



Hortonworks Data Flow Workshop

Thiago Santiago
Solutions Engineer - LATAM

Who is here?



ArcelorMittal



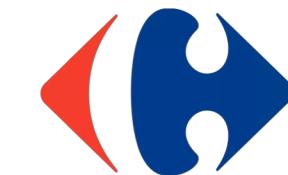
Bradesco



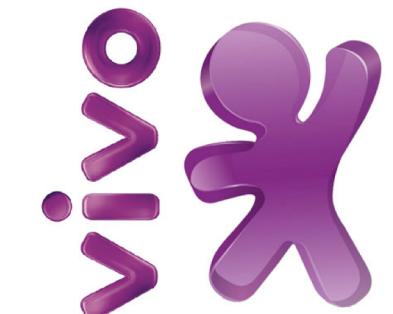
GrupoFleury



via varejo



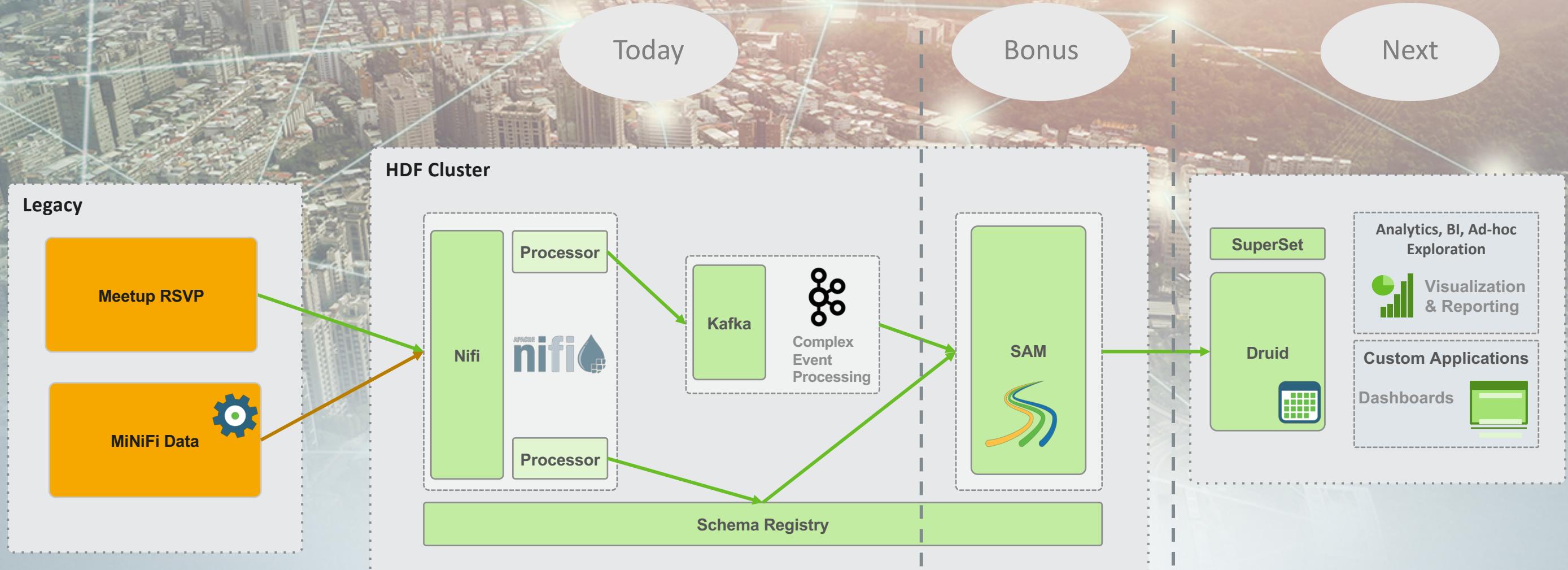
Carrefour



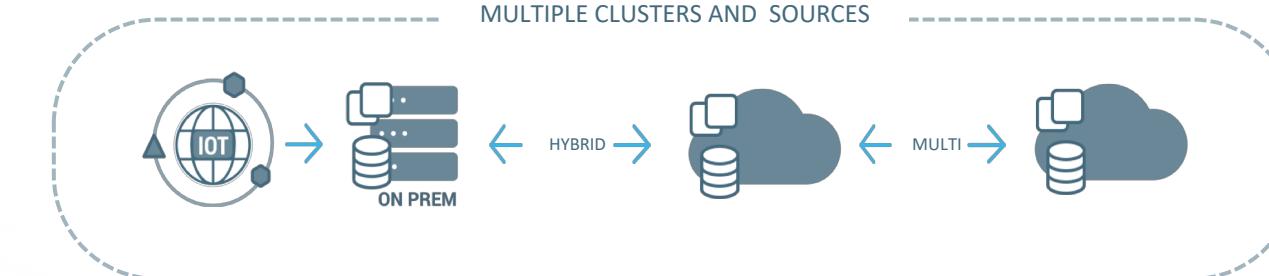
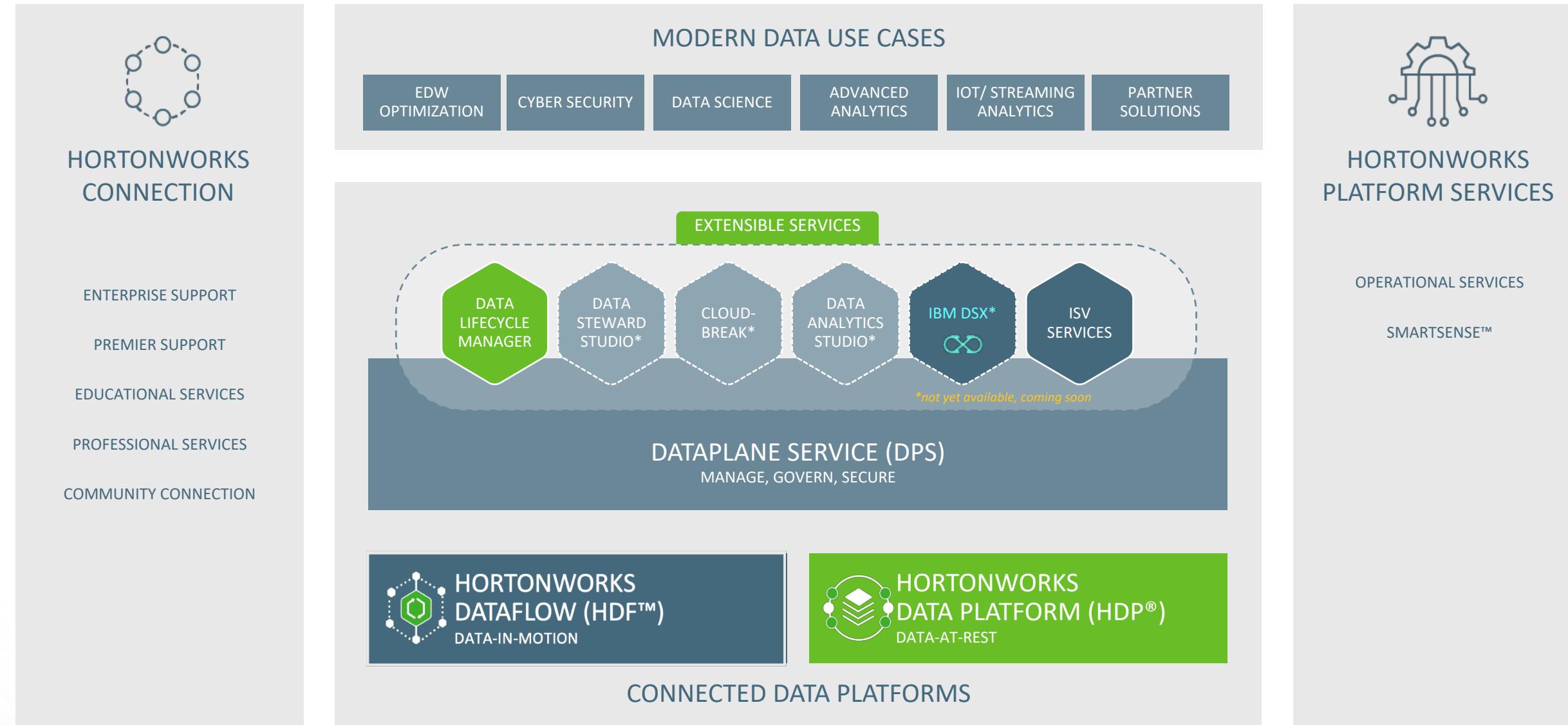
Workshop Agenda

- ◆ Access cluster
- ◆ Getting started with NiFi
- ◆ MiNiFi
- ◆ Kafka Basics
- ◆ Integrating Kafka with NiFi
- ◆ Integrating the Schema Registry
- ◆ Bonus Round : SAM

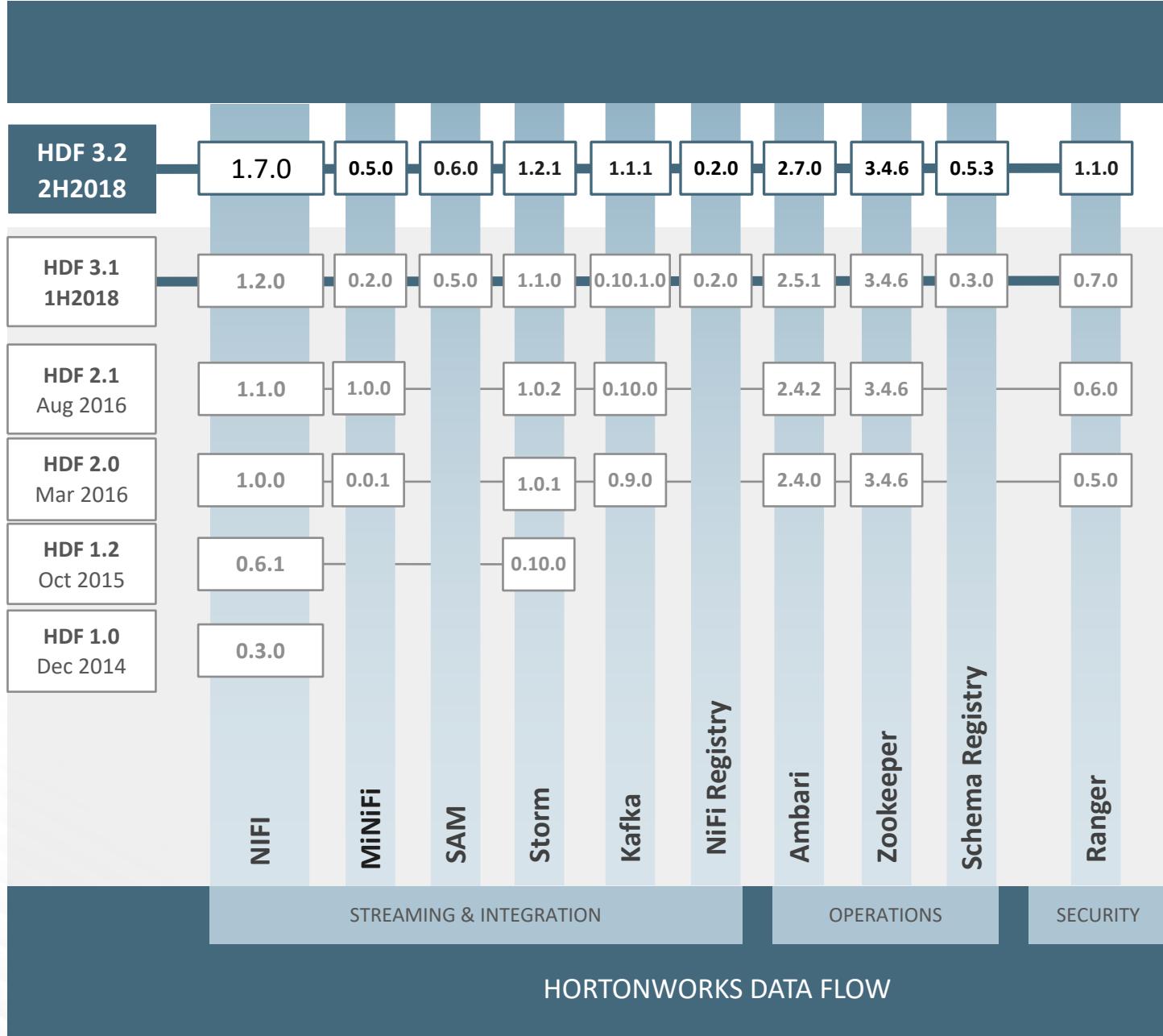
Workshop Agenda



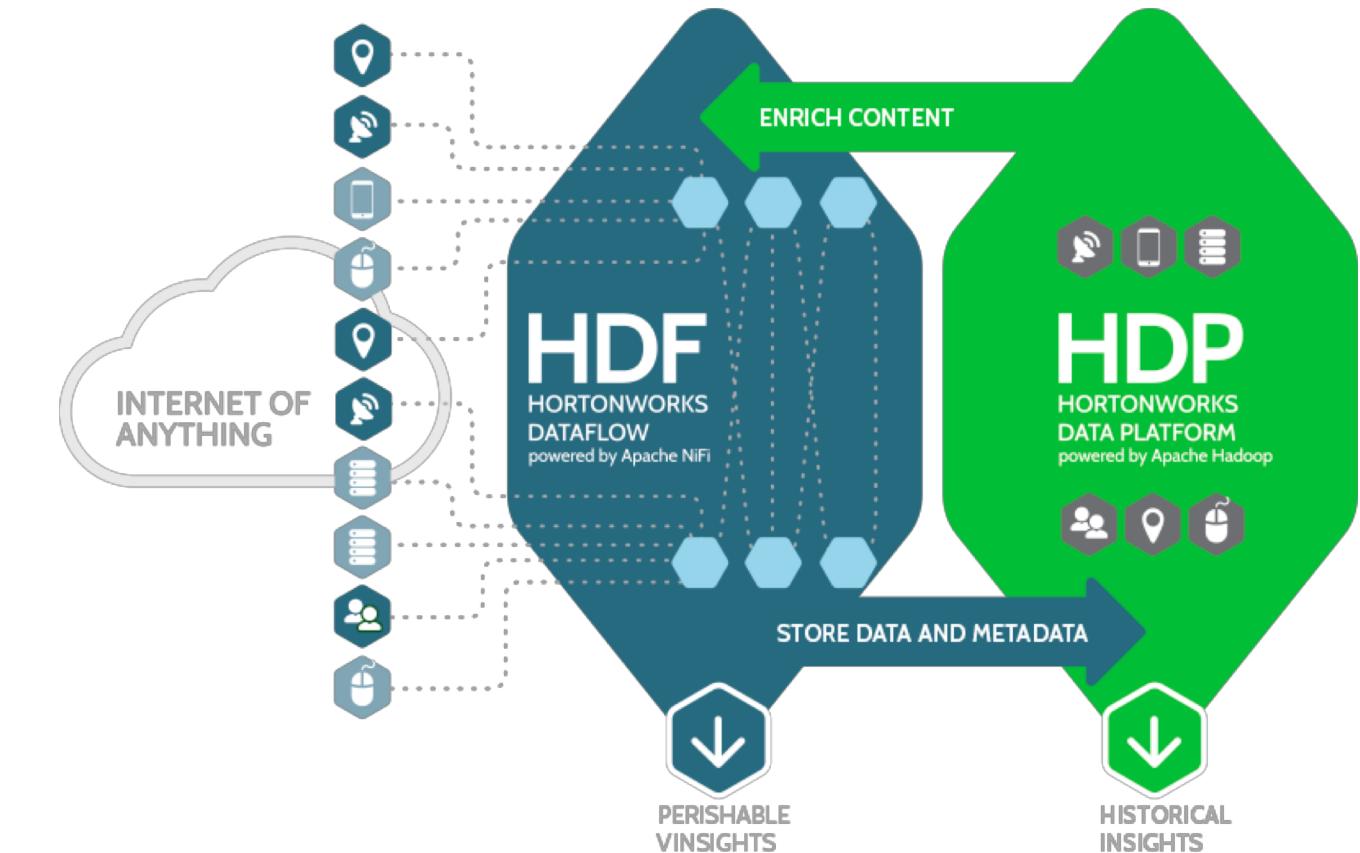
Global Data Management With Hortonworks



Hortonworks Data Flow 3.2



* HDF 3.0 – Shows current Apache branches being used. Final component version subject to change based on Apache release process.

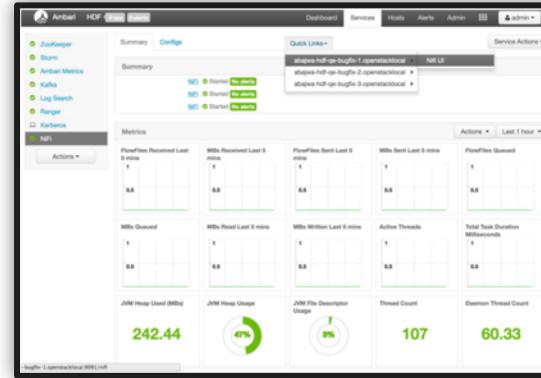


Enterprise Services

Centralized security, governance and operations for HDP and HDF

Apache Ambari

“Manage, Maintain & Monitor”



- Installation, patch and upgrade
- Configuration Management
- Alert & Monitor

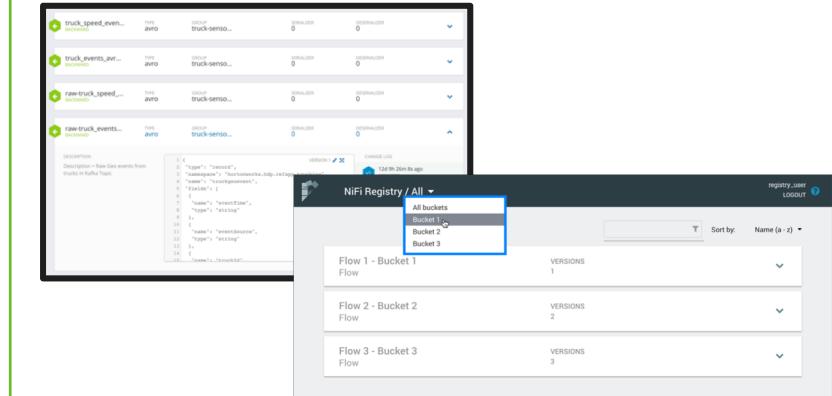
Apache Ranger/ Atlas

“Secure & Govern”

- Centralized authorization
- RBAC & ABAC fine grain authorization
- Key management
- Audit / Lineage

Schema & Flow Registry

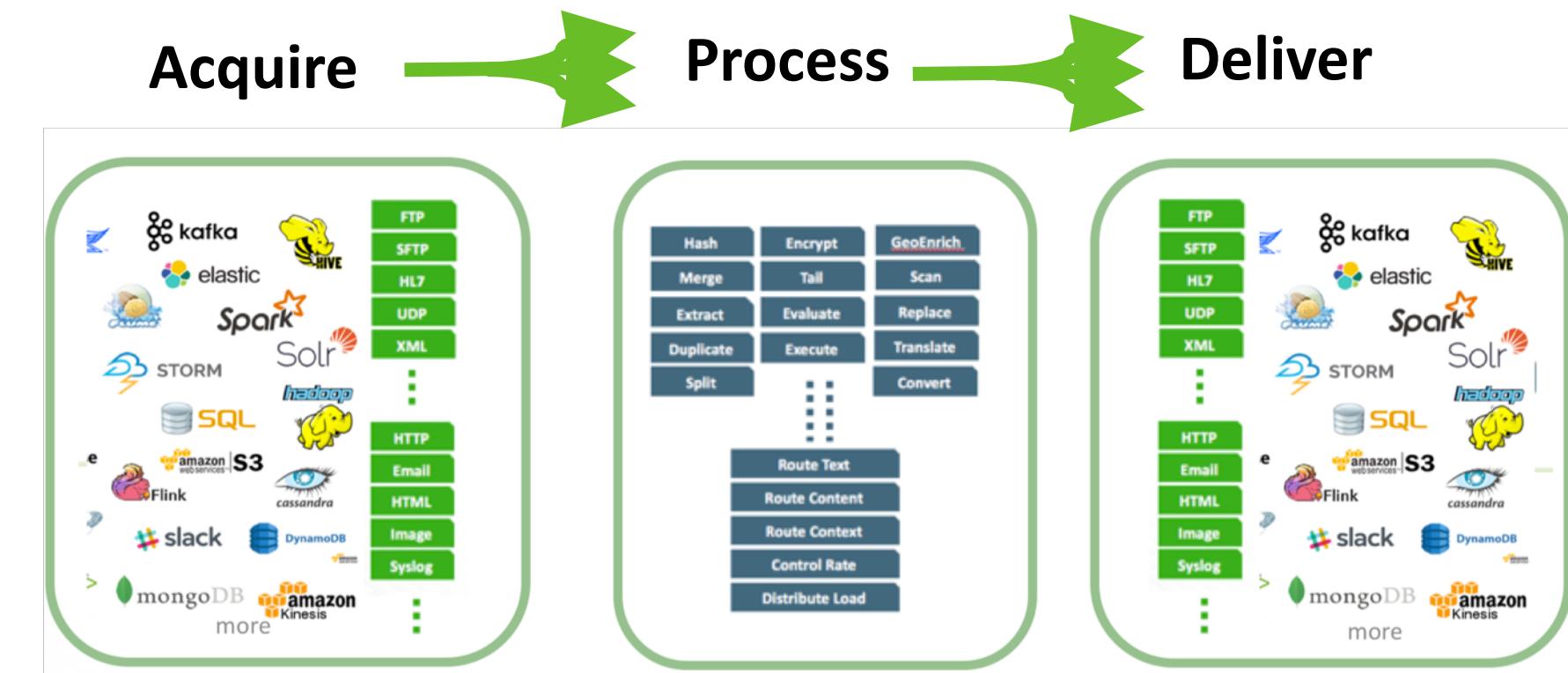
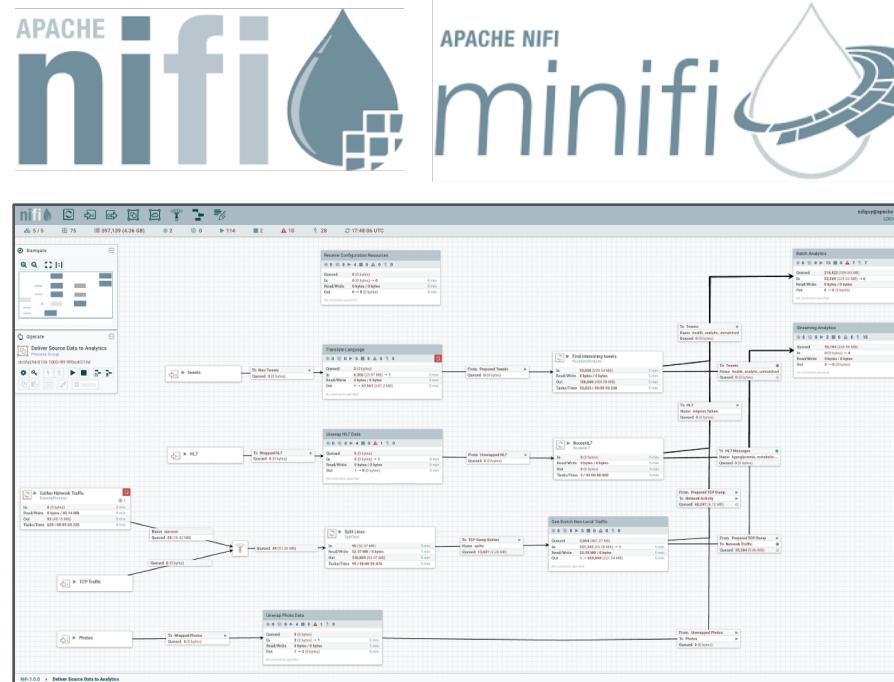
“Govern Schema”



- Provide Reusable Schemas/flows
- Version Management
- Format & Ensure Data Quality
- FDLC

Flow Management

Enable easy ingestion, routing, management and delivery of any data anywhere (Edge, cloud, data center) to any downstream system with built in end-to-end security and provenance



Advanced tooling to industrialize flow development (Flow Development Life Cycle)

- Over 360+ Prebuilt Processors
 - Easy to build your own
 - Parse, Enrich & Apply Schema
 - Filter, Split, Merger & Route
 - Throttle & Backpressure
 - Guaranteed Delivery
 - Full data provenance from acquisition to delivery
 - Diverse, Non-Traditional Sources
 - Eco-system integration

Hands on!