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

Moty Fania
Principle Engineer,
Big Data Analytics @ Intel
Sep 2016
ABOUT US - ADVANCED ANALYTICS @ INTEL

- Specialized in Big Data and Machine Learning

- Team charter
  ✓ Solve strategic high value business problems
  ✓ Improve & enable Intel’s products

- Applied in various domains: Manufacturing, design, Sales and Marketing, Health, Deep learning etc.
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
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.
VERTICAL VS. HORIZONTAL IOT

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

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

"See" the data – visualization/Charts
Let the Data flow from sensor to Cloud Storage
Define rules / monitors
Auto Machine Detection
Build On Top

THE IOT HIERARCHY OF NEEDS 😊

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

USE CASE 1
- Water Leak
- Temperature
- Power

USE CASE 2
- Water Leak
- Temperature
- Power

USE CASE 3
- Water Leak
- Temperature
- Power

COMMON IOT PLATFORM
- MANAGEMENT
- STORAGE
- ANALYTICS
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
- Can be applied at the edge

EASY TO USE
- Easily subscribe to any Stream and focus on logic
- Use familiar development Languages (Java, Scala)
- Easy to deploy, anywhere
The platform leverages Intel's IT Big data assets + Docker & CoreOS containers
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.

Preconfigured containers ready to be loaded.
STREAM PROCESSING MANAGEMENT Layer (“PIGEON”)

• Scale Out Symmetric Architecture (Cluster Sharding)
• Fault tolerance & Persistence
• Back Pressure using reactive streams

1. Code your processing logic in Java or Scala
2. Subscribe to your data stream.
3. Deploy topology to the processing cluster.
AKKA & THE ACTOR MODEL

Producer

Inbox (Queue)

Behavior

onRecieve()

case “sensor 1” =>

case “keep alive” =>

State

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

MICRO-SERVICE (ACTOR) ORIENTED.
PIGEON HL DESIGN: SINGLE NODE

Subscriber Region

Mediator Actor

Topology Region

Scalable Message Queue

kafka

Device Registry

topology/create

REST API

my topic

Storage
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 an issue 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
## Analytical Layer - Four Main Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time series classification</td>
<td>A component that checks for each sensor whether its data is stationary, periodic, or neither</td>
</tr>
<tr>
<td>Periodicity removal</td>
<td>A component that models the period (if one exists) and cancels it out</td>
</tr>
<tr>
<td>Change detection</td>
<td>An ensemble of tests that either monitor a single sensor or a collection of device sensors</td>
</tr>
<tr>
<td>Alert classification</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
SUMMARY

• 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
THANK YOU!