You've got two weeks: Why you shouldn't discuss cool ideas near your manager

Dan Chaffey

Itinerant Hacker and Solutions Engineer, Hortonworks

About Hortonworks

Customer Momentum

- ~1000 customers
- >150 customers added per quarter
- Publicly traded on NASDAQ: HDP

The Leader in Connected Data Platforms

- Hortonworks DataFlow for data in motion
- Hortonworks Data Platform for data at rest
- Powering new modern data applications

Partner for Customer Success

- Leader in open-source community, focused on innovation to meet enterprise needs
- Unrivaled support subscriptions

HORTONWORKS®

Founded in 2011

Original 24 Architects, Developers, Operators of Hadoop from Yahoo!

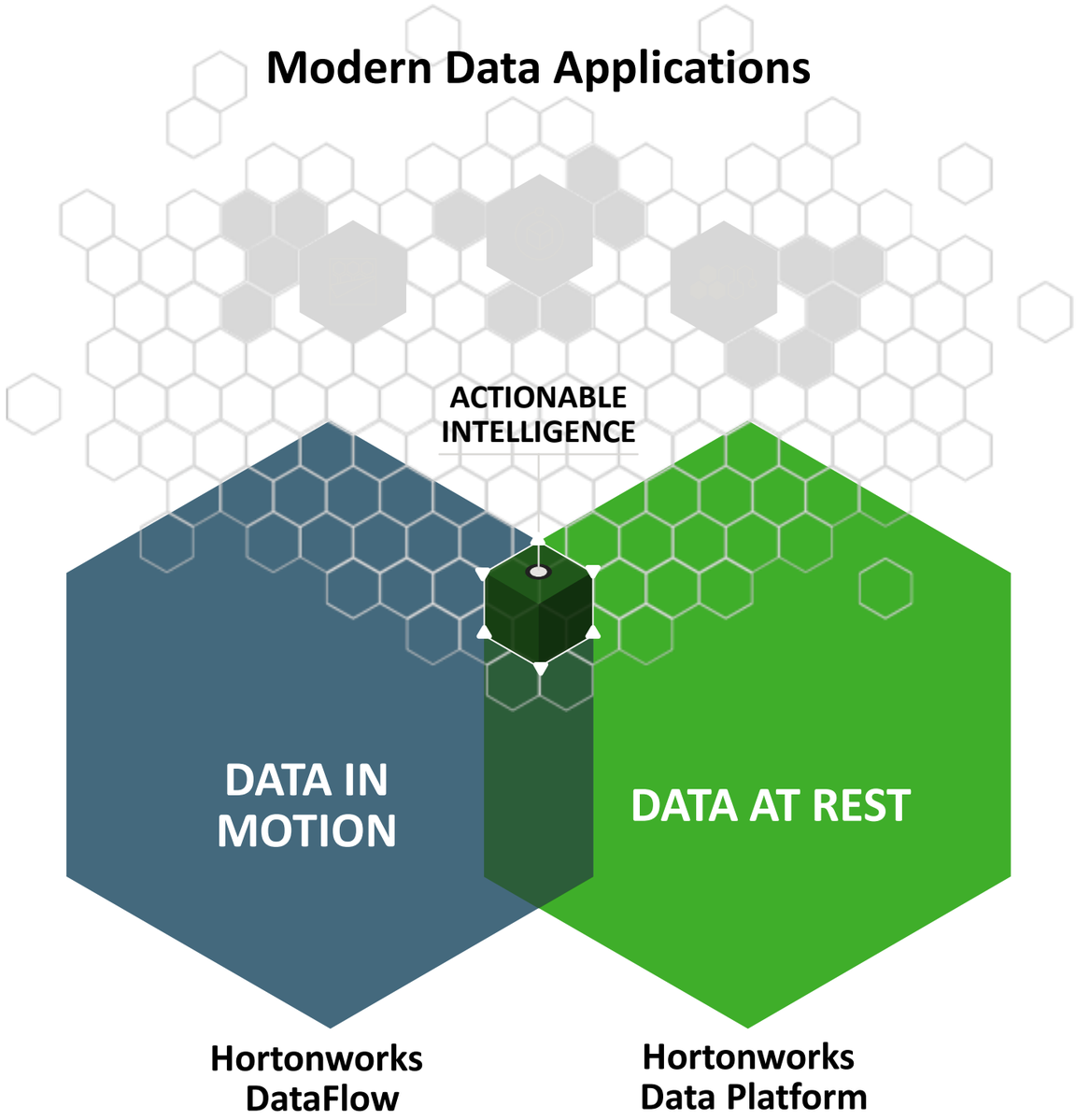
1000+ EMPLOYEES

1500+ ECOSYSTEM PARTNERS

HORTONWORKS®

Actionable Intelligence from Connected Data Platforms

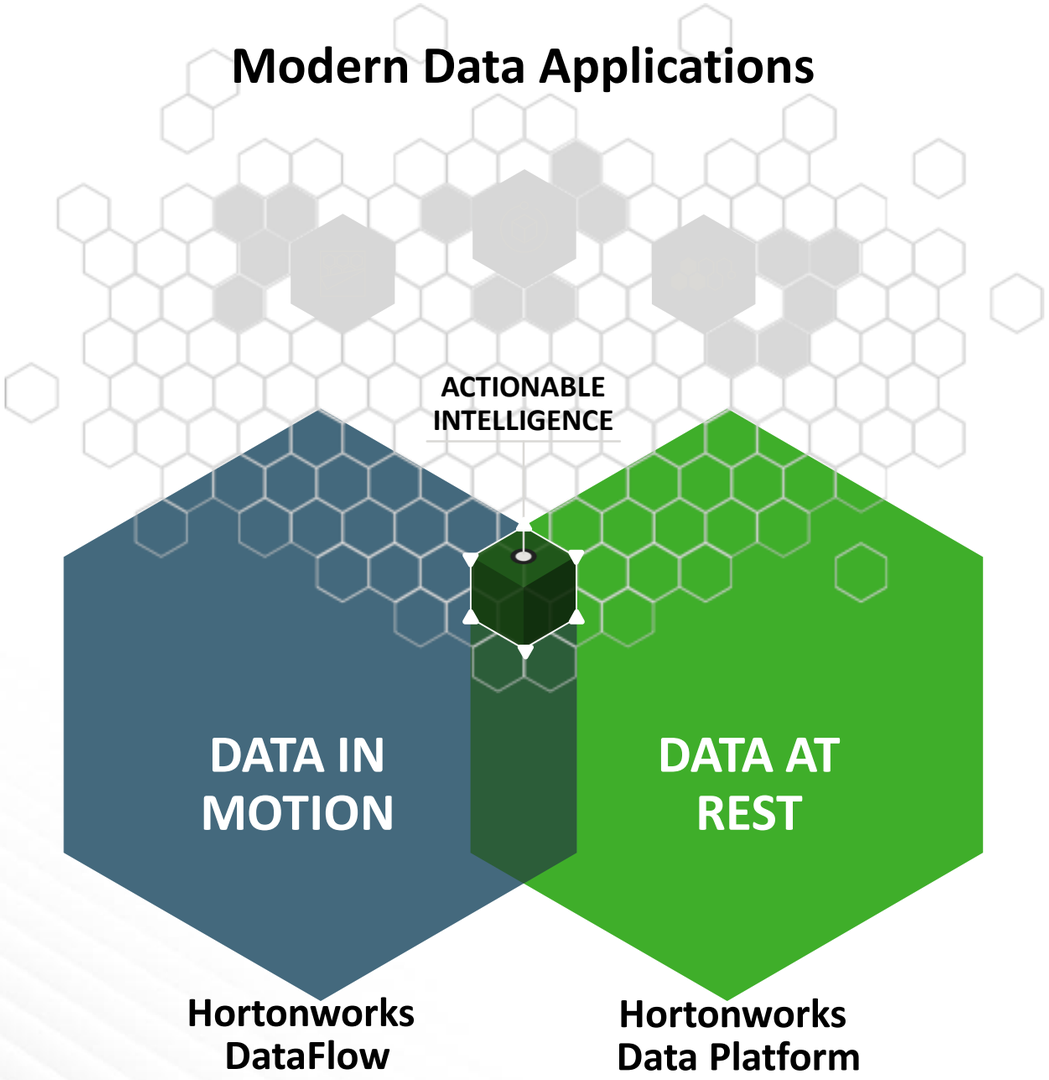
- Capturing perishable insights from data in motion
- Ensuring rich, historical insights on data at rest
- Necessary for modern data applications



HORTONWORKS

Modern Data Apps

Custom or Off the Shelf



Real-Time Cyber Security
protects systems with superior threat detection

Smart Manufacturing
dramatically improves yields by managing more variables in greater detail

Connected, Autonomous Vehicles
drive themselves and improve road safety

Automatic Recommendation Engines
match products to preferences in milliseconds

Credit Fraud Prevention
predict in real-time based on each customer's transaction record

HORTONWORKS

Credit Fraud Prevention on a Connected Data Platform

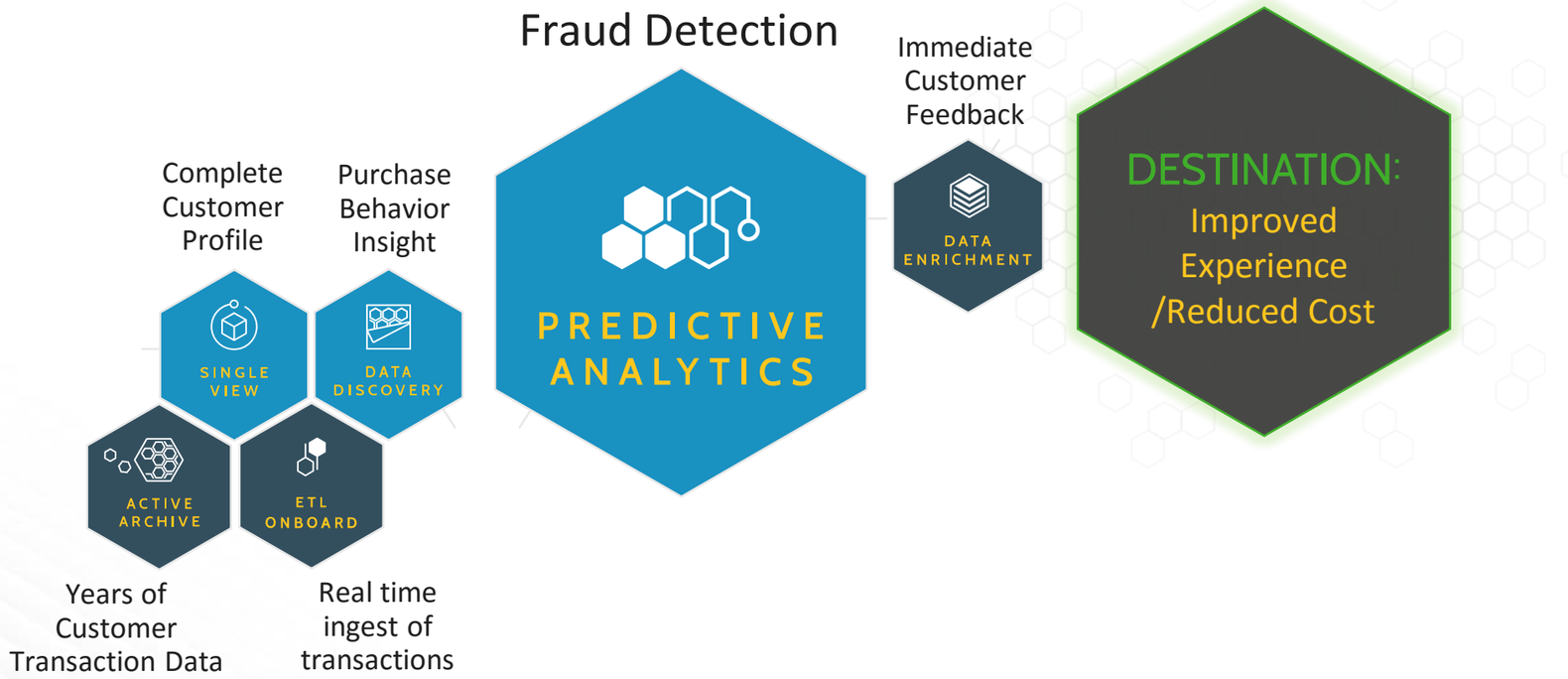
Built by: Kirk Haslbeck & Vadim Vaks

HORTONWORKS®
POWERING THE FUTURE OF DATA™

Journey to Fraud Detection

Innovate

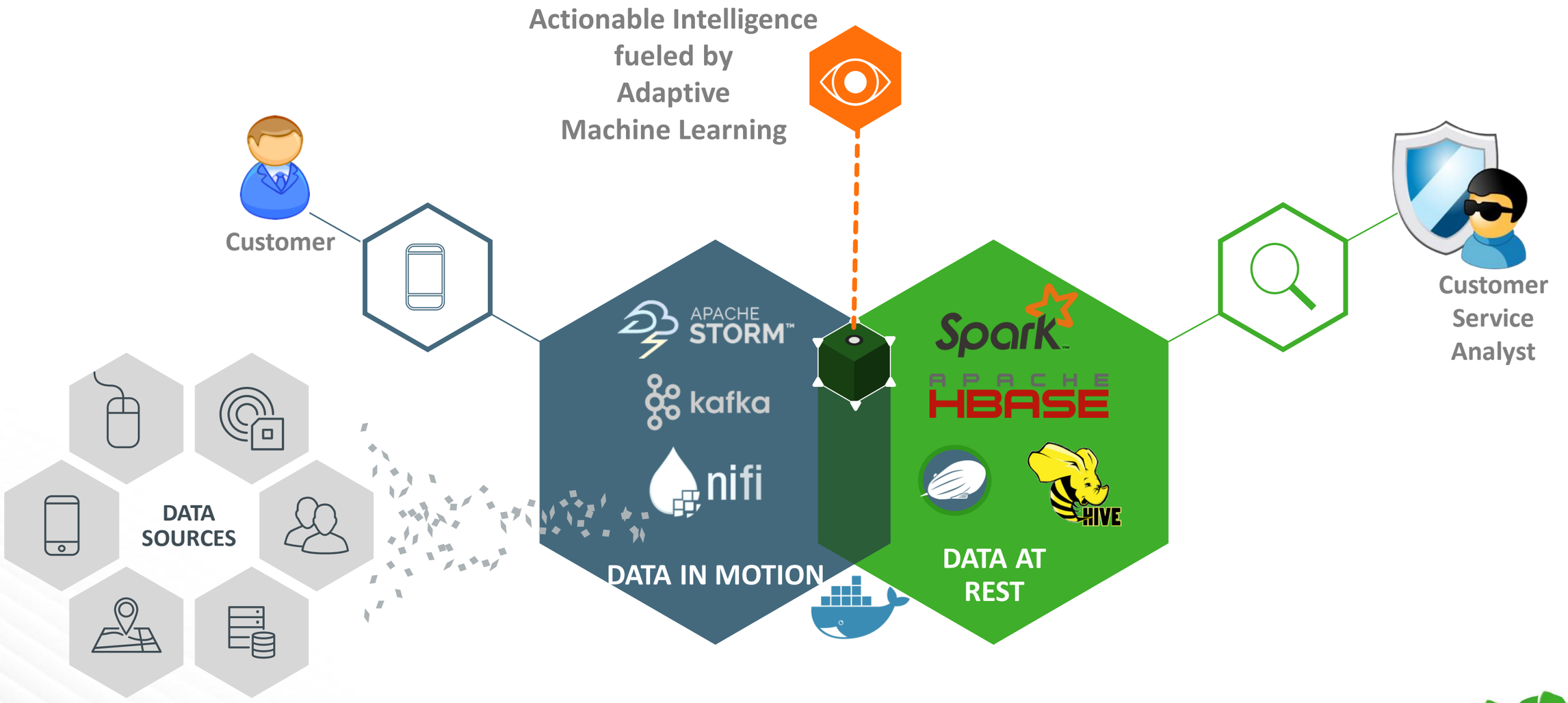
Renovate



Proactively identify potential fraudulent transactions to protect the customer and improve customer experience

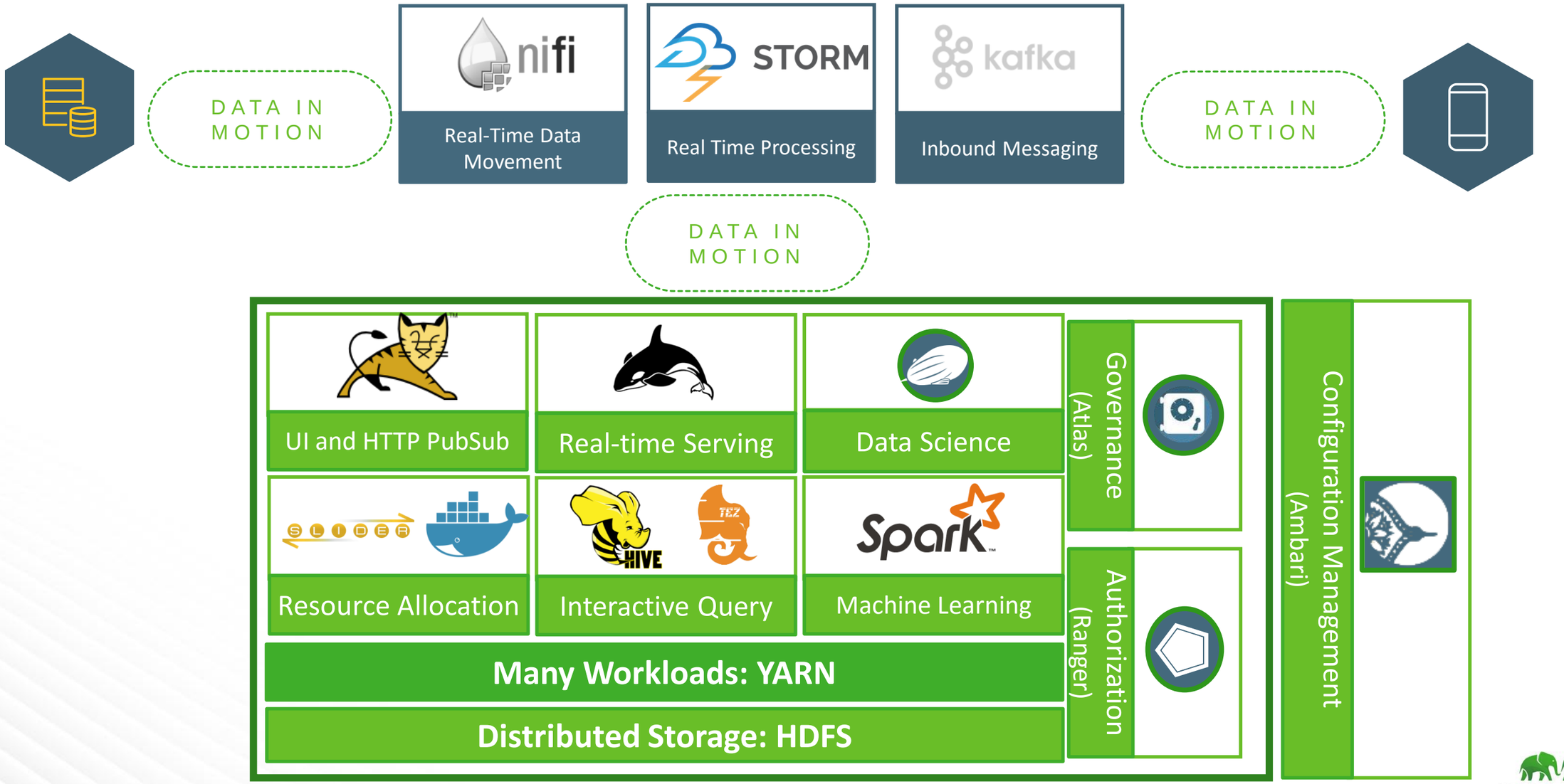
- Proactively monitor every credit card transaction using machine learning to catch potential fraud
- Customer Service Analyst reviews flagged transactions in real time via a next generation application running on the connected platform
- Control real time flow of data in and out of the connected platform to the various source and destination points

Fraud Detection Application Architecture



HORTONWORKS

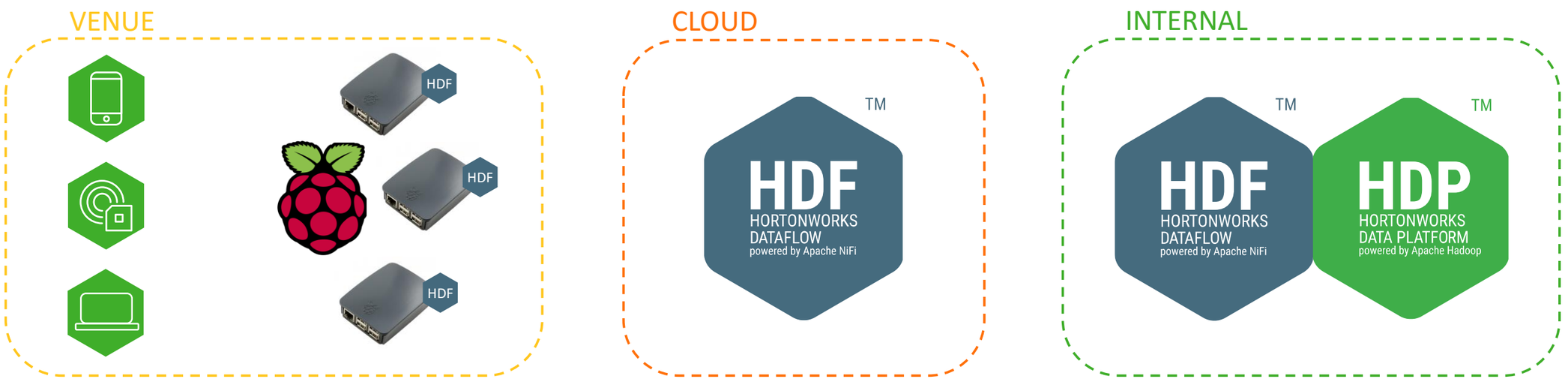
Fraud Detection Demo Architecture



PiWiNiFi: Bi-directional IoT sensor networks

Built by: Simon Elliston Ball & Dan Chaffey

System Architecture

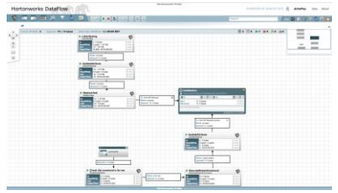


HORTONWORKS

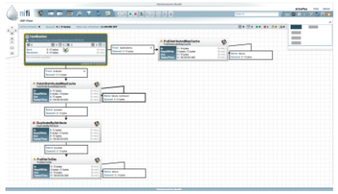
Demo Features

VENUE

Raspberry Pi

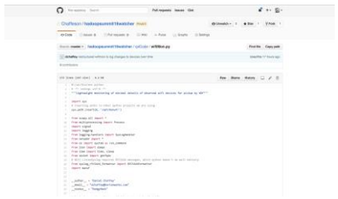


Booth

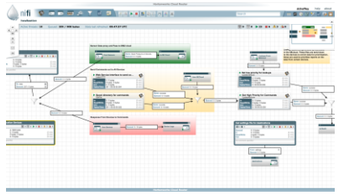


CLOUD

Codebase



DMZ

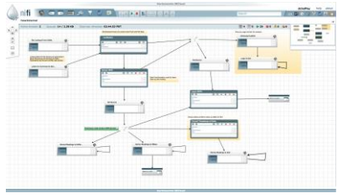


INTERNAL

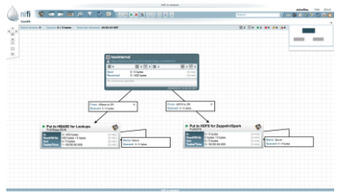
Solr / Banana

Zeppelin/Spark

Prod



DR



http://nifi.apache.org/
http://hortonworks.com/hdf/
http://community.hortonworks.com/

Data Flow

Data in Motion at the core and at the edge.

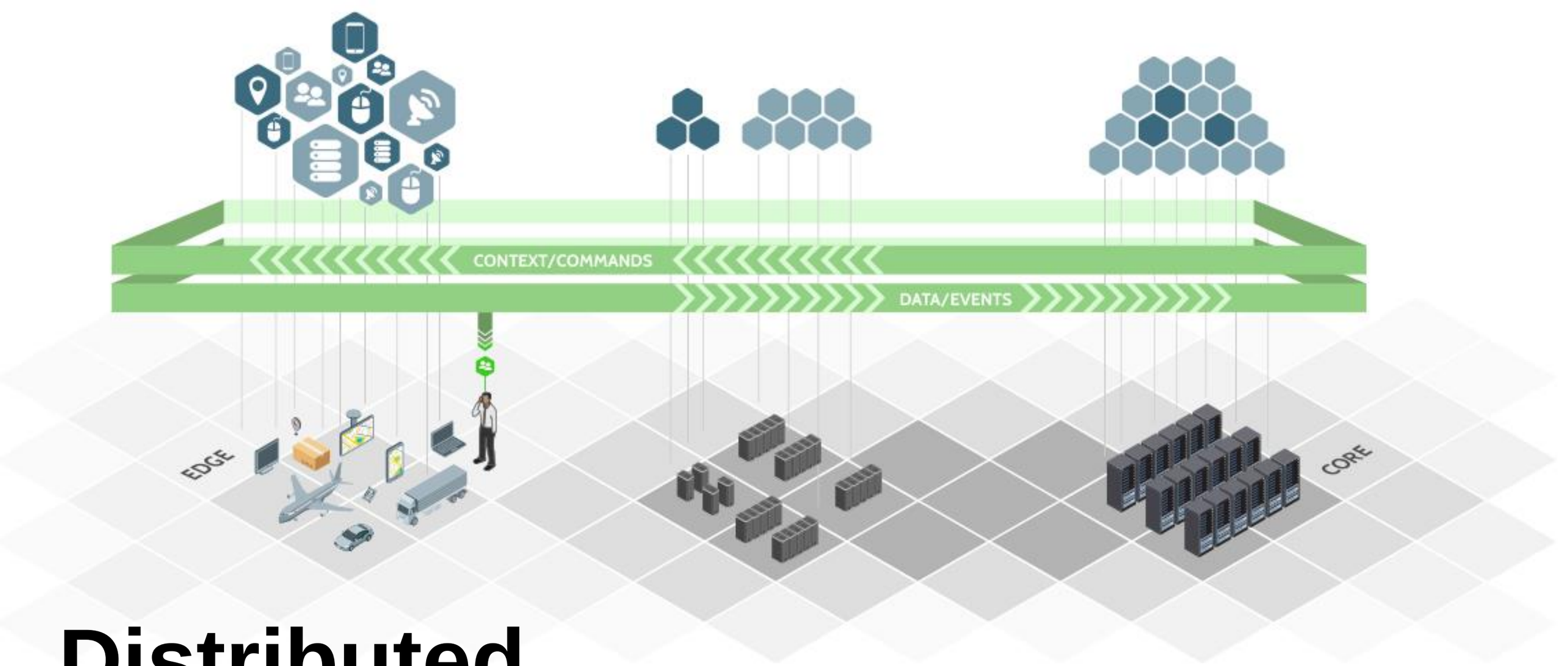
HORTONWORKS

https://www.flickr.com/photos/rosefirerising/3818133872

Resilient

Clustering, Availability, Load Balanced Site-to-Site

HORTONWORKS



Distributed

Site-to-Site takes processing and prioritization right to the edge. Coming soon: **MiNiFi!**

HORTONWORKS

Thank You

References

- http://hortonworks.com/blog/credit-card-fraud-prevention-on-a-connected-data-platform/
- https://community.hortonworks.com/repos/27236/credit-fraud-prevention-demo.html
  - https://github.com/vakshorton/CreditCardTransactionMonitor
  - https://github.com/vakshorton/CreditCardTransactionMonitorMobileApp
- https://github.com/simonellistonball/PiWiNiFi

HORTONWORKS

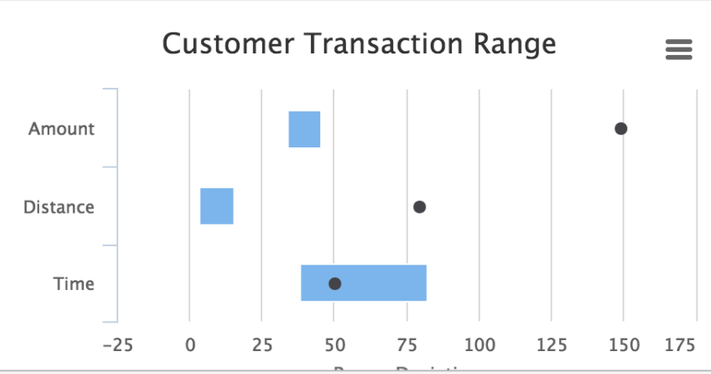
Credit Fraud Analyst Inbox

HORTONWORKS

TransactionId: 20011459123200000 Account Number: 19123
TransactionId: 10011459126800000 Account Number: 19123
TransactionId: 10021459130400000 Account Number: 19123
TransactionId: 20011459134000000 Account Number: 19123
TransactionId: 10041459137600000 Account Number: 19123
TransactionId: 10051459141200000 Account Number: 19123
TransactionId: 20011459144800000 Account Number: 19123

TransactionId: 20011459123200000
Account Number: 19123
Account Type: VISA
Amount: 149
Merchant Id: 2001
Merchant Type: entertainment
Time of Transaction: 1459123200000

Reason Flagged: Amount for vendor type, distance from previous, and/or time between transactions is outside expected range.



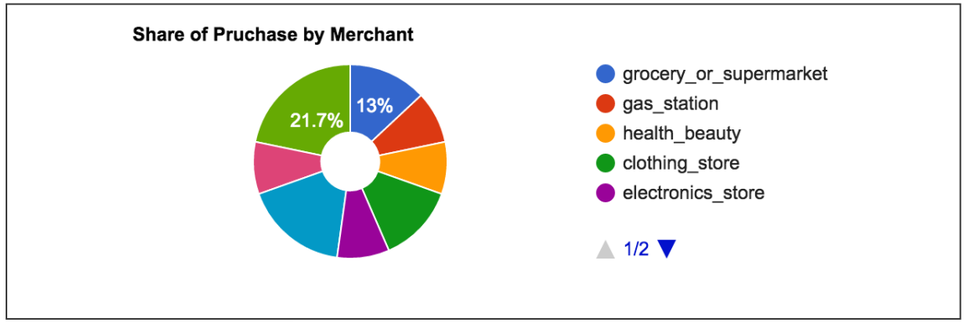
HORTONWORKS

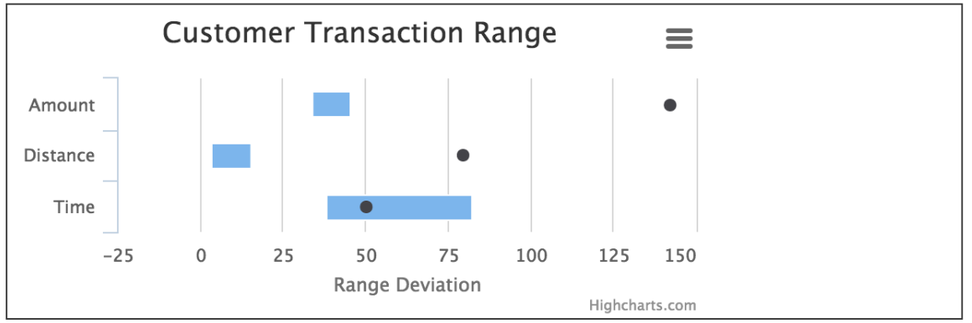
Hortonworks Data Flow

HORTONWORKS

Account Number: 19123
Account Type: VISA
Account Status: Active

Customer Name: Regina Smith
Gender: Female
Age: 32
Address: 1234 Tampa Ave.
Cherry Hill NJ 08003





Notify Customer/Suspend Account

TransactionId	MerchantType	Amount	Distance
10001458950400000	grocery_or_supermarket	35	1.661
10001459036800000	grocery_or_supermarket	142	1.661
10001459123200000	grocery_or_supermarket	118	1.661
10011459040400000	convenience_store	23	0.098
10011459126800000	convenience_store	23	1.623
10021459044000000	restaurant	32	1.173

TransactionId	MerchantType	Amount	Distance
20011458954000000	entertainment	142	79.316
20011458957600000	entertainment	147	79.316
20011458964800000	entertainment	131	79.317
20011458968400000	entertainment	148	79.317
20011458972000000	entertainment	139	79.317
20011458975600000	entertainment	143	79.317

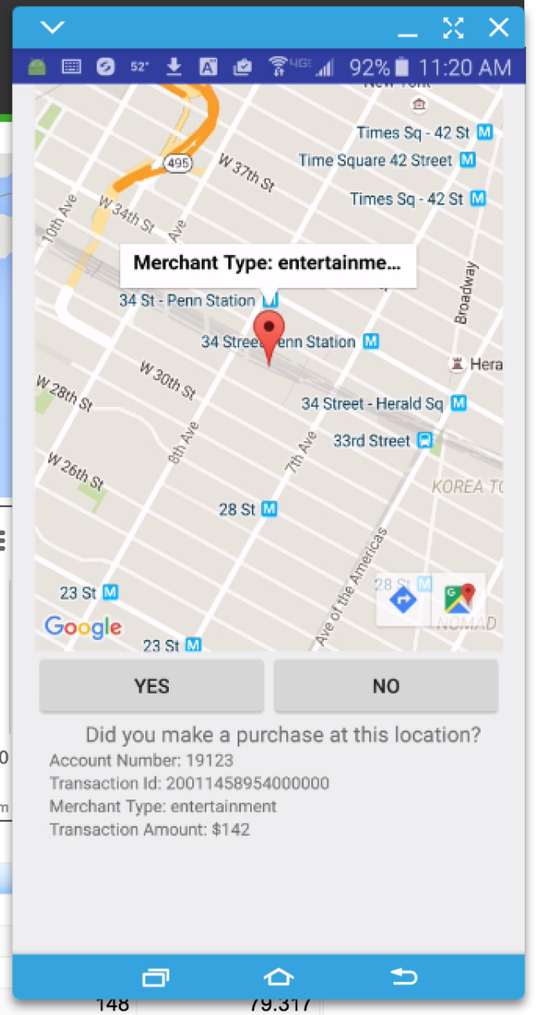
HORTONWORKS

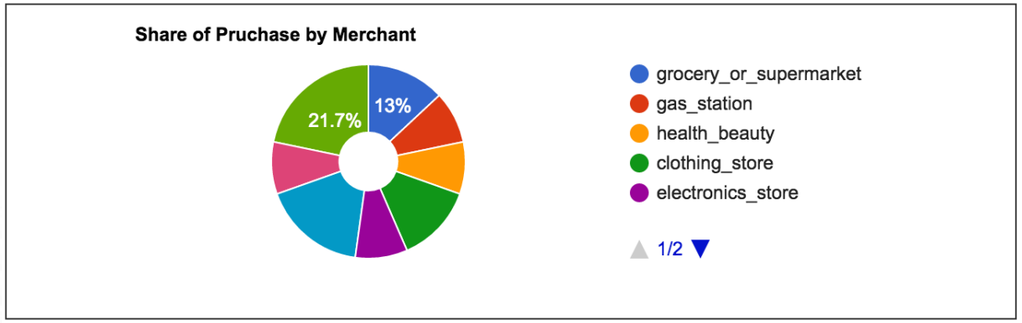
Hortonworks Data Flow

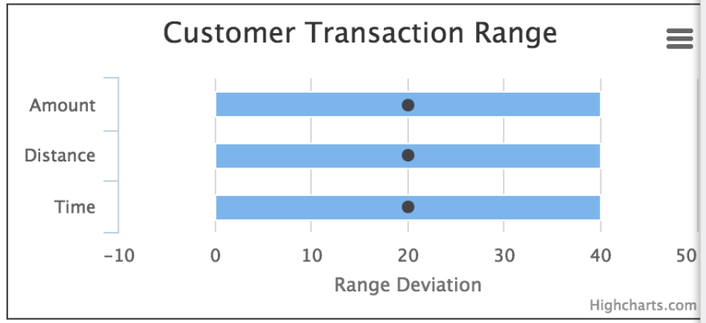
HORTONWORKS

Account Number: 19123
Account Type: VISA
Account Status: Suspended

Customer Name: Regina Smith
Gender: Female
Age: 32
Address: 1234 Tampa Ave.
Cherry Hill NJ 08003





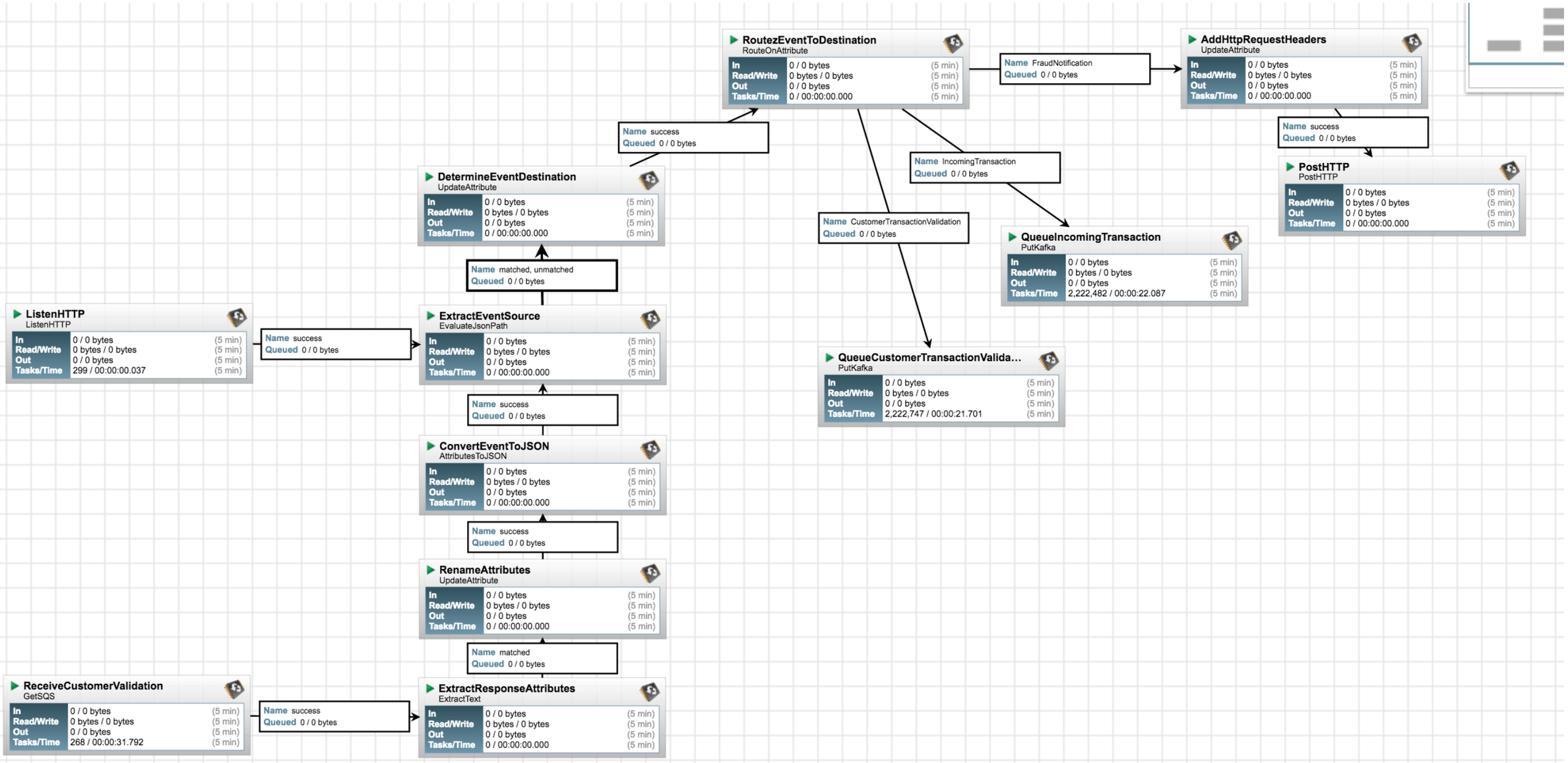


Notify Customer/Suspend Account

TransactionId	MerchantType	Amount	Distance
10001458950400000	grocery_or_supermarket	35	1.661
10001459036800000	grocery_or_supermarket	142	1.661
10001459123200000	grocery_or_supermarket	118	1.661
10011459040400000	convenience_store	23	0.098
10011459126800000	convenience_store	23	1.623
10021459044000000	restaurant	32	1.173

TransactionId	MerchantType		
20011458954000000	entertainment		
20011458957600000	entertainment		
20011458964800000	entertainment		
20011458968400000	entertainment	148	79.317
20011458972000000	entertainment	139	79.317
20011458975600000	entertainment	143	79.317

HORTONWORKS



HORTONWORKS

Zeppelin Notebook Interpreter Configuration Sea

Credit Fraud Detection ML

```
%dep
z.reset()
z.addRepo("Spark Packages Repo").url("http://dl.bintray.com/spark-packages/maven")
z.load("com.databricks:spark-csv_2.10:1.2.0")
```

```
res0: org.apache.zeppelin.dep.Dependency = org.apache.zeppelin.dep.Dependency@41737814
```

Took 18 seconds

```
import org.apache.spark.sql.Row
import org.apache.spark.mllib.regression.LabeledPoint
import org.apache.spark.mllib.linalg.{Vector, Vectors}
import org.apache.spark.ml.classification.ProbabilisticClassifier
import org.apache.spark.ml.classification.LogisticRegression
//import org.apache.spark.mllib.classification.{LogisticRegressionWithLBFGS, LogisticRegressionModel}
import org.apache.spark.mllib.evaluation.MulticlassMetrics
import org.apache.spark.mllib.regression.LabeledPoint
import org.apache.spark.mllib.linalg.Vectors
import org.apache.spark.mllib.util.MLUtils

// Load training data
val data = sqlContext.read.format("com.databricks.spark.csv").option("header", "true").load("/user/zeppelin/training5.csv")
data.dtypes
data.show

val row = data.map(x=> Tuple4( x.getString(0).toDouble, x.getString(2).toDouble, x.getString(3).toDouble, x.getString(4).toDouble))
row.toDF.registerTempTable("row")


// create feature vector
val df = data.map { x=>
  LabeledPoint( x.getString(0).toDouble,
    Vectors.dense( x.getString(2).toDouble, x.getString(3).toDouble, x.getString(4).toDouble ))
}.toDF
```


HORTONWORKS