

Why a data lake alone can't replace your PI system

John de Koning, Customer Success Advisor OSIsoft

Conference Theme & Keywords



Business challenge

Operational Performance

Asset Integrity & Reliability

Disciplined delivery

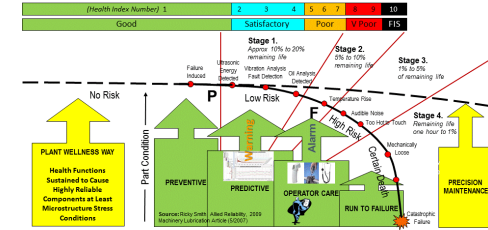
Empowered People

Additional challenges

Top Quartile plants spend
3.5x times less on maintenance

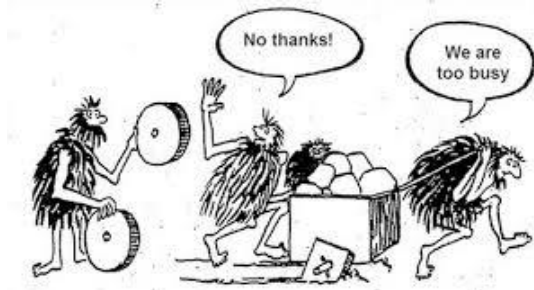


Machine Health & Reliability Measurement

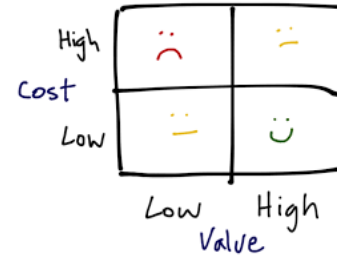


Production reliability not Top Quartile

Continuous drive to improve safety

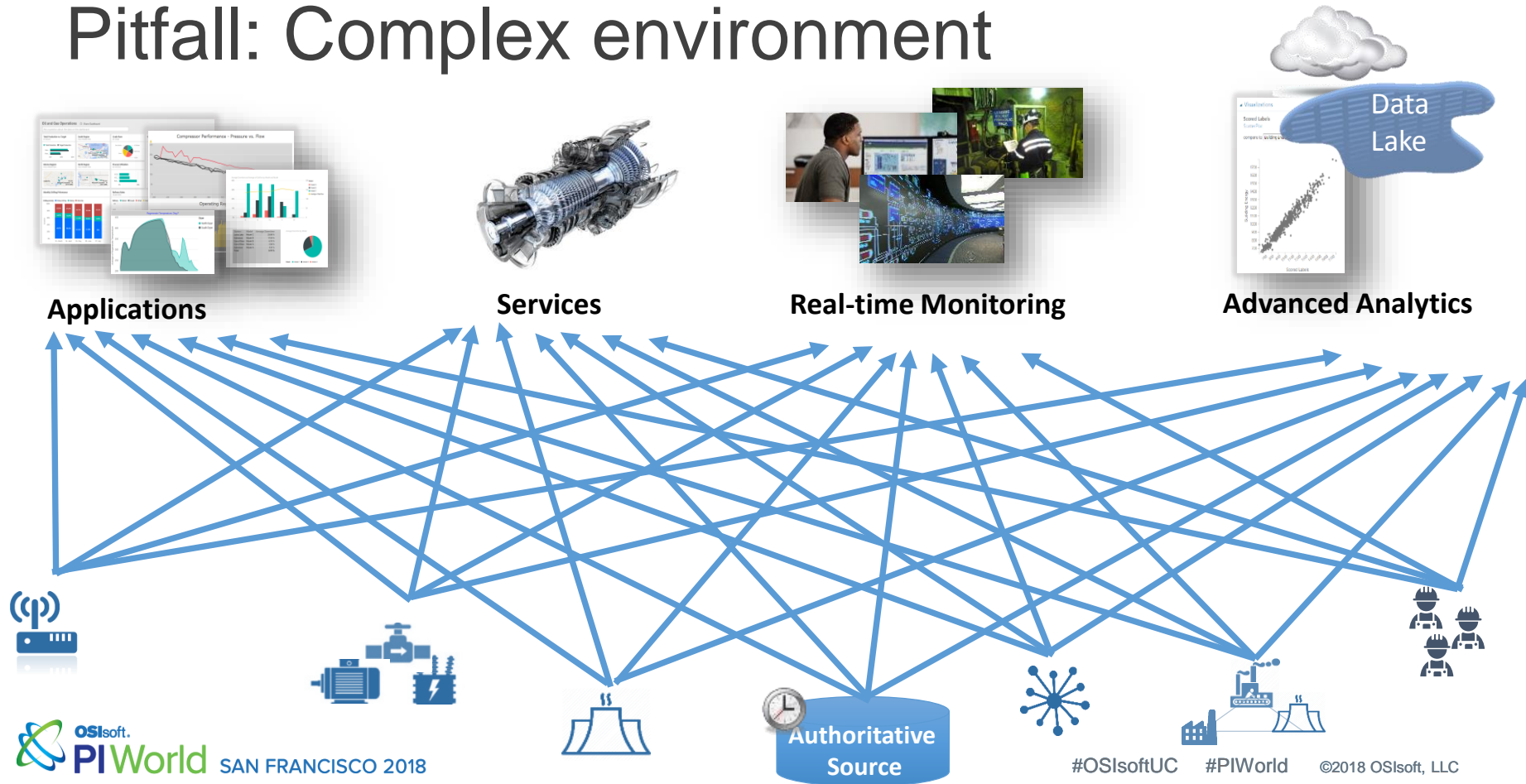


Each facility addressing similar business challenges

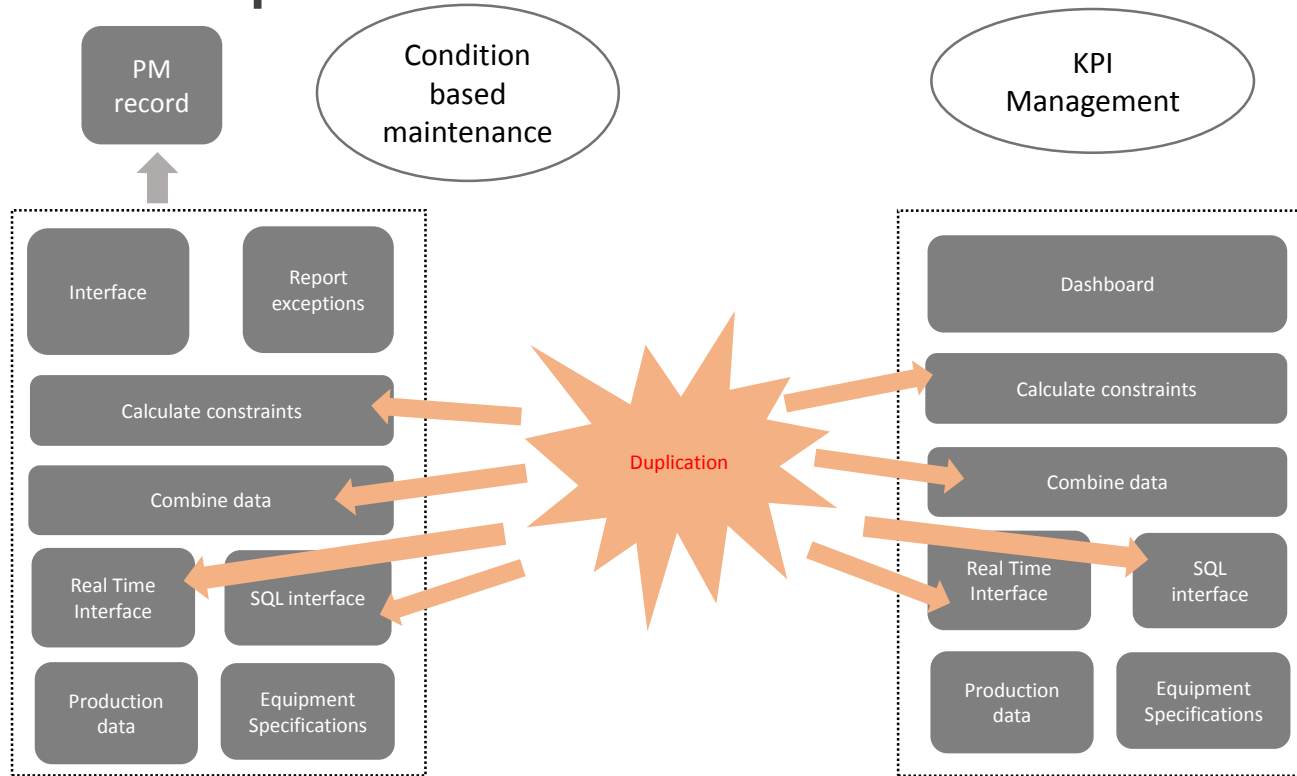


Overall result:
Cost too high,
Value too low

Pitfall: Complex environment

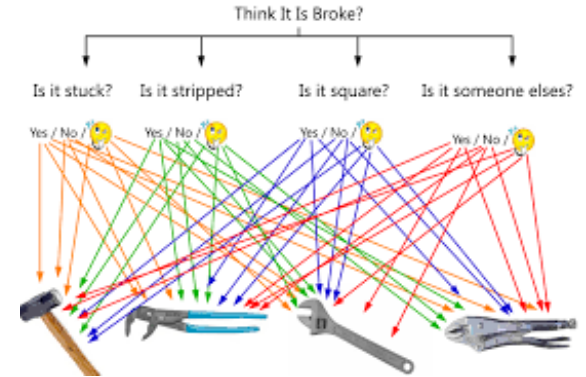


Pitfall: Duplication of Effort



Pitfall: Tools don't match the workflows

- Every tool has a specific job
- In the right hands, the right tool will get the job done
- Be functional not fashionable
- For using not for displaying



Design Principles: Business Impact Assessment

Disruption Risk Scales

| Category | Negligible | Minor | Moderate | Major | Severe |
|-------------------|---|--|--|--|--|
| Financial | Potential loss of < \$10,000 | Potential financial loss of \$10,000-\$50,000 | Potential financial loss of \$50,000-\$200,000 | Potential financial loss of \$200,000-\$500,000 | Potential financial loss of \$500,000 + |
| Human | Potential for minor injury requiring first aid treatment | Potential for injury or illness resulting in medical attention and several days off work | Potential for injury or illness resulting in short-term hospitalisation | Potential serious long-term injury | Potential for death, permanent disability or ill-health |
| Legal/ Compliance | Minor internal dispute that can be remedied without external intervention | Potential for compliance, contractual or regulatory breaches with external implications | Confirmed compliance, contractual or regulatory breaches. Specific activities required to remedy situation | Significant penalties and/or costs to remedy legal and/or compliance breaches | Severe penalties to company and/or staff |
| Operational | No noticeable impact on operational functions | Short-term disruption to operational functions | Significant disruption to operational functions | Extended disruption to operational functions | Collapse of operational functions |
| | No noticeable impact on supply chain | Short-term disruption to supply chain | Significant delays to supply chain | Extended delays to supply chain | Total supply chain failure |
| | Minimal change to work conditions | Short-term increase in working hours | Sustained deterioration in working conditions | Long-term deterioration in working conditions resulting in increased sick leave and potential resignations | Unacceptable working conditions resulting in workplace injuries/illness and resignations |
| Reputational | Adverse impact that can be remedied immediately | Adverse impact that is short term and reversible at minimal cost | Adverse impact with potential for significant damage | Impacts requiring long term remedial attention | Irreversible damage to brand and reputation |
| Strategic/ Market | Localised concern – no impact on long term viability | Detrimental to short-term profitability and/or strategic direction | Detrimental to mid-term profitability and/or strategic direction | Significant long-term impacts. Will required change strategic direction and objectives | Business viability in question |

Design Principles: Recovery objective



Design Principles: High level rules

- All closed loop related control functionality or functionality that could cause a facility shutdown must be in the control system
- All functionality that could jeopardize safety must be in the control system

Business reasons for a PI system in the control environment:

- Control system doesn't have the capability for high fidelity trending
- Control system doesn't have the capability to calculate derived values or run small streaming analytics
- Network connectivity is of low performance (VSAT) or low reliability
- Productions need to run independent from corporate/business network

Find your way in the forest of Technologies

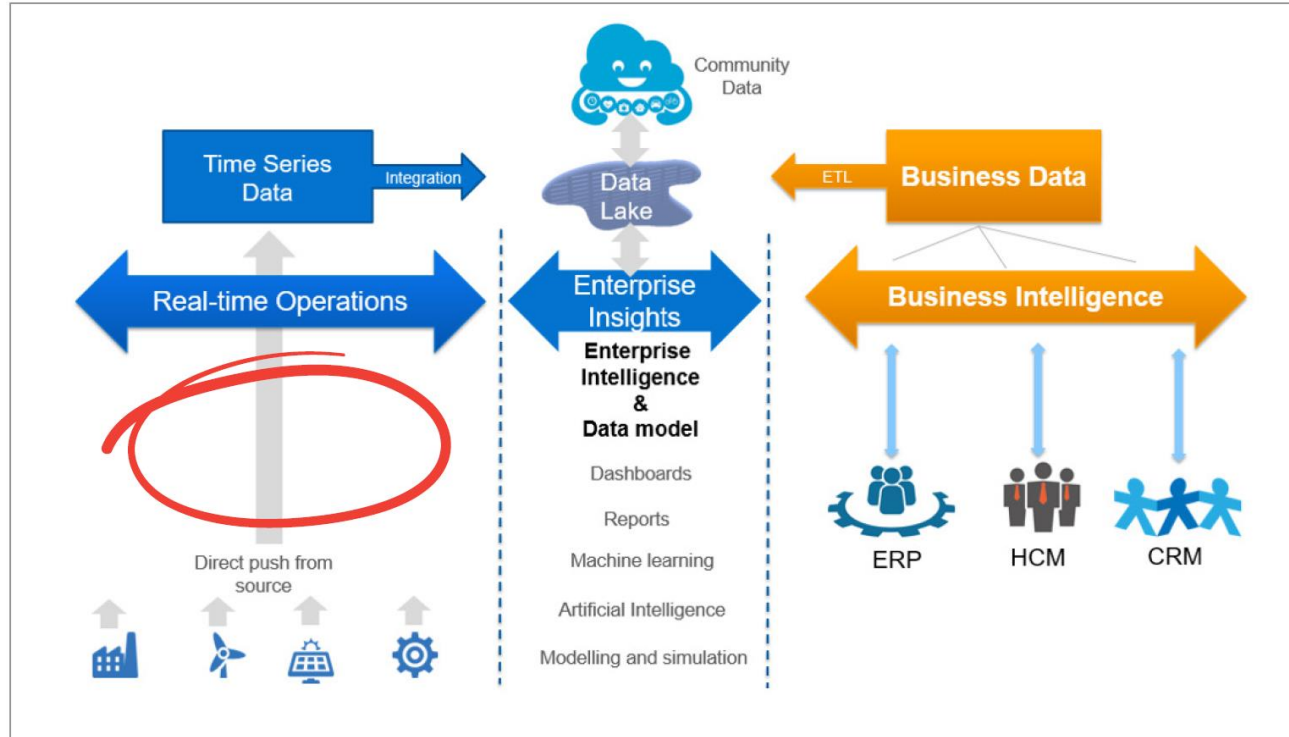
- *Businesses of all kinds are beginning to see their data as an important asset.*
- *Data can help make operations more effective and profitable.*
- *The ability to gather time-series and non time-series data is growing.*
- *More technologies are becoming available to help us make sense of it.*

How do we choose the right technology and approach for our business problems?

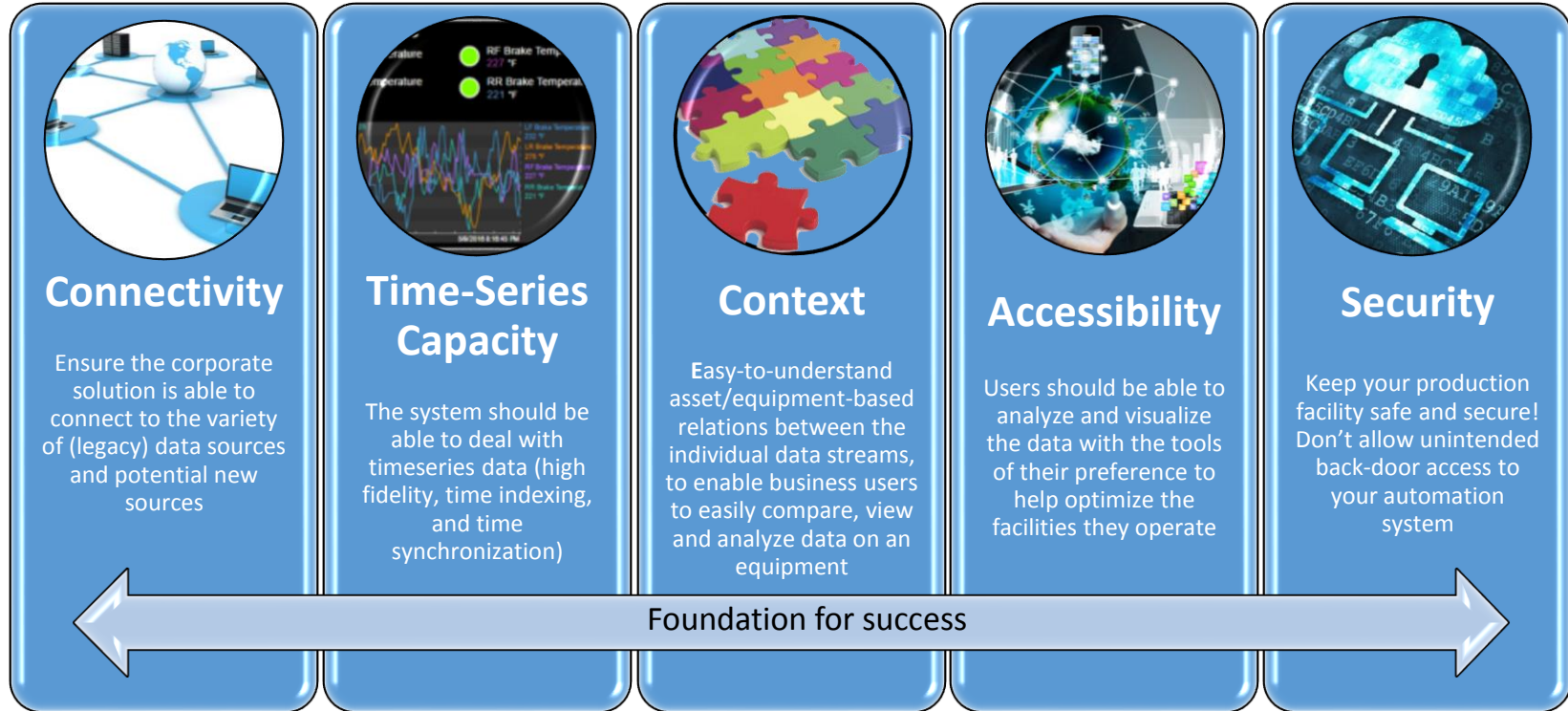
Data Warehouse versus Data Lake

| Data Warehouse | | Data Lake |
|----------------------------------|-------------------|---|
| Structures, processed | Data | structured / semi-structured / unstructured, raw |
| schema-on-write | Processing | schema-on-read |
| expensive for large data volumes | Storage | designed for low-cost storage |
| less agile, fixed configuration | Agility | highly agile, configure and reconfigure as needed |
| mature | Security | maturing |
| business professionals | Users | data scientists et. al. |

Perfect World



Fundamentals for Time Series Architecture



Value of Real-Time Infrastructure

Advanced Real-Time systems

Off-the-shelf interfaces available to support a high variety of (legacy) data sources. Even supporting system of 20+ years old.

Data compression avoids sending similar values and avoids data overloads on low bandwidth links like Satellite. The data buffering avoids data loss in case of network failure.

Grouping data points into asset sets makes equipment data easy accessible

Relationship between the various pieces of equipment makes it easy to find data without knowing the details of the facility

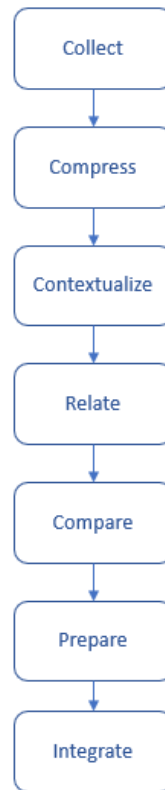
Context and relationships between data makes it very easy for non-data scientists to add related analytics and compare similar types of equipment. Identify events and compare events between data streams

Based on the context and relationships it is easy to select, align in time, cleanse and prepare data to be used in reporting or other systems on a facility and enterprise level

Prepared relational and contextualized data can easily be integrated in data lakes or other cloud based solutions like GIS, ERP or BI

VS

Data lake



Limited interfaces available. For new industry standards only.

Lack of compression results in higher storage cost and overload of remote connections. Data loss in case of network failure.

Context between data points is done after the data landed in the lake leaving the field without context

The concept of a Data Lake is to have unstructured data. This makes it hard to connect data. Knowledge at a data scientists level is needed to find relationships between data streams.

Additional applications and data scientist knowledge is needed to compare the data of similar types of equipment

Additional applications and data scientist knowledge is needed to prepare and select the data for further processing in other tools

Data lakes will easily hold and integrate structured and non-structured data. However this data is without context and relationship.

Power of Context

Visualisations



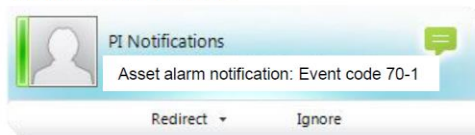
Historical Performance & Future Predictions



Events

| Name | Duration | Start Time | End Time |
|----------------------|----------|---------------------|---------------------|
| TopLevel_EventFrame | 0:01:52 | 4/5/2015 5:57:03 PM | 4/5/2015 5:58:55 PM |
| Unlabeled_EventFrame | 0:01:36 | 4/5/2015 5:57:07 PM | 4/5/2015 5:58:43 PM |
| Operation_EventFrame | 0:01:17 | 4/5/2015 5:57:10 PM | 4/5/2015 5:58:27 PM |
| Phase_EventFrame | 0:00:33 | 4/5/2015 5:57:13 PM | 4/5/2015 5:57:46 PM |
| Operation_EventFrame | 0:00:14 | 4/5/2015 5:57:50 PM | 4/5/2015 5:58:04 PM |
| Phase_EventFrame | 0:00:09 | 4/5/2015 5:58:36 PM | 4/5/2015 5:58:45 PM |

Alerts & Notifications



Pump

Reports



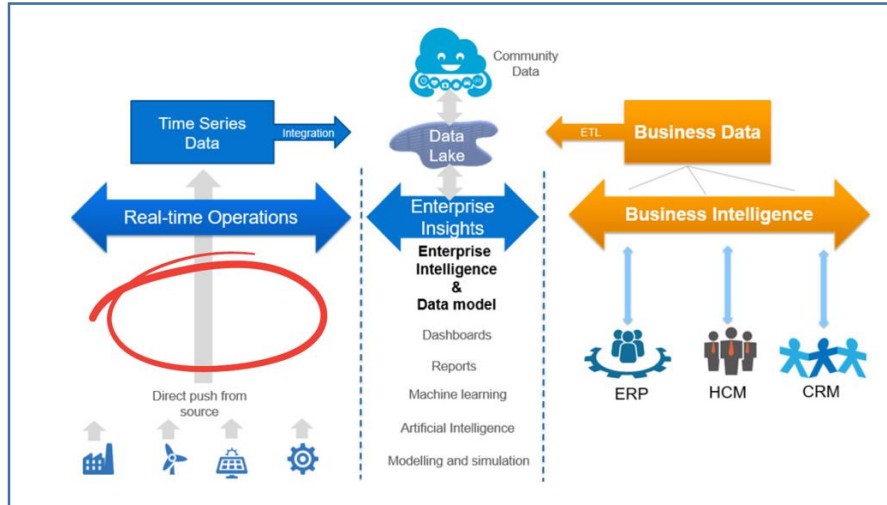
Automated Workflows



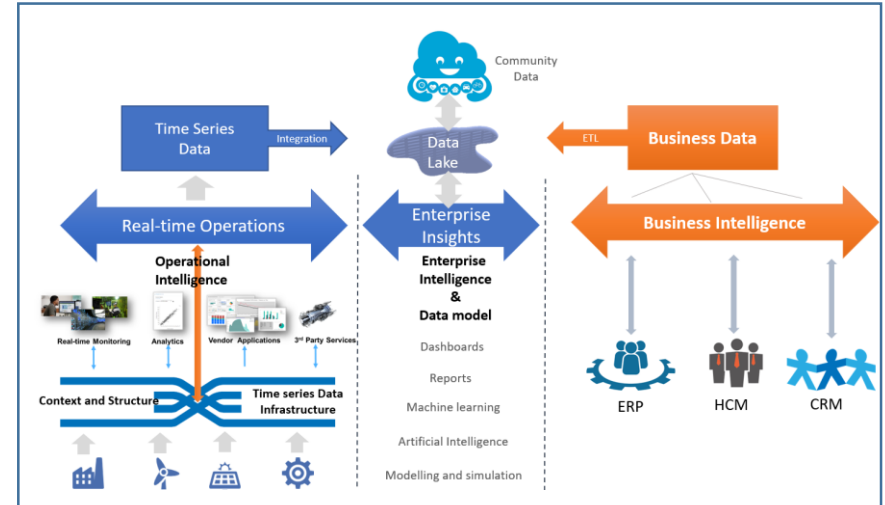
Time Series Data Infrastructure with Data Lake integration

- Automation vendor-based like Honeywell PHD or Yokogawa Exaquantum
- Open source-based like InfluxDB, Graphite, and Prometheus
- Rotating Equipment vendor-based like Siemens XHQ
- Vendor independent systems like the OSIsoft PI System

Hybrid Solution for RT Data in a Data Lake Environment

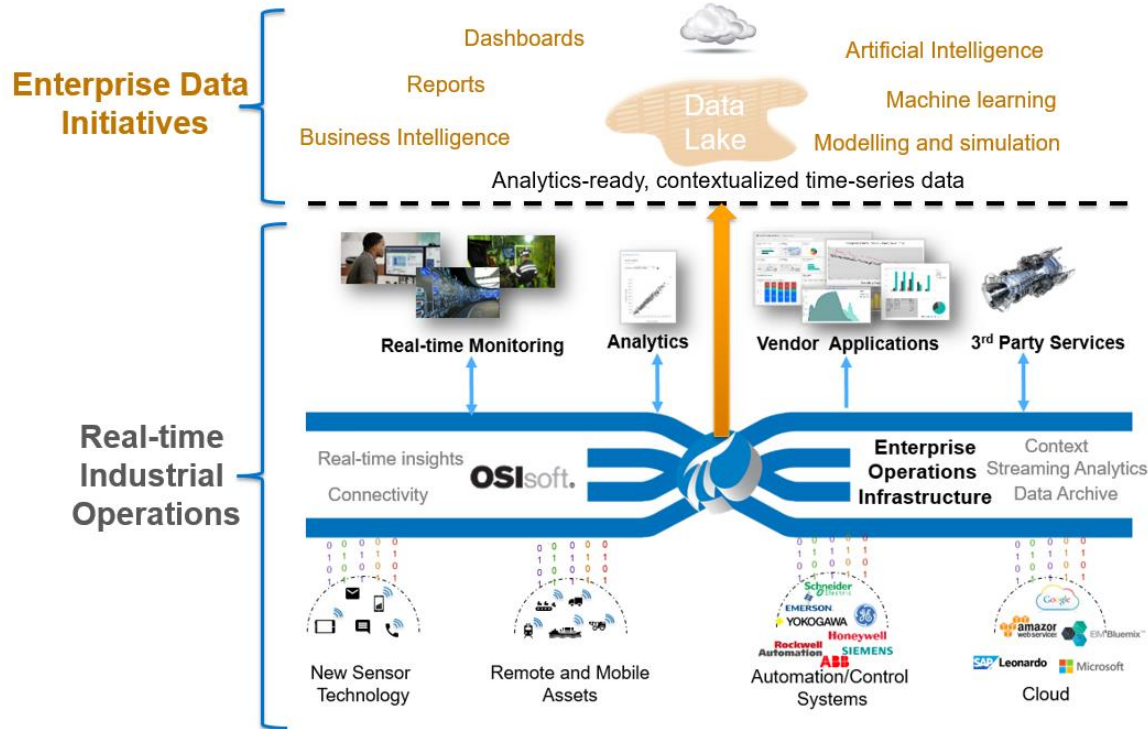


Perfect world

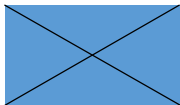


Hybrid reality

How OSIsoft PI System Blends in Perfectly



Why a data lake alone can't replace your PI system



Leveraging Data Lake technology in a Real-Time environment

Fundamentals for Time Series Architecture



CHALLENGE

How to best integrate data in a Data Lake

- Many data sources
- Diversity of technology
- Context of data
- Data Security

SOLUTION

Understand your business challenges, What do you want to fix where, keep the five fundamentals in mind

- Get uniform access to data
- Create Digital Twins to standardize tools and use cases
- Synchronize in time
- Unlock the data to commonly used tools
- Secure the data

RESULTS

Predictable and value driven journey for Digital Transformation

- Company standardization
- Reusable solutions at the right place in the architecture
- Success transformation from Reactive to Predictive

Why a data lake alone can't replace your PI system



- **John de Koning**
- con-jdekoning@osisoft.com
- Customer Success Advisor
- OSIsoft

Questions

Please wait for the **microphone** before asking your questions

State your **name & company**



Please remember to...

Complete the Online Survey for this session



Download the Conference App for OSISOFT PI World Conference 2018

- View the latest agenda and create your own
- Meet and connect with other attendees



search OSISOFT in the app store

Merci

谢谢

Спасибо

Danke

Gracias

Thank You

감사합니다

ありがとう

Grazie

Obrigado

Optional: Click to add a takeaway you wish the audience to leave with.