

Big Data Streaming Architecture on HDP 3.0

IoT - Relevance

- **7.2 Billion active SIM Cards**
- **25 Billion connected things**
- **500 Million tweets per day**
 - Average life of a tweet is 18 minutes
- **2 Zettabytes per year of Global IP Traffic**
 - 80% is unstructured

5 Attributes of a Streaming Platform

- Ingest
- Process
- Analyze
- Respond
- Visualize

Data Flow versus Stream Analytics for IoT

- **Data Flow**

- Ingest and route terabytes data into a "unified firehose"
- Actively performance manage the latency and quality of these data flows –
- Across high variability of data formats, size of data and speed of data

- **Real Time Stream Analytics**

- Sub second event processing with linear scalability to billions of events
- Predictive Analytics at Scale
 - Real time data aggregation across edge nodes while processing 10s of millions of events and 100s of gigabytes per second
- Guaranteed no data loss and events processed in order

6 Focus for Building a Streaming Platform for IoT

1)Common Abstraction Layer

2)Latency

3)Lambda Architecture

- “Orchestrate” over static and real time data

4)Scale-out

5)Rapid Application Development

6)Data Visualization

Focus Areas for Streaming Platform for IoT

1 . Common Abstraction Layer

- Select one or more streaming engines
- Select one or more cloud providers
- Select one or more resource managers
- Select one or more event sources
- ...Future Proof

Focus Areas for Streaming Platform for IoT

2. Latency

- 500 millisecond or less- Dashboards, Security Incidents, Asset Performance
- 20 milliseconds or less - Ad Networks, Preventive Maintenance

Focus Areas for Streaming Platform for IoT

3 . Lambda Architecture

- Integrate static and real time data
- Enrichment
- Orchestration of Batch workflows
- Predictive Classification and Scoring

Focus Areas for Streaming Platform for IoT

4 . Scale Out

- Linear Scale out or scale down
- Resource Management
- Handle Transient workloads

Focus Areas for Streaming Platform for IoT

5 . Rapid Application Development

- Industry vertical specific applications
- Aggregations
- Filters
- Multi Stream Correlations
- Splits
- Joins
- Normalizations
- Business Rules Editor

Focus Areas for Streaming Platform for IoT

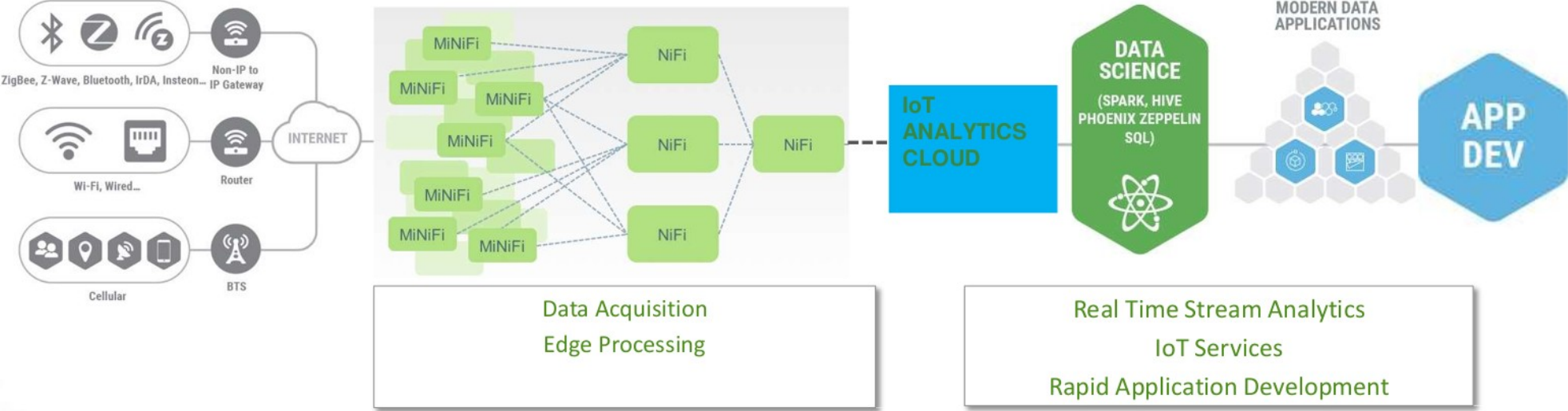
6. Data Visualization

- Time Series Visualizations
- Metrics Dashboarding
- Trends
- Comparisons
- Thresholds
- Custom UI extensions

6 Focus for Building a Streaming Platform for IoT

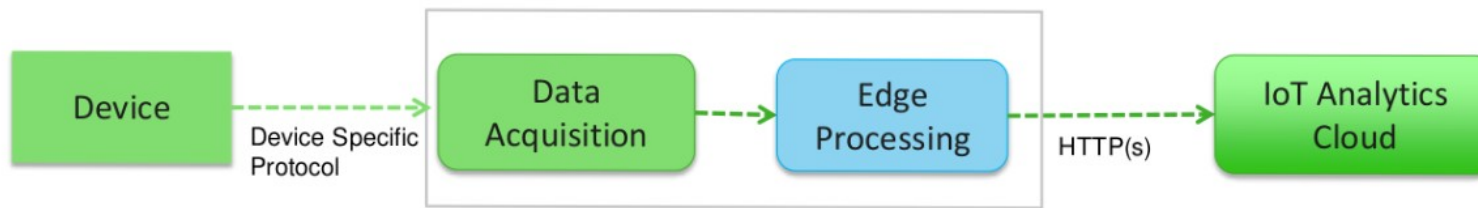
- Abstracts underlying Streaming Engine- Storm, Spark Streaming, Flink..
- OOTB Support for multiple (cloud) event sources - Kafka, AWS Kinesis, Azure Event Hub
- Built-in Operators for Complex Event Processing
- Built-in Real Time Dashboarding- Metrics and Events
- PMML Support
- Pluggable Workflow Management
- Business Rules Editor and Rapid Application Development Framework
- Cloud Deployment
- Scalable Architecture
- Handles different latency requirements

Hortonworks IoT Platform



IoT Edge

Data Acquisition & Processing



Edge Processing

- Reliable Delivery
- Buffering and Flow Control
- Simple Event Processing
- Edge Analytics

High Availability

- Scale out and Clustering
- Multi Data Center Support

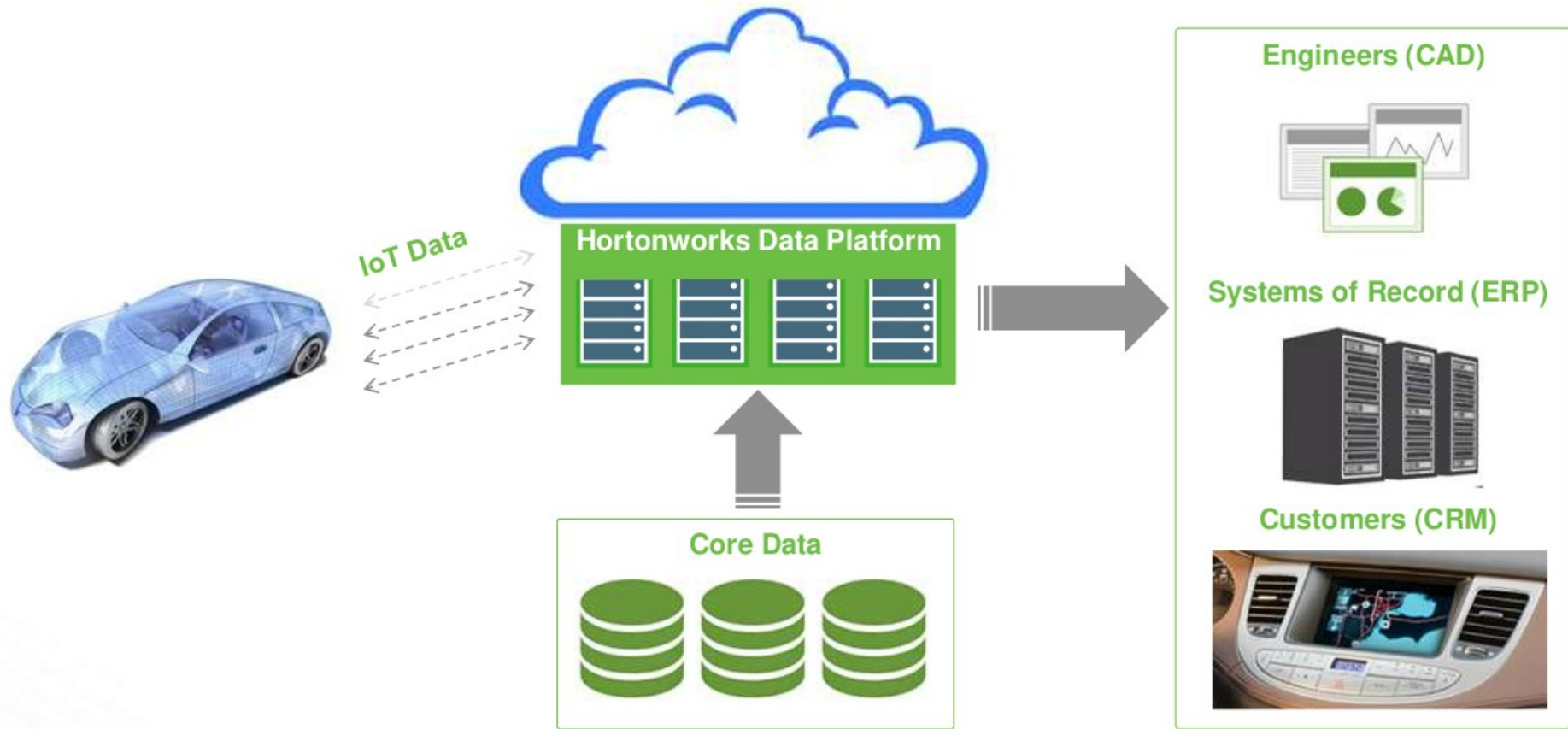
Security

- 2 way SSL
- Data Element Masking
- Flexible Routing based on Data Sensitivity

Kafka Enablement

- Data Ingest and Drain based on Kafka Consumer and Producer Support

Reference Architecture



Demo Architecture by my son

