

GLOBAL  WEBINAR

Leverage the Right Technology to Avoid a Data Swamp

Presenter:

John de Koning, Customer Success Advisor, OSIsoft

November 8, 2017



Our Speaker



John de Koning

Customer Success Advisor, OSIsoft

Email: con-jdekoning@osisoft.com

John de Koning is a Customer Success Advisor who helps OSIsoft customers be more successful in generating value on their Digital Transformation Journey. John is able to leverage 40 years experience in the Oil & Gas industry in various business roles ranging from operations, process control, process technology, and IT. The last 10 years of his career he was leading Digital Transformation at the enterprise level for Shell.

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?

Polling Question



Which technology is your company currently using?

A. Application specific databases

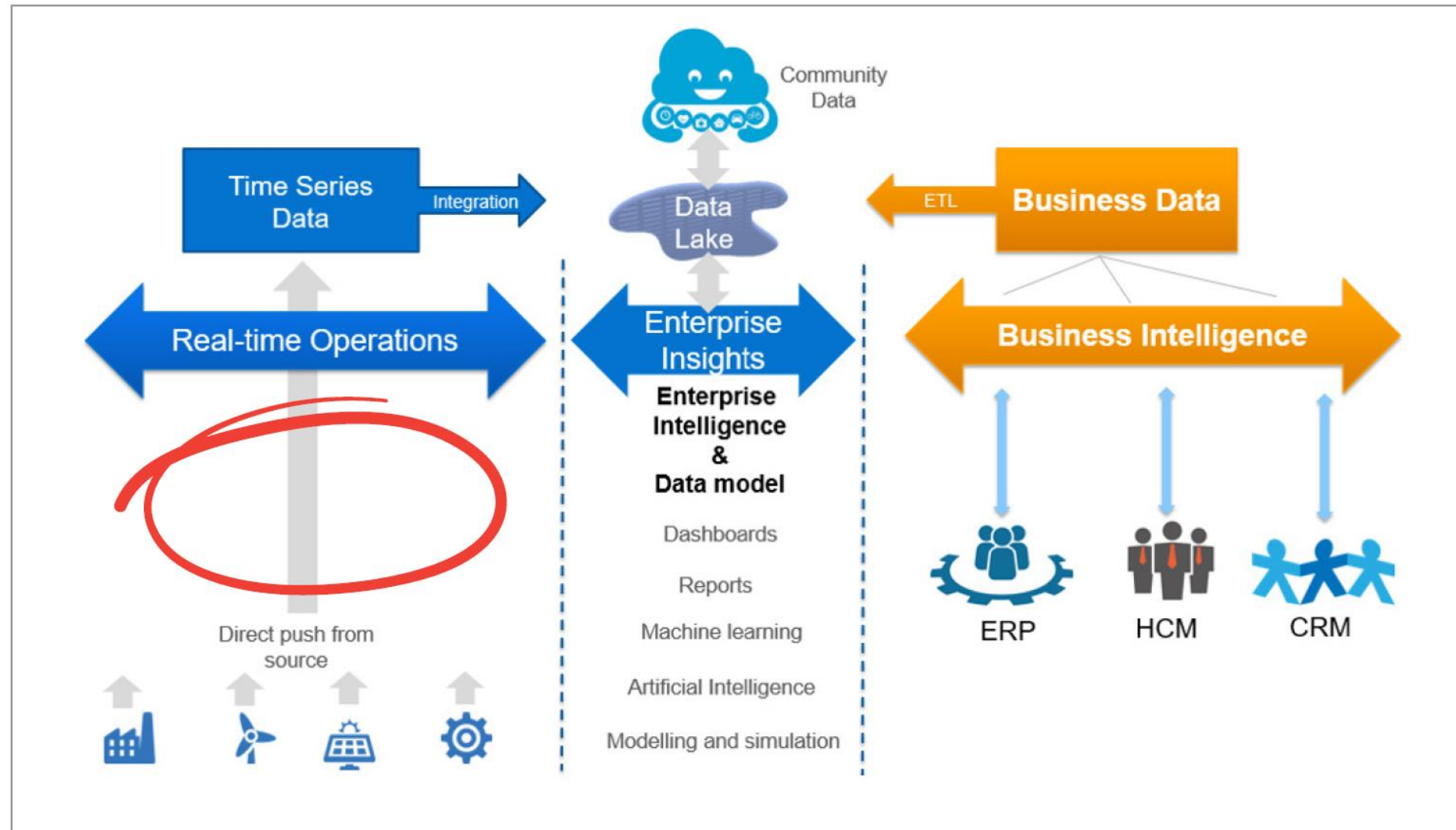
B. Centralized warehouse

C. Data Lake technology (SAP-HANA, Azure, Hadoop, etc.)

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



Polling Question



Is your company using cloud based data storage?

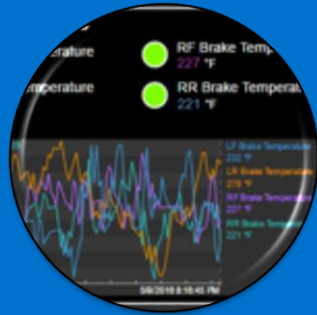
- A. All our data is in the cloud
- B. New development will be based on cloud storage
- C. Only to share some data with others
- D. Only non-critical data is in the cloud
- E. No, not at all

Fundamentals for Time-Series Solution Architecture



Connectivity

Ensure the corporate solution is able to connect to the variety of (legacy) data sources and potential new sources



Time-Series Capacity

The system should be able to deal with timeseries data (high fidelity, time indexing, and time synchronization)



Context

Easy-to-understand asset/equipment-based relations between the individual data streams, to enable business users to easily compare, view and analyze data on an equipment



Accessibility

Users should be able to analyze and visualize the data with the tools of their preference to help optimize the facilities they operate



Security

Keep your production facility safe and secure! Don't allow unintended back-door access to your automation system

Foundation for success

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

Collect

Limited interfaces available. For new industry standards only.

Compress

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

Contextualize

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

Relate

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.

Compare

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

Prepare

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

Integrate

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

The Power of Context

Visualisations



Historical Performance & Future Predictions

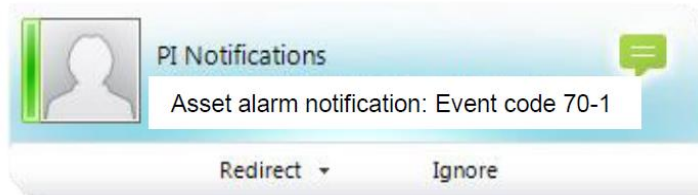


Events

Event Frame Search 1

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
UnitProcedure_EventFrame	0:01:38	4/5/2015 5:57:07 PM	4/5/2015 5:58:45 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
Phase_EventFrame	0:00:33	4/5/2015 5:57:50 PM	4/5/2015 5:58:23 PM
Operation_EventFrame	0:00:14	4/5/2015 5:58:31 PM	4/5/2015 5:58:45 PM
Phase_EventFrame	0:00:09	4/5/2015 5:58:36 PM	4/5/2015 5:58:45 PM

Alerts & Notifications

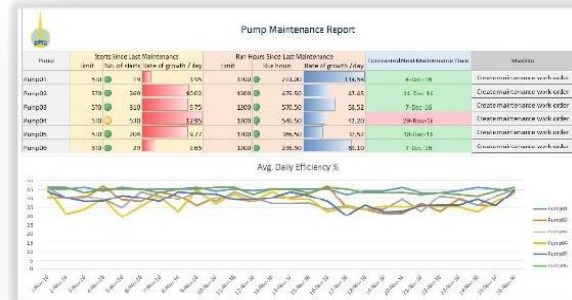


Asset Framework (AF) Data

Category	Property	Value
General	Asset Name	Pump
	Asset Type	Pump
	Asset Location	Plant
	Asset Status	Operational
Performance	Flow Rate	464.66 US gal/min
	Pressure	150 ft
	Temperature	117.05 °F
	Efficiency	85.32%
Maintenance	Last Maintenance Date	10/16/2016 12:00:00 PM
	Next Maintenance Due	11/16/2016 12:00:00 PM
	Maintenance Status	Upcoming
	Maintenance Trigger	Time-based

Pump

Reports



Automated Workflows



Polling Question

Is your company using a standardized corporate model for equipment definitions?

- A. Yes, all of our use cases are based on this
- B. Yes, but it is currently in development and only used in a few cases
- C. No, some standards are available but only at a site level
- D. No, not at all



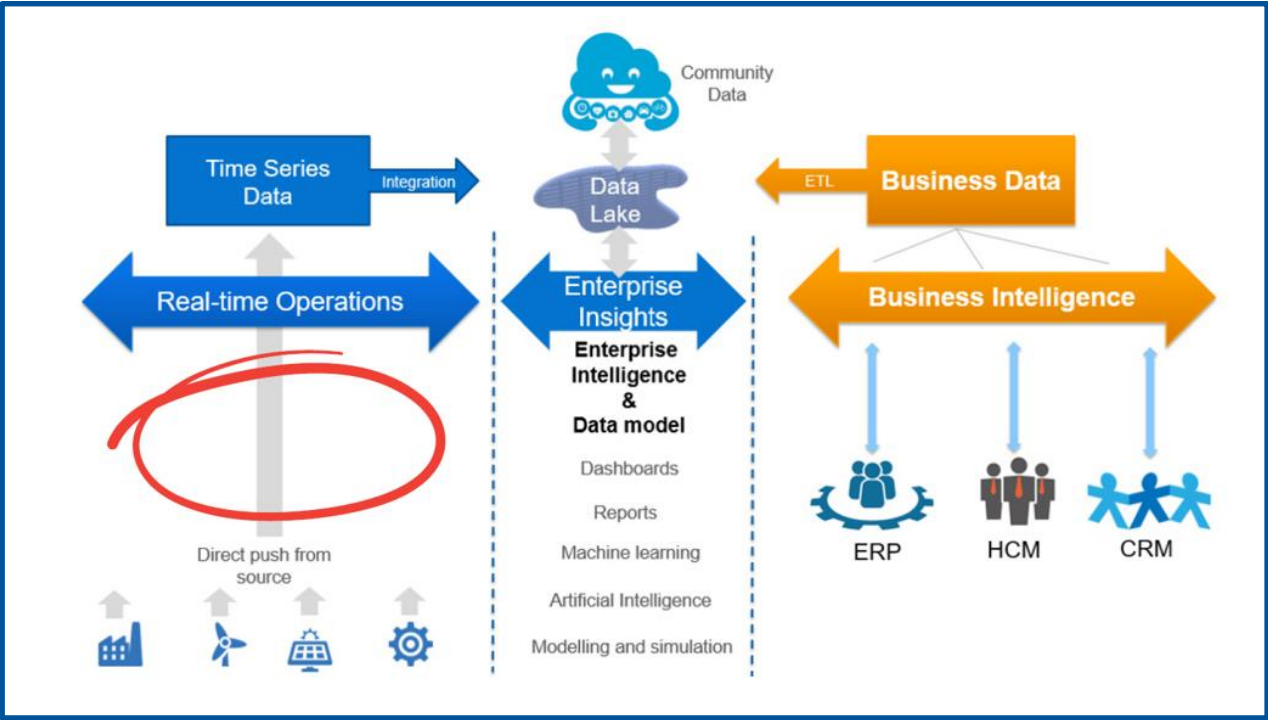
Time Series Data Infrastructure with Data Lake Integration

- 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

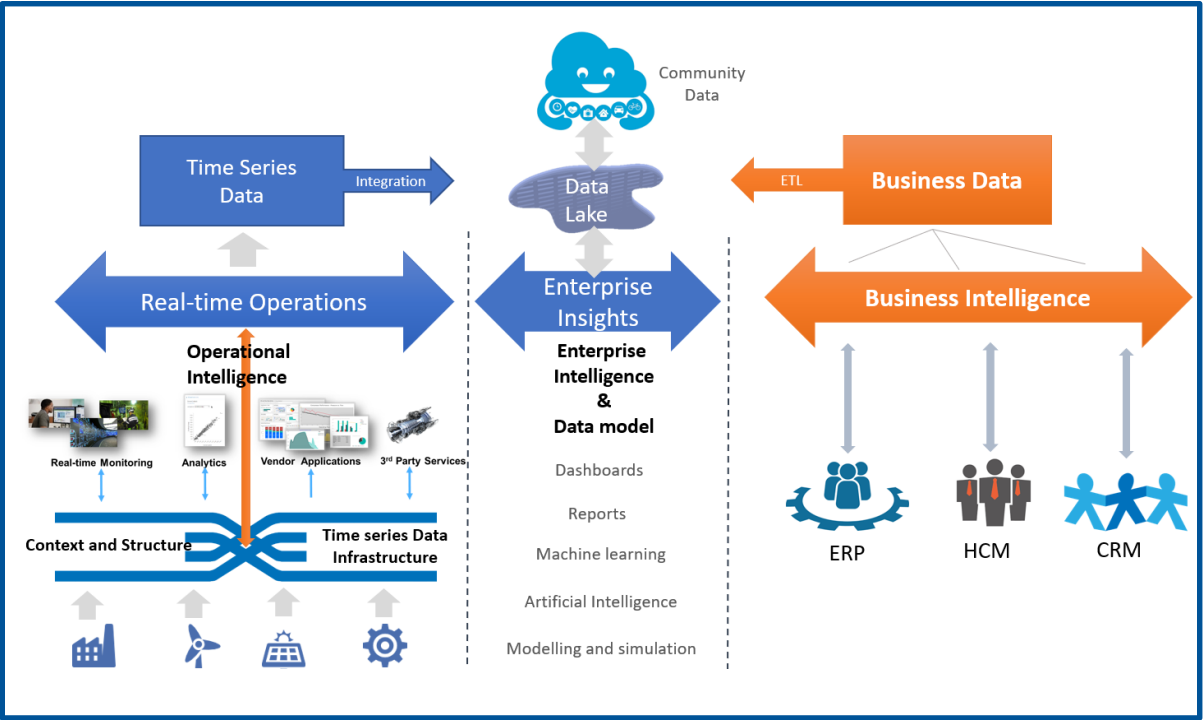
See White Paper for detailed comparison

<https://www.osisoft.com/White-Papers/Data-Lake-or-Data-Swamp/>

Hybrid Solution for RT Data in a Data Lake Environment

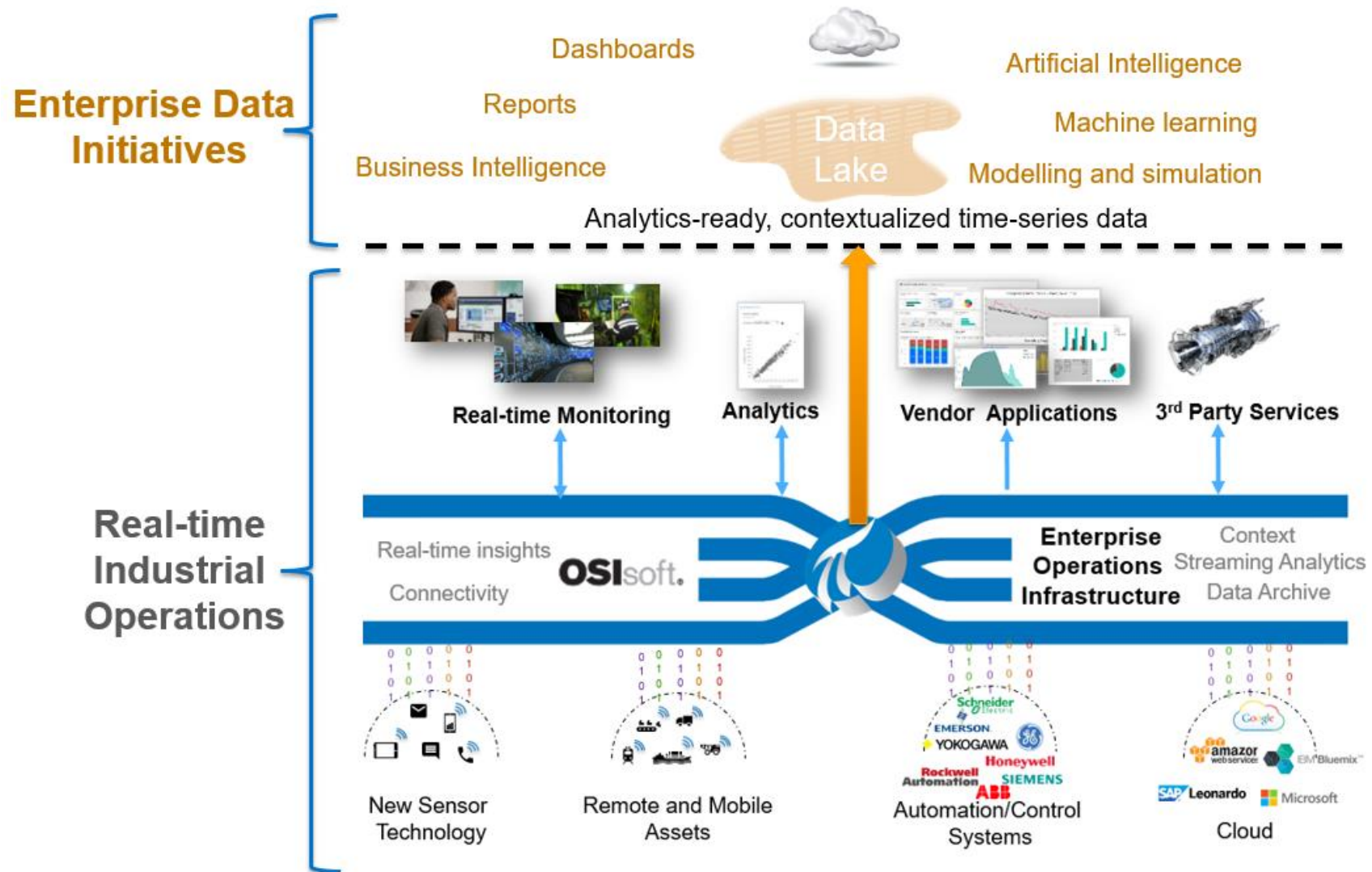


Perfect world



Hybrid reality

How OSIsoft PI System Blends in Perfectly



Polling Question



Will this webinar help you be successful with your Data Lake and Real-time Data?

- A. Yes, this gives me clear direction
- B. Only some parts are useful
- C. This is not the right direction for our Data Lake integration.
- D. No, not at all

Digital Transformation

Data Lake or Data Swamp

GOAL

Maximize the value of Data Lake technologies for integration of corporate data



CHALLENGE

How to avoid a Data Lake becomes a data swamp

- Many (legacy) data sources
- Context of data is not available
- Time synchronization and Data Latency
- Data aggregation and integration

SOLUTION

Leverage technologies designed to bring structure to unstructured real-time data

- Get uniform access to data
- Create standardized Digital Twins
- Unlock the data in a structured way
- Transform in a usable format

RESULTS

Aggregation and Integration of structured and contextualized data at a corporate level

- Company wide standardization
- Enable Powerful Data Analytics
- Accelerate value generation with fit for purpose data
- Successful basis for Digital Transformation

White Paper: Data Lake or Data Swamp



Keeping the Data Lake
from Becoming a
Data Swamp.

<https://www.osisoft.com/White-Papers/Data-Lake-or-Data-Swamp/>

Questions and Answers



Polling Questions



How was the presentation?

- A. Too technical
- B. Just right
- C. Needs more depth

We appreciate any feedback from you.

Would you like to speak to your Account Manager?

- A. Yes
- B. No

Next

- White paper: Data Lake or Data Swamp
- Stayed tuned for follow-up email
 - Link to OSIsoft On-Demand webinars
 - Visit PI Square
- Have conversation with John



John de Koning

Customer Success Advisor, OSIsoft
Email: con-jdekoning@osisoft.com

감사합니다

谢谢

Danke

Merci

Gracias

Thank You

ありがとう

Спасибо

Obrigado

GLOBAL  WEBINAR

Leverage the Right Technology to Avoid a Data Swamp

SPEAKER:

John de Koning

Customer Success Advisor, OSIsoft

LIVE | Wednesday, November 8, 2017 | 8-9AM PDT

