Script Slums to Beam Scrapers

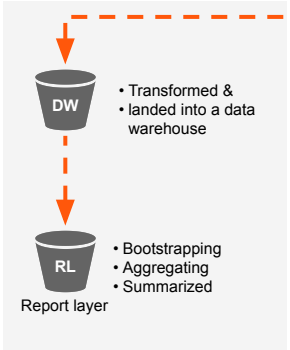
By Shailesh Mangal

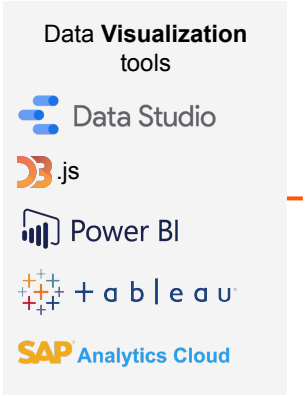
BEAM SUMMIT Austin, 2022

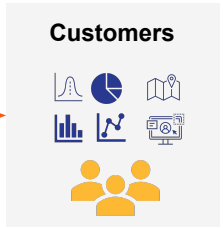
A perfect world...

BEAM SUMMIT Austin, 2022

In a perfect world...



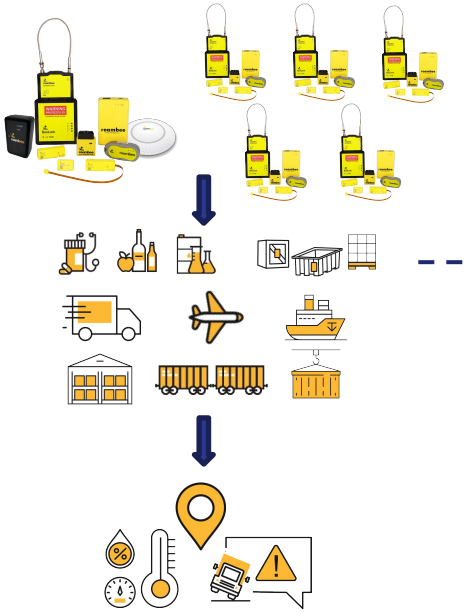




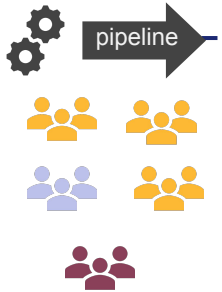
Data movement →

Data access - - →

Data at Scale



1+ billion messages
written

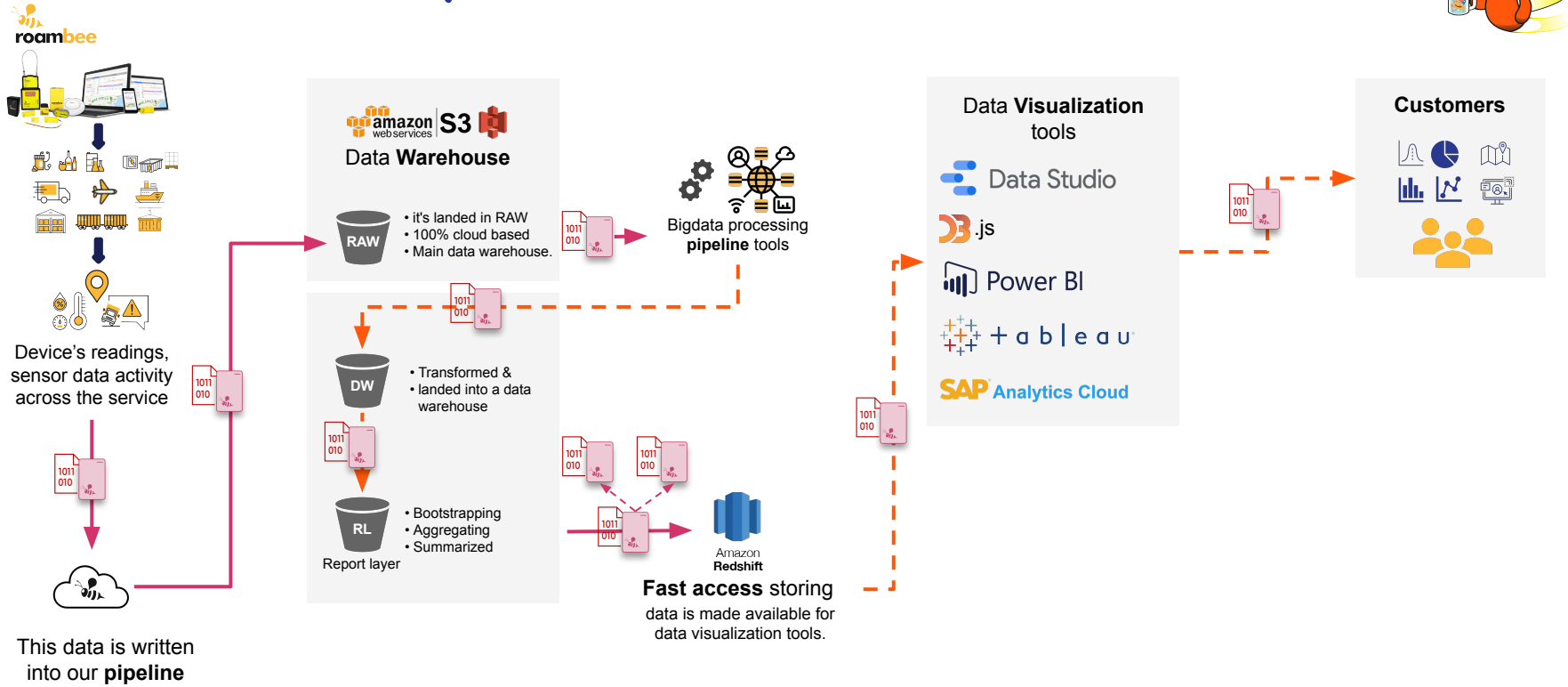


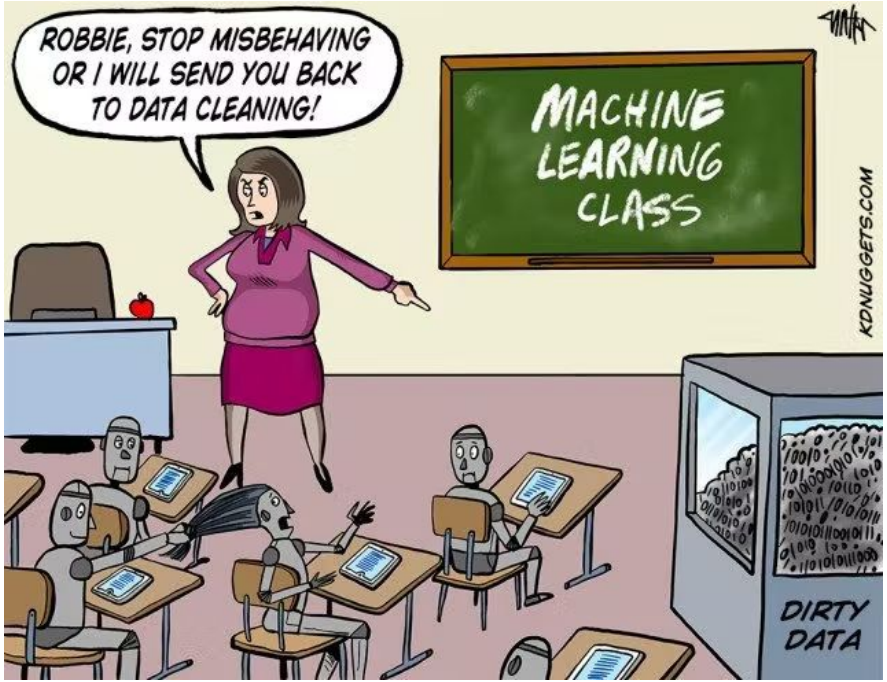
DW **storage**

The reality...

BEAM SUMMIT Austin, 2022

But in reality...





Source :
https://hackernoon.com/_next/image?url=https%3A%2F%2Ffirebasestorage.googleapis.com%2Fv0%2Fb%2Fhackernoon-app.appspot.com%2Fo%2Fimages%252FwMbJVdeGT1X6HWDd10iTNCitlsF3-g73q3u3u.jpeg%3Falt%3Dmedia%26token%3Dc92feb45-cd0a-4954-9bb1-0f78905b4270&w=1920&q=75

About me

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

https://www.linkedin.com/in/shaileshmangal/

@sxm20

Today's agenda

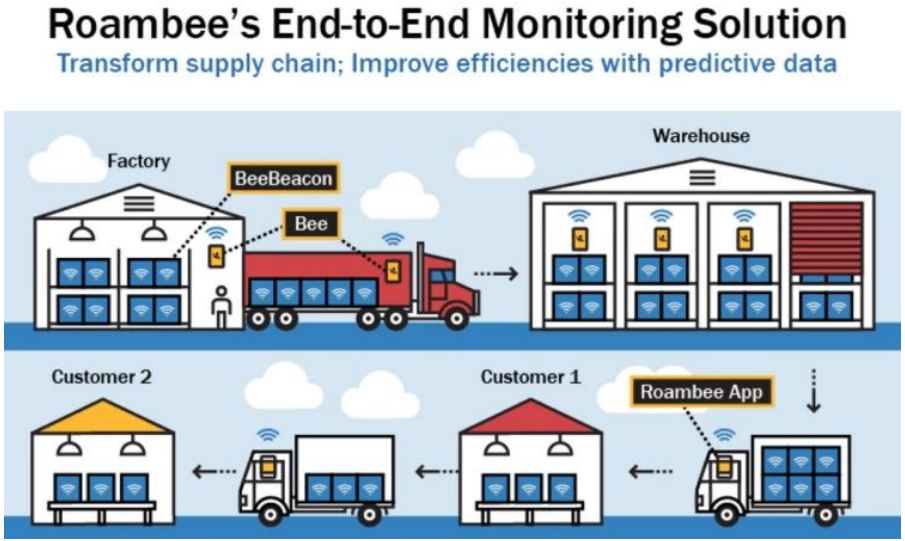
- About Roambee
- Architecture
- Classic Approach
- Pipeline challenges
- Use of Apache Beam
- Road Ahead

Introducing Roambee

BEAM SUMMIT Austin, 2022

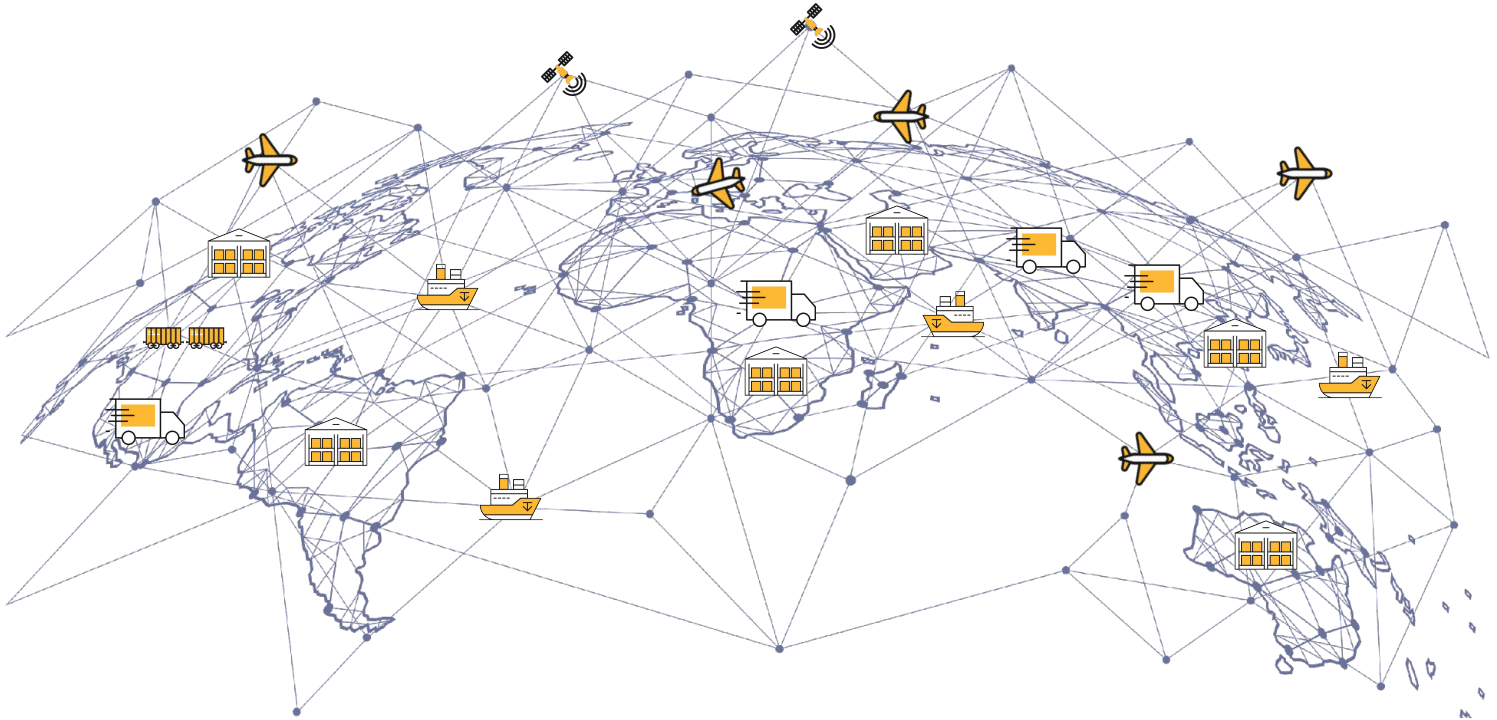
Its all about the Bee

- 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...*

roambee



*almost in anything.

Problems we solve

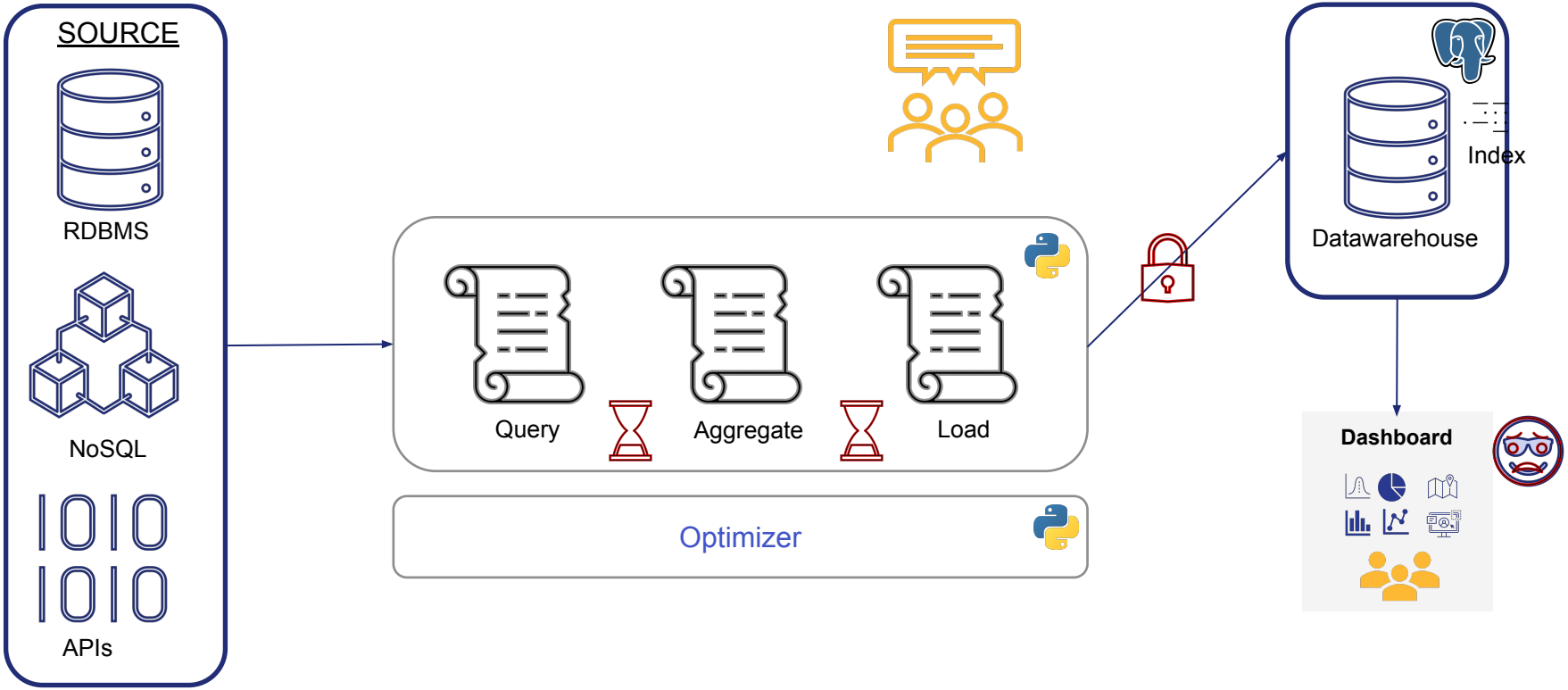
- Real time visibility
- Operational control thro’ Exception signals
- Lane & Transporter Analysis
- Location Intelligence
  - Multimodal ETA
  - Stoppage Analysis
- Outlier detection
- Risk prediction & Prevention
- Demand prediction

Be reliably actionable with Roambee

Leveraging sensor data + curated data to validate every milestone

Freight Forwarder	Freight Forwarder	Ocean/Air Freight	Freight Forwarder	Freight Forwarder
Package missed at dispatch?	Temperature trend seems bad for shelf life?	Package missed or on wrong flight?	Shipment opened or deviated route?	Shipment arrived at right location?
Anomaly signal triggers workflows	Checks for location	Alerts delivery and planning teams	AI identifies if risk is real or not	Shipment arrived in right condition?
Dispatch team gets alerted	Alerts specific driver or airport in-charge	Replacement dispatched	If regular inspection, no action!	Shipment delivered to right custodian?
No response? Control Tower calls them	No response? Control Tower calls them	✓ Issued Resolved	If not regular? Control Tower escalates	If not, Control Tower intervenes
✓ Issued Resolved	✓ Issued Resolved		✓ Issued Resolved	✓ Issued Resolved

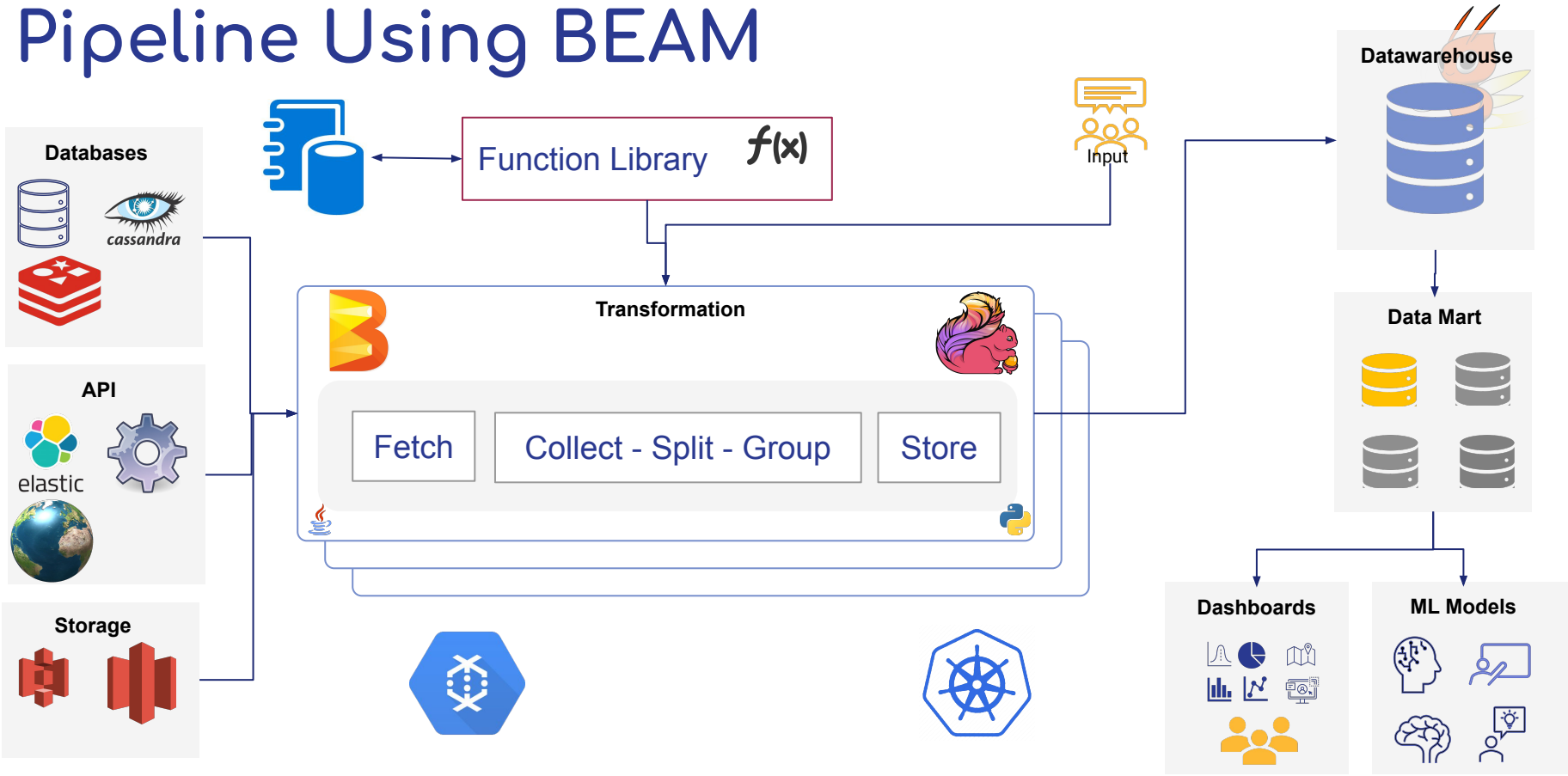
Classic Pipeline Architecture



Technology Exploration

© apatheticlemming

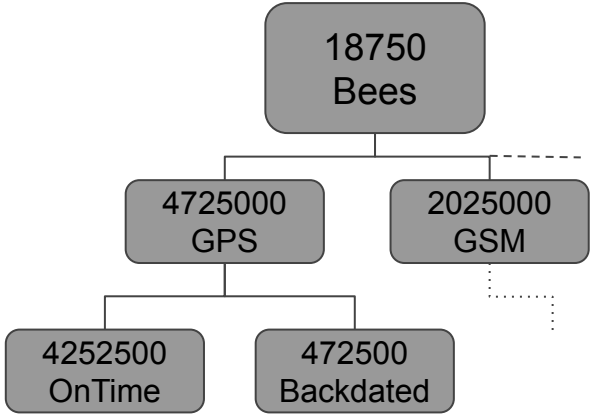
Pipeline Using BEAM



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/

BEAM SUMMIT Austin, 2022