Build your own Real Time Data Processing Platform in the Cloud for Connecting Millions of Things

Philipp Behre, Solutions Architect
Jan Metzner, EMEA Solutions Architect Mobile/IoT

@pbehre     Berlin, April 12th 2016     @janmetzner
What to expect from this session

• How to Collect, store, and analyze data from small things in a big world?

• What tools are there for Data Engineers to build a cloud based data platform with AWS?

• How to Enable your business teams to make data informed decisions?

• How get smart support for people to make decisions with confidence based on near-real time predictions?

Next: Start creating !!
The Person has the context to decide
The Person has the context to decide

Analyze & decide
The Cloud make decisions with smart situational awareness

Monitor & have the final say

Enable smart decisions & act
One Example: Water Pipe
Connect – Secure – Integrate
AWS IoT: How it Works

**AUTHENTICATION**
Secure with mutual authentication and encryption

**DEVICE SDK**
Set of client libraries to connect, authenticate and exchange messages

**DEVICE GATEWAY**
Communicate with devices via MQTT and HTTP

**RULES ENGINE**
Transform messages based on rules and route to AWS Services

**SHADOW**
Persistent thing state during intermittent connections

**REGISTRY**
Identity and Management of your things

**AWS IoT API**

**APPLICATIONS**

**AWS Services**

**3P Services**
“If you can’t measure it, you can’t improve it”

-Lord Kelvin
analyze your data →
make data-informed decisions →
improve your processes
Three Types of Data-Driven Decision Making

- **Retrospective**
  Analyze historical trends to know what's happening in the app

- **Inquisitive**
  Discover latent user behavior to shape product or marketing decisions

- **Predictive**
  Anticipate user behavior to enhance experience
IoT-Data Architectures build out of AWS services
Primitives for IoT – with a focus on collect, store, analyze

AWS IoT
Amazon Kinesis
AWS Lambda

Amazon Machine Learning
Amazon Redshift
Amazon QuickSight
Amazon EMR

Amazon S3
Amazon DynamoDB
Amazon Elasticsearch Service
Understanding your data - People involved

• BI Analysts
• Data Engineers
• Application Developers
• Data Scientists
• ....

Actually ... everyone in your company making a decision!
• Primary tool is SQL
• Data is largely structured with well known data sources
• Primary concern is fast, consistent performance
• Need to extend SQL with custom functions
Data Engineer familiar with Hadoop and Spark

- Amazon EMR
- Spark
- ETL
- New Structured Data
- Amazon Redshift
- Integration
- Existing Structured Data
- Amazon Redshift
- Data Source
- Amazon Redshift
- Enrichment / Transformation
Data Scientist with existing toolsets

- Work with unstructured datasets
- Use existing toolsets to connect to Redshift
Example: Querying Redshift with R Packages

- **RJDBC** – supports SQL queries
- **dplyr** – Uses R code for data analysis
- **RPostgreSQL** - R compliant driver or Database Interface (DBI)

Application Developers can build smart applications using Amazon Machine Learning

- All skill levels
- Machine Learning technology is accessed through APIs / SDKs
- Embed visualizations in applications
Back to our water pipe ...
Instantly React – getting ‘smarter’
Smart Application – supporting people

1. sensors send data
2. Inbound stream (raw data)
3. time-series aggregation*
4. write aggregate & trigger event
5. process event

* https://github.com/awslabs/amazon-kinesis-aggregators
Follow up and capture results

1. follow up
2. Capture result & activity
3. Frequently load to S3
Collect business and contextual data – learn and improve

1. Store additional data
2. Transform and load
3. Let apps query data
4. Let people query data
5. Re-train prediction model
The Cloud make decisions with smart situational awareness
Start building !!!

monitor

make decisions
Resources

AWS IoT Landing Page:  http://aws.amazon.com/iot

YouTube Channels/Playlist:

Start building today!!