Data driven automation solutions and insights

Dr. Josef Papenfort, Head of Product Management TwinCAT, Beckhoff Automation
Christian Petters, Senior Solutions Architect, Amazon Web Services
Agenda

• Why Data Lake
• How to build a Data Lake
• Use case realized by Beckhoff
• Big Data architectural patterns
• Data Lake Quick Starts
Traditionally, analytics look like this

- Almost always “Relational” data
- TBs–PBs scale
- Schema defined prior to data load
- Operational reporting and ad hoc
- Large initial CAPEX + $10K–$50K/TB/Year
## New requirements break the traditional approach

### Customers need to:  
- Do new types of analysis (ML, big data & real-time)
- Capture and store new non-relational data at EB scale
- Secure and combine data from new and existing sources

### Challenges with traditional approach:
- Operational and ad hoc analysis on relational only
- DW is optimized for relational data up to PB scale
- Data exists in silos, ETL does not scale at EB data volumes
Date Lake extends the traditional approach

- Relational and non-relational data
- TBs-EBs scale
- Schema defined during analysis
- Diverse analytical engines to gain insights
- Designed for low cost storage and analytics
Why a Data Lake?

Increase speed to which information is curated, added to the platform, and access is provided to derive business value.

- **Democratize** Data Access to accelerate more insight
- **Collecting and store** any data at scale and at low costs
- **Securing** and protecting all of data stored in the central repository
- **Quickly search** and find the relevant data
- **Easily** perform new types of data analysis and data science
- **Query** the data by defining the data’s structure at the time of use
How does Data Lake look in AWS?
Beckhoff globally provides leading-edge automation technology for...

- machine manufacturers
  factory automation applications
- energy suppliers
- building automation projects
- infrastructure applications
- ... and many more like...
Industrial Competency Partner Beckhoff Automation

Beckhoff: PC-based Automation for machines

Industrial displays

EtherCAT: High-speed Communication

Fieldbus I/O

Servo Drives & Motors

Industrial PC

TwinCAT: Control and Automation Software
### Industrial Competency Partner Beckhoff Automation

**Facts:**

- **Headquarters:** Verl, Germany
- **Employees worldwide:** 3,900
- **Number of engineers:** 1,300
- **Sales/technical offices in Germany:** 22
- **Beckhoff companies worldwide:** 34 countries
- **Subsidiaries and distributors:** > 75 countries
- **Sales worldwide 2016:** 679 million € (+9.5 %)
- **Sales worldwide 2017:** 810 million € (+19 %)

*as of: 11/2017*
Industrial Competency Partner Beckhoff Automation

Sales:

- Harmonic and organic growth
- Consistent advanced technology
- Stable customer base
- Low employee fluctuation
- Financial stability

Million €

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
Corporate goals of end customers

**Machine builders**
- Reduce **machine costs**
- Smart machine **optimization**
  - Optimize production cycle times
  - Optimize energy consumption
- **Efficient** machine maintenance
  - Dedicated and predictable
- Increase machine **attractivity**

→ Increase competitiveness !!

**End customers**
- Reduce **production costs**
- Increase **product quality**
- Efficient **production control**
- Minimize **production losses**

→ Increase competitiveness !!

→ Internet of Things and Cloud based services helps:
## Required bandwidth comparison

<table>
<thead>
<tr>
<th></th>
<th>Structural dynamics of wind turbines</th>
<th>Condition Monitoring of assets in intralogistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor sampling rate</td>
<td>100 Hz</td>
<td>1000 Hz</td>
</tr>
<tr>
<td>Number of sensors</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Raw binary data per sensor (including timestamp)</td>
<td>16 Byte</td>
<td>16 Byte</td>
</tr>
<tr>
<td>Total raw binary data (one timestamp per value)</td>
<td>900 Byte</td>
<td>900 Byte</td>
</tr>
<tr>
<td>JSON encoded data per sensor (including timestamp)</td>
<td>70 Byte</td>
<td>70 Byte</td>
</tr>
<tr>
<td>Total JSON encoded data (one timestamp per value)</td>
<td>3.490 Byte</td>
<td>1.420 Byte</td>
</tr>
<tr>
<td>Required bandwidth binary</td>
<td>0.8 Mbps</td>
<td>7.2 Mbps</td>
</tr>
<tr>
<td>Required bandwidth JSON</td>
<td>2.8 Mbps</td>
<td>11.4 Mbps</td>
</tr>
</tbody>
</table>

⇒ Bandwidth necessary! Edge computing can help!
Edge computing for pre-processing

Aggregated analysis
"at the edge"
- Tx35xx Analytics
- TF3600 CM
- TE1400 Mat./Sim.
- ...

Local real-time analysis
- TF3510 Analytics Lib.
- TF3600 CM
- TE1400 Mat./Sim.
- ...

\( f_1(x) \rightarrow \Psi(f_i(x)) \rightarrow f_2(x) \)

\( \Phi(\Psi(f_i(x))) \)

Aggregated analysis
"in the Cloud"

→ Beckhoff Edge Computer: also with AWS Greengrass
Pre-processing and compression

\[ y = a + b \sin(2\pi f t) \]

- \( a = 3 \)
- \( b = 13 \)
- \( f = 200 \text{ Hz} \)

- Sufficient to send \( a, b \) and \( f \) on change
  Recipient knows mathematic correlation and can re-construct original signal (if desired)

- Send statistical information, e.g. average value
  Recipient does not know original signal but can use compressed, statistical information
  It's a matter of use case
  1. What do I want to achieve ?
  2. What should be analyzed ?
  3. What kind of data do I need where ?
• **TwinCAT IoT product family**
  - Provides easy connectivity for IoT communication
  - Supports several Public and Private Cloud systems
  - Supports standardized communication via MQTT, OPC-UA
  - Enables push technologies to wearable devices
TwinCAT IoT

- TF6720 IoT Data Agent
  - Gateway application
  - Connect third-party devices via OPC UA
  - Retrofitting of existing machines
  - Easy-to-use graphical configuration tool
TwinCAT IoT

- TF6730 IoT Communicator + App
  - Smartphone app with push functionalities
  - Android and iOS
  - PLC library for sending values and push messages
Beckhoff IoT Scenarios overview

- **Beckhoff IoT Gateway**
  - Vendor-specific communication via ADS
  - Standardized communication via OPC UA

- **Office**
  - Beckhoff data analytics for machine analysis
  - End user
  - Machine builder
  - Automation
  - Other

- **New Beckhoff Controller with TC3**
- **Old Beckhoff Controller with TC2**
- **3rd party control**
EK9160 IoT Coupler

- IoT Coupler: EK9160
- Push I/O data directly and easily to a Cloud service
- No TwinCAT System Manager necessary
- Instead use website for configuration
EK9160 IoT Coupler

- IoT Coupler: EK9160
- Automatic I/O terminal detection and configuration
- Simply select I/O channels and enable/disable for a Cloud service
Energy Management Project @Beckhoff

- Energy management @Beckhoff Automation
- German "Energiedienstleitungsgesetz" (EDL-G)
  - Option 1: Energy audit
  - Option 2: Building an energy management system (ISO50001)
- ISO50001 certification - many things to consider
  - energy-political (definition of goals)
  - organizational
  - technical
Where to install energy measuring points?
- Pareto analysis: find 20% consumers that consume 80% of total energy
  - Product testing (climatic chamber)
  - Lighting (old T8 pipes)
  - Motors and compressors

Use cases for measured data?
- Historical and live data, visualization for both
- Simple analysis (data correlation)

Where to host all the data? Which infrastructure to use?
- Self-hosting: cost-intensive, poor scalability (new measurement points in future!)
- Cloud-hosting: the cloud scales and is cost-effective

Energy Management System (EnMS)

Energy data

© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.
Energy Management Project @Beckhoff

- AWS IoT provides gateway to eco-system
- Compatible with TwinCAT IoT
- Security built-in via X.509 client/server certificates and TLS encryption
Energy Management Project @Beckhoff

- AWS IoT provides gateway to eco-system
- Compatible with TwinCAT IoT
- Security built-in via X.509 client/server certificates and TLS encryption
- Easy-to-configure rule engine
Energy Management Project @Beckhoff

Different views to the data:

- Visualization of live data → TwinCAT HMI
- Visualization of historical data → Amazon QuickSight
Energy Management Project @Beckhoff

- Visualization of live data via TwinCAT HMI
  - HMI extensions allow MQTT integration in order to connect to AWS IoT
  - TwinCAT HMI can be hosted in an Amazon EC2 instance (virtual machine)
Energy Management Project @Beckhoff

- Visualization of historical data via Amazon QuickSight
  - QuickSight can perform BI analysis based on RDS data
  - Dashboards can be shared
  - Smartphone app available for mobile access to dashboards
Energy Management Project @Beckhoff

- AWS eco-system perfectly suited for realizing energy management project
- Use cases for EnMS easy to implement and setup
- TwinCAT IoT establishes secure connection to AWS even easier
How to build a Data Lake?
Core Architectural Principles (“tried and tested”)

• Decoupled “data bus” (microservices)
  • Collect → Store → Analyze → Answers

• Use the right tool for the job

• Use event driven serverless wherever possible

• Leverage managed services wherever possible

• Big Data ≠ Big Cost
Key Components of a Successful Data Strategy

Attributes of a Modern Data Architecture

1. Automated And Reliable Data Ingestion
2. Preservation Of Original Source Data
3. Lifecycle Management And Cold Storage
4. Metadata Capture
5. Managing Governance, Security, Privacy
6. Self-Service Discovery, Search, Access
7. Managing Data Quality
8. Preparing For Analytics
9. Orchestration And Job Scheduling
10. Capturing Data Change

Key Pillars of a Data Lake

- API & UI
- Entitlements/Security
- Catalogue & Search
- Storage
- Compute
A Sample Batch Analytics Pipeline

Data Sources → S3 → EMR → S3 → Redshift → QuickSight

- Upload data from multiple sources into S3
- Use Amazon EMR to transform and cleanse the data (ETL)
- Load formatted and cleansed data into S3
- Redshift loads data in parallel optimizing it for fast analytics queries
- Analyze and visualize data with Amazon Quicksight

Ad-hoc access to data using Athena

Athena can query aggregated datasets as well
On-demand Big Data Analytics

Sales orders, inventory & trends data from multiple online and physical locations → Data is uploaded to S3 for staging → Amazon EMR with spot instances is used to sort, aggregate and join datasets → Processed Data is loaded into Amazon Redshift → Reporting, business apps, and business intelligence

Real-time data is loaded and processed with Amazon Kinesis Streams → Real-time data is uploaded to DynamoDB → Updated information immediately available online to business users & customers
Create an Amazon Kinesis stream for receiving data

Use AWS Lambda to coordinate the data flow

Create an Amazon Machine Learning Model to create real-time predictions

Use Amazon SNS to notify customer support agents
Event-driven Extract, Transform, Load (ETL)

1. Online order is placed
2. Order data is stored in operational database
3. Lambda is triggered
4. Lambda runs data transformation code and loads results into data warehouse
5. Analytics generated from data
AWS Solution - Data Lake on AWS

Reference Architecture deployment via CloudFormation

Configures core services to tag, search and catalogue datasets

Deploys a console to search and browse available datasets

http://amzn.to/2nTVjcp
Data Lake Foundation on AWS – Quick Start

Ready to Deploy
https://amzn.to/2p0xaTc
AWS Quick Starts for Data Lakes

Quick Starts are built by AWS solutions architects and partners to help you deploy popular solutions on AWS, based on AWS best practices for security and high availability.

These reference deployments implement key technologies automatically on the AWS Cloud, often with a single click and in less than an hour. You can build your test or production environment in a few steps, and start using it immediately.

https://aws.amazon.com/quickstart/#data_lake
Call to action