Apache NiFi: Ingesting Enterprise Data @ Scale

Timothy Spann

2017 Future of Data – Princeton Meetup
Hosted by TRAC Intermodal

HORTONWORKS®
POWERING THE FUTURE OF DATA™

DATA WORKS SUMMIT
Presented by Hortonworks and Yahoo!
HADOOP SUMMIT
DATAWORKS SUMMIT/HADOOP SUMMIT
JUNE 13–15, 2017
San Jose McHenry Convention Center
REGISTER NOW AND SAVE $1,000
REGISTER NOW! >
dataworkssummit.com

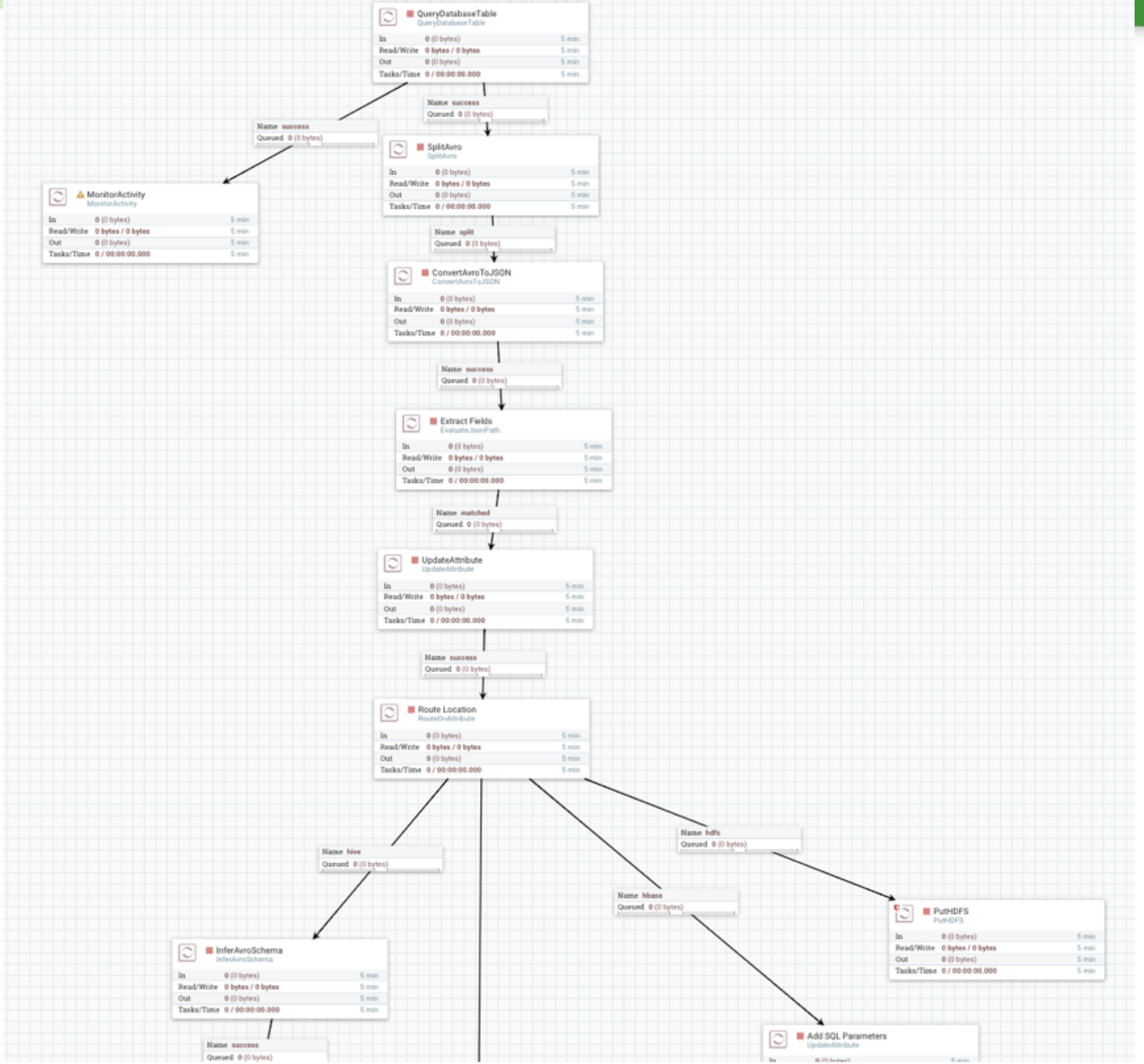
Agenda

- Apache NiFi RDBMS, EDI, JSON, CSV, Sensors

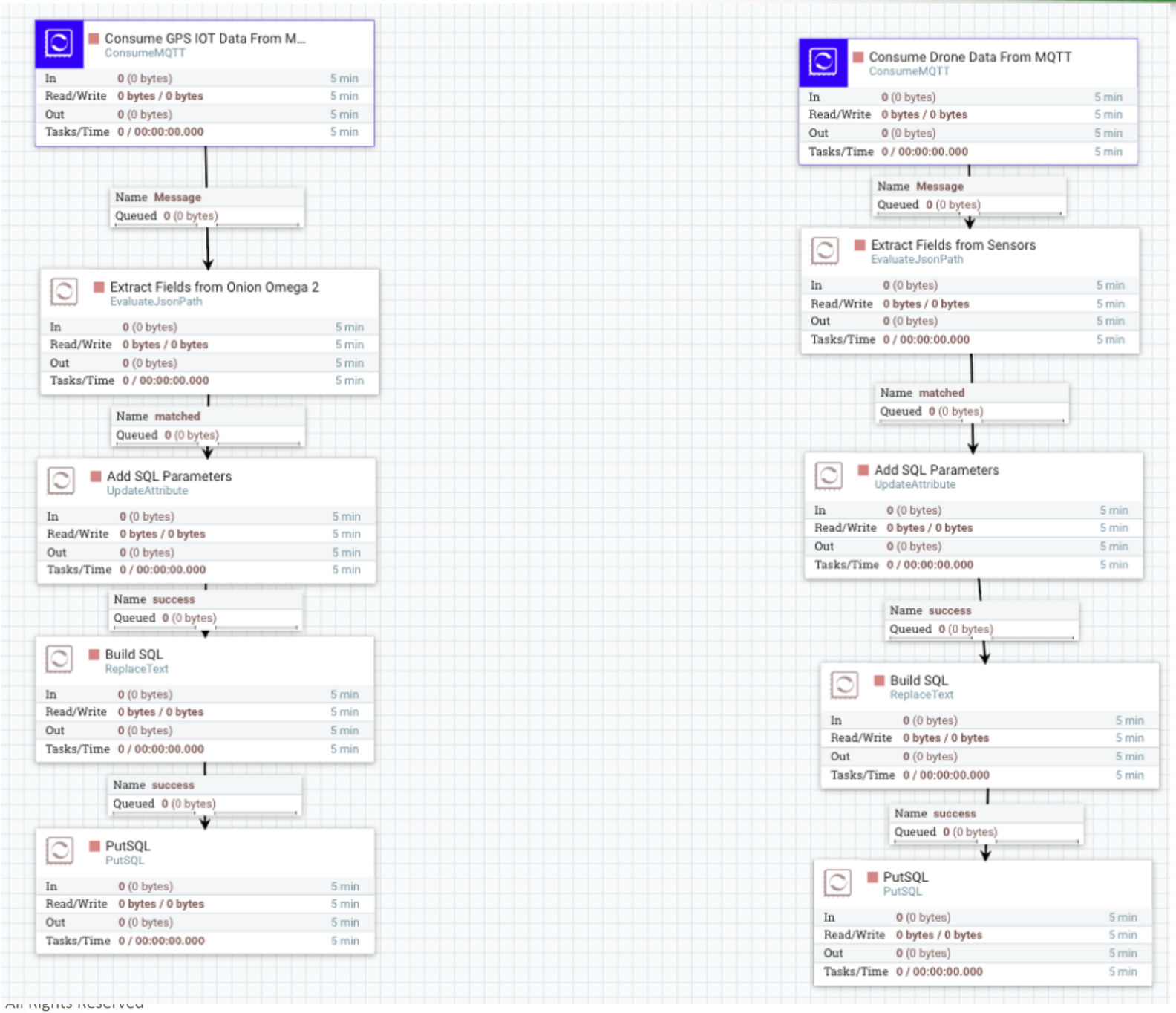
- EDI
- https://community.hortonworks.com/content/kbentry/59975/ingesting-edi-into-hdfs-using-hdf-20.html
- https://github.com/tspannhw/EnterpriseNIFI

© Hortonworks Inc. 2011 – 2017. All Rights Reserved

HORTONWORKS

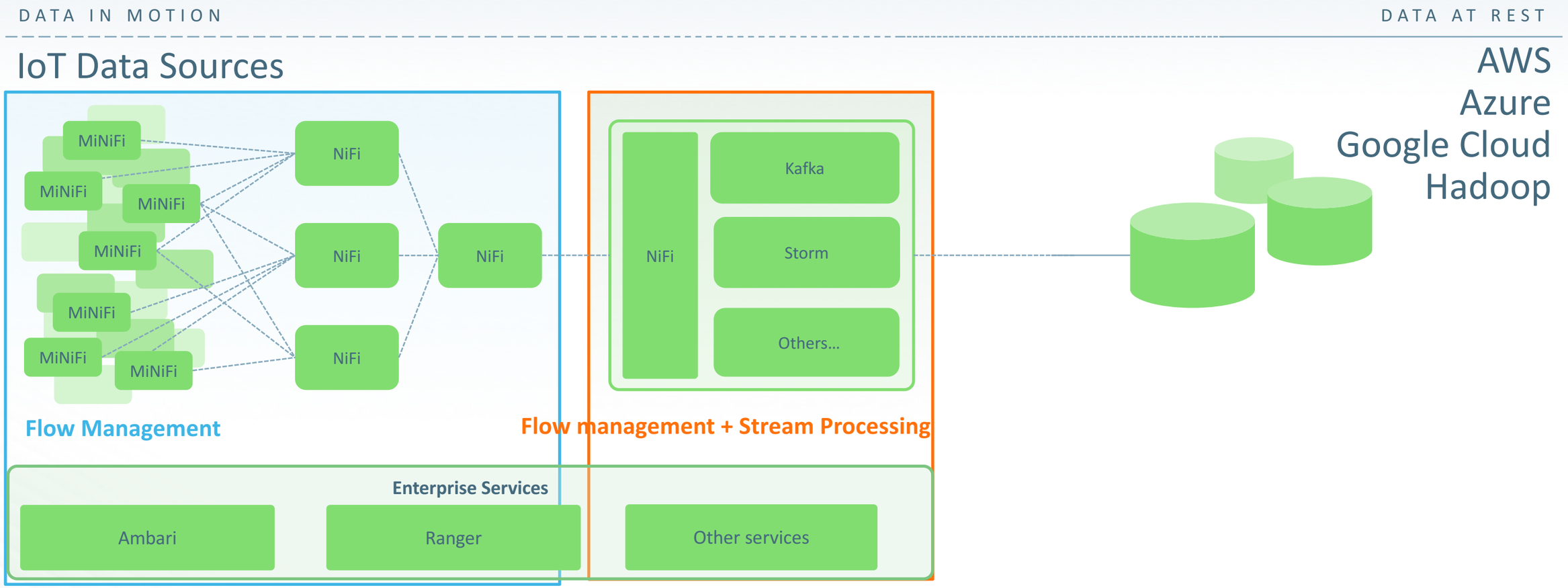


© Hortonworks Inc. 2011 – 2017. All Rights Reserved



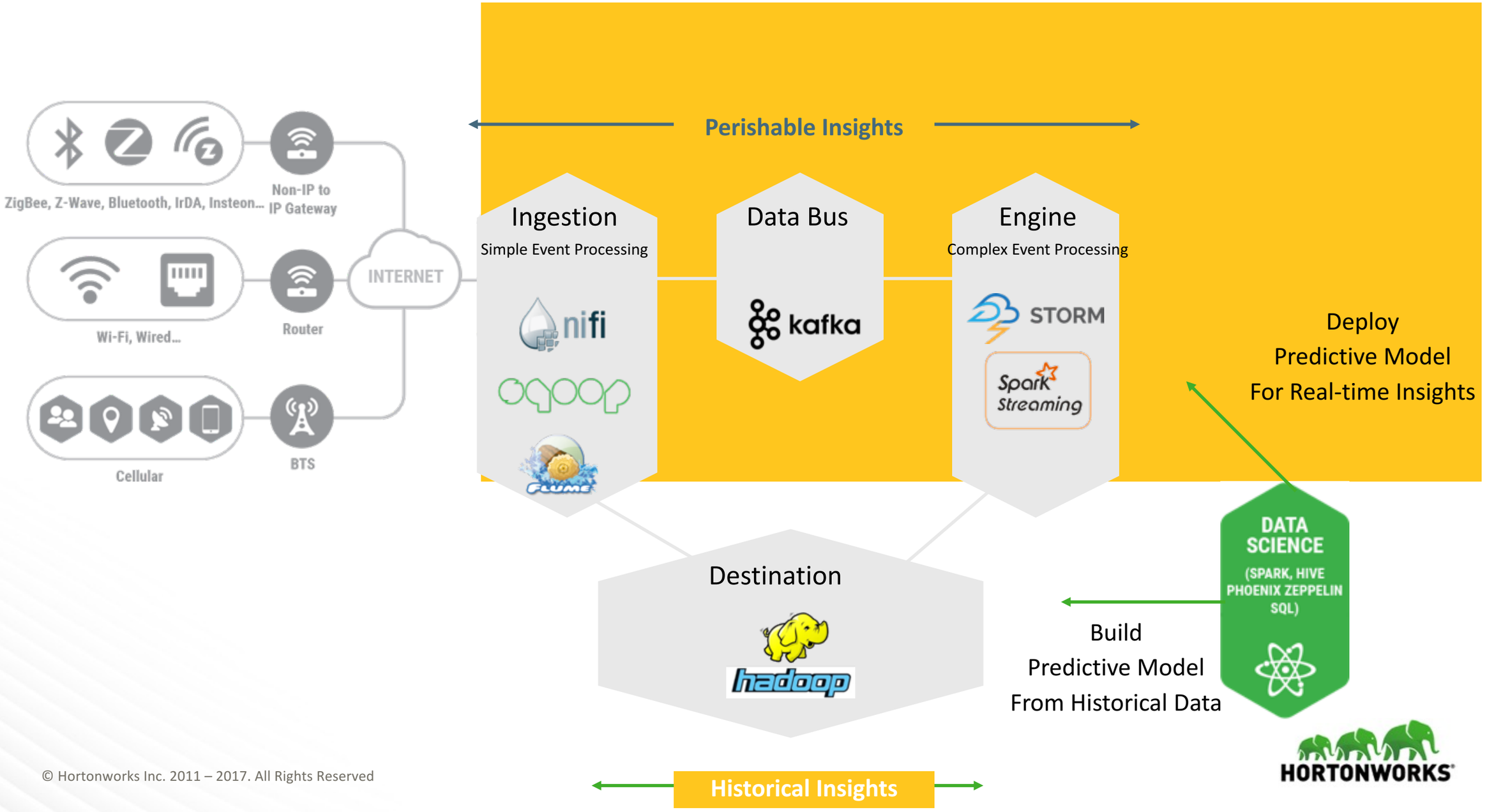
© Hortonworks Inc. 2011 – 2017. All rights reserved

HDF 2.1 – Data in Motion Platform



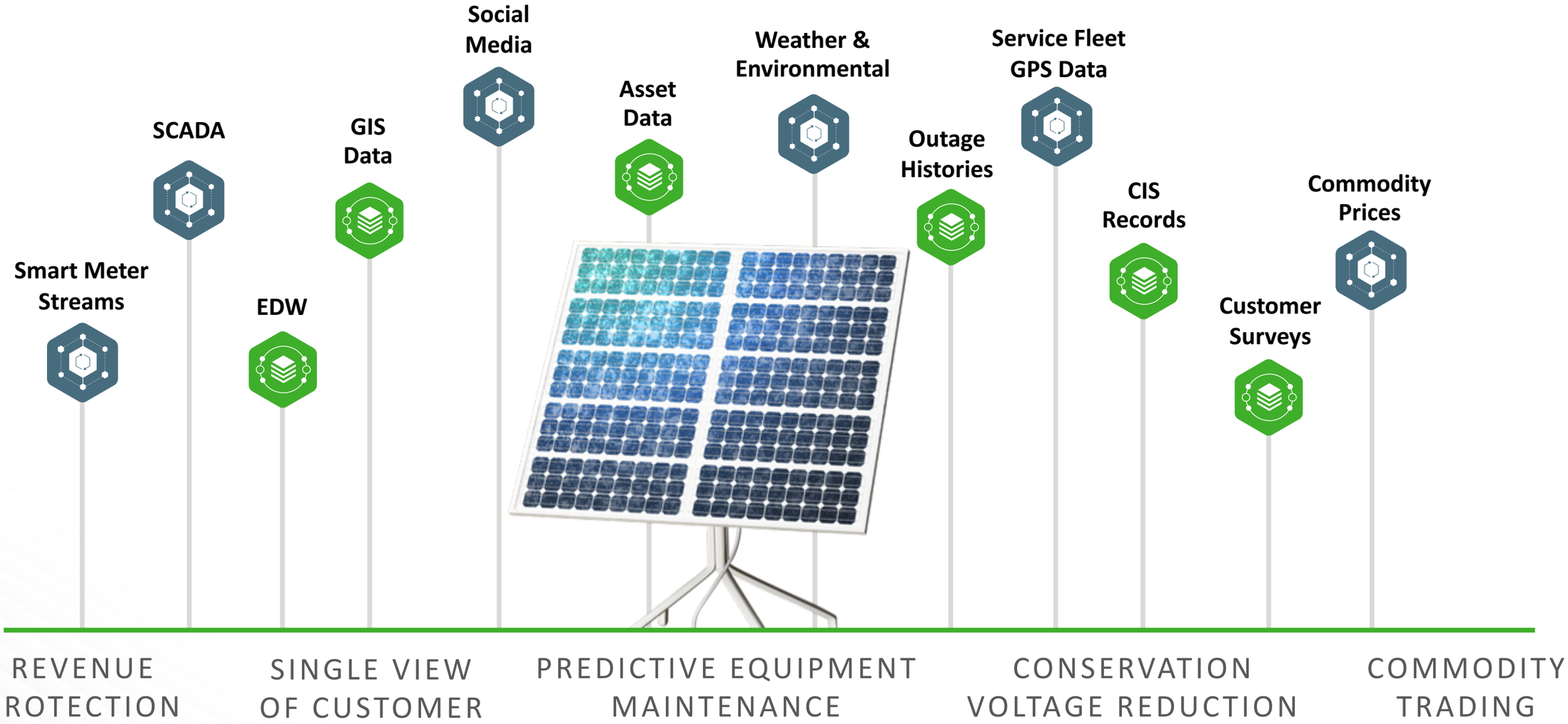
© Hortonworks Inc. 2011 – 2017. All Rights Reserved

Actionable Insights Architecture



© Hortonworks Inc. 2011 – 2017. All Rights Reserved

Actionable Intelligence Transforms Industrial, Transportation & Utilities



© Hortonworks Inc. 2011 – 2017. All Rights Reserved

What is Apache NiFi?

- Created to address the challenges of global enterprise dataflow
- Key features:
  - **Visual Command and Control**
  - **Data Lineage (Provenance)**
  - **Data Prioritization**
  - **Data Buffering/Back-Pressure**
  - **Control Latency vs. Throughput**
  - **Secure Control Plane / Data Plane**
  - **Scale Out Clustering**
  - **Extensibility**

© Hortonworks Inc. 2011 – 2017. All Rights Reserved

HORTONWORKS

Apache NiFi

What is Apache NiFi used for?

- Reliable and secure transfer of data between systems
- Delivery of data from sources to analytic platforms
- Enrichment and preparation of data:
  - Conversion between formats
  - Extraction/Parsing
  - Routing decisions

What is Apache NiFi NOT used for?

- Distributed Computation
- Complex Event Processing
- Complex Rolling Window Operations

© Hortonworks Inc. 2011 – 2017. All Rights Reserved

HORTONWORKS

NiFi Terminology

FlowFile

- Unit of data moving through the system
- Content + Attributes (key/value pairs)

Processor

- Performs the work, can access FlowFiles

Connection

- Links between processors
- Queues that can be dynamically prioritized

© Hortonworks Inc. 2011 – 2017. All Rights Reserved

HORTONWORKS

Contact:

Timothy Spann @PaaSDeV

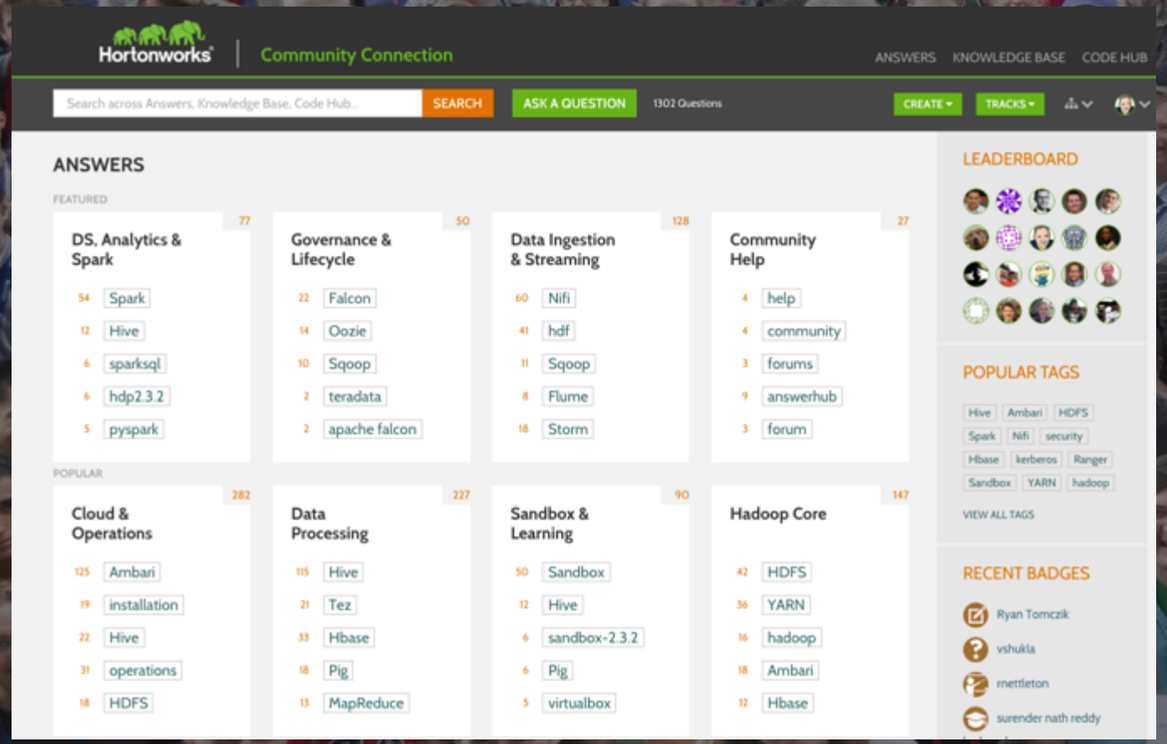
www.meetup.com/futureofdata-princeton

community.hortonworks.com/users/9304/tspann.html

© Hortonworks Inc. 2011 – 2017. All Rights Reserved

HORTONWORKS

Hortonworks Community Connection



- Full Q&A Platform (like StackOverflow)
- Knowledge Base Articles
- Code Samples and Repositories

Read access for everyone, join to participate and be recognized

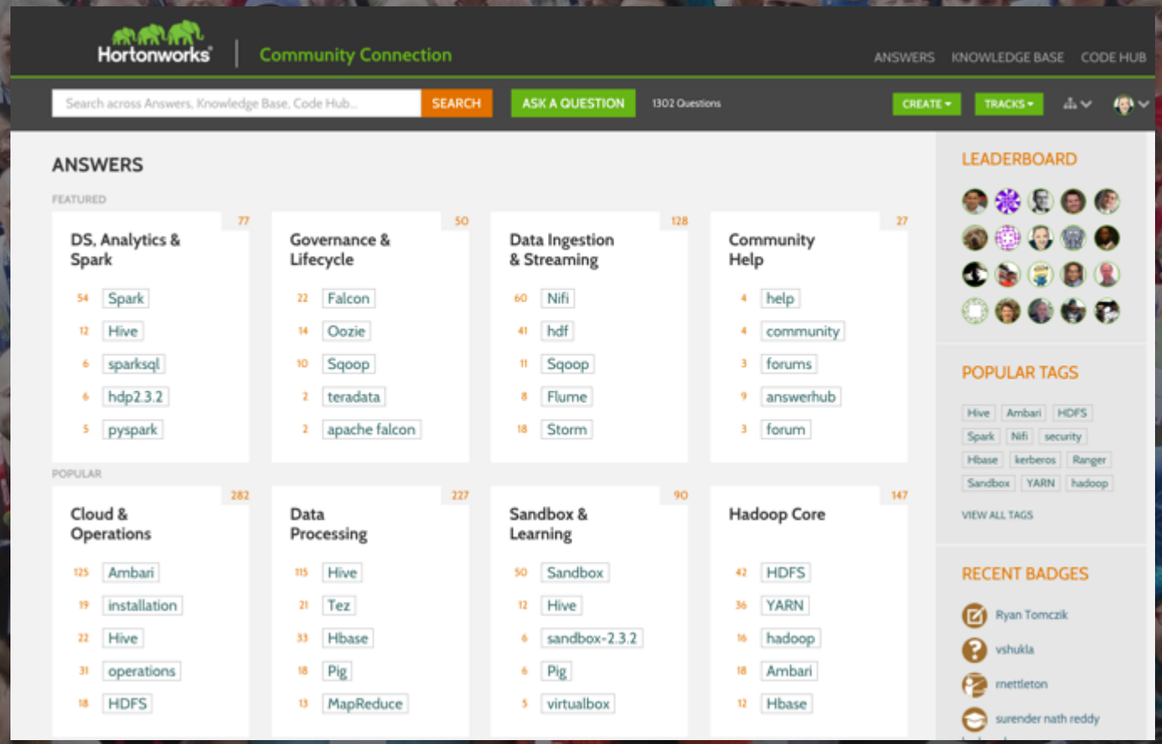
Community Engagement

4,000+
Registered Users

10,000+
Answers

15,000+
Technical Assets

One Website!



© Hortonworks Inc. 2011 – 2015. All Rights Reserved

Participate now at: community.hortonworks.com