PUBLIC
SAP HANA Platform 2.0 SPS 04
Document Version: 1.1 – 2019-10-31

SAP HANA Machine Learning Overview

THE BEST RUN SAP

Content

1.1.7

1 SAP HANA Machine Learning Overview

SAP HANA supports a comprehensive environment for machine learning. On the server side, it offers embedded machine learning libraries as well as capabilities for integrating common machine learning tools. On the client side, its machine learning APIs can be used to directly develop embedded machine learning models in SAP HANA.

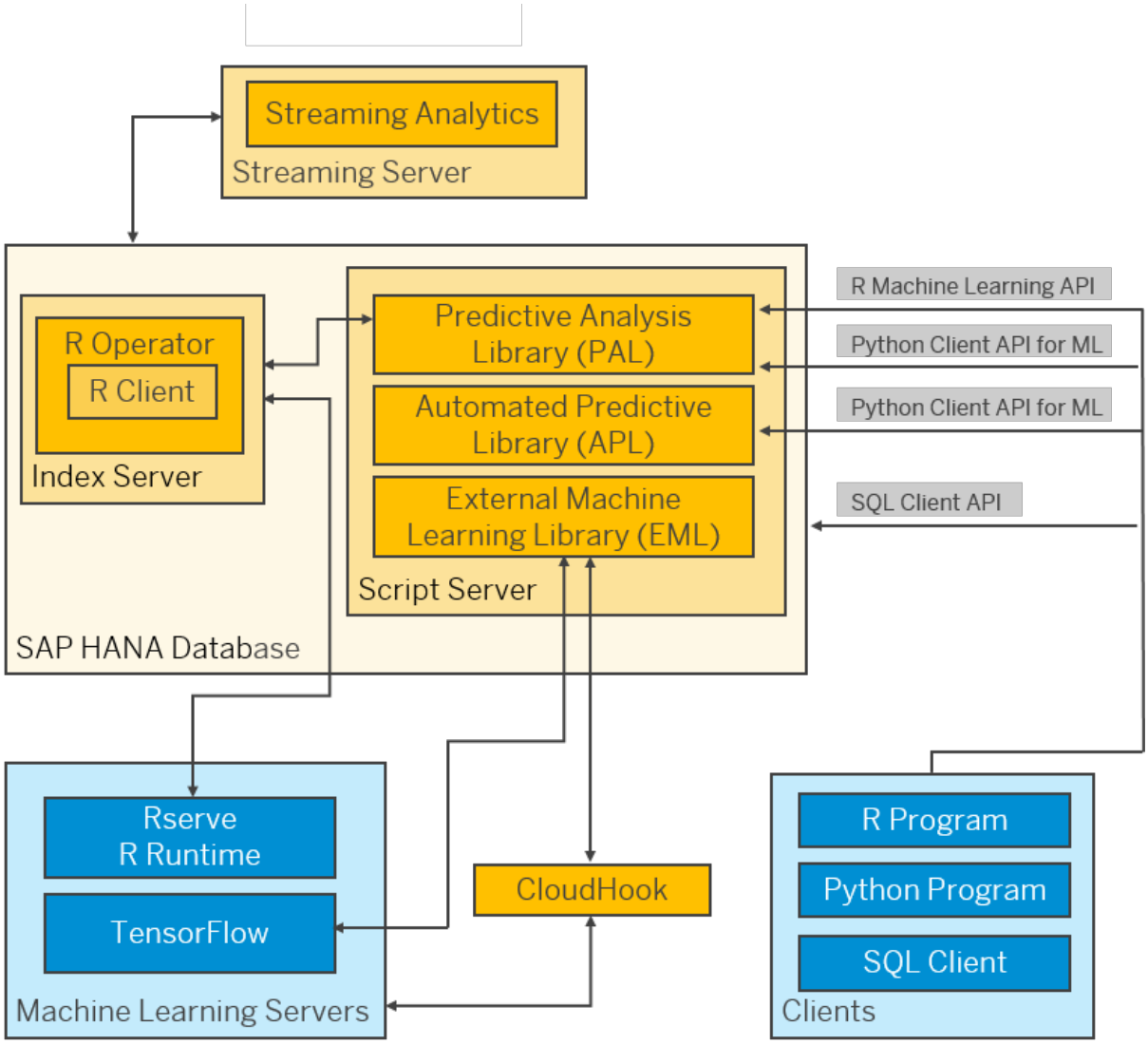
Feature Overview

The table below gives a high-level overview of the main features available:

Embedded Machine Learning Libraries	
Predictive Analysis Library (PAL)	Native in-database machine learning for data scientists and developers
Automated Predictive Library (APL)	Automated in-database machine learning for developers and business analysts
APIs	
SAP HANA Python Client API for machine learning algorithms	For data scientists using Python
R Machine Learning library for SAP HANA database clients	For data scientists using R
Integration Capabilities	
R Integration	Integration with the R software environment
External Machine Learning Library (EML)	Integration with Google TensorFlow

Architecture Overview

The main features are shown graphically below. Hover over an area to display a brief description:

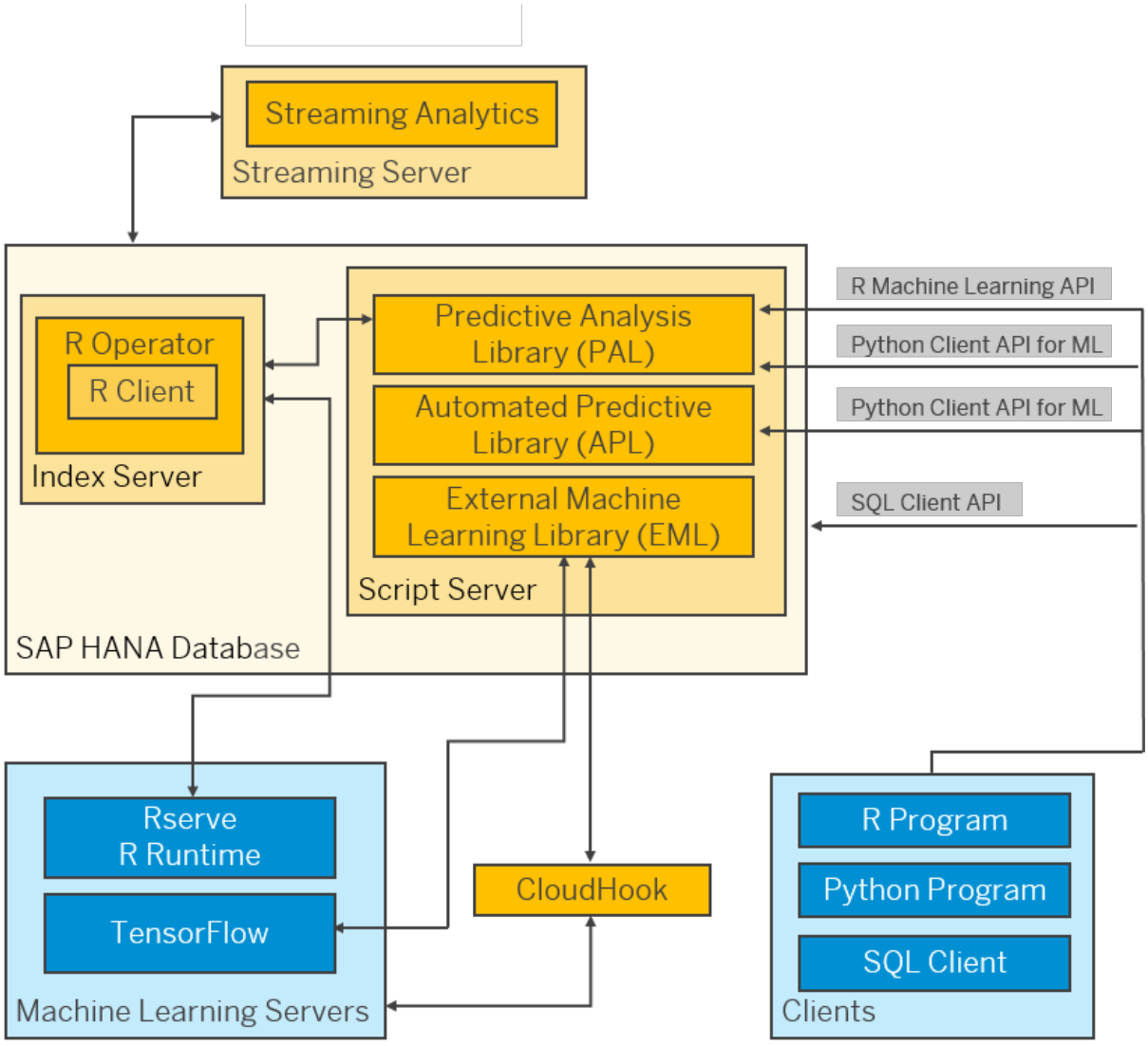


- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]

- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]
- SAP HANA Machine Learning Overview [page 3]

Architecture Overview

The main features are shown graphically below:



8.7

2 Machine Learning Features in SAP HANA

SAP HANA provides two embedded machine learning libraries, the Predictive Analysis Library (PAL) and the Automated Predictive Library (APL). On the client side, the Python Client API and R Machine Learning library address the needs of data scientists using Python and R.

In addition, an integration with Google TensorFlow is supported through the External Machine Learning Library (EML), and an integration with the R software environment through the R integration.

These features are described briefly below:

Machine Learning Feature	Description
Predictive Analysis Library (PAL)	PAL provides functions for predictive analysis and machine learning. It supports algorithms for classification, regression, cluster analysis, time series analysis, association analysis, social network analysis, recommender systems, data preprocessing, and statistics. PAL functions can be called from within SAP HANA SQLScript procedures.
Documentation	SAP HANA Predictive Analysis Library (PAL)
Installation	PAL is part of the Application Function Library (AFL) delivered with SAP HANA. It needs to be installed separately.
PAL with SAP HANA Streaming Analytics	PAL provides incremental machine learning capabilities for SAP HANA streaming analytics. Incremental machine learning algorithms train and update a model on the fly, so that predictions are based on a dynamic model.
Documentation	Machine Learning with Streaming Analytics (*SAP HANA Streaming Analytics: Developer Guide*)
Installation	SAP HANA and SAP HANA streaming analytics can be installed on separate dedicated servers or on the same server using the SAP HANA platform lifecycle management tool.
PAL for Python clients	For clients, a Predictive Analysis Library (PAL) package is available as part of the `SAP HANA Python Client API for machine learning algorithms`. The PAL package provides a set of Python algorithms and functions that allow clients to access the machine learning capabilities in SAP HANA. The PAL Python functions provide additional dataframe management capabilities as well as some visualization and utility features. The SAP HANA Python Client API makes use of the SAP HANA dataframe, which provides a set of methods that allow data to be analyzed directly in SAP HANA without having to transfer it to the client.

Machine Learning Feature	Description
Documentation	SAP HANA Python Client API for Machine Learning Algorithms Configure the SAP HANA Python Client API for Machine Learning Algorithms Python Client API for machine learning algorithms
Installation	The SAP HANA Python Client API for machine learning algorithms is installed as part of the SAP HANA client install. The *SAP HANA Python Client API for Machine Learning Algorithms* option needs to be selected.
PAL for R clients	See the R section below
Automated Predictive Library (APL)	APL provides functions for creating predictive models for classification, regression, clustering, time series forecast, recommendation, and social network analysis. Based on the type of data mining function to be applied to the data, APL automatically identifies the appropriate algorithm and creates the optimal model. In this way, APL allows predictive algorithms to be run on data stored in SAP HANA in an automated manner. It also means that data scientist skills are not needed to use APL (in contrast to PAL). APL is designed primarily for business analysts. APL functions can be called using SQL.
Documentation	SAP HANA Automated Predictive Library Reference Guide
Installation	SAP HANA APL is an Application Function Library (AFL). It needs to be installed separately.
APL for Python clients	For clients, an Automated Predictive Library (APL) package is available as part of the *SAP HANA Python Client API for machine learning algorithms*, providing functions which expose the data mining capabilities of the automated analytics engine in SAP HANA. The APL package contains algorithms for classification (auto classifier) and regression (auto regressor). The SAP HANA Python Client API makes use of the SAP HANA dataframe, which provides a set of methods that allow data to be analyzed directly in SAP HANA without having to transfer it to the client.
Documentation	SAP HANA Python Client API for Machine Learning Algorithms Configure the SAP HANA Python Client API for Machine Learning Algorithms Python Client API for machine learning algorithms

Machine Learning Feature	Description
Installation	The SAP HANA Python Client API for machine learning algorithms is installed as part of the SAP HANA client install. The *SAP HANA Python Client API for Machine Learning Algorithms* option needs to be selected.
R integration with SAP HANA	By integrating R, stored procedures can be created and run in SAP HANA with embedded R script. The R script is automatically transferred to the remote R server, where it is executed. Intermediate tables are transferred directly into and from the vector-oriented data structures of R. Since the internal column-oriented data structure used within the SAP HANA database for intermediate results is very similar to the vector-oriented R data frame, the conversion is very efficient. This offers a performance advantage compared to standard SQL interfaces. This scenario is intended for SAP HANA-based applications that want to use the R environment for specific statistical functions.
Documentation	SAP HANA R Integration Guide
Installation	R and Rserve need to be installed on a separate host to SAP HANA. SAP HANA needs to be configured for use with R.
R integration with SAP HANA Streaming Analytics	The R data service allows R script to be run as part of CCL (Continuous Computation Language) queries on the streaming analytics server. The query is passed to the remote R server. An R data service can be added in the streaming runtime tool and SAP HANA cockpit.
Documentation	Adding a Connection to an R Data Service (*SAP HANA Streaming Analytics: Developer Guide*)
PAL for R clients	The `SAP HANA R Machine Learning library for SAP HANA database clients` provides a set of client-side machine learning algorithms for R. The API can be used from any R IDE, such as RStudio. Since the actual machine learning occurs in the SAP HANA database, the training data does not need to be transferred to the R server or client side.
Documentation	R Machine Learning API Configure the R Machine Learning API R Client API for machine learning algorithms
Installation	The R Machine Learning API is installed as part of the SAP HANA client install. The *R Machine Learning API* option needs to be selected.

Machine Learning Feature	Description
External Machine Learning Library (EML)	EML supports the integration of Google TensorFlow with SAP HANA. EML is an SAP HANA Application Function Library (AFL). It uses Google's gRPC remote procedure call package to remotely invoke predefined TensorFlow models, hosted on the TensorFlow Serving server, through gRPC calls encapsulated inside AFL procedures. In addition, EML supports the use of external management functionality, enabled through an intermediate CloudHook server positioned between EML and the external services.
Documentation	SAP HANA External Machine Learning Library
Installation	SAP HANA EML needs to be installed separately.

3 Software Downloads

You can download the SAP HANA Predictive Analysis Library (PAL), SAP HANA Automated Predictive Library (APL), SAP HANA External Machine Learning Library (EML), and SAP HANA Streaming Analytics from SAP Software Downloads on the SAP Support Portal.

You can access SAP Software Downloads at: https://launchpad.support.sap.com/#/softwarecenter

The table below gives the component names in SAP Software Downloads. Make sure that you read the information about compatibility before choosing a specific component version to download:

For SAP HANA 2.0 SPS 04

	SAP Software Download	Compatibility Information

Predictive Analysis Library (PAL)	SAP HANA AFL 2.0
	For example:
	`SAP HANA AFL Rev 40.0 only for HANA 2.0 Rev 40`
	`SAP HANA AFL Rev 41.0 only for HANA 2.0 Rev 41`
		The AFL should have the same version as the SAP HANA database.
		The revision number of the AFL must match the revision number of SAP HANA. For more information, see SAP Note 1898497 - Versioning and delivery strategy of application function libraries (AFLs).

Automated Predictive Library (APL)	APL 4 FOR HANA 2 SP03+[1] For example: `Predi. Analy. APL 1909 for SAP HANA 2.0 SPS03 and beyond`	For example, SAP Note 2832109 - New APL Release: APL 1909

Streaming Analytics	SAP HANA streaming analytics 2.0 -> SAP HANA STREAMING 2.0
	For example:
	`Revision040.00 (SP04)`
	`Revision041.00 (SP04)`
		SAP HANA streaming analytics should have the same version as SAP HANA.
		For a matrix of compatible SAP HANA and streaming analytics versions, see SAP Note 2659633 - SAP HANA Streaming Analytics 2.0 SP 04 Release Note.

External Machine Learning Library (EML)	SAP HANA EML AFL 1.0
	For example:
	`EML AFL Rev 40.0 only for HANA 2.0 Rev 40`
	`EML AFL Rev 41.0 only for HANA 2.0 Rev 41`
		The revision number of the EML AFL must match the revision number of SAP HANA. For more information, see SAP Note 1898497 - Versioning and delivery strategy of application function libraries (AFLs).

	SAP Software Download	Compatibility Information
R Integration	n/a	The interoperability of the SAP HANA R integration with the R software environment requires compatible versions of R and Rserve. For more information, see SAP Note 2185029 - SAP HANA and R compatibility and support.

[1] Starting from SAP HANA APL 1903 onwards, choose *APL 4 FOR HANA 2 SP03+* to download the SAP HANA APL for SAP HANA 2 SPS03 and beyond. Choose *APL 4 FOR HANA 2 SP3+* for lower SAP HANA APL versions.

The versions of the client APIs that are generally available with the SAP HANA client of SAP HANA 2.0 SPS 04 are as follows:

Client API	Version with SAP HANA client of SAP HANA 2.0 SPS04	Compatibility Information
Python Client API - PAL	1.0.5	SAP HANA 2 SPS03 or higher
Python Client API – APL		SAP HANA APL 1811 or higher
R Client API	1.0.4	SAP HANA 2 SPS03 or higher

4 Additional Information

Additional sources of information about machine learning in SAP HANA include code samples, blogs, and tutorials.

Code Samples

Code samples are available for the SAP HANA predictive analysis and machine learning scenarios. These consist of SQL examples for the Predictive Analysis Library (PAL) as well as Python client API examples.

The code samples simply need to be downloaded and then used within the appropriate editor, for example, the SQL console of the SAP HANA Database Explorer.

The code samples are available here: https://github.com/SAP-samples/hana-ml-samples

Blogs

The following blogs focus on the R and Python client APIs:

- What's New
  - New R and enhanced Python API
- R
  - Machine Learning with SAP HANA – from R Part 1
  - Machine Learning with SAP HANA – from R Part 2
  - Machine Learning with SAP HANA & R – Evaluate the Business Value
- Python
  - Python Client API for machine learning in SAP HANA 2.0, Express Edition SPS 03, Revision 33
  - Getting Started With "SAP HANA Python Client API For Machine Learning Algorithms" SAP HANA 2.0, Express Edition Support Package 03, Revision 035
  - Diving into the HANA DataFrame: Python Integration Part 1
  - Diving into the HANA DataFrame: Python Integration Part 2

Tutorials - SAP HANA, express edition

The following SAP HANA, express edition tutorials cover machine learning topics:

- Predictive Analysis Library (PAL)
  - Learn to Use the SAP HANA Predictive Analytics Library (PAL)
  - Machine Learning in A Box: Get Started with SAP Predictive Analytics on SAP HANA, express edition

- PAL/APL for Python clients
  - Install the SAP HANA Python Client API for Machine Learning Algorithms
- R integration with SAP HANA
  - Set up R integration with SAP HANA, express edition
  - Configure the SAP HANA R Integration with SAP HANA, express edition
- External Machine Learning Library (EML)
  - Discover SAP HANA External Machine Learning Library (aka Google TensorFlow integration)

5 Application Function Library SDK for SAP HANA (SAP HANA AFL SDK)

You can develop your own application function libraries using the SAP HANA AFL software development kit (SDK). AFLs are external libraries that extend the SAP HANA database kernel. They are implemented using the C++ programming language.

For more information, see the following:

SAP Note	Title
2421209	Application Function Library SDK for SAP HANA 2.0 Release Information
2190908	Customer use of the Application Function Library SDK for SAP HANA
2743704	Re-build of AFLSDK for SAP HANA-libraries required before deployment to SAP HANA 2 SPS04
2121474	Building custom Application Function Libraries for production SAP HANA deployment

Important Disclaimers and Legal Information

Hyperlinks

Some links are classified by an icon and/or a mouseover text. These links provide additional information.
About the icons:

- Links with the icon  : You are entering a Web site that is not hosted by SAP. By using such links, you agree (unless expressly stated otherwise in your agreements with SAP) to this:
  - The content of the linked-to site is not SAP documentation. You may not infer any product claims against SAP based on this information.
  - SAP does not agree or disagree with the content on the linked-to site, nor does SAP warrant the availability and correctness. SAP shall not be liable for any damages caused by the use of such content unless damages have been caused by SAP's gross negligence or willful misconduct.
- Links with the icon  : You are leaving the documentation for that particular SAP product or service and are entering a SAP-hosted Web site. By using such links, you agree that (unless expressly stated otherwise in your agreements with SAP) you may not infer any product claims against SAP based on this information.

Beta and Other Experimental Features

Experimental features are not part of the officially delivered scope that SAP guarantees for future releases. This means that experimental features may be changed by SAP at any time for any reason without notice. Experimental features are not for productive use. You may not demonstrate, test, examine, evaluate or otherwise use the experimental features in a live operating environment or with data that has not been sufficiently backed up.
The purpose of experimental features is to get feedback early on, allowing customers and partners to influence the future product accordingly. By providing your feedback (e.g. in the SAP Community), you accept that intellectual property rights of the contributions or derivative works shall remain the exclusive property of SAP.

Example Code

Any software coding and/or code snippets are examples. They are not for productive use. The example code is only intended to better explain and visualize the syntax and phrasing rules. SAP does not warrant the correctness and completeness of the example code. SAP shall not be liable for errors or damages caused by the use of example code unless damages have been caused by SAP's gross negligence or willful misconduct.

Gender-Related Language

We try not to use gender-specific word forms and formulations. As appropriate for context and readability, SAP may use masculine word forms to refer to all genders.

www.sap.com/contactsap

THE BEST RUN SAP