Stream Analytics in the Enterprise
A look at Intel’s internal IOT implementation

Moty Fania
Principle Engineer,
Big Data Analytics @ Intel
June 2016
About us - Advanced Analytics Team @ Intel IT

- Specialized in Big Data and Machine Learning
- Team charter
  ✓ Solve strategic high value business problems
  ✓ Leverage analytics to grow Intel’s revenue
- Main skills are Big Data software engineering and Data Science
- Open Source technologies
Internet of Things

- Adoption of IoT within the enterprise is lagging
  - Challenges of on-premises deployment
  - Hard to determine & realize value
  - Lack of infrastructure & skills to enable Analytics
THE APPLICABILITY OF IOT WITHIN THE ENTERPRISE

- Stream Analytics can be applied effectively in the enterprise to enable near-real-time actuation and closed loop systems.

- Use cases identified at Intel:
  - Manufacturing
  - Smart Buildings
  - Supply Chain
  - Preventive maintenance

100k+ measurements points read manually to ensure regulatory compliance, factory operations, and production ramps.
**COMMN PROBLEMS IN THE ENTERPRISE**

- Large number of E2E products
- Non-interoperable solutions
- Increased security risk footprint
- Conflicting wireless radios
- Duplicated infrastructures
- Duplicated support models

**VERTICAL VS. HORIZONTAL IOT**

- **USE CASE 1**
  - Water Leak
  - Temperature
  - Power

- **USE CASE 2**
  - ZigBee

- **USE CASE 3**
  - WiFi

#StrataHadoop
ARE THERE ANY COMMONALITIES IN IOT PROJECTS?

- Different use cases have different requirements
- However most IOT projects have a basic set of common needs
- These basic needs can be addressed with one reusable platform

THE IOT HIERARCHY OF NEEDS

1. Let the Data flow from sensor to Cloud Storage
2. “See” the data – visualization/Charts
3. Define rules / monitors
4. Auto Machine Detection
5. Build On Top

#StrataHadoop
A common horizontal platform for open connectivity, better security, centralization, and support simplicity.

Breaking Down the Silos:

- **USE CASE 1**: Water Leak, Power, Temperature
- **USE CASE 2**: Water Leak, Power, Temperature
- **USE CASE 3**: Water Leak, Power, Temperature

**COMMON IOT PLATFORM**

- **GW**: Gateway
- **MANAGEMENT**: Management
- **STORAGE**: Storage
- **ANALYTICS**: Analytics

Water Leak
Temperature
Power
**SMART INGESTION CHARACTERISTICS**

**SCALABILITY**
- Linear scalability (scale Out)
- Extremely High concurrencies
- High Throughput

**FAULT TOLERANCE**
- No Single point of failure
- Seamless recovery
- Persistent

**SMART DATA PIPE**
- Apply analytics on the Stream
- Trigger actions (close the feedback loop) in timely manner

**PERSONALIZED**
- Per single device or user
- Maintain state and required data for ML

**EASY TO USE**
- Easily subscribe to any Stream and focus on logic
- Use familiar development Languages (Java, Scala)
- Easy to deploy, anywhere
**HIGH LEVEL PLATFORM ARCHITECTURE**

The platform leverages Intel’s IT Big data assets + Docker & CoreOS containers.
AKKA & THE ACTOR MODEL

Producer

Inbox (Queue)

Behavior

onRecieve()

tell case "sensor 1" =>
case "keep alive" =>

State

- Message Driven
- Lock-free
- Location-transparent
- High performance
- Fault Tolerant
- Scales linearly

MICRO-SERVICE (ACTOR) ORIENTED.
STREAM PROCESSING MANAGEMENT Layer ("Pigeon")

- Scale-Out, Symmetric Architecture (Cluster Sharding)
- Fault tolerance
- Persistence
- Back Pressure (reactive streams)

1. Code your processing logic in Java or Scala
2. Subscribe to your data stream.
3. Deploy topology to the processing cluster.
PIGEON HL DESIGN: SINGLE NODE

Subscriber Region -> Mediator Actor -> Topology Region

Scalable Message Queue: kafka
Device Registry

my topic -> topology/create -> REST API -> Storage
PIGEON HL DESIGN: CLUSTER

Scalable Message Queue

Seed 1
Node 1
Node 2
Node 3
Seed 2

Raw Data

Enriched Data

Distributed Persistence (Apache Cassandra)

Storage

Raw Data

#StrataHadoop
Core OS & Docker containers enable portability and ease of deployment anywhere.

Enables the flexibility of choosing a set of desired containers based on a given use case requirements.
SELF-SERVICE DATA MONITORS

- Spark Streaming based capability
- Allows users to define their own monitors / rules through UI
  - Self- Service
  - Avoid the need to monitor charts manually
  - Near Real time– address any issues in a timely manner
  - Leverage human domain expertise
  - Automated actuation

**Example:** Trigger action A when temperature of device X is above 40 degrees for more than 10 min
Advanced Predictive Analytics Building Blocks

The platform includes out-of-the-box Machine learning capabilities to automatically react to machine or sensor data:

- Reusable building blocks
- Reduces the need for manual rules or domain expertise
- Near real-time actuation
- Can enable preventive maintenance use cases
<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time series classification</strong></td>
<td>A component that checks for each sensor whether its data is stationary, periodic, or neither</td>
</tr>
<tr>
<td><strong>Periodicity removal</strong></td>
<td>A component that models the period (if one exists) and cancels it out</td>
</tr>
<tr>
<td><strong>Change detection</strong></td>
<td>An ensemble of tests that either monitor a single sensor or a collection of device sensors</td>
</tr>
<tr>
<td><strong>Alert classification</strong></td>
<td>A set of nonparametric statistical tests that enable pointing out the most significant changes causing each alert</td>
</tr>
</tbody>
</table>
OTHER OPPORTUNITIES

• Many operational IT activities can be “translated” into IOT Kind, stream analytics scenarios

• Will allow a higher level of proactivity and a shift from manual monitoring and fire fighting to higher value work
We deployed Internal, multi-tenant IoT platform to enable stream analytics use cases.

Platform leverages previous Big data infrastructure investments.

Core OS & Docker enable ease of deployment on-premises.

Smart Data pipes & stream analytics are key to derive insights in timely manner.

Summary
THANK YOU!