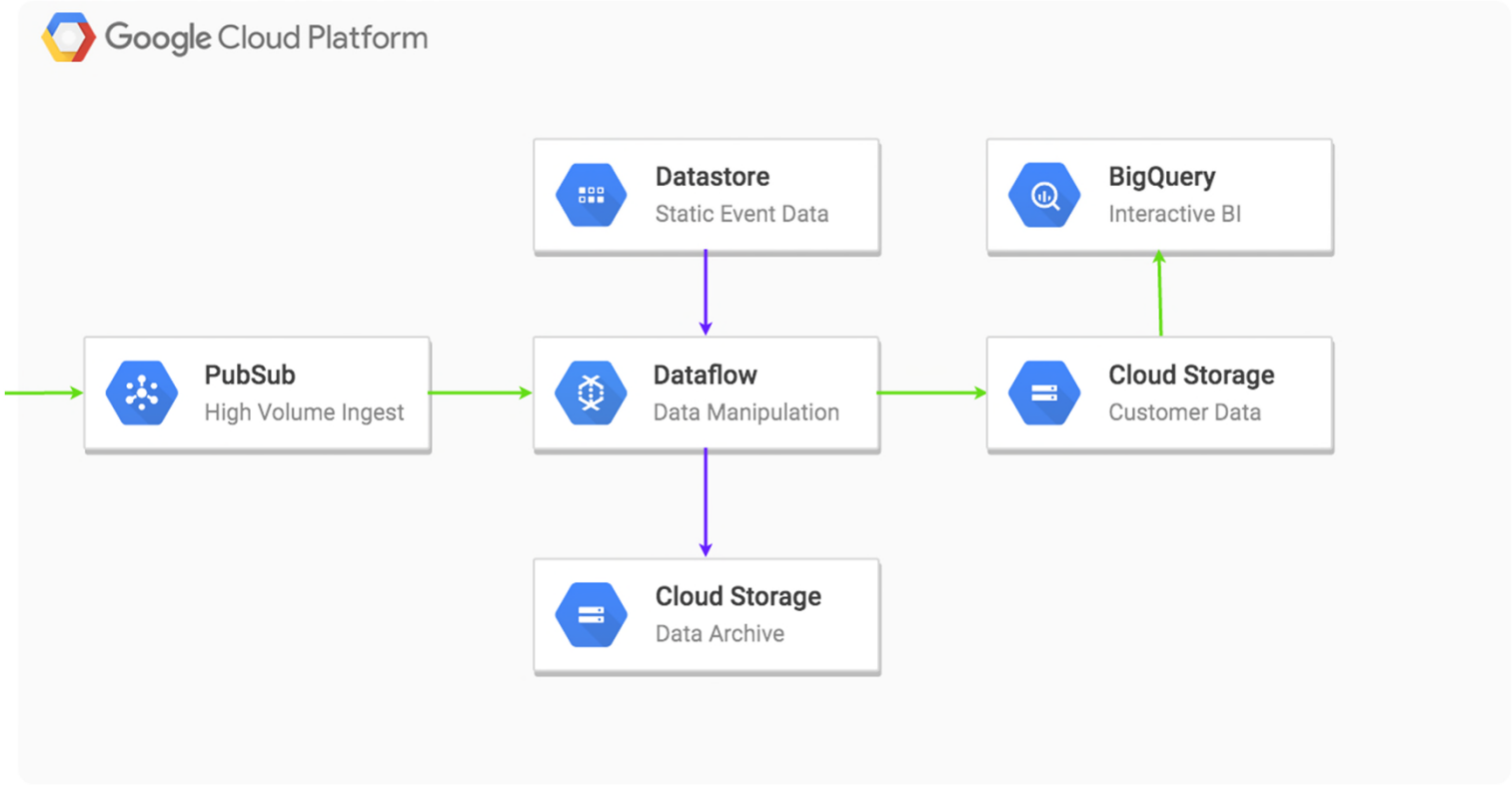
Volume: Depends on crowd size, let’s assume 10s of thousands. Reads are fairly frequent, with a combination of area reads and short range “check ins”.

Complexity: The Analytics will require aggregation and combination of data. Ideally we should be able to track individual paths through the event as well as group averages.

Security: Like many use cases that reveal movement of personnel, we want to protect this information.

Courtesy Impinj

Attendee Tracking: High Volume, Low Immediacy, High Complexity, Medium Security



Pharmaceutical Inventory

Customer needs to account for high-value inventory for financial and compliance reasons. They also want to detect presence of personnel near storage, and track movement of the drugs through the building.

Immediacy: Because we want to detect theft or mis-application of the drugs, any suspect movement should be detected quickly.

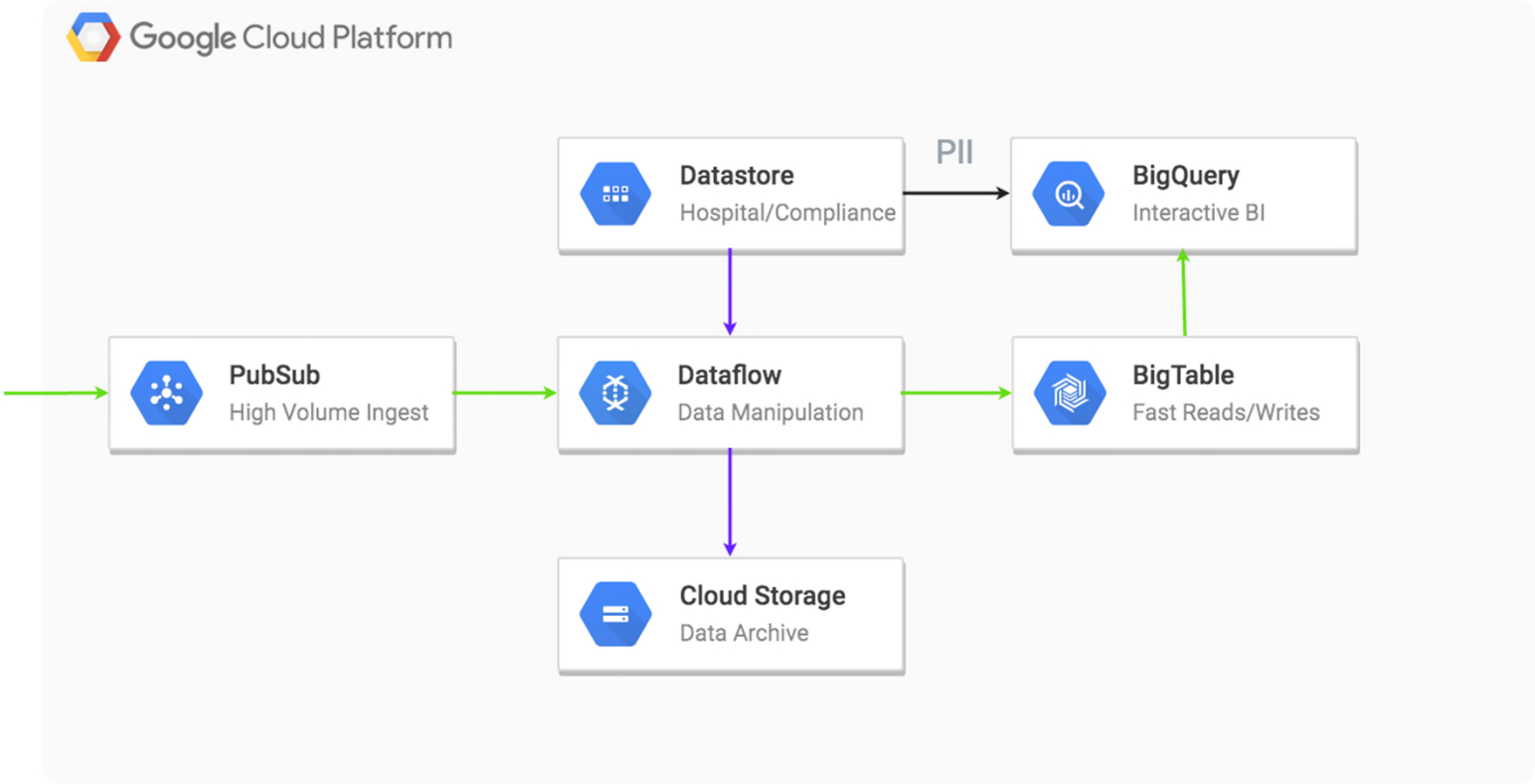
Volume: Readers will scan 1000s to 10's of thousands of items. Movement of items adds to volume, as well as the scanning of personnel badges.

Complexity: Builds on the Smart Shelf use case, but also requires "history" of item movement and cross-index with personnel/security databases. Can also track movement for potential mis-application. Compliance may also be an issue.

Security: Like many use cases that have the capability to expose personal medical information or movement of personnel, we want to protect this information.

Courtesy Impinj

Pharmaceutical Tracking: High Volume, High Immediacy, High Complexity, High Security



Outline

Summary

1. **Start with Use Cases**

1. **Characterize the Data**

1. **Find Patterns**

1. **Experiment and Iterate**

Thank You!

Google Cloud Home:
https://cloud.google.com

How to Use Google Cloud:
https://cloud.google.com/docs/tutorials

stevejohn@google.com