



SAP
Innovation
Awards 2019



SAP Innovation Awards 2019 Entry Pitch Deck

Co-Innovation of BASF & SAP on SAP Data Hub for agile data orchestration & machine learning model integration

BASF SE

THE BEST RUN



Intro Video



<https://youtu.be/EOC-6iht0ck>



**WHAT ARE THE MAIN CHALLENGES IN BASF'S
JOURNEY INTO THE DATA STRATEGY?**

Co-Innovation of BASF & SAP on SAP Data Hub for agile data orchestration & machine learning model integration

BASF SE



“Quote”

BASF utilizes data science for a competitive edge in a digital economy. Therefore, a flexible environment to address rapidly evolving data science techniques is key to leverage opportunities of digitalization. To achieve this developers of SAP and BASF showed co-innovation at its best.

Frank Strohmaier,

Director Advanced Business Analytics – IT Solutions

Challenge

BASF data scientists needed a consistent and flexible data pipeline solution to apply machine learning models to solve use cases for several entities (business units, regions) without having to rebuild unique data pipelines for each use case.

Solution

BASF and SAP co-innovated a SAP Data Hub solution to address the BASF need for consistent and flexible data orchestration to cover multiple data sources, data pipelining and data governance that BASF machine learning models could easily be applied to solve business challenges.

Outcome

SAP Data Hub provided a stable data orchestration platform that connected data siloes, integrated machine learning models and averted the need to rebuild a data pipeline for each use case/business units.

BASF has a more **agile** environment that will easily embrace new data science techniques.

Data pipeline is transparent and consistent. Time and money saved on not rebuilding pipeline for each use case (**months to weeks**).

Data Scientists can leverage their respective machine learning models skillset (R, Python, TF) and changing models takes **days v. weeks** for productivity boost.



Business Challenge & Objectives

BASF data scientists develop state-of-the-art machine learning models in R, Python, TensorFlow and in the future potentially other programming languages. To date, there was the need to build an unique data pipeline for each specific use case and manually adjust data pipelines to support upscaling of solution into the different business areas and regions of the company.

SAP Data Hub would provide a flexible and consistent data orchestration solution and data pipeline to apply machine learning models to solve use cases for several entities (business units, regions). SAP Data Hub is the platform to orchestrate data loading, data preparation, machine learning model execution and visualization using cloud and on-premises services.



Project / Use Case Details

BASF is the largest chemicals company in the world with \$75B in revenue, subsidiaries and joint ventures that span 80 countries and 115K employees. BASF business segments include Chemicals, Performance Products, Functional Materials & Solutions, Agricultural Solutions and Oil & Gas. BASF has six Verbund sites (Ludwigshafen is world's largest chemical complex owned by single company) and 347 additional production sites. This highly-integrated manufacturing operation amounts to over 15.5 petabytes of data. BASF is interested in utilizing said data to improve services, plant maintenance and use of plant assets.

In 09/2017, FJ Schwenke, Head of BASF's Advanced Business Analytics Team attended a SAP Big Data Event @ Hudson Yards NYC, USA whereby SAP Executives presented Big Data innovations that will dramatically accelerate the ability to drive business value (<https://events.sap.com/sap-big-data-new-york/en/home>). BASF and SAP afterwards agreed on a **co-innovation project to develop SAP Data Hub to specifically address a BASF challenge**: BASF data scientists develop state-of-the-art machine learning models in R, Python, TensorFlow and in the future potentially other programming languages. To date, there was the need to build a unique data pipeline for each specific use case and manually adjust data pipelines to support upscaling of a solution into the different business areas and regions of the company.

From 11/2017-02/2018 **Design Thinking** workshops were held to better understand the needs of BASF and potential use of SAP Data Hub within that environment. BASF worked on data and internal infrastructure where SAP worked on SAP Data Hub development to prepare a foundation for a POC. During that phase **developers from both companies collaborated** directly to solve the given challenges. The POC environment was built by 02/2018. **The first POC use case involved predicting Customer Churn** and ran from 04/2018-07/2018. The POC target was to understand if different models for prediction can be implemented and run in SAP Data Hub without rebuilding a data pipeline for each use case. Specifically, SAP Data Hub had to process data from 1.1B sales orders from SAP HANA and apply different machine learning models from R and Python. **The result of the POC was that SAP Data Hub enabled BASF to switch between different models instantaneously and allows for formerly not possible scalability of solutions.**

Since the POC was successful, further advanced analytics solutions of BASF will be transferred to SAP Data Hub. **The next use case is a BASF Product Recommender for sales people** that will be prepared to be orchestrated by SAP Data Hub. The flexibility of SAP Data Hub will allow sales people to get the product recommendation results in many formats as required by their respective business needs (e.g. via dashboards, write-back into other systems, etc.), at the same time allowing BASF to efficiently upscale the solution. **The implementation is planned to be rolled out company wide** by early, 2019.

The SAP Data Hub architecture allows BASF to easily change machine learning models using different data science techniques in R, Python, TensorFlow, etc. while using the same data pipeline (same source, same visualization). This supports not only breaking down internal data silos but also systematically including non-SAP data and enabling BASF to apply state-of-the-art data science models. **Instead of rebuilding a pipeline for a specific data source / modelling language which can take months, BASF is now able to reduce this effort to weeks.**

SAP Data Hub intuitive environment empowers different skillsets to be more productive and makes **BASF more agile to embrace newer technologies** that are coming particularly in the rapidly evolving data science field. For example, **BASF is also in the implementation phase of a voice controlled virtual assistant to support sales representatives in their daily tasks using SAP Cloud Platform (SCP)**. Data generated by the virtual assistant could be fed into SAP Data Hub for further analysis.



Benefits and Outcomes

Business / Social

BASF has a more agile environment that will easily embrace new data science techniques and quickly respond to changing business needs.

Developers can focus on applying state-of-the-art modelling techniques, and no longer have to spend time on building unique data pipelines.

IT

Scalable, enterprise-ready data science solutions can be orchestrated efficiently.

Data pipeline is transparent and consistent. Time and money saved on not rebuilding pipeline for each use case (months to weeks).

Break down data siloes to gain better business view.

Human Empowerment

Data Scientists can leverage their respective machine learning models skillset (R, Python, TF) and changing models takes days v. weeks for productivity boost.

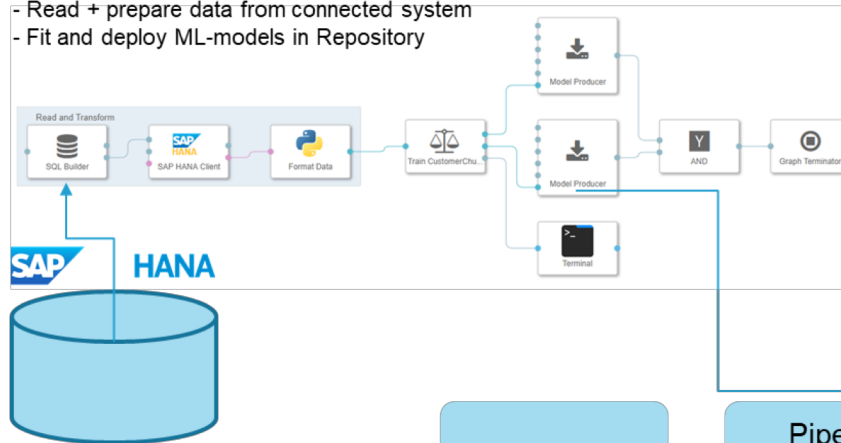
SAP Data Hub drag and drop user interface empowers different skillsets to be more productive.



Architecture

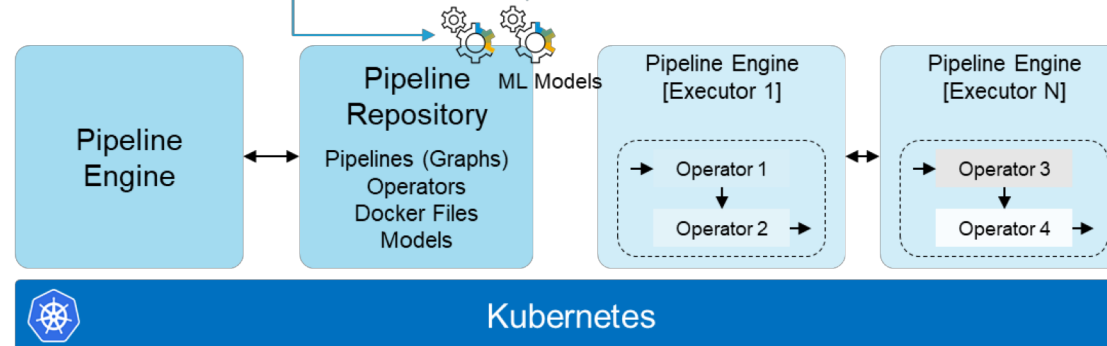
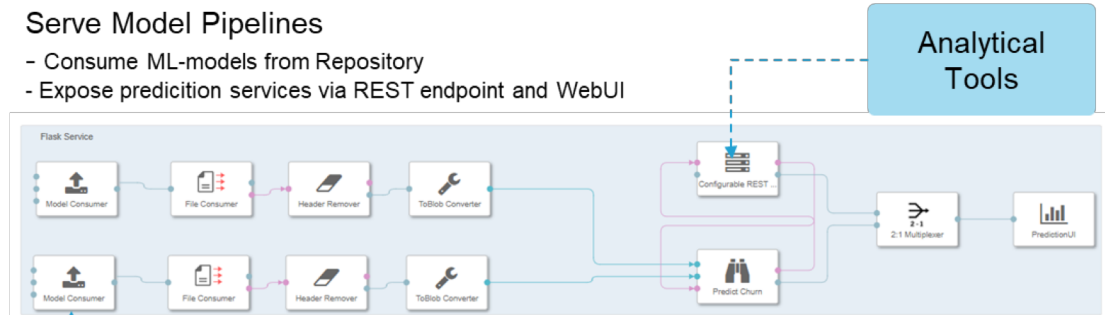
Train Model Pipelines

- Read + prepare data from connected system
- Fit and deploy ML-models in Repository



Serve Model Pipelines

- Consume ML-models from Repository
- Expose prediction services via REST endpoint and WebUI





Deployment

Date of Deployment or POC: 04/2018

Number of live users: 20

SAP Technologies Used:

SAP Data Hub

SAP HANA

Server Processor: Intel X86_64 provisioned on MS Azure Cloud

Linux Distribution: Linux Ubuntu Version 16.04.3 LTS



Emerging Technologies and Use Cases

The following Emerging Technologies and use-cases are part of the project and describe the contribution

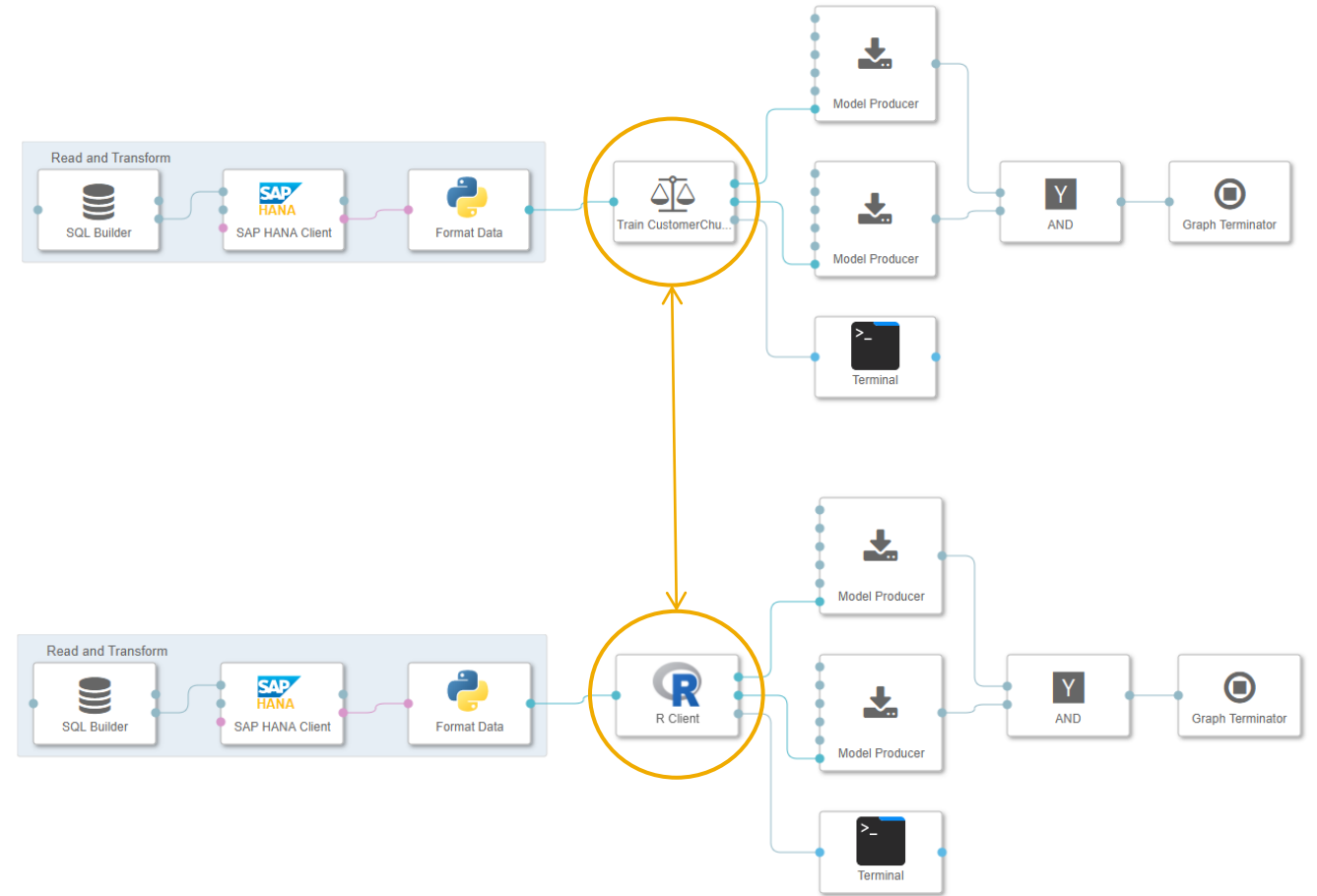
	Technology or Use Case	Yes/No	Contribution to Project
1.	Machine Learning / Artificial Intelligence	Yes	R and Python models used to predict customer churn.
2.	IoT		
3.	3D printing		
4.	Blockchain		
5.	API Economy / Integrate the Intelligent Enterprise		
6.	Cloud Native / Event Based Architectures	Yes	Docker and Kubernetes architecture in MS Azure.
7.	Extending the digital core with SAP CP / ABAP in SAP CP		
8.	SAP Leonardo Application (extending SAP application, using Industry Innovation Kits or result of Design Thinking workshop)	Yes	Co-innovation with SAP Data Hub via Design Thinking workshops.

Additional Information:

Customer Churn Prediction with SAP Data Hub Pipelines

Modularity / Extensibility

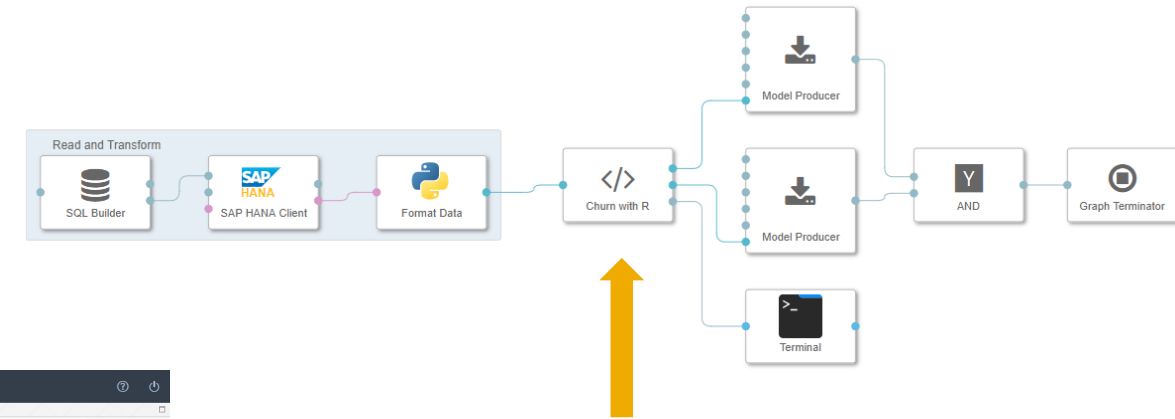
- Re-use Pipelines for different scenarios by exchanging operators
- Run any number of Pipeline instances in parallel for scalability



Customer Churn Prediction with SAP Data Hub Pipelines

Modularity / Extensibility

- Embed any code or library into custom operators, e.g. Python Pandas, Lifetimes or R BTYD
- Store operators in repository for re-use in other pipelines or scenarios



Create operator

Name: RChurn

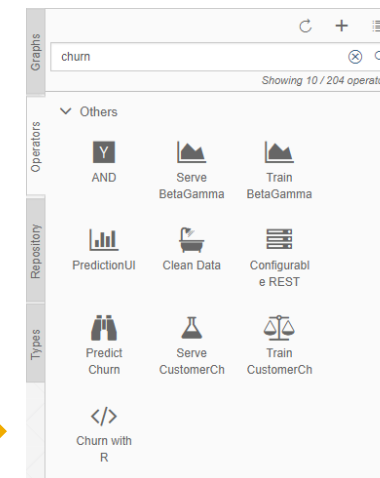
Display name: Churn with R

Base Operator: rClient

OK Cancel

The screenshot shows the 'Churn with R' operator configuration in the SAP Data Hub Pipeline Modeler. The 'Script' tab is active, displaying R code for fitting a BTYD model. The code includes library imports, data reading from SAP HANA, model fitting, and saving the model to a repository. The 'switchAffectTimerCallbacks' and 'switchStartOff' properties are set to 'True'.

```
1 library(dplyr)
2 library(devtools)
3 devtools::load_all(c("/Rchurnpackage"))
4
5 onInput <- function(lines) {
6
7   config <- list(fit_params = list(
8     predict_date = "2017-08-01",
9     predict_date_from = "2014-01-01",
10    from_date = "2014-01-01",
11    calibration_end = "2017-01-01",
12    units="day" #day or week as unit of analysis, all transactions within a unit of time is treated as a single interactions with the customer
13  ))
14
15  data_product <- read.csv(text=lines, header=TRUE, sep=";", colClasses=c("integer","date", "factor"))
16
17  # The table input requires 3 columns, cust, date, sales
18  transaction_data <- rename(data_product, cust=cust, date=sales)
19  transaction_data <- mutate(transaction_data,sales=0)
20
21  fitted_model <- fit_BTVD_with_transactions(transaction_data, config)
22
23  conn <- rawConnection(raw(0), "w")
24  saveRDS(fitted_model, conn)
25  model_blob <- rawConnectionValue(conn)
26  close(conn)
27
28  numRows <- paste("Number of rows in data: ", nrow(transaction_data))
29  list(model=model_blob, debug=toString(numRows))
30 }
31
32 #setPortCallback(c("indata"), c("model", "debug"), "onInput")
33
34
```



Customer Churn Rate Prediction Pipeline

R / BTYD

