Синергия Open Source и SAP в области машинного обучения

Буслов Дмитрий, SAP
Октябрь, 2018

EXTERNAL

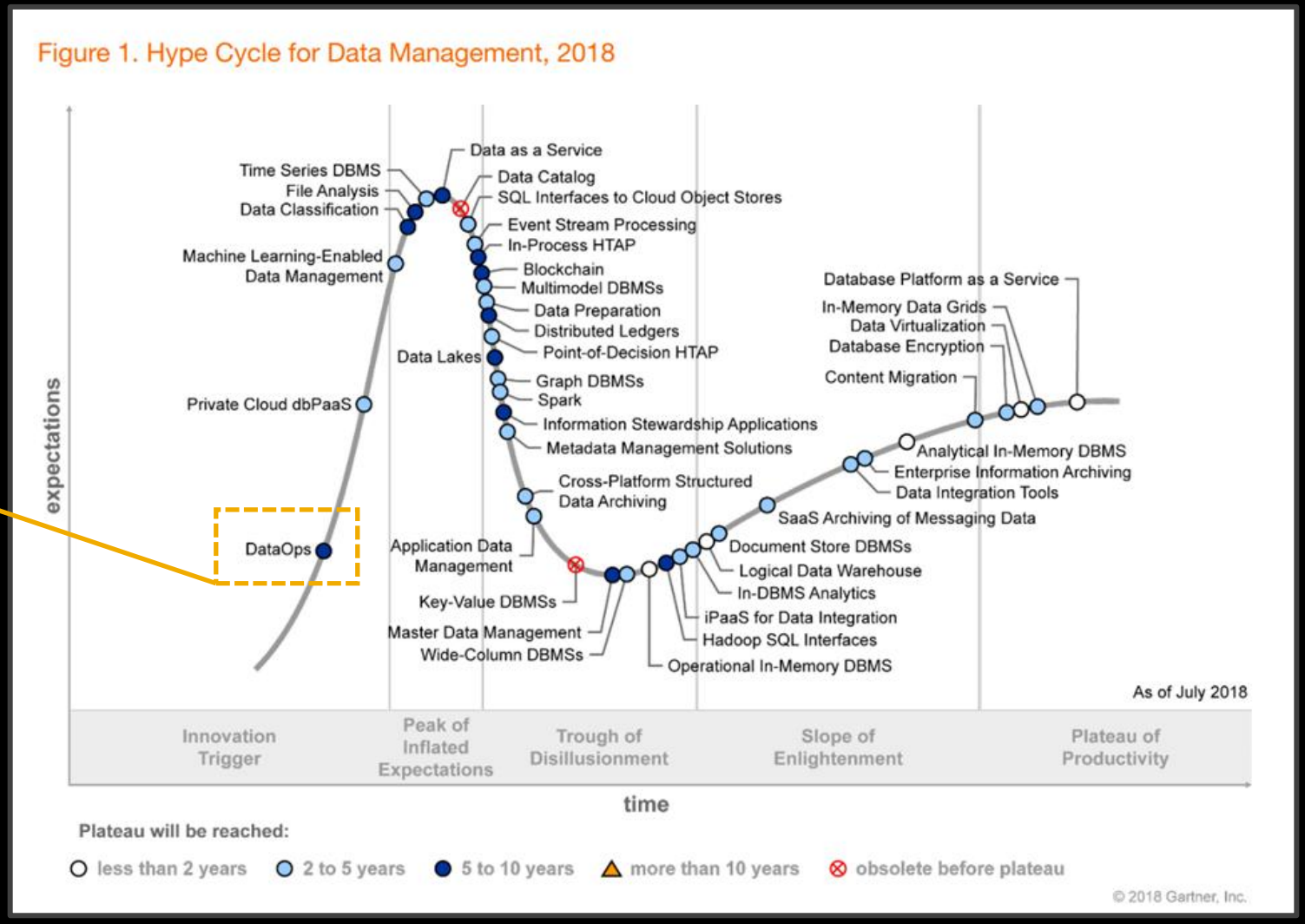
THE BEST RUN SAP

- Data Operations
- Интеллектуальное предприятие
- Open source в Enterprise среде
- Встроенный ML
- Инструменты
- HANA PAL
- Различные подходы
- Приложение

© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

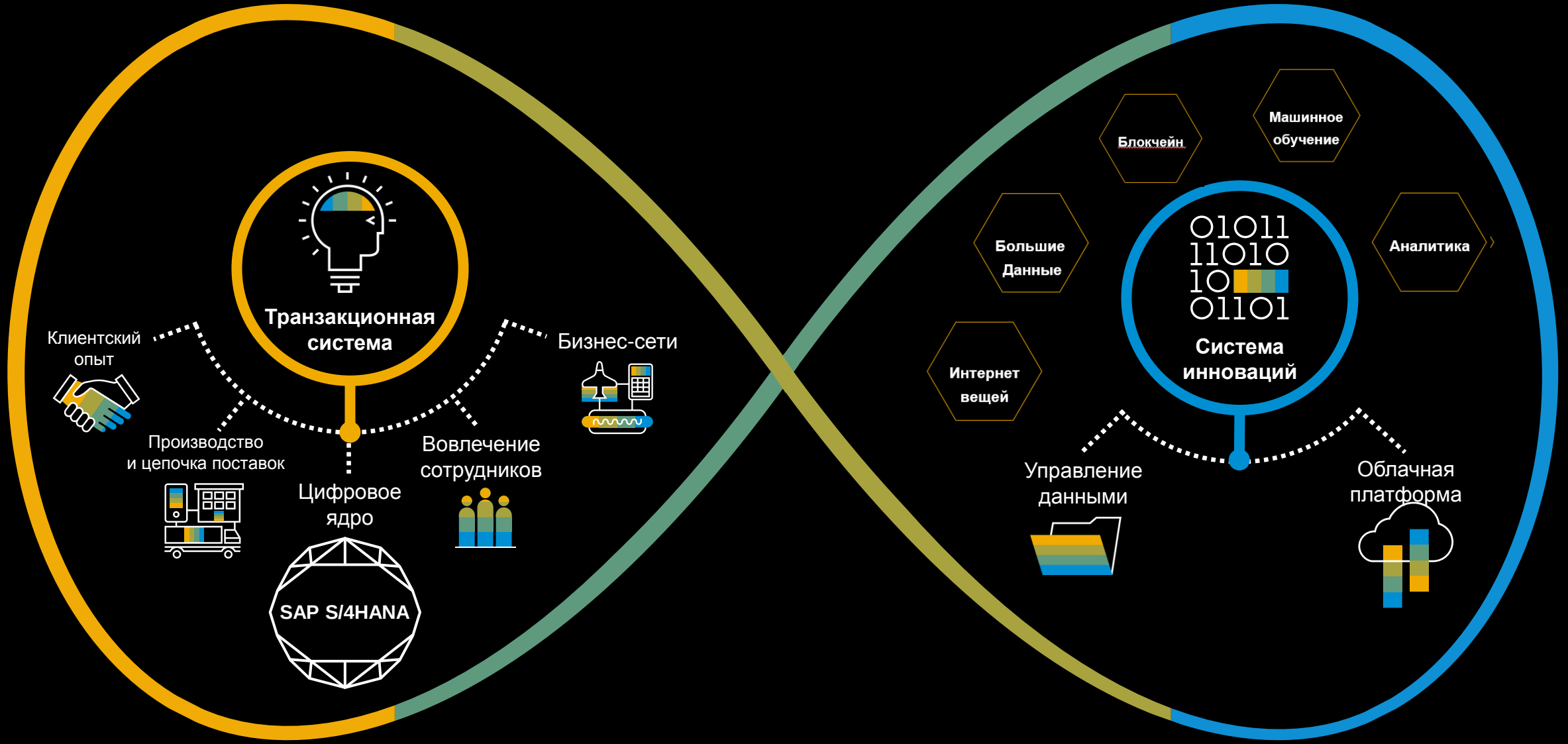
Data Operations (DataOps)

Data Ops – это подход дизайна, внедрения и управления разнородными данными в распределенной архитектуре, которая поддерживает широкий охват инструментов open source, различных фреймворков и корпоративных систем.



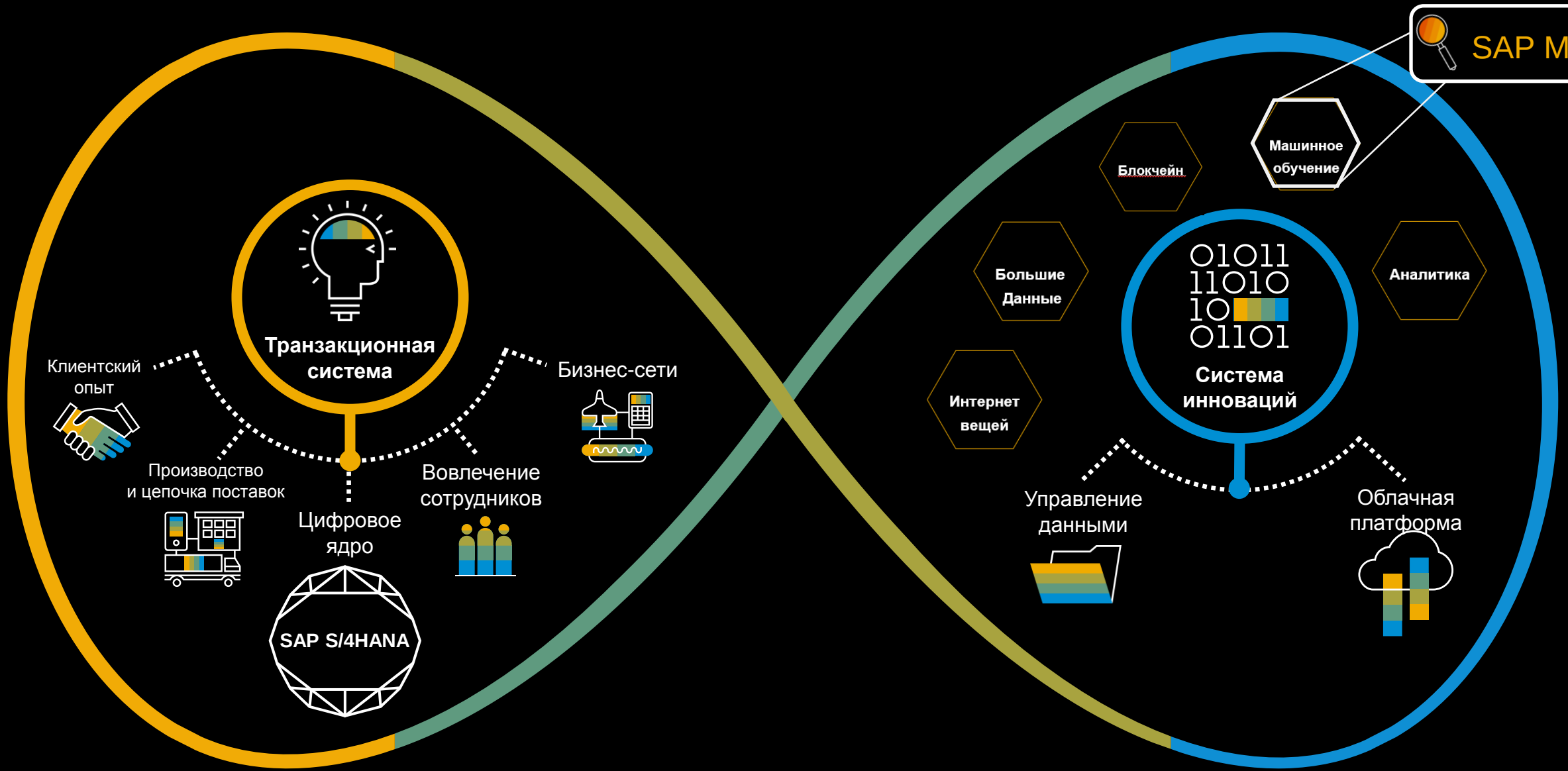
© 2018 SAP SE or an SAP affiliate company. All rights reserved.

Интеллектуальное предприятие



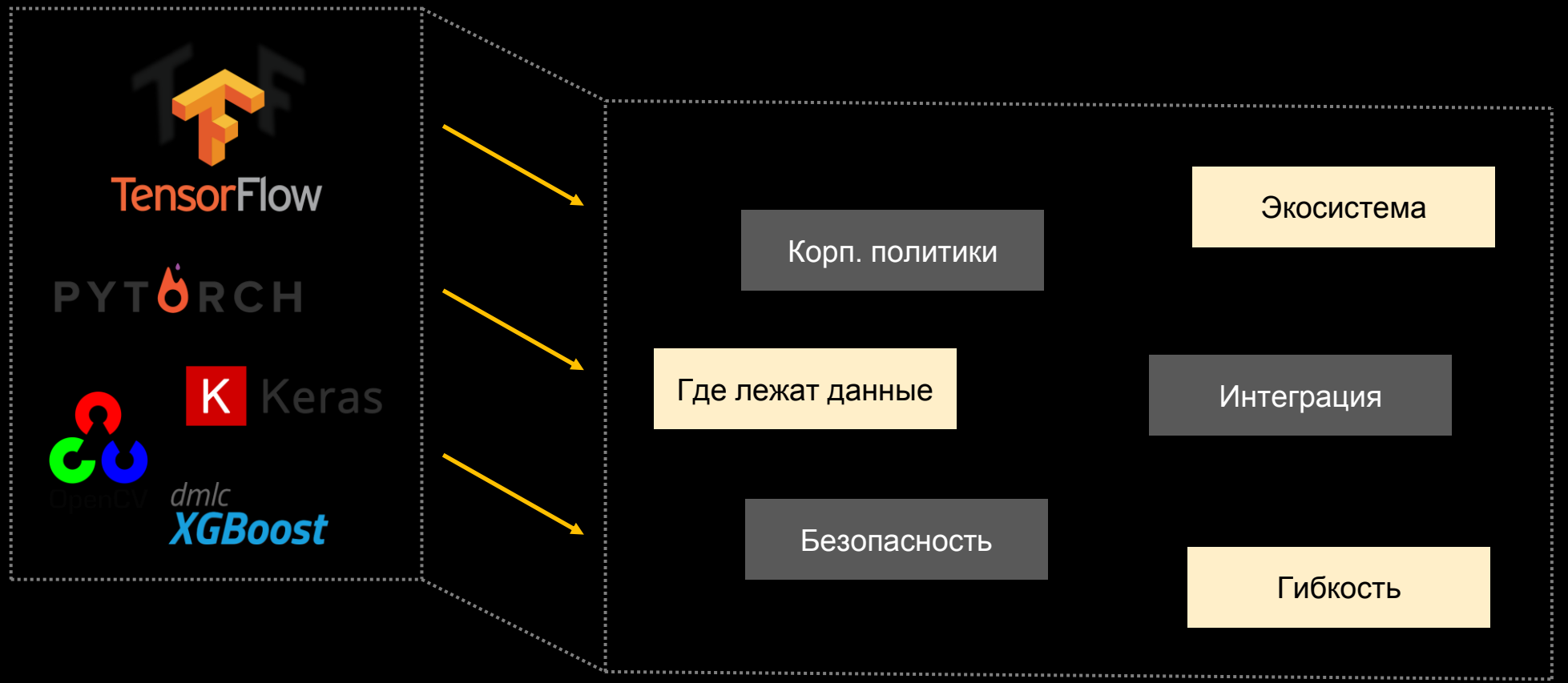
© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

Интеллектуальное предприятие



© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

Open source в Enterprise среде



© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

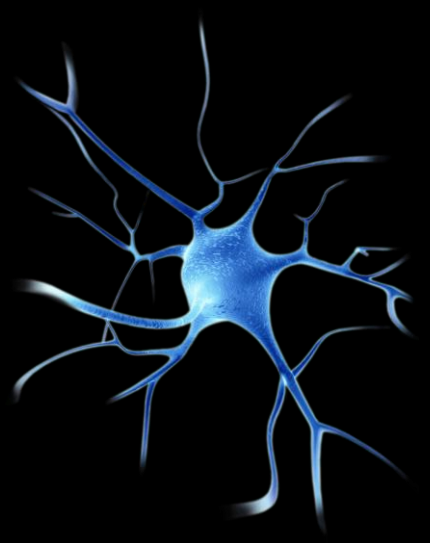
Встроенный ML (Embedded ML)

Predictive Analytics Integrator

CE 1705 OP 1709	SAP S/4HANA Sourcing and Procurement	+	SAP Leonardo SAP Predictive Analytics	=	**Contract Consumption** Proactively renew contracts by predicting their expiration and consumption
CE 1708 OP 1709	SAP S/4HANA Material Management	+	SAP Leonardo SAP Predictive Analytics	=	**Stock in Transit** Mitigate production risks by forecasting stock in transit delays
CE 1802 OP 1809	SAP S/4HANA Sales Analytics	+	SAP Leonardo SAP Predictive Analytics	=	**Sales Quotation Conversion** Improve Sales Forecasts
CE 1805 OP 1809	SAP S/4HANA Procure-to-Pay	+	SAP Leonardo SAP Predictive Analytics	=	**Payment Block – Cash Discount at Risk** Avoid loss of cash discounts
CE 1805 OP 1809	SAP S/4HANA Finance	+	SAP Leonardo SAP Predictive Analytics	=	**Smart Alerts for Profit & Loss Analysis** Get informed of hidden unusual costs and revenue opportunities
CE 1805 OP 1809	SAP S/4HANA Idea	+	SAP Leonardo SAP Predictive Analytics	=	**Project cost forecast** Mitigate project budget overruns

© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

Разные инструменты – разные цели



Data Hub

HANA Predictive Analysis Library

Predictive Analytics

HANA EML

SAP Cloud Platform

HANA R integration

SAP Analytics Cloud

SAP Leonardo MLF

Использование Open Source

© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

HANA Predictive Analysis Library

SQL

```
Truncate table; #PAL_PARMTBL
INSERT INTO #PAL_PARMTBL VALUES ('CLASS_MAP0', null, null, 'no');
INSERT INTO #PAL_PARMTBL VALUES ('CLASS_MAP1', null, null, 'yes');
INSERT INTO #PAL_PARMTBL VALUES ('METHOD', 0, null, null);
INSERT INTO #PAL_PARMTBL VALUES ('EXIT_THRESHOLD',null,0.000001,null);
INSERT INTO #PAL_PARMTBL VALUES ('THREAD_RATIO',null,0.9,null);
INSERT INTO #PAL_PARMTBL VALUES ('MAX_ITERATION',10000,null,null);
INSERT INTO #PAL_PARMTBL VALUES ('STAT_INF', 1, null, null);
INSERT INTO #PAL_PARMTBL VALUES ('PMML_EXPORT', 2, null, null);
INSERT INTO #PAL_PARMTBL VALUES ('HAS_ID', 0, null, null);
INSERT INTO #PAL_PARMTBL VALUES (‘ENET_LAMBDA', null, 0.01, null);
INSERT INTO #PAL_PARMTBL VALUES (‘ENET_ALPHA', null, 0.1, null);

CALL _SYS_AFL.PAL_LOGISTIC_REGRESSION (
                        DBM2_TRAINING_TBL,
                        "#PAL_PARMTBL",
                        DBM2_LOGISTICR_RESULT_TBL,
                        DBM2_LOGISTICR_PMMLMODEL_TBL,?,?)
                         WITH OVERVIEW;

Truncate table; #PAL_PARMTBL
INSERT INTO #PAL_PARMTBL VALUES ('THREAD_RATIO',null,0.1,null);

Truncate table DBM_SCORED_TBL;
CALL _SYS_AFL.PAL_LOGISTIC_REGRESSION_PREDICT(
                        DBM2_VALIDATION_TBL,
                        DBM2_LOGISTICR_PMMLMODEL_TBL,
                         "#PAL_PARMTBL",
                        DBM_SCORED_TBL) with overview;
```

Python

```
ctxt = DataFrame.ConnectionContext(<connect details>)
select_statement = 'select * from "DEVUSER"."DBM2_TRAINING_TBL“'
training_set = DataFrame.DataFrame(ctxt, select_statement)

validation_set=ctxt.table(‘DBM2_VALIDATION_TBL’, schema=’MySchema’)

features = ['AGE','JOB',…,'NREMPLOYED']
label = "LABEL"
enet_lambda = 0.01
enet_alpha = 0.1

mylr = hana_ml.LogisticRegression(connection_context, solver='Cyclical',
tol=0.000001, max_iter=10000, stat_inf=1,pmml_export=2, has_id=1,
lamb=enet_lambda, alpha=enet_alpha)

mylr.fit(training_set, features, label, "no","yes")

accuracy_val = mylr.score(validation_set, features, label)
```

© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

Различные подходы

Различные контексты

Машинное обучение

Данные

- В SAP Hana: PAL, APL, R, TensorFlow
- В Data Hub: использование ML для Больших Данных, TensorFlow, весь стэк Open Source
- SAP Analytics Cloud: (гибридная модель)

Данные

Машинное обучение

- SAP Cloud Platform
- SAP Leonardo Machine Learning Business and Functional services

* PAL: Predictive Analytics Library. APL: Automated Predictive Library

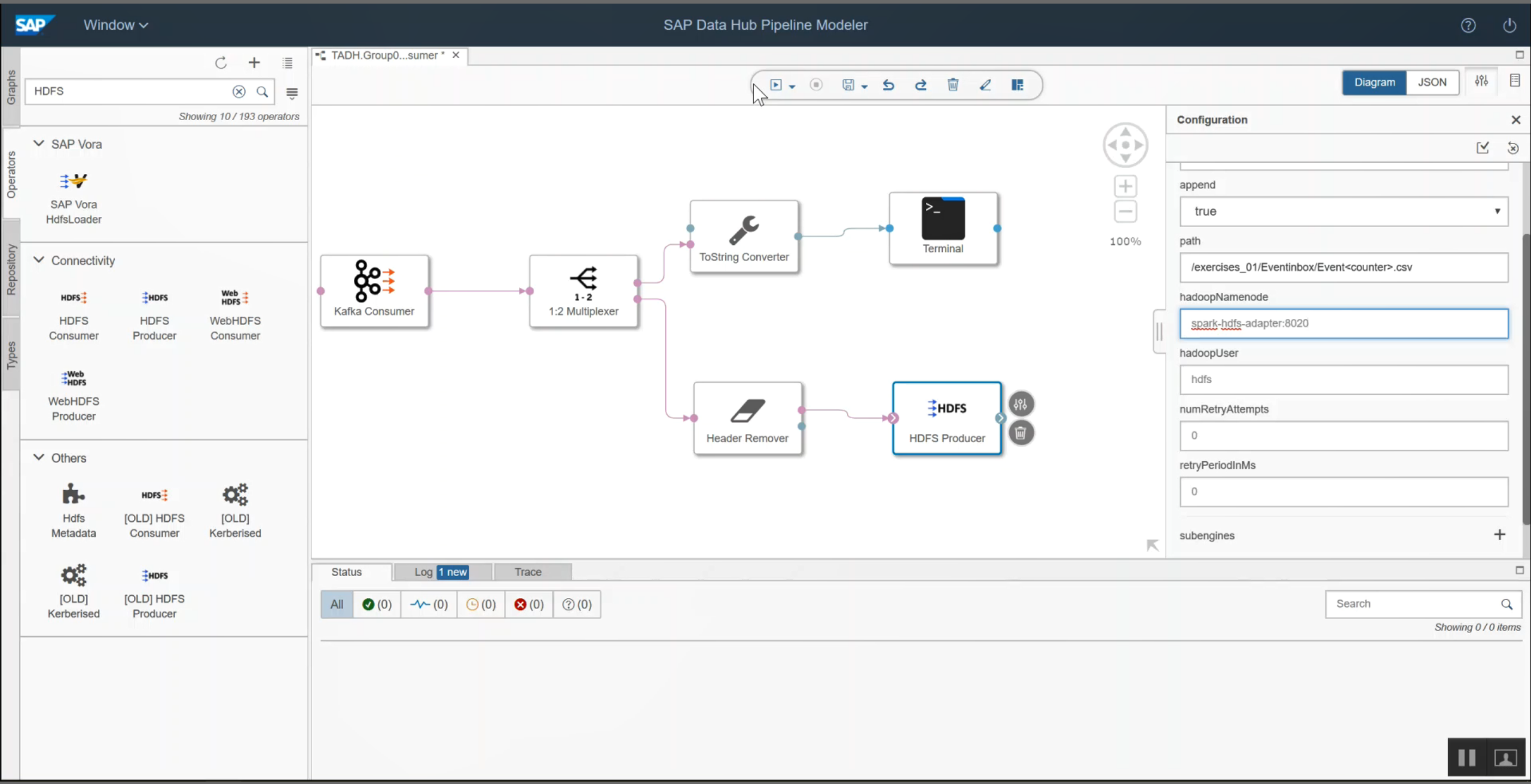
© 2018 SAP SE or an SAP affiliate company. All rights reserved. I INTERNAL

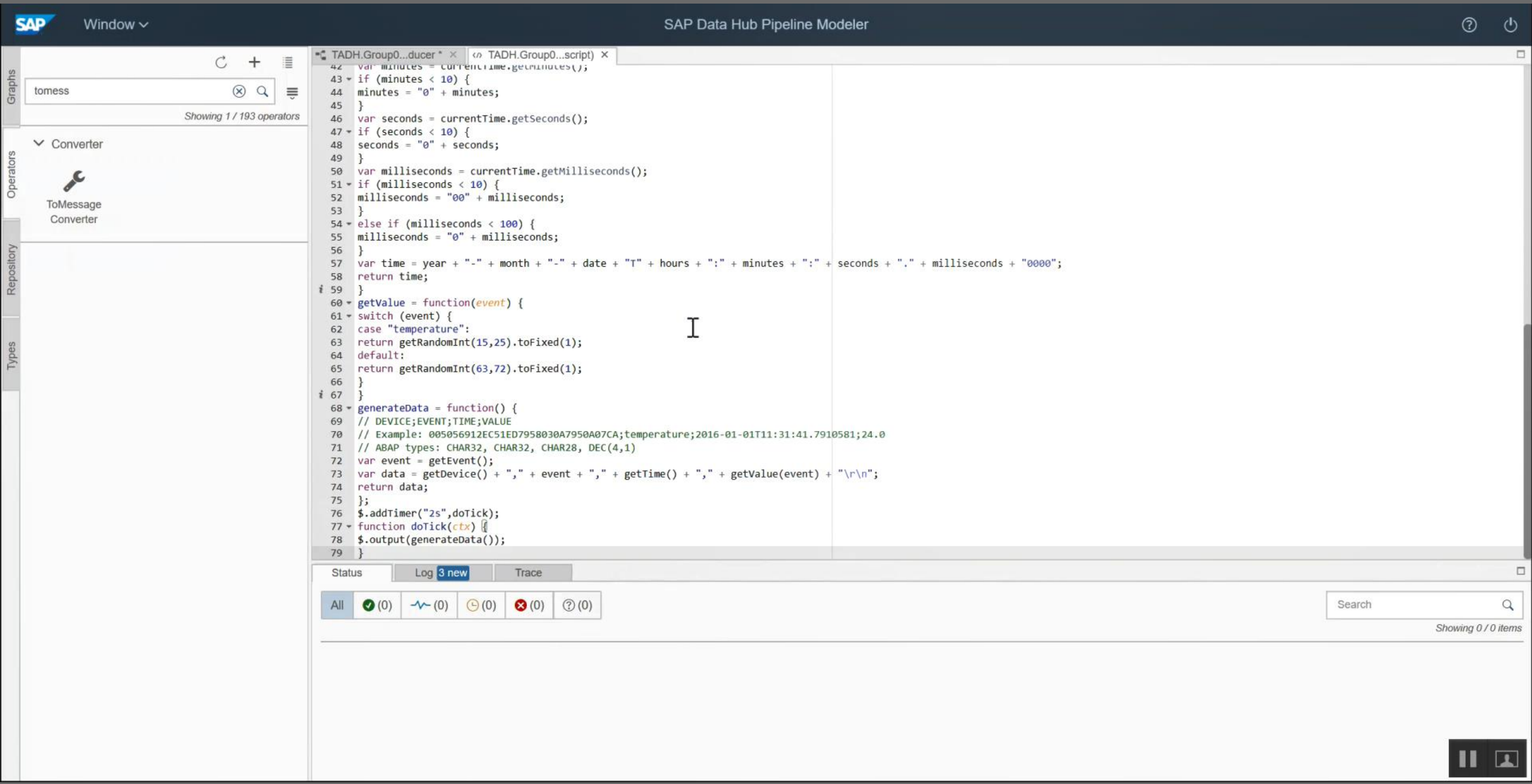
Спасибо за внимание.

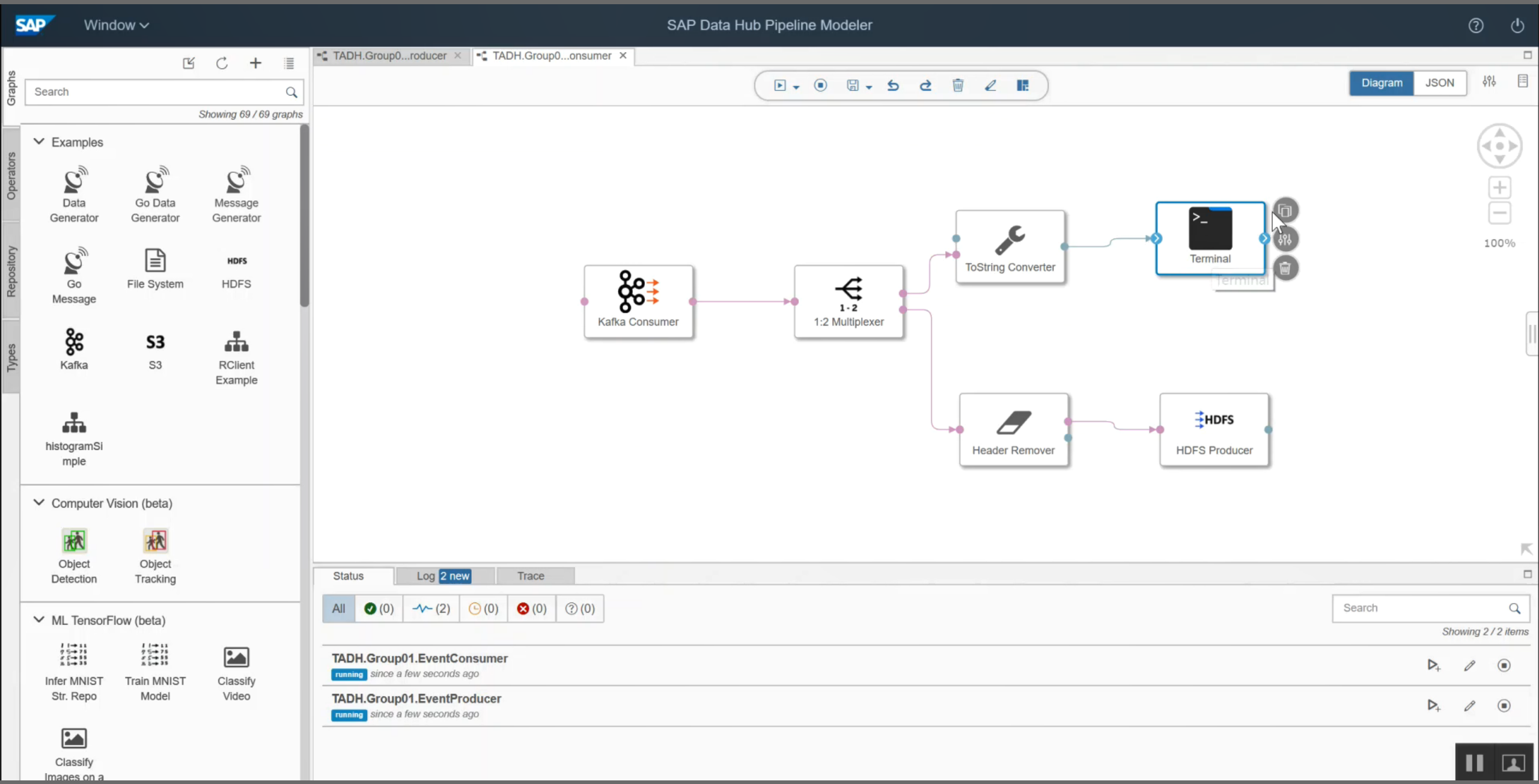
Контактная информация:

Дмитрий Буслов
Senior Presales Specialist
SAP CIS, Москва
M +7/929/545-73-74
dmitry.buslov02@sap.com

THE BEST RUN SAP®


SAP
Window
SAP Data Hub Pipeline Modeler
Graphs
Operators
Repository
Types
HDFS
Showing 10 / 193 operators
SAP Vora
SAP Vora HdfsLoader
Connectivity
HDFS Consumer
HDFS Producer
WebHDFS Consumer
WebHDFS Producer
Others
Hdfs Metadata
[OLD] HDFS Consumer
[OLD] Kerberised
[OLD] Kerberised
[OLD] HDFS Producer
TADH.Group0...sumer *
Diagram
JSON
Kafka Consumer
1 - 2
1:2 Multiplexer
ToString Converter
Terminal
Header Remover
HDFS Producer
100%
Configuration
append
true
path
/exercises_01/Eventinbox/Event<counter>.csv
hadoopNamenode
spark-hdfs-adapter:8020
hadoopUser
hdfs
numRetryAttempts
0
retryPeriodInMs
0
subengines
Status
Log 1 new
Trace
All
(0)
(0)
(0)
(0)
(0)
Search
Showing 0 / 0 items


Window
SAP Data Hub Pipeline Modeler
Graphs
Operators
Repository
Types
tomess
Showing 1 / 193 operators
Converter
ToMessage Converter
TADH.Group0...ducer *
TADH.Group0...script)
var minutes = currentTime.getMinutes();
if (minutes < 10) {
minutes = "0" + minutes;
}
var seconds = currentTime.getSeconds();
if (seconds < 10) {
seconds = "0" + seconds;
}
var milliseconds = currentTime.getMilliseconds();
if (milliseconds < 10) {
milliseconds = "00" + milliseconds;
}
else if (milliseconds < 100) {
milliseconds = "0" + milliseconds;
}
var time = year + "-" + month + "-" + date + "T" + hours + ":" + minutes + ":" + seconds + "." + milliseconds + "0000";
return time;
}
getValue = function(event) {
switch (event) {
case "temperature":
return getRandomInt(15,25).toFixed(1);
default:
return getRandomInt(63,72).toFixed(1);
}
}
generateData = function() {
// DEVICE;EVENT;TIME;VALUE
// Example: 005056912EC51ED7958030A7950A07CA;temperature;2016-01-01T11:31:41.7910581;24.0
// ABAP types: CHAR32, CHAR32, CHAR28, DEC(4,1)
var event = getEvent();
var data = getDevice() + "," + event + "," + getTime() + "," + getValue(event) + "\r\n";
return data;
};
$.addTimer("2s",doTick);
function doTick(ctx) {
$.output(generateData());
}
Status
Log 3 new
Trace
All (0) (0) (0) (0) (0)
Search
Showing 0 / 0 items


SAP
Window
SAP Data Hub Pipeline Modeler
Graphs
Search
Showing 69 / 69 graphs
Operators
Repository
Types
Examples
Data Generator
Go Data Generator
Message Generator
Go Message
File System
HDFS
Kafka
S3
RClient Example
histogramSimple
Computer Vision (beta)
Object Detection
Object Tracking
ML TensorFlow (beta)
Infer MNIST Str. Repo
Train MNIST Model
Classify Video
Classify Images on a
TADH.Group0...roducer
TADH.Group0...onsumer
Diagram
JSON
Kafka Consumer
1-2
1:2 Multiplexer
ToString Converter
Terminal
Header Remover
HDFS
HDFS Producer
100%
Status
Log 2 new
Trace
All
(0)
(2)
(0)
(0)
(0)
Search
Showing 2 / 2 items
TADH.Group01.EventConsumer
running since a few seconds ago
TADH.Group01.EventProducer
running since a few seconds ago

SAP Data Hub

TA_01

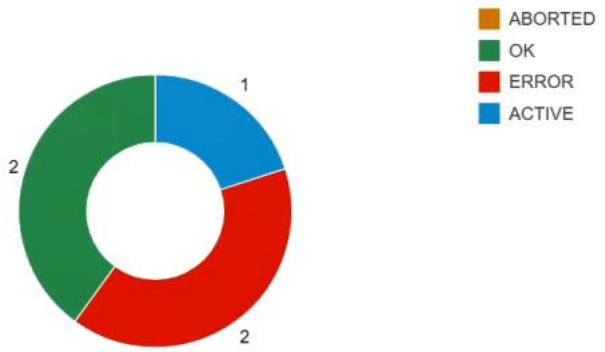
Monitoring Dashboard

Year | Month | Week | Day | Hour | Custom

TASK WORKFLOW | DATA PIPELINE | SLT CONFIGURATION

Showing executions started between 4/7/2018 and 4/13/2018

Summary

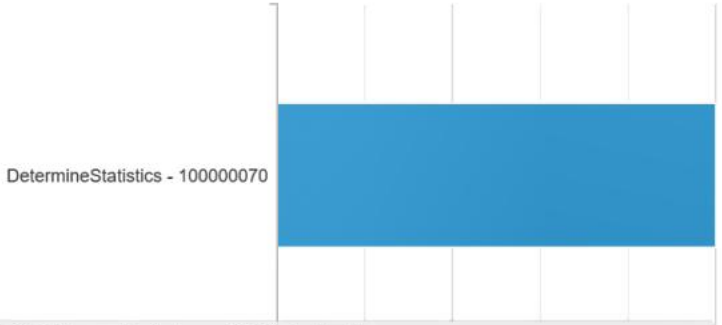


Task Workflow vs Time

Most Executed

DetermineStatistics DEMO	3 times
DetermineStatistics TA_DH_Solution	2 times

Longest Running



Recently Executed

View All

- DetermineStatistics — DEMO
- DetermineStatistics — TA_DH_Solution
- DetermineStatistics — DEMO
- DetermineStatistics — DEMO

https://ip-18-184-38-115.ta.sapdatahub.com:9052/index.html#

Task Workflows and Task Executions

Search By Task Workfl...

Task Workflows and Task Executions
All

DEMO

DetermineStatistics
DEMO

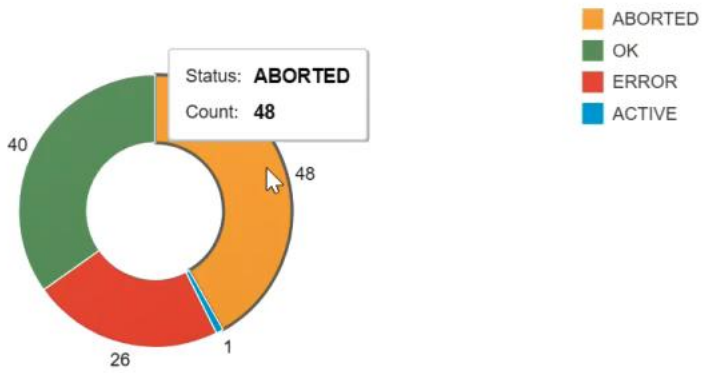
TA_DH_Solution

DetermineStatistics
TA_DH_Solution

Filtered by: Date(ALL) Status(ALL)

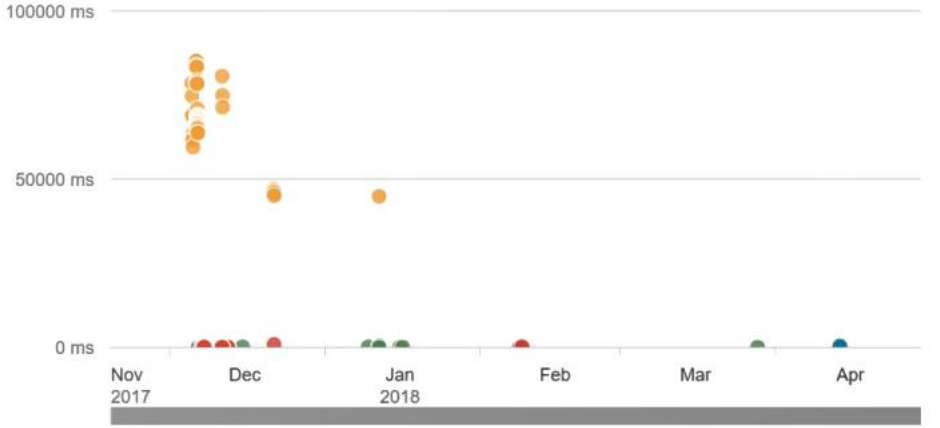
Charts | Filters

Summary (115)



Task Workflow vs Time

ms | s



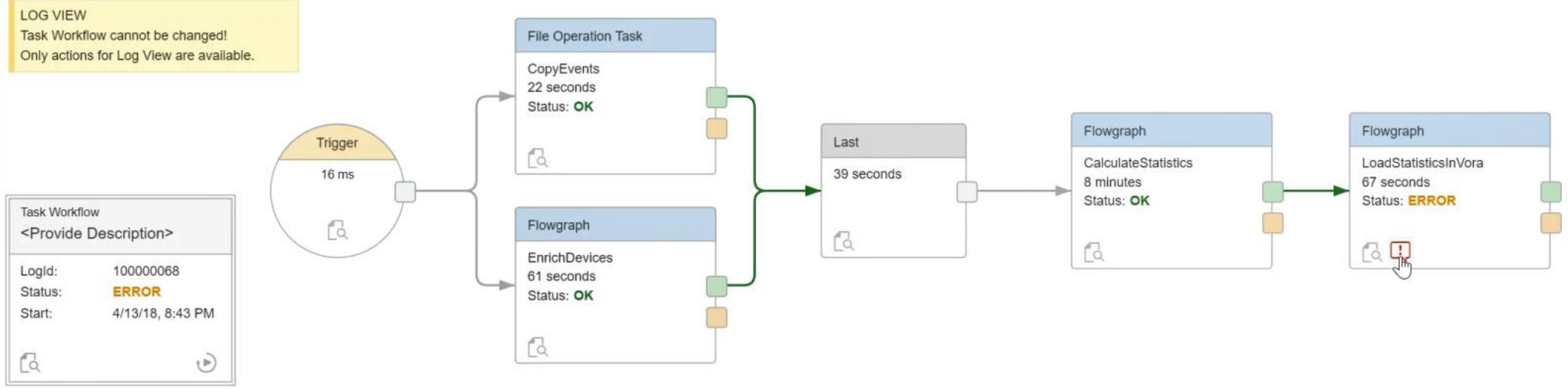
Log Id	Task Workflow Id	Description	Schedule Description	Executed By	Start Time	End Time	Status
100000070	**DetermineStatistics** DEMO		Direct Execution	TA_01	4/13/2018, 9:33:20 PM		△
100000069	**DetermineStatistics** TA_DH_Solution		Direct Execution	TA_01	4/13/2018, 8:54:51 PM	4/13/2018, 8:58:46 PM	✓
100000068	**DetermineStatistics** DEMO		Direct Execution	TA_01	4/13/2018, 8:43:44 PM	4/13/2018, 8:54:01 PM	!
100000067	**DetermineStatistics** DEMO		Direct Execution	TA_01	4/13/2018, 8:32:12 PM	4/13/2018, 8:39:05 PM	!
	DetermineStatistics		Direct Execution	TA_01	4/13/2018, 8:26:34 PM	4/13/2018, 8:30:11 PM	✓

DetermineStatistics | Log Id : 100000068

Show Nodes | All (6) | Processed (6) | Error (1)

Properties | Resume Run

LOG VIEW
Task Workflow cannot be changed!
Only actions for Log View are available.



Logs

Log Id	Start Time	End Time	Status
100000070	4/13/2018, 9:33:20 PM		
100000068	4/13/2018, 8:43:44 PM	4/13/2018, 8:54:01 PM	
100000067	4/13/2018, 8:32:12 PM	4/13/2018, 8:39:05 PM	