Implementation Guide

SIMCA-online 16.1

Ensure Manufacturing Success

©1992-2021 Sartorius Stedim Data Analytics AB, all rights reserved

Information in this document is subject to change without notice and does not represent a commitment on part of Sartorius Stedim Data Analytics AB. The software, which includes information contained in any databases, described in this document is furnished under license agreement or nondisclosure agreement and may be used or copied only in accordance with the terms of the agreement. It is against the law to copy the software except as specifically allowed in the license or nondisclosure agreement. No part of this user guide may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose, without the express written permission of Sartorius Stedim Data Analytics AB.

Guide edition date: May 19, 2021

SARTORIUS

Sartorius Stedim Data Analytics AB
Östra Strandgatan 24
SE-903 33 Umeå
Sweden
Phone: +46 (0)90 18 48 00
Email: umetrics@sartorius.com

Contents

1 Introduction 2
1.1 Additional documentation 2
1.2 What is SIMCA? 3
1.3 What is SIMCA-online? 3
1.4 SIMCA-online is designed for process data 3
1.5 Batch processes and continuous processes 3
1.6 Two types of SIMCA-online project configurations 4
1.7 SIMCA-online is licensed on the server 4
1.8 SIMCA-online compared to SIMCA-Q 4
1.9 Getting data out from SIMCA-online 4

2 Technical requirements and best practices 4
2.1 A stable process 4
2.2 Good models 4
2.3 Low latency networking 5
2.4 Server computer 5
2.5 Data source 6
2.5.1 A SimApi for the data source 6
2.5.2 Low data acquisition latency: real-time data must be current 6
2.5.3 Historical data access is strongly recommended 6
2.5.4 Real-time (current) data must match data read as historical data 6
2.5.5 Batch processes 6

3 Implementation 8
3.1 People 8
3.2 Implementation process 8
3.2.1 Preparations 8
3.2.2 Outline of a SIMCA-online installation 8
3.2.3 Next steps 9

1 Introduction

SIMCA-online enables real-time multivariate process monitoring and control using SIMCA models and data from a data source, such as a process historian.

This document outlines the core functionality of SIMCA-online, puts it in context with the Umetrics Suite of Data Analytics Solutions, describes the requirements and best practices for implementing the SIMCA-online solution, and finally shows, step by step, how to install and set up SIMCA-online.

1.1 Additional documentation

The document you are reading now is one of a set of related documents, each with different focus and target audience:

Source	What	Where
SIMCA-online web page	Introductory information and downloads	sartorius.com/umetrics-simca-online
SIMCA-online ReadMe and Installation.pdf	Installation and how to get started with SIMCA-online demo data	In the installation zip file
SIMCA-online Implementation Guide	Outlines SIMCA-online functionality, puts it in context with other Umetrics Suite software, describes requirements and best practices for successful deployment, and step-by-step installation instructions.	sartorius.com/umetrics-simca-online
SimApi Guide	Preparing for and performing SimApi installations, including troubleshooting. Also contains technical details on SimApis for developers.	sartorius.com/umetrics-simapi
SimApi User Guides	One for each published SimApi. Lists the features, installation instructions, and configuration specifics of each individual SimApi.	sartorius.com/umetrics-simapi
SIMCA-online Technical Guide	Technical reference for SIMCA-online server installation planning, troubleshooting, and in-depth how SIMCA-online works.	sartorius.com/umetrics-simca-online
SIMCA-online help	How to use SIMCA-online and how SIMCA-online works.	In the software itself, and on sartorius.com/umetrics-simca-online
SIMCA-online Web Client Installation Guide	Describes the installation of the SIMCA-online Web Client.	sartorius.com/umetrics-simca-online
Umetrics knowledge base	Searchable database with articles about each released software version, technical articles, and known issues on Umetrics Suite products.	sartorius.com/umetrics-kb
SIMCA help / user guide	How to use SIMCA for creating projects.	In SIMCA and on sartorius.com/umetrics-simca

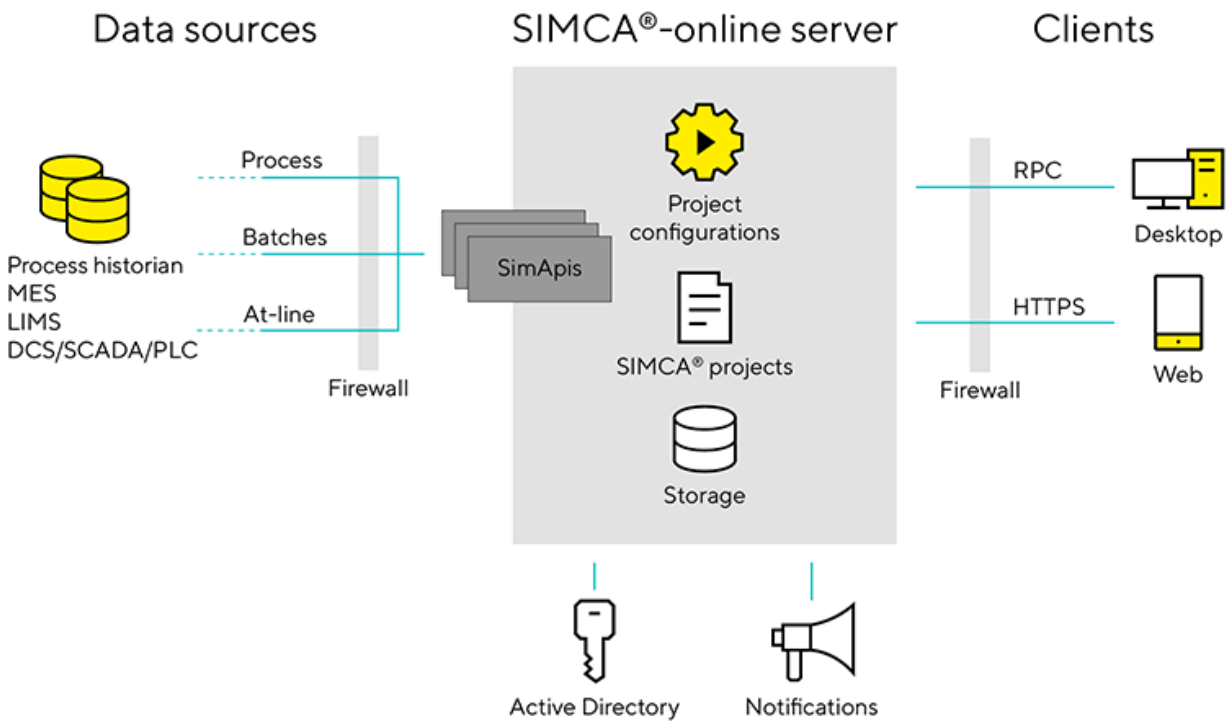
1.2 What is SIMCA?

SIMCA is Sartorius Stedim Data Analytics AB software for all-purpose multivariate data analysis of any data that can be turned into numeric data. SIMCA is run on a desktop computer and performs the data analysis using models which are saved in SIMCA project files (.usp files).

1.3 What is SIMCA-online?

A SIMCA-online system is a client-server solution consisting of:

- **desktop clients** connect to the SIMCA-online server to monitor the process, change settings, and interact with charts
- optional **Web Clients** connecting securely with https can be used to monitor production in mobile or desktop browsers. Web clients use the public web API of the SIMCA-online server.
- a **SIMCA-online server** that
  - uses copies of SIMCA project files containing models
  - stores all data collected and produced by the server in its own storage
  - keeps **project configurations** containing settings that tie a project's variables to tags in the data source
  - uses an execution algorithm that uses the above to execute projects
- one or more **SimApi** software interfaces used by the server to pull data from one or more **data sources** such as process historians. Data can optionally be written back to the data source through the SimApis.
- optional connection to **Active Directory** for user authentication
- optional delivery of email **notifications** and/or logging to Windows Application event log
- SIMCA-online works well with firewalls between data sources, server and clients



1.4 SIMCA-online is designed for process data

SIMCA, the offline desktop software, can be used to model and analyze almost any data, while **SIMCA-online** is designed to work with **process data** where a process is monitored over time. SIMCA-online supports two types of processes; continuous and batch.

1.5 Batch processes and continuous processes

A **continuous process** is measured at regular intervals. The process is assumed to be stable around a finite number of ideal states.

In a **batch process** data is aggregated into batches. Each batch has a start time, an evolution, and an end. An example of a batch process from real life is frying an egg. It starts out raw, and you monitor the frying process (batch evolution). Then the batch is ended when the egg reaches the desired maturity.

You decide which type of process you model when you create the project. A project can be created in SIMCA, or directly in SIMCA-online using File | New.

1.6 Two types of SIMCA-online project configurations

In SIMCA-online, continuous processes and batch processes are handled by two types of project configurations

- **Continuous configuration** – execute and monitor a single model where data is read at regular intervals. The model can be a hierarchical top model.
- **Batch configuration** – execute and monitor a batch model group (BM) that can contain many models, each of which may be a hierarchical model:
    - one or more batch evolution models (BEMs). Each phase is executed at regular intervals and models the batch maturation.
    - optional batch level models (BLMs) that execute at pre-defined levels of batch maturity, and optionally at regular intervals after the batch has finished to be able to analyze late entry data.

Batch configurations are more complex, but also more capable than continuous configurations. For example, more than one batch can execute simultaneously in different units of a batch project configuration.

1.7 SIMCA-online is licensed on the server

The SIMCA-online server license together with your agreement with Sartorius controls which features are available in SIMCA-online. The license is activated during installation on the server and is typically locked to the server hardware. A license can be valid for a limited time, locked to a specific SimApi, or have other limitations. Talk to your sales representative to learn more about licensing.

1.8 SIMCA-online compared to SIMCA-Q

SIMCA-online is an off-the-shelf system that pulls data from a data source and executes models.

SIMCA-Q is the Sartorius Stedim Data Analytics AB product for embedded multivariate data analysis. Like SIMCA-online, SIMCA-Q can be used to execute models stored inside SIMCA project files. But unlike SIMCA-online it is not a standalone system but needs to be integrated in another system that handles the data acquisition, execution logic, charting, and user interface. This requires software development by an OEM integrator.

1.9 Getting data out from SIMCA-online

There are two ways to obtain data from SIMCA-online to be used in another system:

- Many SimApis support write back to the data source. Data from project execution, such as predictions or alarm status can be written back to a data source. Learn more in the SIMCA-online help.
- SIMCA-online has a built-in web server that exposes a web API. It provides information about configurations, projects and models, and output data including predictions. The web API is used by the SIMCA-online Web Client. Learn more in Using the SIMCA®-online web API for developers (Q907) (at sartorius.com/umetrics-kb)

2 Technical requirements and best practices

In order to implement SIMCA®-online, several requirements need to be fulfilled. This section lists these requirements prefixed with a checkbox, ☐. We also list best practices. For a deeper discussion of many topics, refer to the SIMCA-online Technical Guide.

2.1 A stable process

A stable process without uncontrolled drift over time is desirable. Otherwise models will be less applicable as the process drifts and will require re-modelling, or for batch processes the use of local centering.

2.2 Good models

☐ Multivariate monitoring models should be built on representative data when the process is in the desired state. Design of Experiments (DOE) can be used to find the design space of the process.

☐ Multivariate prediction (regression) and forecasting (predictive) models should be built on data with representatively sized variation to get good predictions. Design of Experiments (DOE) can be used to ensure representative data.

☐ Multivariate advised future (prescriptive) models should be built on data with representatively large variation, specifically at those times when values for manipulated variables are computed. Design of Experiments (DOE) can be used to ensure representative data and causal relationship between the manipulated variables and dependent variables. Bump tests should be used to verify the causality in the process.

☐ Perform model validation in SIMCA before deploying to SIMCA-online. Test the models with out-of-specification data to ensure that anomalies are detected.

The execution interval (data sampling frequency) for data in a model should be selected to match how fast the process changes based on the sampling theorem. For example, to be able to see changes once a minute, the sampling interval should be 30 seconds.

☐ The execution interval must match the sampling frequency used to build the model if it uses time-dependent variables, such as lagged variables[1], or generated variables that depend on earlier observations.

Batch project specifics:

- ☐ Evolution maturity must be monotonically increasing or decreasing.
- Prefer using a natural maturity variable from the process, rather than automatically generated maturity, $Time.
- ☐ If you do use automatic time, $Time, as maturity, make sure that the SIMCA-online execution interval matches the sampling frequency used when building the model.
- Do not configure maturity to be normalized.
- ☐ Use projects with a **single** batch model group (BM). If you have more than one, simply delete the extra ones in SIMCA.
- You can have different sampling intervals for different evolution phase models (BEMs).

How to build models is out of scope of this document. Sartorius Stedim Data Analytics AB provides training and offer services from data scientists to help you.

2.3 Low latency networking

The network should be high bandwidth, low latency to not affect performance negatively.

- ☐ Locate the SIMCA-online server close to the data source to ensure fast access to data.
- ☐ Locate the SIMCA-online clients close to the server to ensure responsive clients. Tip: desktop virtualization such as remote desktop can be used to increase performance of clients.

2.4 Server computer

Server hardware, and operating system settings affect SIMCA-online:

- ☐ Decide on a suitable windows server computer to install the SIMCA-online server on.
- ☐ The computer system clocks should be synchronized between the server computer, data source and clients.
- ☐ The SIMCA-online server service account, or the account configured for use by a SimApi, must have access to the data source.
- ☐ The SIMCA-online server service account must have access to Active Directory if it is to be used for user-authentication and single sign-on. If the server computer is a domain member it will automatically have access to Active Directory.
- ☐ A SMTP server reachable from the server computer is required to use email-based notifications.

The SIMCA-online Technical Guide, chapter 2, gives more information on these topics.

[1] An exception to this rule is a model that uses **dynamic lags,** introduced in SIMCA 15, which **does** support a different execution interval in SIMCA-online.

2.5 Data source

A commercial data source for process systems such as OSIsoft PI is well suited for SIMCA-online because it provides the necessary logic and data structures to deal with process data and batches.

A high-availability data source is recommended, because errors in the connection or other data source problems can lead to problems with batch project execution[2].

2.5.1 A SimApi for the data source

☐ A **SimApi** for the process historian data source is required for SIMCA-online to be able to access data.

The SimApi is a software layer that translates the requests from SIMCA-online to a protocol that the data source understands. There are SimApis available for various data sources such as ODBC for relational databases like SQL server, OSIsoft PI, and OPC.

We provide a software development kit for developers wishing to implement a SimApi to connect to their own data source. However, if the data source is low level such as an instrument in the lab and don't provide historical data, we recommend that you instead of developing a SimApi, you connect your data source to a process historian, such as OSIsoft PI, and let it take care of obtaining, historizing and keeping data.

Learn more about SimApis at sartorius.com/umetrics-simapi.

2.5.2 Low data acquisition latency: real-time data must be current

When a data source is used by SIMCA-online in real-time, it is important that the data in the data source is current. There should be no unnecessary delays in the data acquisition in the data source. For SIMCA-online this means;

☐ The maximum delay for data acquisition inside the data source should be half the execution (sampling) interval in SIMCA-online. Otherwise, real-time data won't match the same data requested as historical data at a later time (see 2.5.4).

☐ Continuous process data for **all** variables must be available at the same time for every observation. Data that come in late for some variables won't be picked up by SIMCA-online.

To re-read past data, an operator has to manually (re-)predict past data. Note: late entry discrete data and batch condition data **can** be picked up automatically (learn more in help).

2.5.3 Historical data access is strongly recommended

☐ We strongly recommend a data source and SimApi that provide not only current data for real-time execution, but also *historical* data to be able to predict and catch-up past data. SIMCA-online automatically switches between real-time data and historical data as needed.

A data source that only provides current data but not historical data can work for continuous projects in SIMCA-online, but for batch projects historical data is required.

2.5.4 Real-time (current) data must match data read as historical data

☐ Data read in real-time must match what can later be read as historical data for the same time. Otherwise data in SIMCA-online will look differently in real-time compared to when catching up or re-predicting past batches. Specifically, this means:

☐ Compression and interpolation in the data source must be tuned correctly to minimize the difference between real-time and historical data. Learn more in Historical data is different from data in real-time (Q612) (at sartorius.com/umetrics-kb).

2.5.5 Batch processes

The following requirements and recommendations apply to batch configurations:

- ☐ A batch identifier is a (string) value uniquely identifying a batch in SIMCA-online. Batch identifiers must be unique over time. Re-use of batch identifiers is not supported.

[2] Tip: search the built-in help for how the Control tag cache can help with this.

- ☐ Each continuous process node (data table) must have a column (variable) holding the batch identifier for each observation. This batch identifier is used by the server to know to which batch the observation belongs.
- ☐ Batch level data (such as initial conditions or quality measurements for whole batches) are organized per batch identifier.
- ☐ There can be only one observation of batch data per batch at batch level. Batch conditions in a SIMCA project must be read as batch data in SIMCA-online.
- ☐ SIMCA-online requires a structure called a batch node to be able to execute batch configurations. The batch node contains the batch identifier and defines the start- and end time of batches. Tip: Batch Context Generator can be used if you don't have a batch node.
- Avoid complex logical expressions in phase execution conditions and sleep conditions. Instead set up tags for those purposes in the data source and populate them with easily understandable values such as 1 for sleeping and "purification" to mean that the model for purification should run.
- SIMCA-online supports **discrete data** which sometimes is referred to as at-line IPC data. Discrete data consist of many observations per batch that are re-read by SIMCA-online many times during batch evolution. A typical usage of discrete data is for laboratory data, where results come in later than the continuous real-time data.

3 Implementation

This section is about the people involved in a SIMCA-online project, their roles and responsibilities, and gives an example of the process of implementing SIMCA-online.

Note that the actual people involved in a project can vary. Discuss project planning for your specific project with your Sartorius sales representative.

3.1 People

Implementing SIMCA-online involves many people with different areas of expertise, both inside the customer's organization and inside Sartorius. Here are the most important roles:

- **Customer project owner / project manager.** Responsible for running the project at the customer.
- **Customer process expert**. A person with knowledge of the process that can work with the Sartorius data scientist to build and validate models.
- **Customer data source experts.** A person with knowledge about the data source; specifically, how to connect to it, add tags (variables) if needed, tune compression so that real-time and historical data match.
- **Customer IT.** Manages networking infrastructure, provides a suitable Windows server for the SIMCA-online server, performs installation and configuration of servers and clients.

- **Sartorius project manager.** Coordinates the different people and tasks for SIMCA-online deployment, together with the customer project owner.
- **Sartorius data scientist.** A data scientist is a multivariate expert and helps with model building and guides in industry best practices. Also knows how SIMCA-online should be deployed, and in many cases handles support after deployment.
- **Sartorius validation manager.** Assist with documentation and procedures during validation at the customer site.
- **Sartorius SIMCA-online support team.** Handles technical questions regarding installation and SimApi setup, and following deployment, technical support and troubleshooting.
- **Sartorius delivery**. Delivers download links for software, passwords, and activation keys for the SIMCA-online server installation.
- **Sartorius sales representative.** Handles sales and licensing questions.
- **Sartorius product manager.** Handles requests and communicates news for the product.

3.2 Implementation process

This section gives an outline of a typical deployment process.

3.2.1 Preparations

1. Review and verify the requirements and recommendations listed in this document. All items prefixed by ☐ are requirements and must be fulfilled.
2. Discuss your environment with your sales representative, project manager and data scientist to create a plan for the installation. We recommend a staged, iterative deployment, where you start small and add features as you learn about SIMCA-online.

3.2.2 Outline of a SIMCA-online installation

The following major steps are performed when installing SIMCA-online at a customer site. The client handles most steps, working together with Sartorius when needed.

Tip: steps 2-3 can be performed by anyone, and on any PC for testing SIMCA-online.

1. Decide on a server computer to use and a SimApi to use, using the information found in this document. For a test installation, any PC works fine.
2. Download the SIMCA-online software from sartorius.com/umetrics-simca-online. Obtain the password to unzip the file from your Sartorius sales representative, or in the delivery email following an order.
3. Install the SIMCA-online server as described in the Readme and Getting started.pdf included in the product zip file. The PDF also contains instructions how to get started with demo data.

You now have an installation that only works with demo data but is otherwise fully functional. Use it to get acquainted with SIMCA-online.

Next, continue the installation to activate your license for the server, enabling the server to connect to your data source:

4. Follow the instructions in the electronic delivery email from Sartorius to activate the license for your server.
5. Configure a SimApi to connect to the data source. Refer to the user guide of the specific SimApi, and the SimApi Guide.
    a. On the server computer, install the SimApi using its installation program.
    b. Add the SimApi to the server in SIMCA-online Server Options, on the SimApi tab. Configure the SimApi, working with the data source experts from your organization and Sartorius SIMCA-online support or Data Scientist as needed.
    c. Remove or uncheck the DBMaker SimApi from the same tab (it is only required for demo data).
    d. Test the SimApi using the SIMCA-online client's File | Extract functionality. Tip: consider enabling debug logging in the SimApi to help troubleshooting any issues.
6. Upload a SIMCA project and create a configuration of it in SIMCA-online. Or use the File | New functionality directly in SIMCA-online to quickly create a project using data from tags in your data source. Your Sartorius data scientist can assist with this.

Finally, finish the IT setup of the server:

7. Set the SIMCA-online server service to start automatically with Windows.
8. Optionally, configure Active Directory authentication in SIMCA-online to use single-sign on and a group-authenticated user.
9. Optionally, configure SIMCA-online server with a SMTP server so email-based notifications can be used.
10. Optionally, configure the SIMCA-online Web Client. See its installation guide.
11. Set up backups of the SIMCA-online environment.

3.2.3 Next steps...

Continue creating projects in SIMCA for your process, working with Sartorius data scientists. Upload and configure them in SIMCA-online to connect them to your process.

Learn more about SIMCA-online in the documentation listed in the beginning of this document. In the built-in help you can learn about new features you might want to try.

Contact SIMCA-online support if you run into technical issues. sartorius.com/umetrics-support.