SIEMENS
Ingenuity for life

XHQ Operations Intelligence

System Architectures for Advanced Analytics

August 2019

XHQ Operations Intelligence Software is used for the aggregation, integration, analysis, and presentation of information from multiple back-end data sources.

XHQ brings together engineering, manufacturing, operations, and equipment data into a cohesive system for in-context reporting, analysis, and visualization. It enables a process plant or enterprise to quickly understand the true state of the business, process and assets, troubleshoot problems, and manage routine conditions.

Advanced analytics uses sophisticated mathematical tools to generate models from data and make predictions and recommendations. This paper describes ways to implement XHQ and advanced analytics to make the best use of your data and help process plants operate better.

Contents

1. Introduction

The process industries are undergoing a digital transformation. The use of **digital twins**—virtual, digital replicas of physical assets, processes, and systems—is increasing. Advanced analytics help implement the digital twin, by using sophisticated mathematical tools to generate models from data. These models make predictions and recommendations, which help process plants operate better.

This paper describes general system architectures for three ways to implement XHQ and advanced analytics: with MindSphere Predictive Learning, with the XHQP Performance Analytics modules, and self-service analytics with Seeq from Seeq Corporation. A companion paper describes use cases for applying analytics, in the areas of fixed asset health, rotating equipment health, maintenance costs, and unstructured data.

A successful analytics system requires four things. One is data: having a wide variety of current and historical data to train and run analytics is important. A second requirement is to develop specialized models—the digital twin—of real facilities that take into account the available data and unique aspects of the operation. A third requirement is to make regular predictions automatically. A fourth requirement is to present predictions and recommendations in a way that makes sense to users, and leads them to take the proper actions.

In general, XHQ provides the first and fourth items: access to data and visualizing information. The three analytics systems described here take different approaches for the second and third items, developing models and making regular predictions.

2. Background

2.1 The Digital Twin

The process industries are undergoing a digital transformation. The use of **digital twins -** virtual, digital replicas of physical assets, processes, and systems - is increasing. Advanced analytics help implement the digital twin, by using sophisticated mathematical tools to generate models from data. These models make predictions and recommendations. Predictions about the future life of assets can help determine how best to operate in the present, and comparison between predicted and actual conditions can help identify issues. More succinctly, analytics makes predictions from data, which can help operate better.

2.2 Analytics

Analytics is the discovery, interpretation, and communication of meaningful patterns in data. There are many techniques. Advanced analytics make use **of machine learning**, which "is a data science technique that allows computers to use existing data to forecast future behaviors, outcomes, and trends. Using machine learning, computers learn without being explicitly programmed. ... Predictive analytics uses math formulas called algorithms that analyze historical or current data to identify patterns or trends in order to forecast future events."[1] Typically an expert uses large data sets to develop, or **train**, algorithms, for example, by using a large collection of historical data about pump seals to come up with methods to predict pump seals failures. Once developed, algorithms can be run on new data to make predictions.

2.3 Types of Analytics

Self-service analytics enable engineers and other subject matter experts do their own investigation and analysis. Tools in this category are intended for people who know their area but are not necessarily trained statisticians or data scientists. This category includes **business intelligence** (BI) tools such as Tableau, Qlik, and Power BI. It also includes specialized tools such as Seeq, which is optimized for time series process history.

The **data science** approach refers to tools used by data scientists to do machine learning and other types of advanced analytics. Typically data scientists are experts who develop methods for general use, such as devising methods to predict the remaining useful life of an asset based on millions of historical records. Typically data scientists use multiple tools that together provide libraries of methods, data sets management, scripting, and visualization. Commonly used languages and tools include R, Hadoop, SAS, and many others. **MindSphere Predictive Learning** falls in this category.

When deployed for use, analytics can be run **centrally** or at the **edge**. Centralized server systems work well when a lot of computer resources are needed, algorithms need to be updated, and some delays in response time are acceptable. Edge systems work well when the resource needs are less, systems can run without attention, and fast response times are useful.

2.4 Analytics Lifecycle

There is a lifecycle to analytics. Typically the cycle starts with an **investigation** phase where a data scientist, subject matter expert, or power user uses tools to investigate data in search of insights or a solution to a problem. For example, an expert may analyze pump seal failures to look for root causes or ways to reduce failures. Analysis may be followed by a **development** phase where an expert develops tools for others to use, such as machine learning algorithms that predict the likelihood of failure. Once tools are developed, they must be **applied** in the field, for example, configuring an algorithm that predicts pump seal failure to predict the likelihood of failure for hundreds of real pumps. Finally, algorithms need to be **executed** to make predictions.

These names - **investigation, development, application,** and **execution** - will be used in this document to refer to these distinct activities.

The figure shows these stages operating sequentially, and shows typical roles of XHQ and MindSphere Predictive Learning in these activities. There are many other possible flows and alternatives. For instance, XHQ can integrate with other analytics engines besides MindSphere.

---

[1] Introduction to Machine Learning in the Azure cloud. Microsoft Azure. 2017, July 12. Retrieved 2017, October 18 from https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-machine-learning

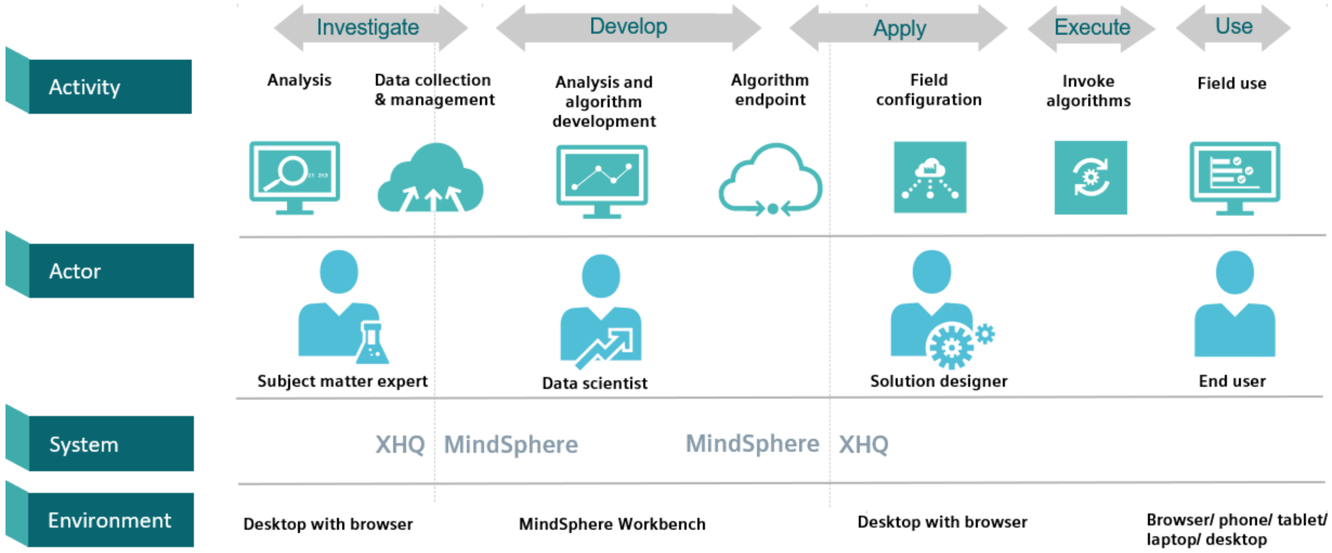


Figure 1. Analytics lifecycle

2.5 XHQ

XHQ is software for operations intelligence in the process industries, bringing together manufacturing, operations, and equipment data into a cohesive system for in-context reporting, analysis, and visualization. With XHQ software, a process plant or enterprise can create a rich set of dashboards and visual displays that combine information from many sources, which helps staff quickly understand the true state of the business, process and assets; troubleshoot problems; and manage routine conditions.

Most people interact with XHQ through its rich web pages, which are customized to suit local needs. Web pages are built from widgets that include schematics, pictures, tables, charts, trends, and GIS.

XHQ connects to any data source - databases, historians, maintenance, ERP, quality, supply chain, and many others - organizes and contextualizes the data into a cohesive information model, and visualizes the data. Displays are driven by the XHQ information model, which organizes and contextualizes data as the business is organized - by function, by facility, by production line.

Performance Analytics is an optional XHQ module for Extract, Transform, and Load (ETL) functions. It is useful to extract information from multiple XHQ systems, cleanse and reformat the data, and store it for long term reporting and analysis.

3. Data Science with MindSphere

3.1 MindSphere Predictive Learning

Some problems are complex and require sophisticated tools for data scientists. MindSphere Predictive Learning is a complete system for data scientists. **MindSphere Predictive Learning** is the analytics part of MindSphere; MindSphere is the cloud-based, open IoT operating system from Siemens that connects products, plants, systems, and machines, enabling you to harness the wealth of data generated by the Internet of Things (IoT) with advanced analytics.

MindSphere Predictive Learning provides two main things: an environment for data scientists to develop algorithms, and an environment for running those algorithms.

3.2 Idea

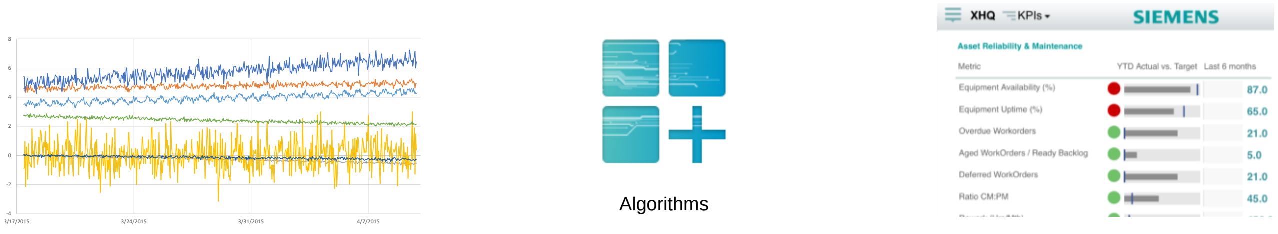
The goal is to make interesting predictions available to end users, by having experts apply machine learning. Machine learning is a discipline for handling problems where three things are true: a pattern exists, we cannot pin it down mathematically, and we have data about it. The essence of machine learning is to develop (train) an algorithm to fit one set of data, and to use that algorithm to make predictions from other data.

To apply machine learning, we need data scientists to develop algorithms capable of making predictions. This means we need an offline, development environment, and an online runtime environment to execute those algorithms. We also need a way to visualize the results.

This approach uses **MindSphere Predictive Learning** to develop algorithms. This is an offline activity that results in algorithms being published and made available to authorized users. The online, runtime system combines XHQ and MindSphere Predictive Learning. XHQ coordinates the activities, schedules jobs, gathers input data, sends the data to MindSphere, invokes the algorithms, fetches the results, and displays the new information. From the point of view of XHQ and XHQ users, MindSphere Predictive Learning is just another data source.

OFFLINE: From data, a data scientist investigates and develops algorithms

PRODUCTION: XHQ calls the algorithms to get predictions, for use in the XHQ solution



3.3 Architecture

The XHQ server is installed on a dedicated Windows server, as usual. This server can be on premises or in the cloud. XHQ is configured with the usual connectors, model, and views.

MindSphere Predictive Learning is a cloud service available on a subscription basis. Siemens provides complete system support – security, authorization and authentication, data management, and so on. More specifically, Predictive Learning has these characteristics:

- MindSphere is designed to be secure, to scale, and to operate as a true service.
- Input and output are handled as text files, in the MindSphere data lake.
- Algorithms are implemented in Zeppelin notebooks. Zeppelin (https://zeppelin.apache.org/) is software from the Apache Software Foundation that is meant for data scientists to do data ingestion, discovery, analytics, and visualization. It supports a variety of languages and system, such as Python, JDBC, and Hadoop.
- MindSphere provides various helper tools for use in the Zeppelin notebooks.
- MindSphere APIs are used for all operations, including copying files to and from the data lake, running the Zeppelin notebooks, and managing the system.

The art of creating an algorithm in MindSphere Predictive Learning boils down to a data scientist creating a Zeppelin notebook and checking it into MindSphere. Once checked it, other MindSphere applications can call the MindSphere APIs to learn about the algorithm, including what inputs it takes, what output it produces, and what it does, and to run the algorithm.

The main integration point between XHQ and Predictive Learning is a planned[2] new XHQ connector, called the "MindSphere Analytics Connector".[3] This new connector has a number of features and supporting tools. One supporting tool is used to map items in the XHQ model to algorithm inputs.

MindSphere Predictive Learning works in the background. End users work in XHQ and data scientists work in MindSphere. From the point of view of XHQ, MindSphere Predictive Learning is just another data source.

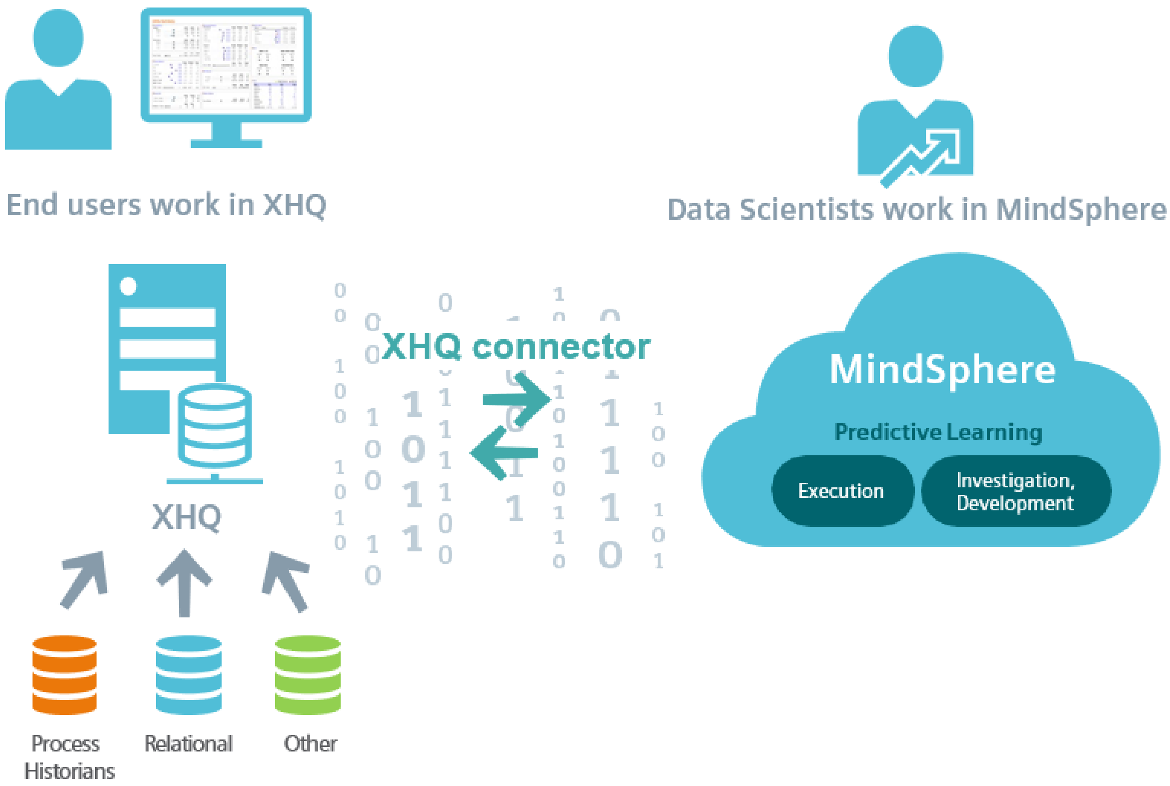


Figure 1. Integrating XHQ and MindSphere Predictive Learning

[2] The MindSphere Analytics connector is part of the XHQ road map plans. As always, plans can change and this document is not a commitment for delivery.

[3] There is also a MindSphere IoT connector, which deals with IoT data and is not involved in analytics.

4. Data Science with Performance Analytics

4.1 Performance Analytics

XHQ is software for operations intelligence in the process industries, bringing together manufacturing, operations, and equipment data into a cohesive system for in-context reporting, analysis, and visualization. Performance Analytics is part of XHQ, an optional module that is useful to support in-depth analysis of complex data sets. For example, a company that wants to build specialized reports across multiple systems and heterogeneous data sources can use standard XHQ functions to connect to the data sources at each system, and use Performance Analytics to pull selected data from the various XHQ systems into a single, consolidated data set, cleansing and normalizing the data along the way.

4.2 Idea

The goal is to make interesting predictions available to end users, by having experts apply machine learning. Machine learning is a discipline for handling problems where three things are true: a pattern exists, we cannot pin it down mathematically, and we have data about it. The essence of machine learning is to develop (train) an algorithm to fit one set of data, and to use that algorithm to make predictions from other data.

To apply machine learning, we need data scientists to develop algorithms capable of making predictions. This means we need an offline, development environment, and an online runtime environment to execute those algorithms. We also need a way to visualize the results.

This approach uses XHQ's Performance Analytics module to gather, organize, structure, and store data. Performance Analytics is designed to do several things:

- Extract, calculate, transform, and save XHQ data in new data sets.
- Perform sophisticated calculations and data cleansing operations.
- Consolidate data from multiple XHQ systems into a single data set.
- Handle large data sets and long term storage.
- Run jobs.

Algorithms can be developed by customers, by Siemens for customers, and by Siemens partners such as Cyient Ltd (www.cyient.com). R, Python, and other common data science tools are used to write algorithms. Standard XHQ techniques are used to read the results from SQL Server. From the point of view of XHQ and XHQ users, Performance Analytics is just another data source.

4.3 Architecture

One or more XHQ servers are installed on dedicated Windows servers, as usual. These servers can be on premises or in the cloud. XHQ is configured with the usual connectors, model, and views.

Performance Analytics is installed on its own dedicated Windows server, along with Microsoft SQL Server and Microsoft's SQL Server Integration Services (SSIS) package. Performance Analytics is configured to extract data from XHQ through calls to the XHQ API, transform the data, and load the resulting data into SQL Server. This extraction process is scheduled, perhaps hourly.

Algorithms are developed offline in R or Python by data scientists and subject matter experts. They are deployed to the Performance Analytics server and wrapped in a SQL Server stored procedure. Depending on the algorithm, Performance Analytics can schedule jobs that run algorithms and write results to SQL Server, or expose stored procedures that return a data set of results.

Generally speaking, the SQL Server database is designed to store all input data (collected by Performance Analytics) and all results from the analytics (from the algorithms). XHQ uses its standard connector for SQL Server to get data from SQL Server tables or run the stored procedures when needed.

Performance Analytics works in the background. End users work in XHQ and data scientists work with tools on SQL Server. From the point of view of XHQ, Performance Analytics is just another data source.

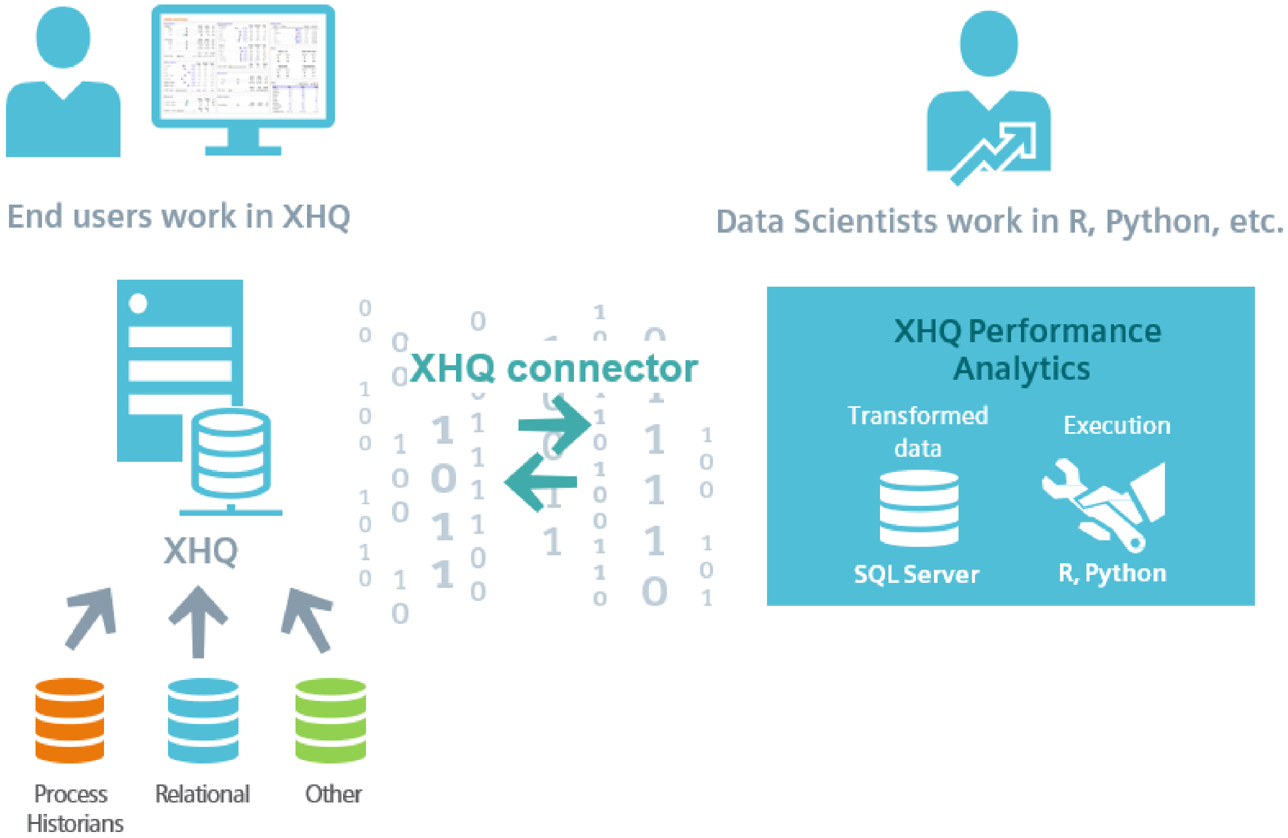


Figure 3. XHQ used with Performance Analytics

4.4 SQL Server and R

The following example[4] executes a SQL Server stored procedure (sp_execute_external_script), passing a complete R script in the second parameter. This particular example uses a data set of car speed and the distance required to stop a car and creates a linear regression model that describes some relationship between these variables.

[4] *Quickstart: "Hello World" R Script in SQL Server.* Microsoft Corporation. 14 July 2018, docs.microsoft.com/en-us/sql/advanced-analytics/tutorials/rtsql-using-r-code-in-transact-sql-quickstart?view=sql-server-2017. Accessed 4 Sept. 2018.


```
DROP PROCEDURE IF EXISTS generate_linear_model;
GO
CREATE PROCEDURE generate_linear_model
AS
BEGIN
    EXEC sp_execute_external_script
    @language = N'R'
    , @script = N'lrmodel <- rxLinMod(formula = distance ~ speed, data = CarsData);
        trained_model <- data.frame(payload = as.raw(serialize(lrmodel, 
connection=NULL)));'
    , @input_data_1 = N'SELECT [speed], [distance] FROM CarSpeed'
    , @input_data_1_name = N'CarsData'
    , @output_data_1_name = N'trained_model'
    WITH RESULT SETS ((model varbinary(max)));
END;
GO
```

Once the model is developed and available in SQL Server, the following script makes use of the model on a new data set to make predictions about predicted stopping distance.

```
DECLARE @speedmodel varbinary(max) = (SELECT model FROM 
[dbo].[stopping_distance_models] WHERE model_name = 'latest model');
EXEC sp_execute_external_script
    @language = N'R'
    , @script = N'
            current_model <- unserialize(as.raw(speedmodel));
            new <- data.frame(NewCarData);
            predicted.distance <- rxPredict(current_model, new);
            str(predicted.distance);
            OutputDataSet <- cbind(new, ceiling(predicted.distance));
            '
    , @input_data_1 = N'SELECT speed FROM [dbo].[NewCarSpeed]'
    , @input_data_1_name = N'NewCarData'
    , @params = N'@speedmodel varbinary(max)'
    , @speedmodel = @speedmodel
WITH RESULT SETS (([new_speed] INT, [predicted_distance] INT))
```

In this architecture, Predictive Analytics gathers the input data from XHQ and provides the environment and scheduling. Typically the stored procedure output is stored in SQL Server, which XHQ reads like any other data source.

5. Self-service Analytics with Seeq®

5.1 Self-Service Analytics

Sometimes engineers and other subject matter experts need tools more than they need expert help. Self-service analytics enable engineers do their own investigation and analysis of assets and operations for which they are responsible.

Self-service analytics refers to systems where business users are able to access relevant data, investigate data history and relationships, perform statistical and other types of analysis, and prepare their own visualizations. The idea is to let business analysts and knowledgeable users do their job without requiring (much) support from the IT organization. ("Self-service" doesn't mean "self-sufficiency", as there is still a need for training, governance, ensuring high quality data sets, security and controls, sharing, and system management – all IT functions that continue to be required.)

5.2 Idea

The goal is to enable engineers and other experts to do their own investigation of process data. Traditional BI tool such as Tableau and Power BI are not ideal for this task, and a potentially better option is to use a product called Seeq® from Seeq Corporation (seeq.com), a privately held company that sells software for self-service analytics of data in process historians. "Seeq" is both the name of the company and the name of its flagship product. Seeq is a Siemens business partner.

Figure 4. Seeq® enables engineers to analyze their data [Image courtesy of Seeq Corporation © 2018]

Seeq is designed specifically to work with data from process historians. Seeq works with OSIsoft® PI, Honeywell Uniformance® PHD, and other historians, and can work with XHQ.

This approach uses XHQ to connect to data sources, organize information, and provide common visualization, and adds Seeq as an option for engineers to do their own analysis.

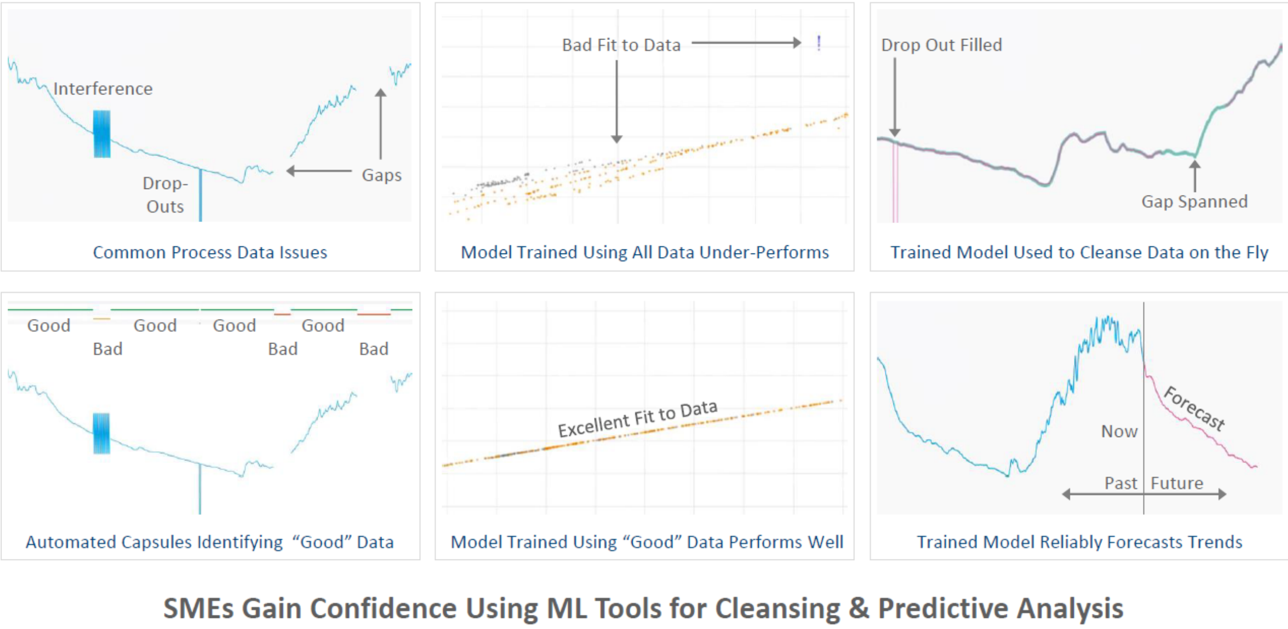


Figure 5. Seeq Machine Learning features. [Image courtesy of Seeq Corporation © 2018]

5.3 Architecture

The XHQ server is installed on a dedicated Windows server, as usual. This server can be on premises or in the cloud.

The Seeq server is installed on another dedicated server. Seeq can be configured to connect directly to process historians such as OSIsoft PI and/or to XHQ. For many purposes it will be useful to have Seeq connect to XHQ, and have XHQ handle connecting to the backend systems. Going through XHQ provides advantages such as a single namespace and access to laboratory data, MindSphere IoT, and other systems.

XHQ and Seeq are both web applications, and users can connect to either application. XHQ views are configured

XHQ is configured with the usual connectors, model, and views.

with options to launch Seeq, passing the context to Seeq. In addition, Seeq can be configured to do pattern matching and calculations, and to expose those results to XHQ. If this is done, XHQ connects to Seeq like any data source, using the XHQ REST connector.

Seeq works in the foreground. End users work in both XHQ and Seeq, and data scientists are not required. It is possible for XHQ to use Seeq in the background for calculations and predictions, but end users are likely to want to go to Seeq to understand the predictions.

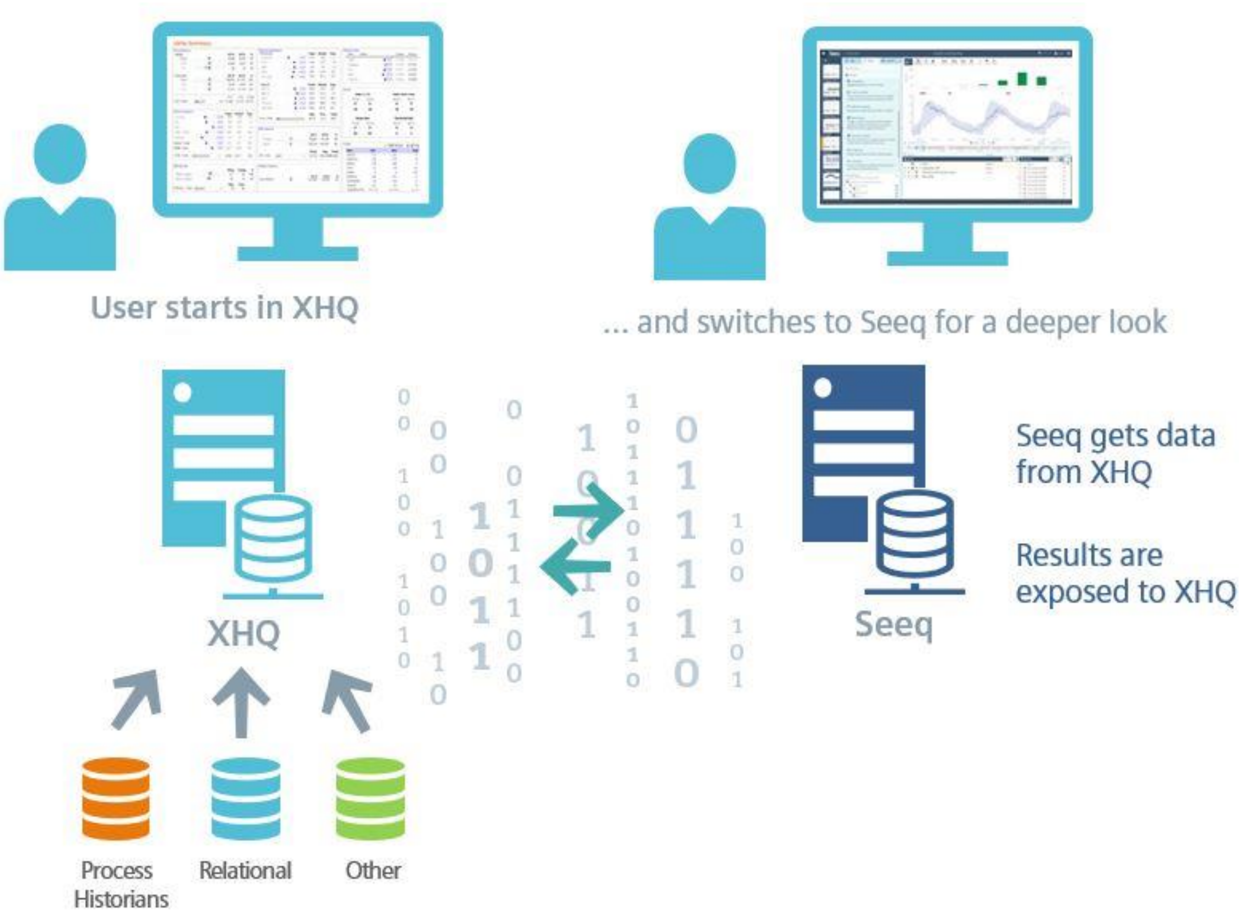


Figure 6. Integrating XHQ and Seeq

6. Comparison

Attribute	Data Science with MindSphere	Data Science with Performance Analytics	Self-Service Analytics with Seeq
Target users	Data scientists	Data scientists	Engineers, subject matter experts
End users interact with	XHQ	XHQ	XHQ and Seeq
Data scientists use	MindSphere Predictive Learning	SQL Server tools	Seeq
Types of algorithms	Rich	Rich	Less complete (appropriate for target users)
Types of data	Any	Any	Process history
Investigation/development environment	MindSphere, Zeppelin	SQL Server, R, Python, others	Seeq
Runtime environment	XHQ + MindSphere	Performance Analytics	Seeq
Getting results into XHQ	MindSphere Analytics connector	SQL Server connector	REST connector
Cloud or on premise	Cloud	Both	Both
Server or edge	Server	Server	Server
Licensing model	SaaS/subscription	License or subscription	License or subscription
Service requirements	Develop algorithms, map XHQ data to algorithms	Develop algorithms, design data warehouse, configure transformations	End user training

XHQ Operations Intelligence Software
Siemens Product Lifecycle Management Software, Inc.
DI PA AE CIS XHQ
6 Journey, Suite 200
Aliso Viejo, CA 92656, USA

info.xhq@siemens.com

siemens.com/xhq