# SIMCA®-online Implementation Guide

September 5, 2024

# Contents

# 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 shows, step by step, how to install and set up SIMCA-online.

## 1.1 Additional Documentation

This document 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 | Documentation for each published SimApi with features, installation instructions, and configuration specifics. | 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 | Web-based 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 in Umetrics Suite products. | sartorius.com/umetrics-kb |
| SIMCA help / user guide | How to use desktop SIMCA for creating projects and modelling data. | In SIMCA and on sartorius.com/umetrics-simca |
| Support web page | How to obtain support technical support. | sartorius.com/umetrics-support |

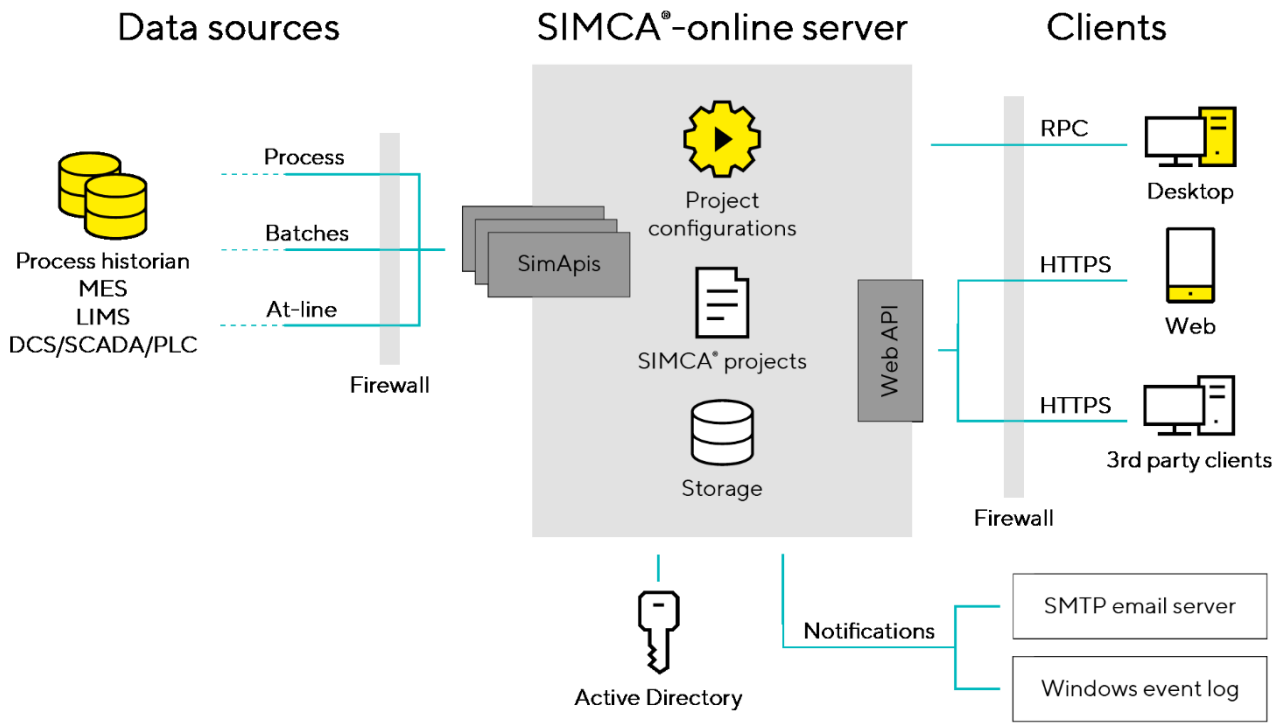
## 1.2 What is SIMCA?

SIMCA is 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 analysis using models and data in SIMCA project files (.usp files).

## 1.3 What is SIMCA-online?

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

- **desktop clients** running on Windows PCs, connected to the SIMCA-online server to monitor the process, change settings, and interact with charts.
- 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 self-contained storage (an external database engine such as SQL server is not needed)
  - keeps **project configurations** containing settings that tie a SIMCA project's variables to tags in the data source.
  - uses an execution algorithm to execute projects.
  - exposes an optional web API.
- optional **Web Clients** connected securely with https, used to monitor production in mobile or desktop browsers. Web clients use the public web API of the SIMCA-online server.




The system

- has 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.
- is designed with security in mind with for example a password policy, strong hashing, and rate-limiting of failed authentication attempts.
- provides security by authenticating users, using group permissions to grant permissions specifying what users can do, and uses folder access rights to prevent unauthorized access to individual projects.

- supports connection to **Active Directory** for user authentication including single sign-on and group memberships.
- has audit trails to record user interaction with the server and project configurations.
- supports secure and authenticated email **notifications** over TLS connections and logging to Windows Application event log.
- works with firewalls between data sources, server, and clients.

## 1.4 Data Integrity and SIMCA-online

Data read and the results from project execution are stored in the SIMCA-online server. The server is robust and uses database transactions and other techniques to ensure data integrity. Revisions of project configurations are kept and there is a built-in comparison and roll back of revisions.

Learn more in How to Ensure Data Integrity and Compliance of Your Data Analytics Systems (sartorius.com)

## 1.5 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.

## 1.6 Batch Processes and Continuous Processes

SIMCA-online supports two types of processes; continuous and batch.

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.

## 1.7 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.
    - a phase can also be a steady-state phase, created in SIMCA 18 or later, to monitor a part of the process where the process is similar to a continuous process.
    - 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 project 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.

## 1.8 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 an embedded multivariate engine that integrates seamlessly into your preferred software running on different hardware platforms. 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. SIMCA-Q consists of a C or Windows COM interface. Supported platforms include Windows and Linux.

## 1.9 Getting Data Out from SIMCA-online

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

- 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)
- Many SimApis support write back to the data source. Data from project execution, such as predictions, alarm status, Control Advisor set points and meta data can be written back to a data source. Learn more in the SIMCA-online help.

# 2 Technical Requirements and Best Practices

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

☐ Build multivariate monitoring models with 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.

☐ Prescriptive models for Control Advisor's Advised-future mode should be built on data with representatively large variation, specifically at the control maturities 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. An unnecessarily short execution interval increases the CPU load of the server and RAM and storage requirements.

☐ 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, the SIMCA-online execution interval must equal 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 desktop SIMCA before uploading to SIMCA-online.
- You can have different sampling intervals for different evolution phase models.

[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.

How to build models is out of scope of this document. Sartorius 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:

- ☐ The computer system clocks should be synchronized between the server computer, data source and clients. Different time zones are supported.
- ☐ 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 Windows 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.

## 2.5 The Latest SIMCA-online Software

Always deploy the latest SIMCA-online software which contains the latest functionality and fixes for reported issues.

Information on each release, including system requirements is found in its knowledge base article.

## 2.6 SIMCA-online License

The SIMCA-online server license together with your agreement with Sartorius control 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 and be limited to allow only certain features.

## 2.7 Security Overview and Guidelines

This is an overview of SIMCA-online from a security perspective with IT best practices and specific SIMCA-online settings.

☐ The SIMCA-online server is installed on a Windows computer. Physical access to the server should be restricted.

In general use strong passwords and manage them properly.

☐ Set the local SIMCA-online Administrator account to a strong password.

☐ The network where server and clients are located should be restricted to prevent unauthorized devices. Firewalls can be used to segment networks. VPNs (Virtual Private Network) can be used to allow remote access for clients.

☐ Protect the network between data sources and server with firewall rules.

☐ SimApis typically use authentication to connect to data sources. Use strong passwords.

☐ The server is a Windows service and runs as a service account in Windows. Use an account with only the necessary privileges.

☐ Windows login rights to the server computer should be restricted to relevant administrators. If non-administrators have access to the server computer, file access *Database directory* on the server disk should be restricted by access control lists in the file system.

☐ Restrict remote access for management of the server Windows computer to prevent unauthorized access.

Access to SIMCA-online requires authentication (log in). Use Active Directory authentication to simplify management and rely on Active Directory security.

☐ Configure the SIMCA-online server to connect to Active Directory using TLS (encryption).

☐ Add Active Directory groups in SIMCA-online and grant permissions to those. Use restrictive permissions for each group that allow users to do just what they need to be able to.

Desktop clients use encrypted and authenticated Windows RPC to connect to the server if the computers are members of the same Windows domain. Otherwise, unencrypted connections are used as fall back. The state of the connection is visible in the client.

Enable the server Web API only when it is needed. Configure and use HTTPs (TLS) by installing a TLS certificate in Windows for the Web API. Follow Qualys SSL Labs - Projects / SSL/TLS Deployment Best Practices and similar guidance. By default, the server Web API is restricted to 'Allowed CORS origins' only pages hosted by the server itself.

SIMCA-online APIs used by clients require authentication and authorization. Each API grants access only to users that are members of a group that has the required permissions.

Folder access rights are used on the server computer to specify which groups have access to folders. A user that is not a member of a group with access does not see nor can change the folder or its contents. Folder access rights are also respected by all API endpoints.

Projects that use Python preprocessing can run Python code as part of execution. When an administrator uploads a project with Python scripts, he or she is presented with a security warning and can also look at the contents of the script before allowing the project.

The server Audit trails records user actions on the server, such as when projects are uploaded and configured, groups are added, modified or permissions are granted. Each project configuration also has an audit trail that records changes made to that project configuration.

The server rate-limits failed login attempts, effectively preventing credential stuffing attempts.

The server log files show error and warning messages relating to security, such as when a user is not allowed to log in or when rate-limiting occurs.

Notifications can be used by administrators to get alerted for security events such as failed logins or rate-limiting.

☐ Do regular backups of server data and settings. Keep logs of configuration management so that you can handle a recovery. Practice a disaster recovery on a different computer.

## 2.8 Data Source Requirements

A commercial data source for process systems such as Aveva PI System (formerly OSIsoft) 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.

### 2.8.1 A SimApi for the Data Source

☐ A **SimApi** for the 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 Aveva PI System, OPC UA, ODBC for relational databases like SQL server.

Sartorius provides a software development kit for developers wishing to implement a SimApi to connect to a data source. However, if the data source is low level such as an instrument in the lab and do not provide historical data, we recommend that you instead of developing a SimApi, connect the data source to a process historian, such as Aveva PI System, and let it take care of obtaining, historizing and keeping data.

Learn more about SimApis in the SimApi Guide at sartorius.com/umetrics-simapi.

### 2.8.2 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 will not match the same data requested as historical data later (see 2.8.4).

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

To re-read past data in SIMCA-online, an operator must manually (re-)predict past data.

### 2.8.3 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.

☐ For batch project execution, historical data is required.

### 2.8.4 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 different 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).

### 2.8.5 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. Avoid using numeric-only identifiers because they are stored as numbers and may lose precision when they become too large.
- ☐ 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: SIMCA-online Batch Context Generator with Generator-nodes can be used to get started if you do not have a batch node, but we always recommend that the batch node is in the data source itself.

- ☐ Each continuous process node (data table) must have a column (variable) holding the batch identifier for each observation. This batch identifier is used to know to which batch the observation belongs.
- ☐ Optional 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 are normally read as batch data in SIMCA-online. They can also be read as Discrete data where the server will use the average value for the variables.
- Avoid complex logical expressions in phase execution conditions and sleep conditions to simplify troubleshooting. 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 indicate that purification model should run.
- SIMCA-online supports phases with 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. Data can also be combined at the batch level.
- Baseline filters can be used to ignore variables with unchanged data at beginning of phase.
- Use the bundled Python scripts that greatly simplifies creating and testing Control Advisor models in desktop SIMCA before deploying to SIMCA-online.

# 3 Global Installations

As you saw in chapter two, low latency and a fast network is crucial for a distributed client-server program like SIMCA-online: both between clients and SIMCA-online server, and between server and its data sources.

This text has so far focused on a **single-site local scenario** where you have the data source, the SIMCA-online server, and clients in the same location. In this case network latency is typically not an issue and does not affect performance.

You can also **scale up the single-site scenario to many sites**, each with a local data source and SIMCA-online installation.

If you want to run clients that are located far away from the SIMCA-online servers, you can improve performance by services like **Remote Desktop or desktop/app virtualization**. Here you run the clients on the network close to the server, but users on the other side of the globe essentially remote control the client.

To create a centralized **global SIMCA-online installation** you need to carefully consider network latency and how it affects performance:

☐ Locate the central SIMCA-online server close to its data sources. Otherwise, performance will suffer when data is obtained, causing a slow server.

☐ Run clients close to the SIMCA-online server. Use Remote Desktop or another desktop/app virtualization software as described above to improve client performance.

☐ Start small and compare performance with a local single-site installation to make sure your centralized environment does not cause performance degradation. Measure and compare performance both of clients, and things like predicting a full batch, but also how fast data source access is (measured by File | Extract and by debug server logs that provide timing measurements for data access calls).

If a centralized server is getting overloaded, consider splitting it into two servers.

# 4 Implementation

This section is about the people involved in a SIMCA-online project, their roles, and responsibilities, and shows an example of the process of implementing SIMCA-online. The people involved in a project can vary. Discuss project planning for your specific project with your Sartorius contact.

## 4.1 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:

| Role | |
|---|---|
| **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 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 methods expert and helps with model building and guides in industry best practices. In many cases handles support. |
| **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 IDs for the licenses. |
| **Sartorius sales representative** | Handles sales and licensing questions. |
| **Sartorius product manager** | Handles requests for new features and communicates news for the product. |

## 4.2 Implementation process

This section gives an outline of a typical deployment process.

### 4.2.1 Preparations

1. Review and verify the requirements and recommendations listed in this document. All items prefixed with a checkbox, ☐, 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.

### 4.2.2 Outline of a SIMCA-online Installation

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

A good practice is to document all settings.

Steps 1-3 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 my.sartorius.com when you have received an Activation ID from Sartorius following an order. Contact your sales representative to obtain an evaluation version.
3. Install the SIMCA-online server as described in the Readme and Installation.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 fully functional. Use it to get acquainted with SIMCA-online.

Next, continue the installation to activate a license for the server and install a SimApi to enable the server to connect to your data source (without a license a SimApi cannot be used):

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 pdf. On the server computer:
    a. 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. Learn more in help. 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 configure one Active Directory group in SIMCA-online for each class of users (operators, project admins, administrators for example).
9. Optionally, configure SIMCA-online server with a SMTP server so email-based notifications can be used.
10. Optionally, enable the SIMCA-online Web Server with its Web API and configure the SIMCA-online Web Client. See its separate installation guide.
11. Set up backups of the SIMCA-online environment. See the Technical Guide.
12. As for other services in your organization, you may want to monitor it using your regular IT software for things like CPU, RAM, and storage usage.
13. The SIMCA-online server log is a good resource to monitor the health of a server.

### 4.2.3 Next Steps...

Continue creating or updating 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.

Sartorius Stedim Data Analytics AB
Östra Strandgatan 24
903 33 Umeå
Sweden

Phone: +46 90-18 48 00
www.sartorius.com

The information and figures contained in these instructions correspond to the version date specified below.
Sartorius reserves the right to make changes to the technology, features, specifications and design of the equipment without notice.
Masculine or feminine forms are used to facilitate legibility in these instructions and always simultaneously denote all genders.