SAS® Analytics for Containers

Take advantage of container-enabled infrastructures to easily deploy SAS® Analytics to the cloud

What does SAS® Analytics for Containers do?
It provides the ability to easily deploy SAS Analytics within container-enabled infrastructures, including Docker and Kubernetes, which are often run in the cloud. Users can interact with their customized container of SAS software using SAS Studio, a browser-based interface, or Jupyter Notebook - an open source notebook-style interface. SAS Analytics for Containers eases software management and deployment for IT.

Why is SAS® Analytics for Containers important?
In many organizations, containers are the preferred method for deploying software to cloud infrastructures. For those moving their data processing to the cloud, this solution simplifies the deployment, availability and configuration of SAS to data scientists and analytical teams. They can quickly build ready-to-run SAS applications that take advantage of flexible and scalable container-enabled environments.

For whom is SAS® Analytics for Containers designed?
It’s designed for individual programmers, statisticians, citizen data scientists and data scientists who need to run SAS in a container-enabled environment.

Benefits

• **Simplify and speed up deployment.**
  The SAS container radically simplifies the deployment of your app, no matter where you are running it. It provides the build once, run anywhere flexibility of the container environment, making it easier and faster to use SAS Analytics.

• **Run SAS easily in the cloud.**
  Putting SAS into a container like Docker enables users to easily run SAS Analytics on public and private clouds. It basically provides a sandbox environment for data scientists and others to work with SAS in the cloud.

• **Reproduce an experiment at any stage in the research process.**
  The ability to capture an experiment’s state, including data, code, results, package versions, parameters, etc., at a single point in time is one of the real values for data scientists who are working to solve large, complex problems.

• **Make applications and workloads more portable and distributed.**
  Because SAS applications can be isolated into containers with instructions for exactly what is needed for deployment, SAS applications can be easily ported from machine to machine for more efficient use of resources.

• **Take advantage of Hadoop processing power.**
  When SAS Analytics for Containers is located within or near a Hadoop cluster, it can be combined with SAS solutions for Hadoop, including the scoring accelerator, code accelerator and data quality accelerator. You can put your SAS solutions to work on the data stored in Hadoop and run SAS Analytics inside the Hadoop cluster, taking advantage of Hadoop’s distributed processing power.

A container is a mechanism that encapsulates only the necessary system libraries and functions needed to run a given piece of software. This guarantees it will always run the same, regardless of the environment. Applications are isolated from each other and underlying infrastructures, providing an added layer of security protection.

One reason for the surging popularity of software packaged in containers is that it makes managing and deploying software much easier. Cloud adoption, big data and the need to quickly build and experiment with more predictive models are also driving increased use.

Using SAS Analytics for Containers, data scientists, statisticians and business analysts can build SAS code, models and applications that run in container-enabled infrastructures.

FACT SHEET

A container is a mechanism that encapsulates only the necessary system libraries and functions needed to run a given piece of software. This guarantees it will always run the same, regardless of the environment. Applications are isolated from each other and underlying infrastructures, providing an added layer of security protection.

Applications are isolated from each other and underlying infrastructures, providing an added layer of security protection.

One reason for the surging popularity of software packaged in containers is that it makes managing and deploying software much easier. Cloud adoption, big data and the need to quickly build and experiment with more predictive models are also driving increased use.

Using SAS Analytics for Containers, data scientists, statisticians and business analysts can build SAS code, models and applications that run in container-enabled infrastructures.
Capabilities

SAS Analytics for Containers delivers a powerful set of data access, analysis and graphical tools in a containerized package.

With this proven SAS toolset, users can access data from nearly any source, perform sophisticated analyses and transform raw data into meaningful, usable information. Then, present the information and insights obtained using visuals that help decision makers quickly understand critical issues. All in modular containers for simplified deployment.

Best-in-class SAS Analytics

With SAS Analytics for Containers, you get one of the broadest sets of statistical techniques in the software industry. Regressions, decision trees, Bayesian analysis, spatial point pattern analysis, missing data analysis and many more techniques are included.

SAS has 40 years of experience developing statistical analysis software and a proven reputation for delivering superior, tested and reliable results.

SAS® Studio web-based programming interface

The intuitive interface, SAS Studio, lets you access SAS programs, data files and libraries from a web browser wherever you are. Write and submit SAS code on any device with a web browser. Programming tools offer interactive feedback, prompts and coding auto-completes that enable you to work more efficiently. Code snippet libraries are available for frequently performed jobs, and code-generating tasks promote consistent programming. You can then view and modify the code that was generated.

Jupyter Notebook-based access

Jupyter Notebook is a notebook-style interface to multiple languages. It is open source and available as a free download. SAS has contributed a Jupyter language kernel so that SAS users who prefer the Jupyter interface can use it to directly interact with SAS.

Powerful fourth-generation programming languages

Base SAS provides highly flexible and extensible 4GLs with easy-to-learn syntax. The SAS DATA step and SAS PROCs make up the traditional SAS programming language. The SAS DS2 programming language combines the power of the SAS DATA step with the flexibility of SQL and the ability to run in a distributed environment.

Prebuilt SAS procedures (SAS PROCs) handle many common tasks, including data manipulation and management, information storage and retrieval, sophisticated statistical and predictive analysis, and report writing.
Also part of the SAS programming language, the SAS macro facility encapsulates code into units with names. You can then work with the names instead of longer code strings, reducing the amount of coding required to perform common tasks or extend your SAS applications. Programmers can modularize their work for easy reuse and maintenance.

Hadoop data access
SAS/ACCESS interfaces pull data directly from source systems into the container for manipulation and modeling. SAS can push processing to the data source and interact with Hadoop, and RDBM systems, efficiently and seamlessly.

Powerful integration with Hadoop
SAS supports integration with various Hadoop technologies, including a variety of Hadoop file types and the Hive metastore HCatalog.

The SAS HADOOP procedure lets you submit HDFS commands, MapReduce programs and Pig language code against Hadoop data from within SAS.

Easily store large SAS data sets on Hadoop clusters, and add the benefits of Base SAS security features such as encryption and password protection to Hadoop implementations.

With optional add-on SAS Hadoop solutions (accelerators), users can both pull data to the container for SAS processing and push SAS processing to the Hadoop cluster, taking advantage of the distributed processing model.

Key Features

Container-ready SAS* Analytics
- Provides container-ready SAS software that is easy for data scientists to use and easy for IT to deploy.
- Provides integration with Hadoop in an IT-friendly way.

Best-in-class SAS* statistical analysis techniques
- Analysis of variance.
- Bayesian analysis.
- Categorical data analysis.
- Cluster analysis.
- Descriptive statistics.
- Discriminant analysis.
- Distribution analysis.
- Exact methods.
- Group sequential design and analysis.
- Market research.
- Mixed models.
- Multiple imputation.
- Multivariate analysis.
- Nonparametric analysis.
- Post-fitting inference.
- Power and sample size.
- Psychometric analysis.
- Regression.
- Spatial analysis.
- Structural equations.
- Survey sampling and analysis.
- Survival analysis.
- Transformations.
- Multithreaded procedures.
- Statistical graphics.
- Ability to add drill-down capabilities so users can visually explore analyses.

SAS* Studio web-based programming interface
- Access all of your SAS programs, data files and libraries from your desktop, Mac® and iPad® through a web browser.
- Auto-complete feature displays a list of SAS procedures when you begin typing a procedure name. When a procedure is selected, it displays the parameter list and pop-up syntax.
- Automatically generates SQL queries and lets you access the SQL code generated behind the scenes.
- Create and add your own code snippets to the snippet library.
- Point-and-click interface guides you through analytical or data manipulation processes.

Powerful 4GLs with support for SQL
- Intuitive 4GLs with easy-to-learn syntax.
- Includes DS2, a modern programming language, which allows advanced data manipulation to be performed on data where it resides in-database.
- SAS macro facility reduces coding for common tasks and lets you modularize work.
- Runs interactively or in batch mode.
Data visualization, presentation and delivery

Users can visually present their ideas and findings using a huge variety of colorful business maps, charts, plots and 3-D relationship graphs. Extensive map data sets are included.

The SAS Output Delivery System (ODS) can be used for data capture and report formatting with numerous choices for displaying analytical results. Programmers can create and deliver accurate and visually appealing reports with less programming.

To learn more about SAS Analytics for Containers, view screenshots and see other related materials, please visit sas.com/analytics-containers.

Key Features (continued)

Hadoop data access
- Includes the option to access data without detailed knowledge of Hadoop or HiveQL. The appropriate HiveQL queries are generated from the interface and passed to the data source for execution.
- Includes the option to use native HiveQL.
- Has the ability to create federated queries that span multiple data sources.
- Mapping SAS statements or functions to database-specific statements or functions enables all appropriate HiveQL statements to be processed directly inside Hadoop.
- Data read and writes are highly optimized with buffering, threading, etc.
- Supports Hive’s external tables interface, including external HDFS files.
- Has the ability to control join-processing capabilities.
- Loads data at very fast HDFS streaming speeds.
- Supports query interrupt; supports binary and decimal data types.
- Provides support for native storage options, including external tables.
- Translates native Hadoop data to the appropriate SAS data type for SAS processing.

Integration with Hadoop
- Support for Pig, MapReduce and HDFS commands from the SAS execution environment.
- Supports external file references from within any SAS component. Delimiters are externalized, making it easy to work with Hadoop files.
- Several Base SAS procedures support SQL push-down optimization to Hive on Hadoop.
- SAS format catalogs can be published and compiled inside Hive on Hadoop so that formats can be applied to the actual data values during query execution.

Data visualization, presentation and delivery
- Built-in map data sets (for countries, and US states and counties).
- Broad range of charts and plots:
  - Scatter, line, area, bubble, multiple axis and overlay.
  - Bar, pie, donut, star and block.
  - Customized colors, line styles and symbols.
  - 2-D and 3-D plots with tilting and rotation.
- Anti-aliased lines for smooth plot lines.
- Generate static or dynamic interactive (Java or ActiveX) charts and graphs with drill-down capabilities.
- Embed interactive graphics in web pages or Microsoft Office documents.
- Graphs are integrated with tables with all output displayed in the same HTML file.