금융권 엔터프라이즈 머신러닝 활용을 위한 분석 플랫폼
: How to solve the issues of machine learning?

Dec 2017

KeunTae Kim
Advisory Consultant / SAS Korea

Copyright © SAS Institute Inc. All rights reserved.

sas
THE POWER TO KNOW.

Age of “Analytic Platform"

Copyright © SAS Institute Inc. All rights reserved.

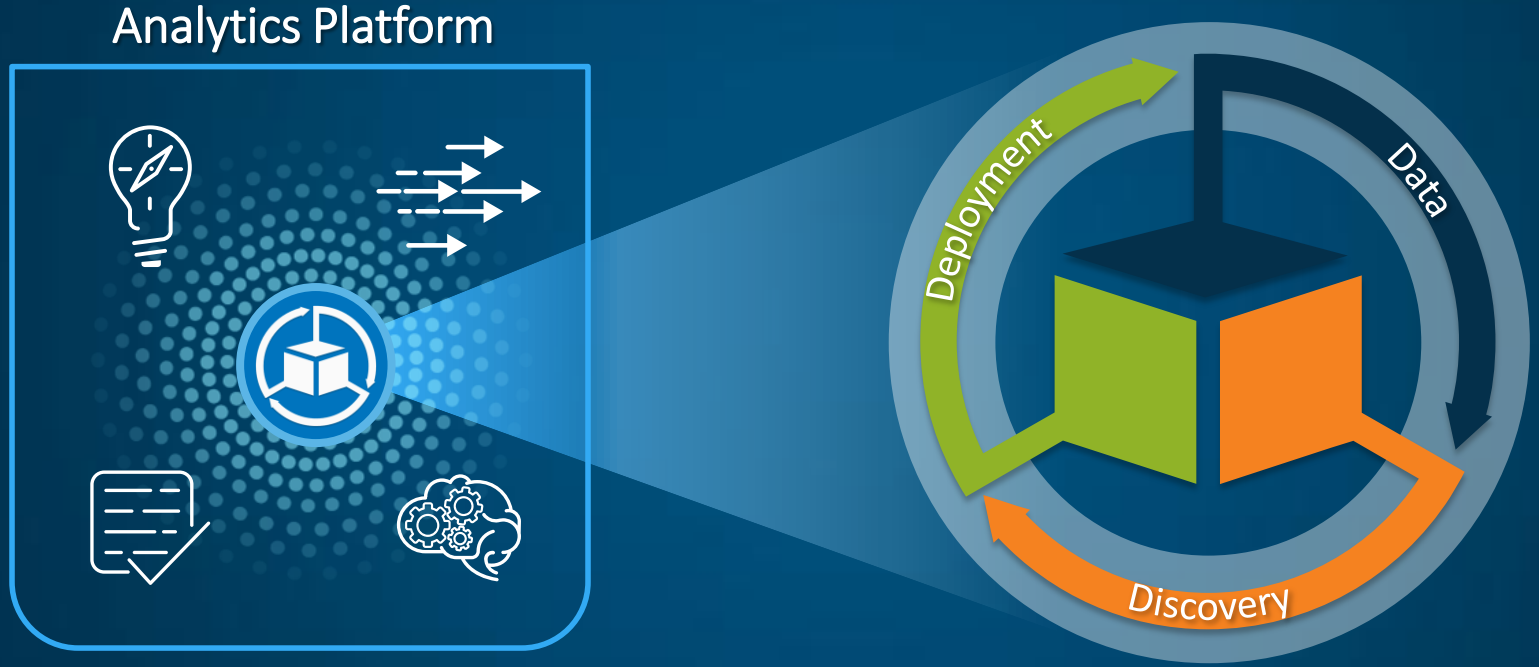
Analytics Trend

Copyright © SAS Institute Inc. All rights reserved.

Enterprise platforms

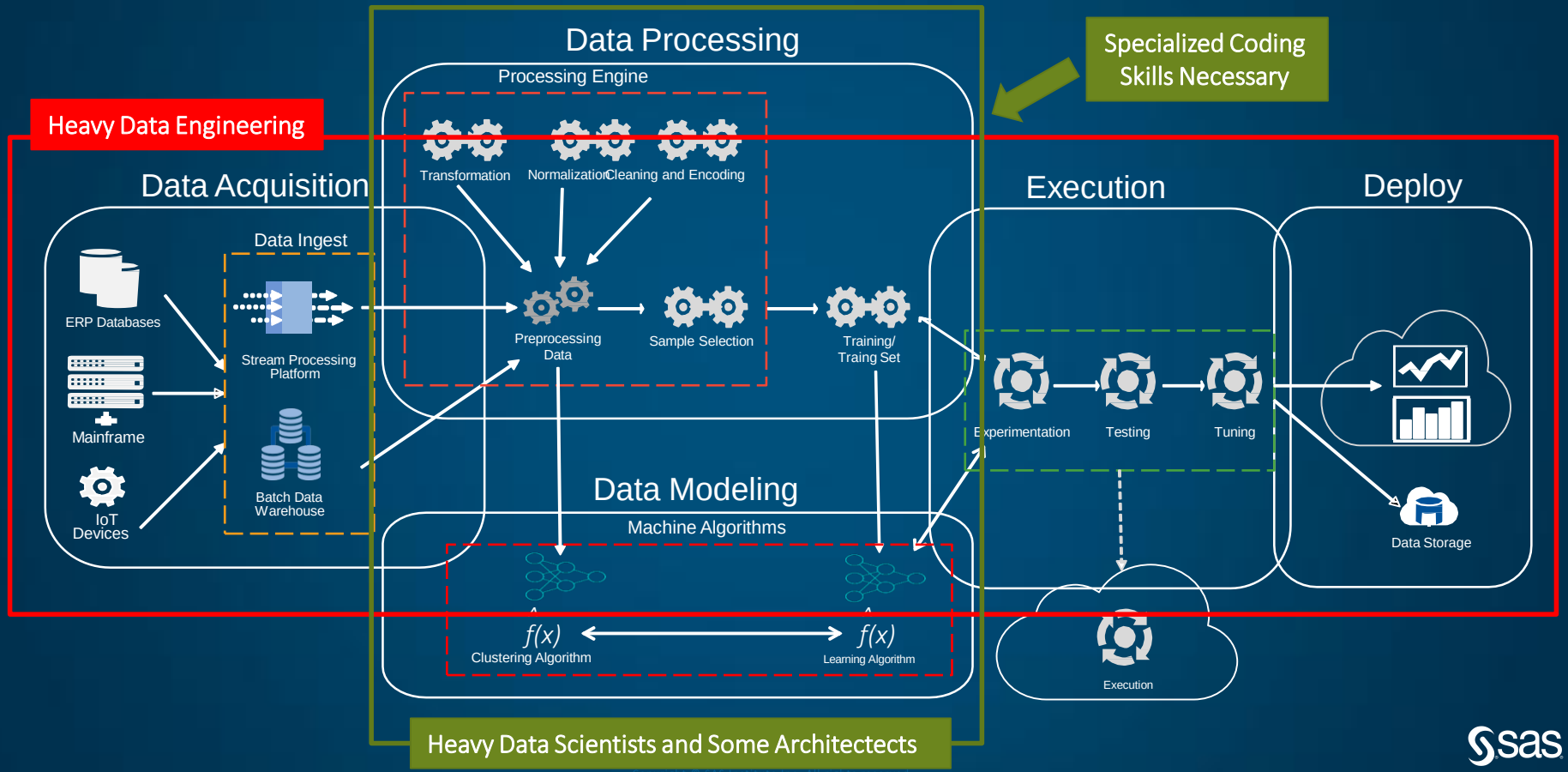
Copyright © SAS Institute Inc. All rights reserved.

Analytics Lifecycle – Data, Discovery, Deployment



Basic Machine Learning Architecture

Gartner



Copyright © SAS Institute Inc. All rights reserved.

sas

Open Source Ecosystem Issues

1. Silos

2. Lack of oversight

3. Delay in execution

Copyright © SAS Institute Inc. All rights reserved.

1. Silos : Best of Breeds ?

- MapReduce, SQL, ETL
- Machine Learning
- Graph Processing
- Streaming

ModelDB

Copyright © SAS Institute Inc. All rights reserved.

2. Lack of oversight -> Hidden Technical Debt

Analytics is more than the sum of its parts

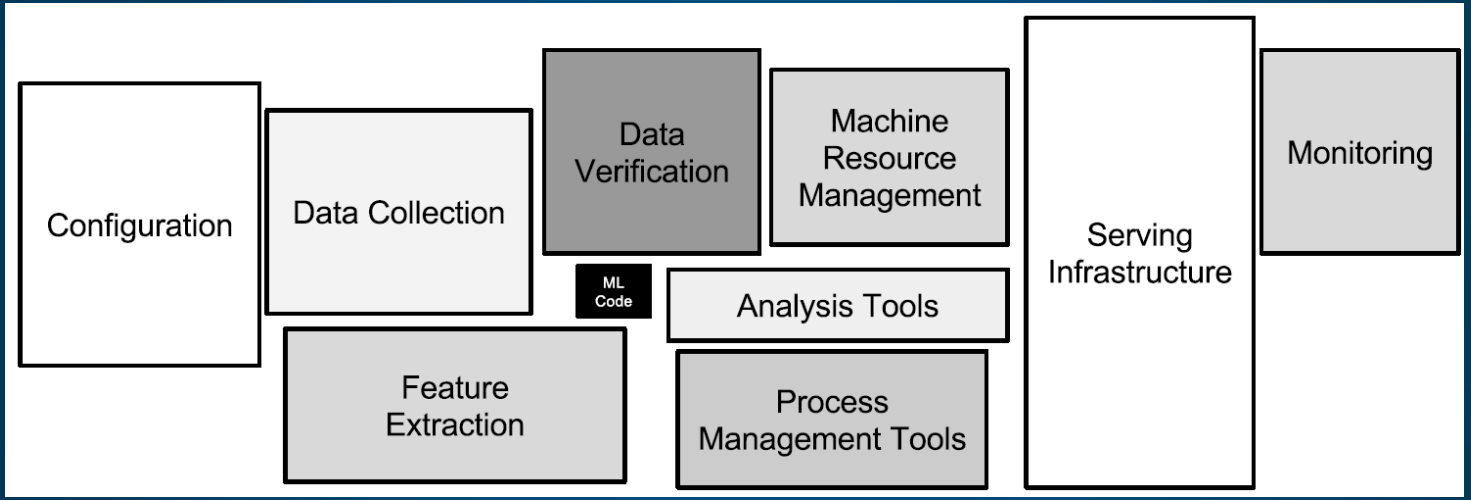


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

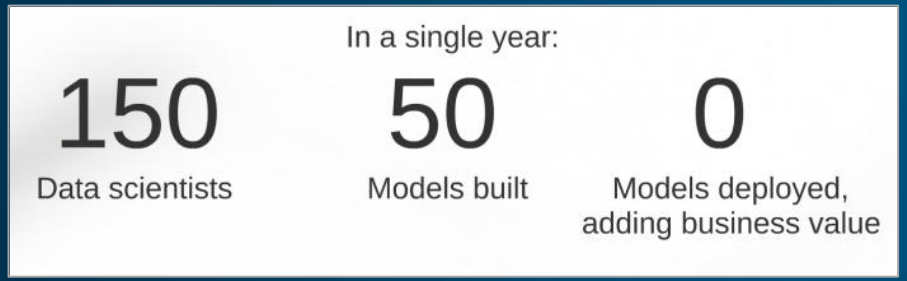
Source : "Hidden Technical Debt in Machine Learning Systems " , Google NIPS 2015

Copyright © SAS Institute Inc. All rights reserved.

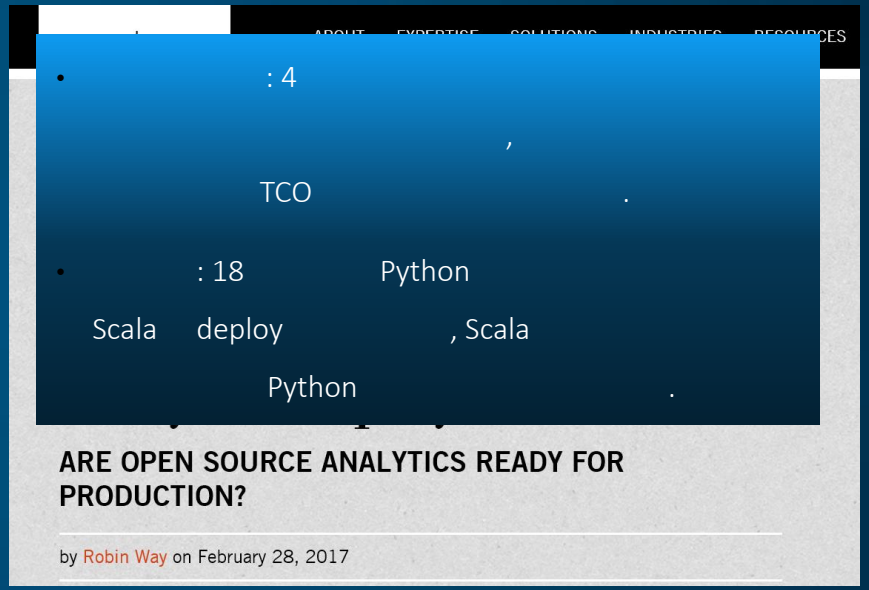
3. Delay in execution -> Time to Value

Are open source analytics ready for production?

It's common that both analytics teams and technology teams forget to spend time working out their plans for deploying these assets in the field, before they complete their exploratory analytics and model development.



Source : "Why You're Not Getting Value from Your Data Science" , HBR, 7 Dec. 2016
https://hbr.org/2016/12/why-youre-not-getting-value-from-your-data-science

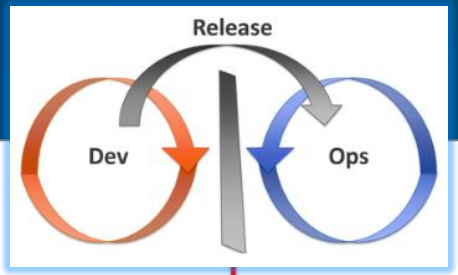


https://coriosgroup.com/financial-services-cdo-feedback-on-open-source-analytics-deployment/

* Chief Data Officers (CDOs)

Copyright © SAS Institute Inc. All rights reserved.

Apache Spark Model Deployment



Data Science

Prototype (Python/R)
Create Pipeline

- Extract raw features
- Transform features
- Select key features
- Fit multiple models
- Combine results to make prediction

Software Engineering

Re-implement Pipeline for production (Java)
Deploy Pipeline

- Extra implementation work
- Different code paths
- Synchronization overhead

✓ 개발 생산성 이슈

✓ 코드 변환 이슈

✓ 운영 환경 및 성능 이슈

* databricks, 2017

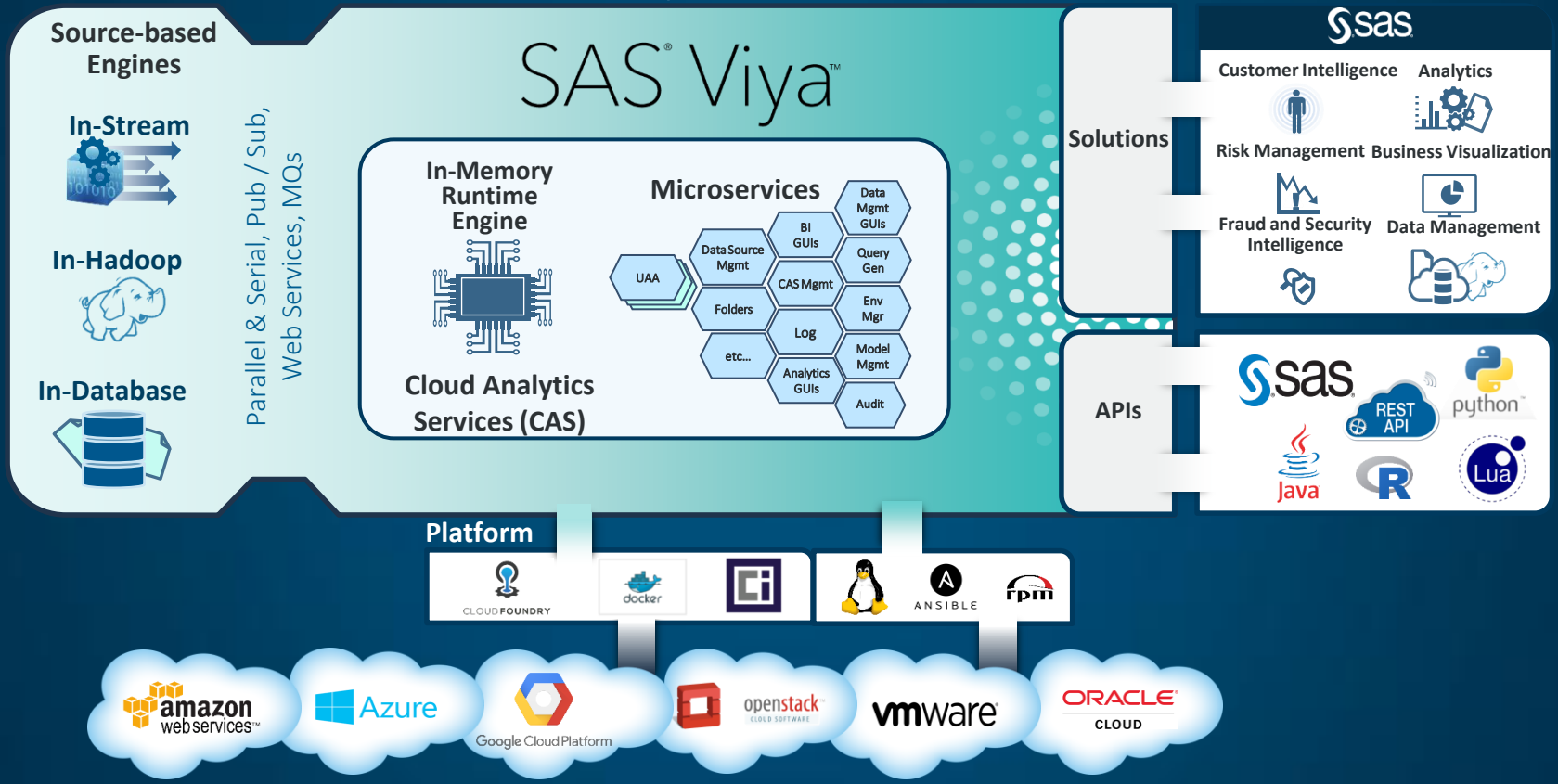
Copyright © SAS Institute Inc. All rights reserved.

How does SAS Viya solve the issue?

Copyright © SAS Institute Inc. All rights reserved.

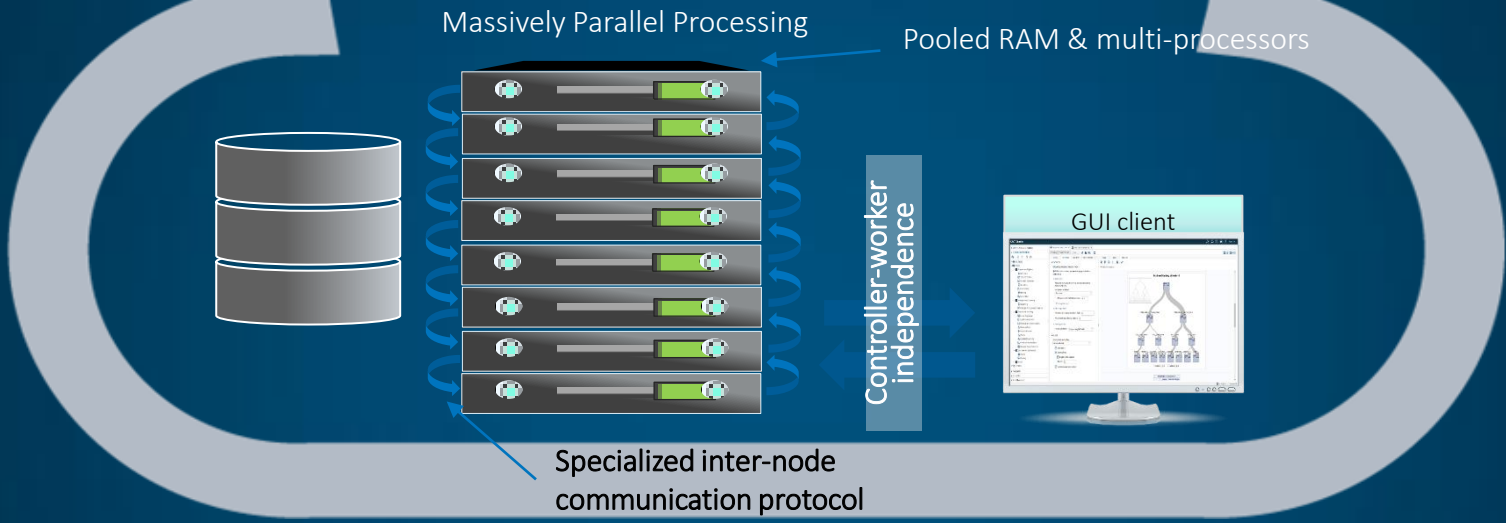
SAS® Viya™
One Environment
One User Experience

SAS Viya Architecture



Copyright © SAS Institute Inc. All rights reserved.

Performance Acceleration – MPP, GPU



Fast

- Multi-threaded
- Distributed In-Memory
- Inter-node Communication

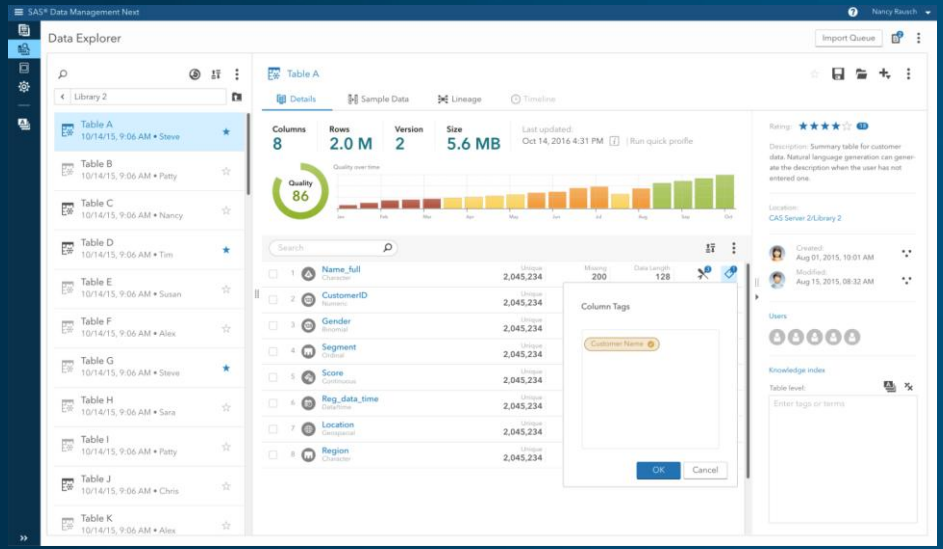
Scalable

- Single Machine to Distributed MPP
- Scale-out
- On-Premise to Cloud

Copyright © SAS Institute Inc. All rights reserved.

Data Management

SAS Viya



Point and click GUI based

SAS Data Preparation

Non-Technical User

Coding interfaces

Data Scientist

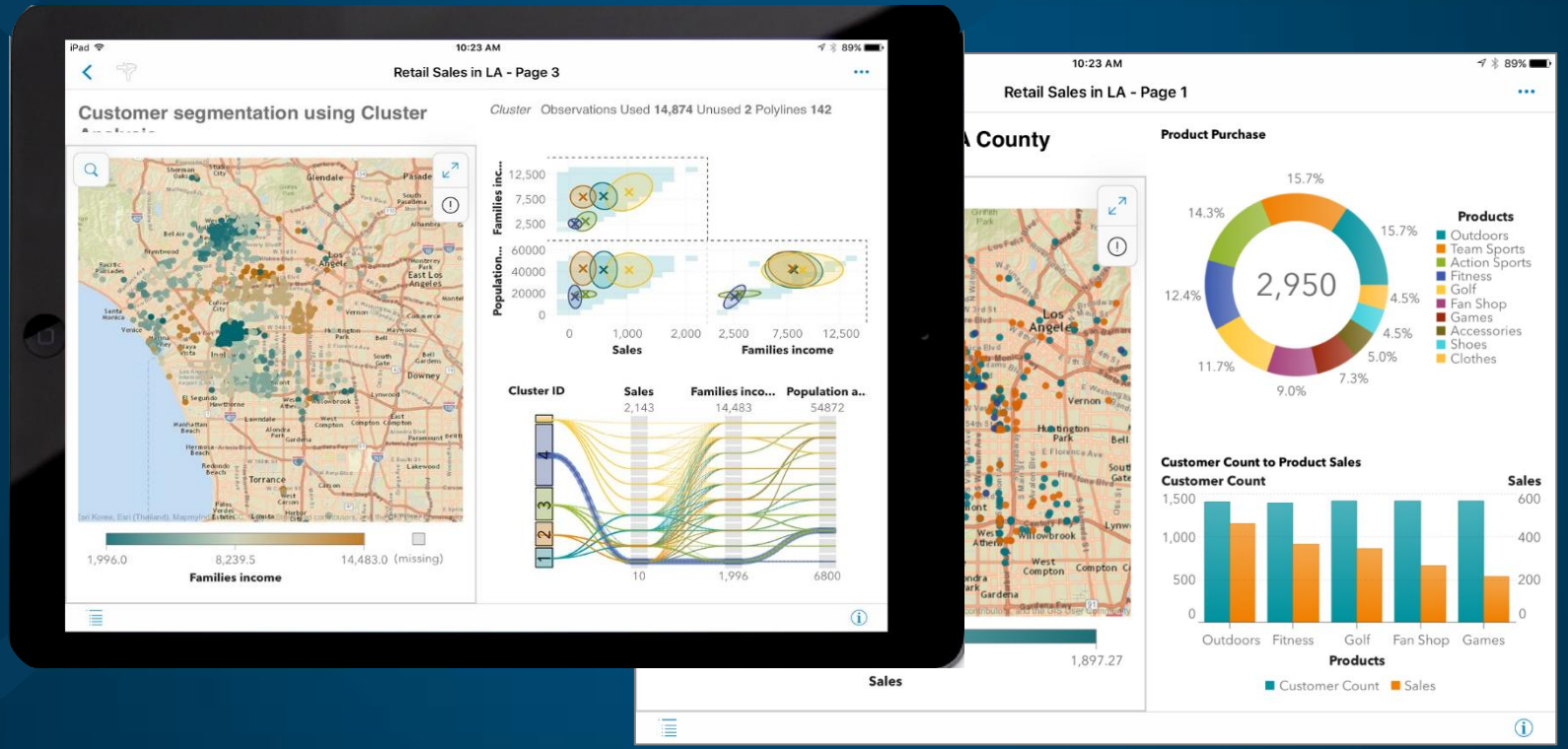
SAS
R
Python
Lua
REST

MPP 분산 서버 In-Memory Engine

- ✓ SQL
- ✓ Data Step
- ✓ Transpose …

Visual Analytics

SAS Viya



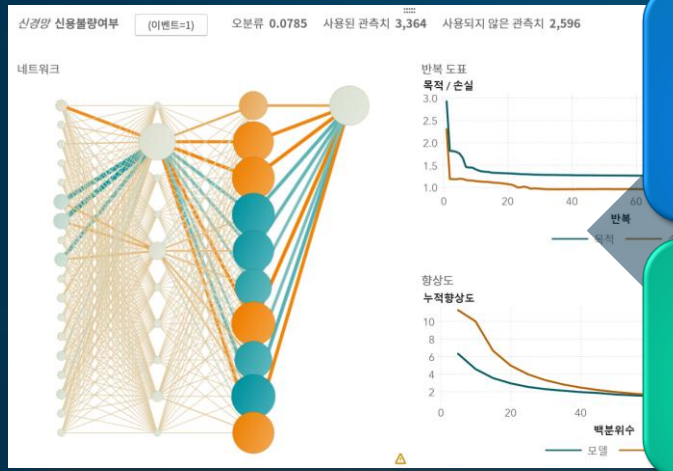
Copyright © SAS Institute Inc. All rights reserved.

Machine Learning / Deep Learning

SAS Viya

Ability to analyze various machine Learning on a single platform

Visual Modeling



- Algorithm : Neural Network
- Hidden Layer : 2

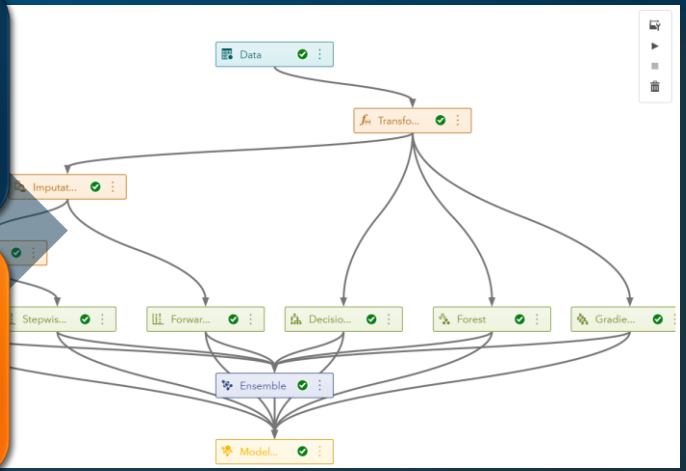
Machine Learning

Deep Learning

Text Analytics

Image Processing

Visual Pipeline

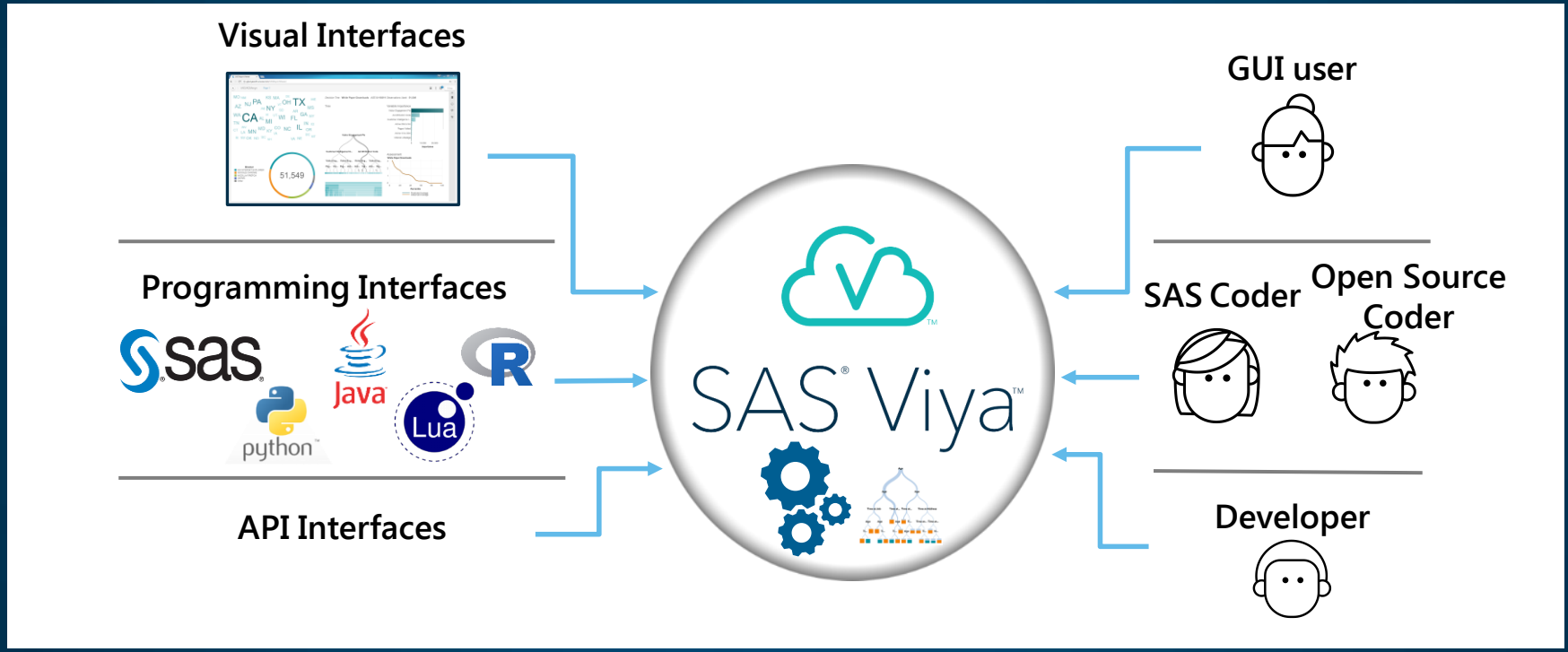


- Model build process
  Feature engineering →Gradient Boosting , Forest
  →Model comparison
- Visual pipeline template

Copyright © SAS Institute Inc. All rights reserved.

Open & Collaboration

API's Expand the SAS ecosystem to other coding languages



Copyright © SAS Institute Inc. All rights reserved.

Viya : 4 types of Deployment

SMP

MPP

Hadoop Co-located

Cloud Native

Copyright © SAS Institute Inc. All rights reserved.

How does SAS Viya solve the issue?

1. Silos

2. Lack of oversight

3. Delay in execution

Copyright © SAS Institute Inc. All rights reserved.

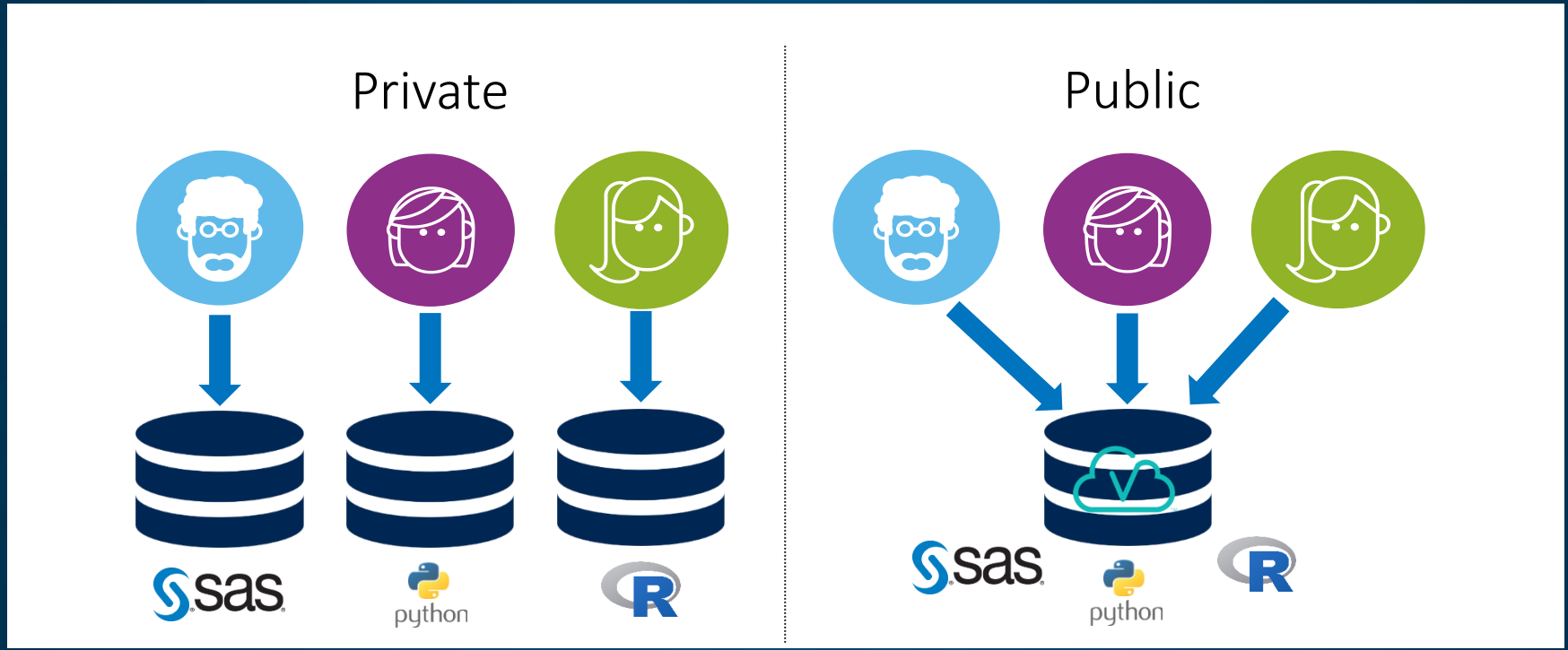
1. Silo --> One User Experience

SAS Viya

Copyright © SAS Institute Inc. All rights reserved.

1. Silo --> Collaboration

SAS Viya



Copyright © SAS Institute Inc. All rights reserved.

sas

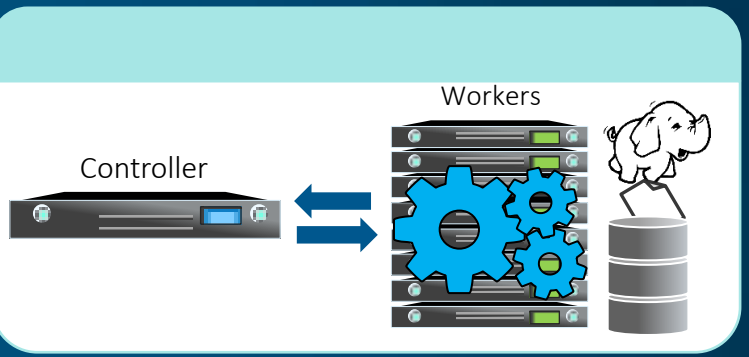
APIs

```
proc print data = x.hmeq (obs = 10);
run;
```

```
df = s.CASTable('hmeq')
df.head(10)
```

```
df <- defCasTable(s, 'hmeq')
head(df, 10)
```

SAS Viya



CAS Action

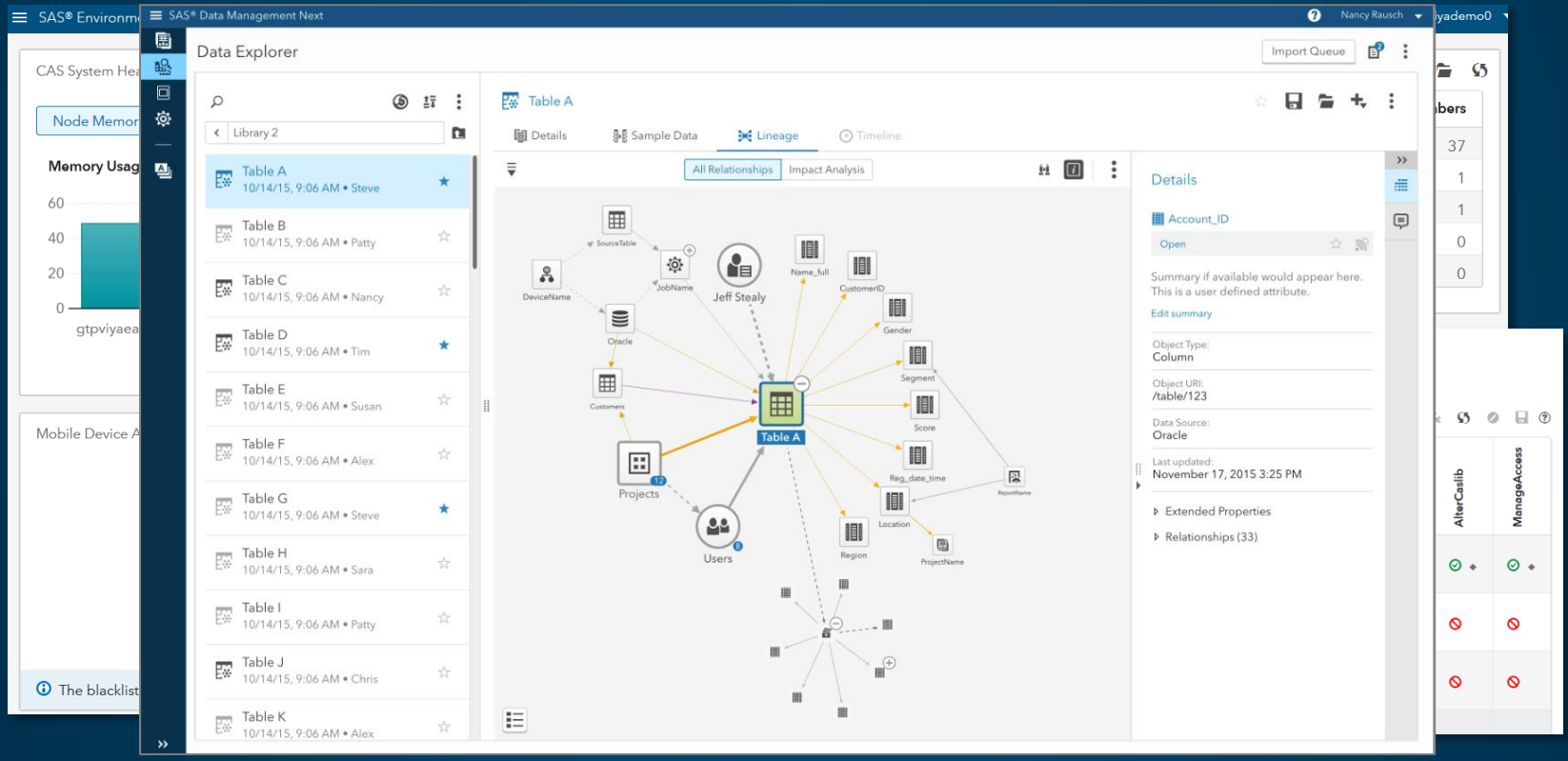
```
[table.fetch]
    table.name = "hmeq"
    from = 1 to = 10
```

Copyright © SAS Institute Inc. All rights reserved.

2. Lack of oversight --> Integration & Governance

SAS Viya

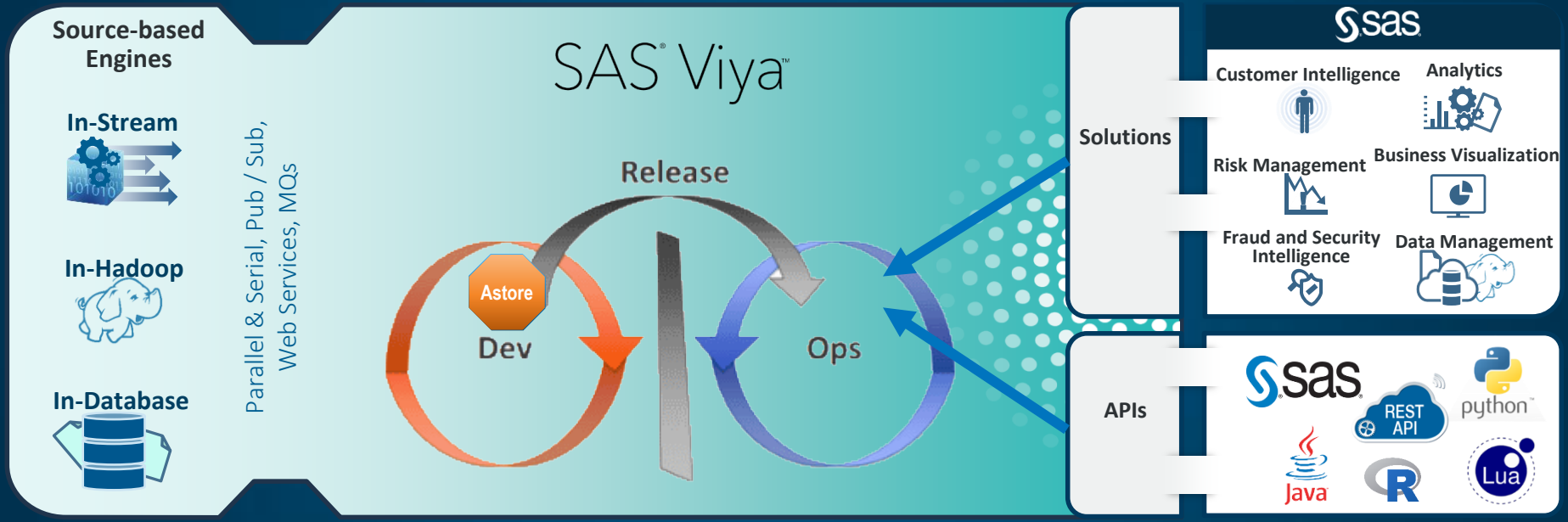
Environment Manager & Lineage



Copyright © SAS Institute Inc. All rights reserved.

3. Fast in execution --> DevOps

SAS Viya

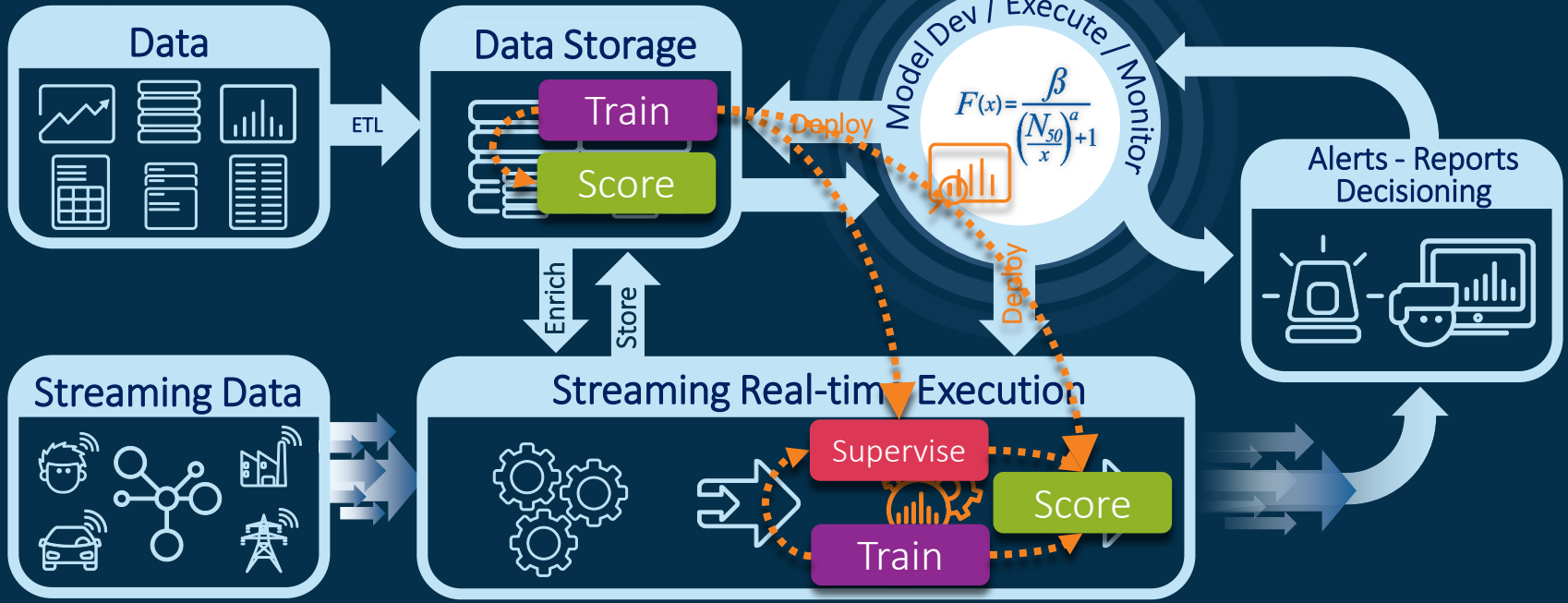


- Performance
- High Quality
- Low cost to deploy
- Time to Value

Copyright © SAS Institute Inc. All rights reserved.

sas

3. Fast in execution --> Streaming Analytics



Copyright © SAS Institute Inc. All rights reserved.

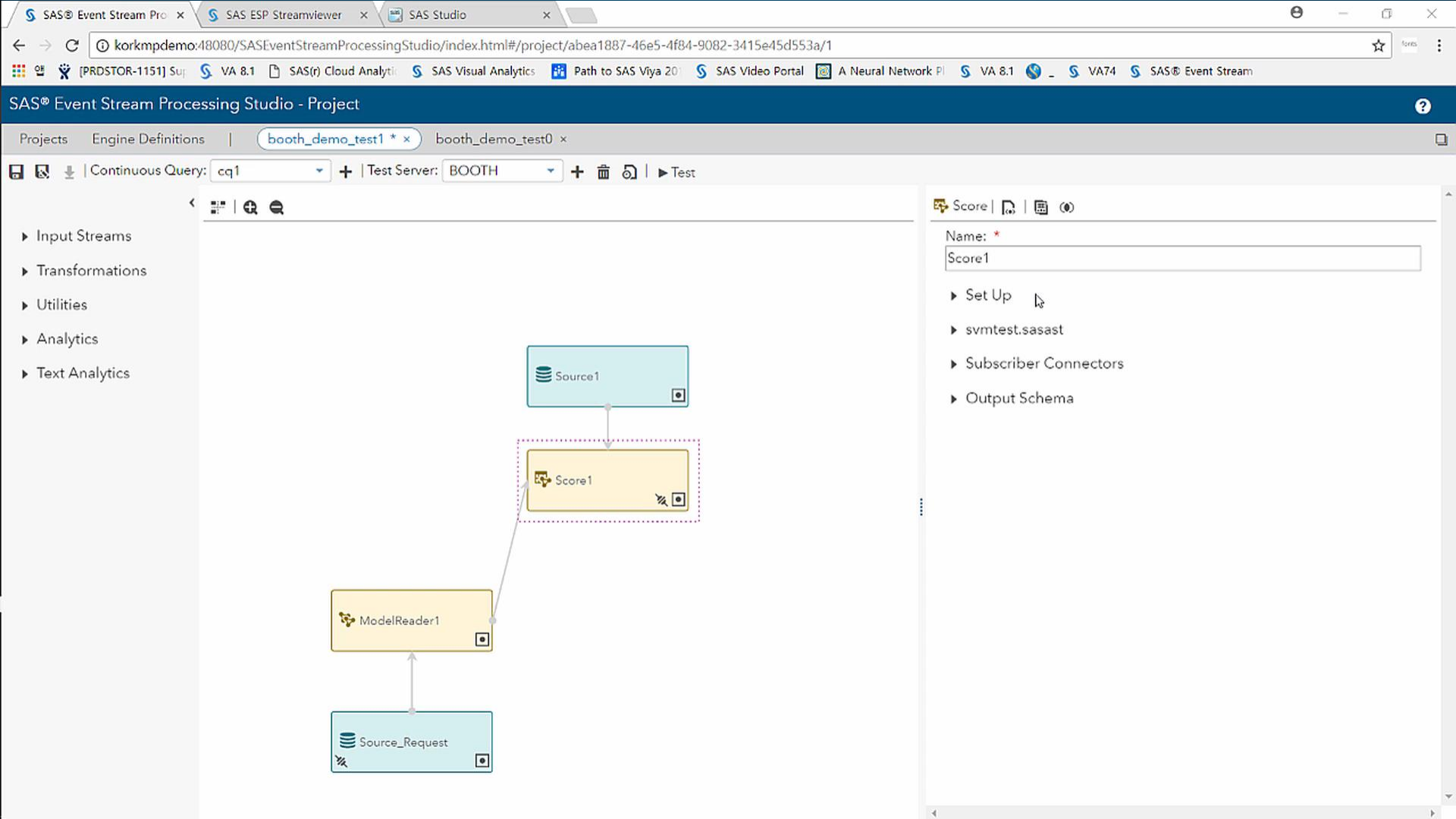
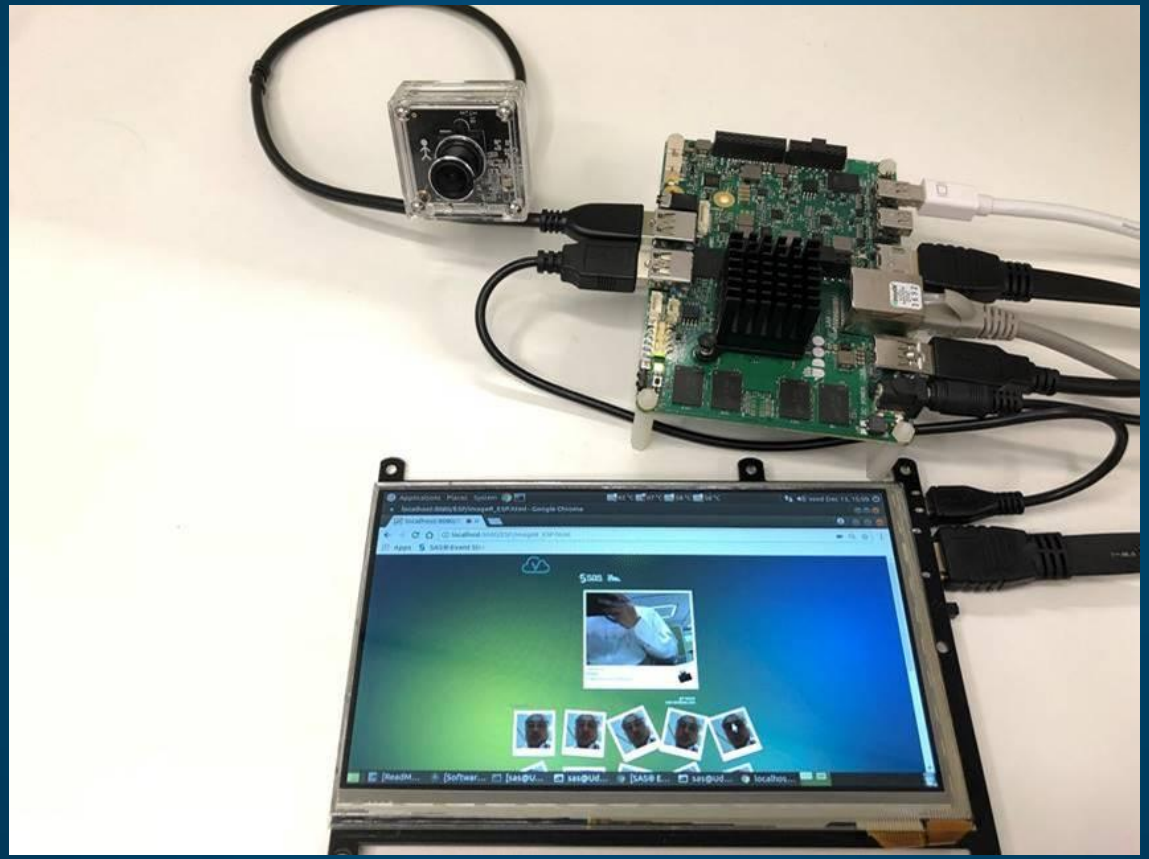

SAS® Event Stream Pro
SAS ESP Streamviewer
SAS Studio
korkmpdemo:48080/SASEventStreamProcessingStudio/index.html#/project/abea1887-46e5-4f84-9082-3415e45d553a/1
SAS® Event Stream Processing Studio - Project
Projects
Engine Definitions
booth_demo_test1 *
booth_demo_test0
Continuous Query: cq1
Test Server: BOOTH
Test
Input Streams
Transformations
Utilities
Analytics
Text Analytics
Source1
Score1
ModelReader1
Source_Request
Score
Name: *
Score1
Set Up
svmtest.sasast
Subscriber Connectors
Output Schema

Image categorization on an edge device



Copyright © SAS Institute Inc. All rights reserved.

3. Fast in execution --> DevOps

SAS Viya

DevOps for Machine Learning

DevOps

Communication

Collaboration

Integration

Agile Development

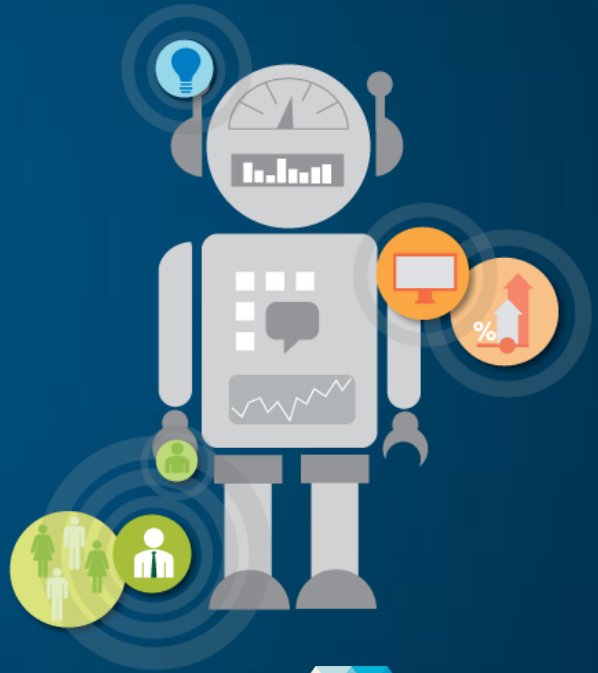
Efficient Operations

Copyright © SAS Institute Inc. All rights reserved.

좋은 머신러닝 플랫폼이 되려면?

What's required to create good machine learning platform?

- ✓ User Environments.
- ✓ Data preparation capabilities.
- ✓ Algorithms – basic and advanced.
- ✓ Performance & Scalability.
- ✓ Governance.
- ✓ Data – Discovery – Deployment (Time to Value)



sas

The SAS Platform supports *diverse* capabilities that enable exemplary *scale* for the enterprise.

SAS is the most *trusted* name in analytics.