Script Slums to Beam Scrapers

By Shailesh Mangal
A perfect world...
In a perfect world...

Data Warehouse
- It's landed in RAW
- 100% cloud based
- Main data warehouse.

Device's readings, sensor data activity across the service

This data is written into our pipeline

Data Visualization tools
- Data Studio
- Power BI
- Tableau
- SAP Analytics Cloud

Customers

Data movement

Fast access storing
data is made available for data visualization tools.
Data at Scale

1+ billion messages written

pipeline

DW storage
The reality...
But in reality...

Device's readings, sensor data activity across the service

This data is written into our pipeline

Data Warehouse
- It's landed in RAW
- 100% cloud based
- Main data warehouse

RAW

Bigdata processing pipeline tools

DW
- Transformed & landed into a data warehouse

RL
- Bootstrapping
- Aggregating
- Summarized

Fast access storing data is made available for data visualization tools.

Customers

Data Visualization tools
- Data Studio
- Power BI
- Tableau

SAP Analytics Cloud

Data movement
Data access

Austin, 2022
ROBBIE, STOP MISBEHAVING OR I WILL SEND YOU BACK TO DATA CLEANING!

MACHINE LEARNING CLASS

DIRTY DATA
About me

● Head of Engg @ Roambee
● Can write code
● Run for fun

https://www.linkedin.com/in/shaileshmangal/
@sxm20
Today’s agenda

- About Roambee
- Architecture
- Classic Approach
- Pipeline challenges
- Use of Apache Beam
- Road Ahead
Introducing Roambee
Its all about the Bee

- 10s Device Types
- 100s of Services
- 100,000s of Bees
- 1,000,000s of Sensors
- 1,000,000,000 of Messages
Data anywhere, at anytime…*

*almost in anything.
Problems we solve

- Real time visibility
- Operational control thro’ Exception signals
- Lane & Transporter Analysis
- Location Intelligence
  - Multimodal ETA
  - Stoppage Analysis
- Outlier detection
- Risk prediction & Prevention
- Demand prediction
Be reliably actionable with Roambee

Leveraging sensor data + curated data to validate every milestone

Shipper | Origin Warehouse | Origin Port/Airport | Freight Forwarder | Ocean/Air Freight | Destination Port/Airport | Freight Forwarder | Consignee
---|---|---|---|---|---|---|---

- **Package missed at dispatch?**
- **Temperature trend seems bad for shelf life?**
- **Alerts delivery and planning teams**
- **Replacement dispatched**
- **Issued Resolved**
- **Shipment opened or deviated route?**
- **Shipment arrived at right location?**

- **Anomaly signal triggers workflows**
- **Checks for location**
- **Package missed or on wrong flight?**
- **AI identifies if risk is real or not**
- **Shipment arrived in right condition?**

- **Dispatch team gets alerted**
- **Alerts specific driver or airport in-charge**
- **If regular inspection, no action!**
- **If not regular? Control Tower escalates**
- **Shipment delivered to right custodian?**

- **No response? Control Tower calls them**
- **No response? Control Tower calls them**
- **Replacement dispatched**
- **Issued Resolved**
- **If not, Control Tower intervenes**

---

Austin, 2022
Classic Pipeline Architecture

SOURCE
- RDBMS
- NoSQL
- APIs

Optimizer

Datawarehouse
- Index

Query
Aggregate
Load

Dashboard

Austin, 2022
Technology Exploration

- Hadoop
- Apache Spark
- Apache Storm
- Flink
- Beam
- Trino
- Dagster
PROTESTING AGAINST NEW TECHNOLOGY - THE EARLY DAYS

© apatheticlemming
Pipeline Using BEAM
Communication Recency

Pipeline

| “Get Latest Bees”, | JdbcIO.<String>read() |
| “Get Msg for Each Bee”, | ParDO(CassandraIO.<BeeMsg>read()) |
| "Flattering all messages", | FlatMapElements.via(InferableFunction<KV<String, List<BeeMsg>>, Iterable<BeeMsg>>()) |
| "Partitioning by Geo Tech", | Partition.of(n, new Partition.PartitionFn<BeeMsg>()) |
| “Group by Msg Recency”, | WithKeys.of("Location Tech") ⇒ GroupByKey |
| “Write count”, | FileIO.Write(options.getOutput() /*Count*/) |
Why we chose BEAM

From python to Beam
>2d → < 1hr

- Simple Programming Model
- Clean documentation
- Native support for Batch and Streaming
- Choice of dev Language
- Testability
- Choice of deployment
Analysis

- Movement Pattern Analysis
- Geofence Elasticity
- Stoppage Patterns
  - Stop Time
  - Stop Frequency
- Travel mode
  - Airline Performance
  - Transporter performance
- Dwell Time
  - Port Dwell
  - Warehouse Dwell
Questions?

https://www.roambee.com
@sxm20
https://www.linkedin.com/in/shaileshmangal/