Microsoft

Cloud and Machine Learning in real world business

Dae Woo Kim (daewoo.kim@)
Technology Evangelist
Microsoft

http://aka.ms/cs443

Presentation & code

Github repo.

Agenda

Objective

Global cloud service vender

Machine Learning & Hadoop in real world business

Cloud and Machine Learning solution architecture debrief with “code”

Microsoft

Global “Cloud” venders

Azure
AWS
GCP
…

Google Cloud Platform

Microsoft

Azure and AWS


Microsoft
Azure

Azure Momentum

100,000 New Azure customer subscriptions/month	**20 Million** SQL database hours used every day	**> 50 Trillion** Storage objects in Azure
425 Million Azure Active Directory Users	**60 Billion** Hits to Websites run on Azure Web App Service	**57%** Of Fortune 500 Companies use Microsoft Azure
> 5 Trillion Storage transactions every month	**1 Trillion** Messages delivered every month with Event Hubs	**1.4 Million** SQL Databases Deployed In Azure

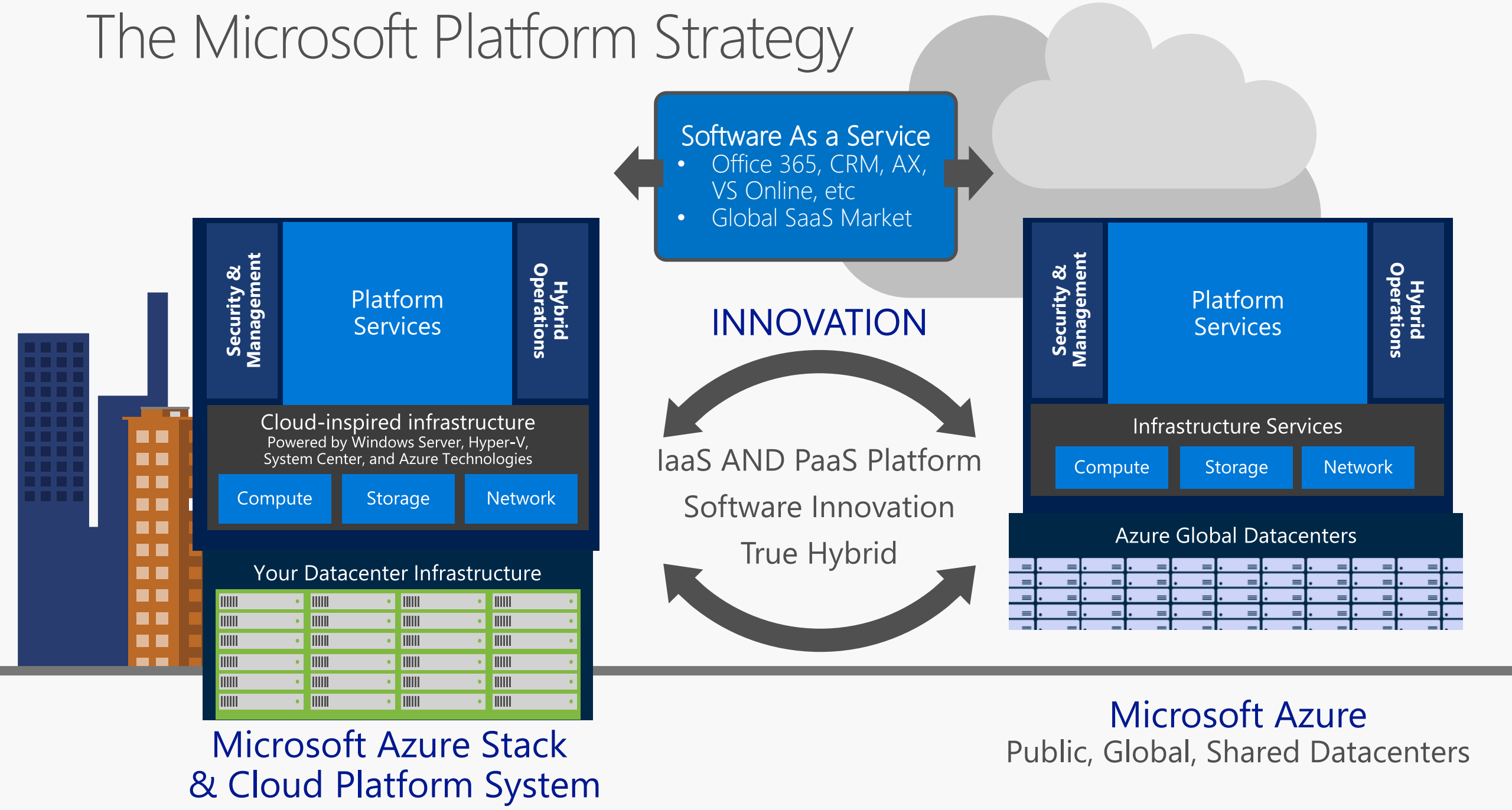
"Microsoft is growing its cloud revenue faster than Amazon" – Business Insider 2016

AWS revenue grew about 69% but Microsoft Azure revenue grew by 127%

>80%

of Fortune 500 use the Microsoft Cloud

The Microsoft Platform Strategy



Microsoft Azure Stack
& Cloud Platform System

Microsoft Azure
Public, Global, Shared Datacenters

What do others say..?

A look inside Gartner Magic Quadrants…

- Microsoft leads in core cloud technologies, IaaS, PaaS, Private and Public Clouds
- Microsoft leads in 19 Gartner MQ's

Topic Area	Magic Quadrant	Last Release	Microsoft
Application Development	Application Development Life Cycle Management	Feb-15	Leader
Business Applications	Business Intelligence and Analytics Platforms	Feb-15	Leader
Software Infrastructure	Client Management Tools	May-15	Leader
Software Infrastructure	Cloud Infrastructure as a Service	May-15	Leader
Business Applications	CRM Customer Engagement Center	Apr-15	Leader
Software Infrastructure	Data Warehouse Database Management Systems	Feb-15	Leader
Software Infrastructure	Enterprise Application Platform as a Service	Mar-15	Leader
Business Applications	Enterprise Content Management	Oct-15	Leader
Software Infrastructure	Horizontal Portals	Sep-15	Leader
Application Development	Integrated Software Quality Suites	Aug-14	Leader
Software Infrastructure	On-premises Application Integration Suites	Jul-14	Leader
Software Infrastructure	Operational Database Management Systems	Oct-15	Leader
Software Infrastructure	Public Cloud Storage Services	Jun-15	Leader
Business Applications	Sales Force Automation	Jul-15	Leader
Software Infrastructure	Secure Email Gateways	Jun-15	Leader
Business Applications	Social Software in the Workplace	Oct-15	Leader
Communications Equipment	Unified Communications	Aug-15	Leader
Software Infrastructure	Web Conferencing	Dec-15	Leader
Software Infrastructure	x86 Server Virtualization Infrastructure	Jul-15	Leader

Magic Quadrant for Cloud Infrastructure as a Service Worldwide (May 2015) LEADER

Magic Quadrant for Public Cloud Storage Services (June 2015) LEADER

Magic Quadrant for Enterprise Application Platform as a Service (March 2015) LEADER

Magic Quadrant for Business Intelligence and Analytics Platforms (Feb 2015) LEADER

Magic Quadrant for X86 Server Virtualization Infrastructure (July 2015) LEADER

Magic Quadrant for Operational Database Management Systems (Oct 2015) LEADER

*Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

The Azure Platform

Platform Services

Security & Management
- Portal
- Azure Active Directory
- Azure AD B2C
- Multi-Factor Authentication
- Automation
- Scheduler
- Key Vault
- Store/ Marketplace
- VM Image Gallery & VM Depot

Services Compute
- Cloud Services
- Service Fabric
- Batch
- RemoteApp

Integration
- Storage Queues
- BizTalk Services
- Hybrid Connections
- Service Bus

Media & CDN
- Media Services
- Content Delivery Network (CDN)

Web and Mobile
- Web Apps
- API Apps
- Mobile Apps
- Logic Apps
- API Management
- Notification Hubs

Developer Services
- Visual Studio
- Azure SDK
- VS Online
- App Insights

Data
- SQL Database
- Data Warehouse
- DocumentDB
- Redis Cache
- Azure Search
- Storage Tables

Analytics & IoT
- HDInsight
- Machine Learning
- Stream Analytics
- Data Lake
- Data Factory
- Event Hubs
- Data Catalog
- IoT Hub
- Mobile Engagement

Hybrid Operations
- Azure AD Health Monitoring
- AD Privileged Identity Management
- Domain Services
- Backup
- Operational Analytics
- Import/Export
- Azure Site Recovery
- StorSimple

Infrastructure Services

OS/Server Compute
- Virtual Machines
- Container Service

Storage
- BLOB Storage
- Azure Files
- Premium Storage

Networking
- Virtual Network
- Load Balancer
- DNS
- Express Route
- Traffic Manager
- VPN Gateway
- App Gateway

Datacenter Infrastructure (30 Regions, 22 Generally Available)


Microsoft
All your choice

Back to the,

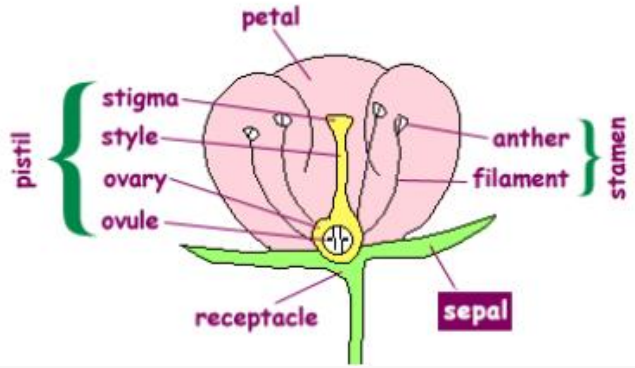
Machine Learning
in real world business

Iris

Iris setosa

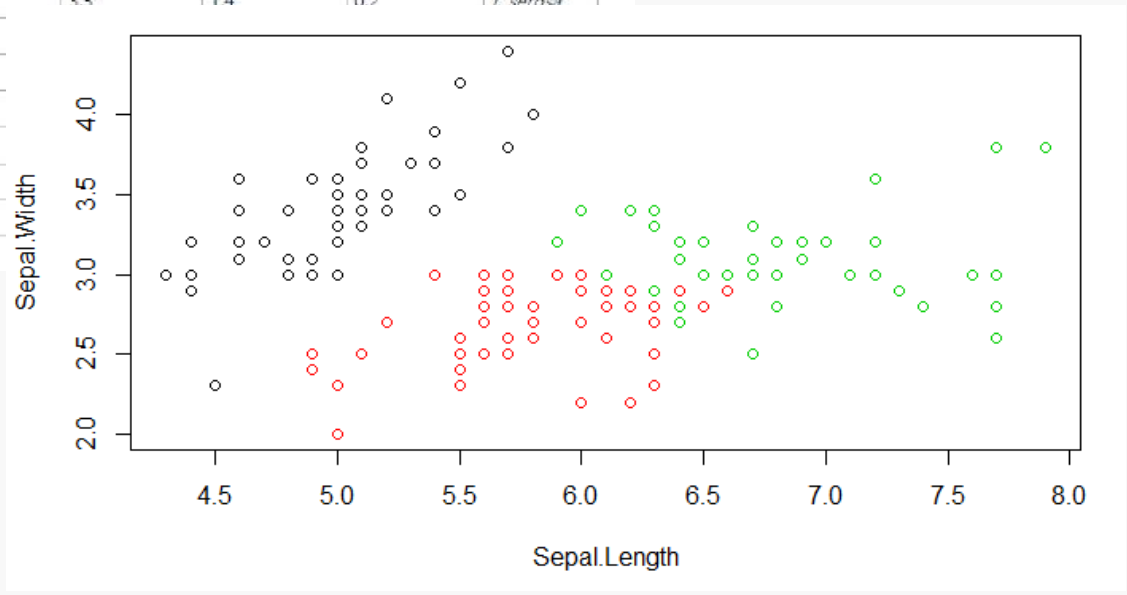
Iris versicolor

Iris virginica



Fisher's *Iris* Data

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	*I. setosa*
4.9				
4.7				
4.6				
5.0				
5.4				
4.6				
5.0				



Machine Learning

Game user churn prediction
In-Game item suggestion

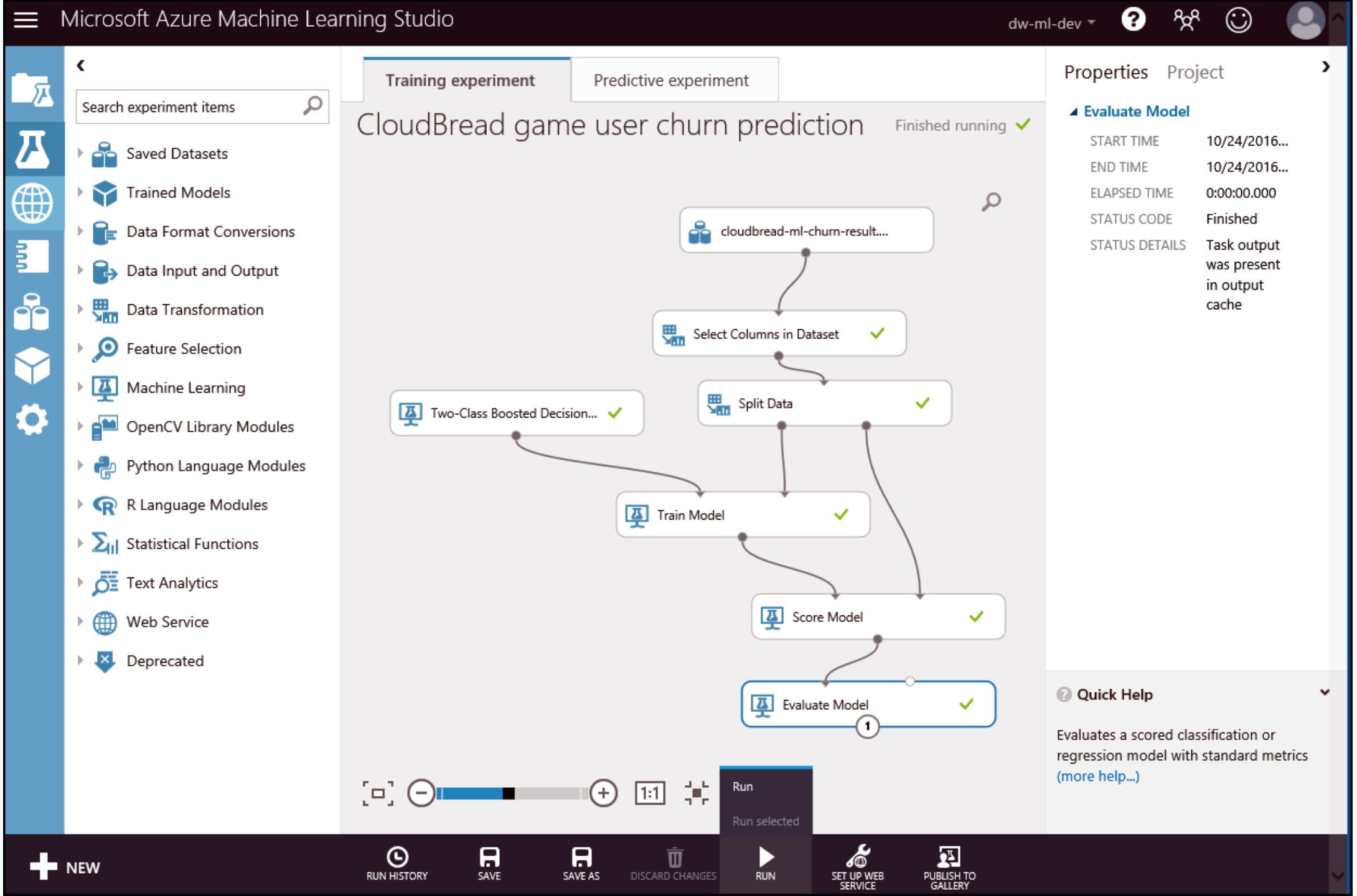
Machine Learning

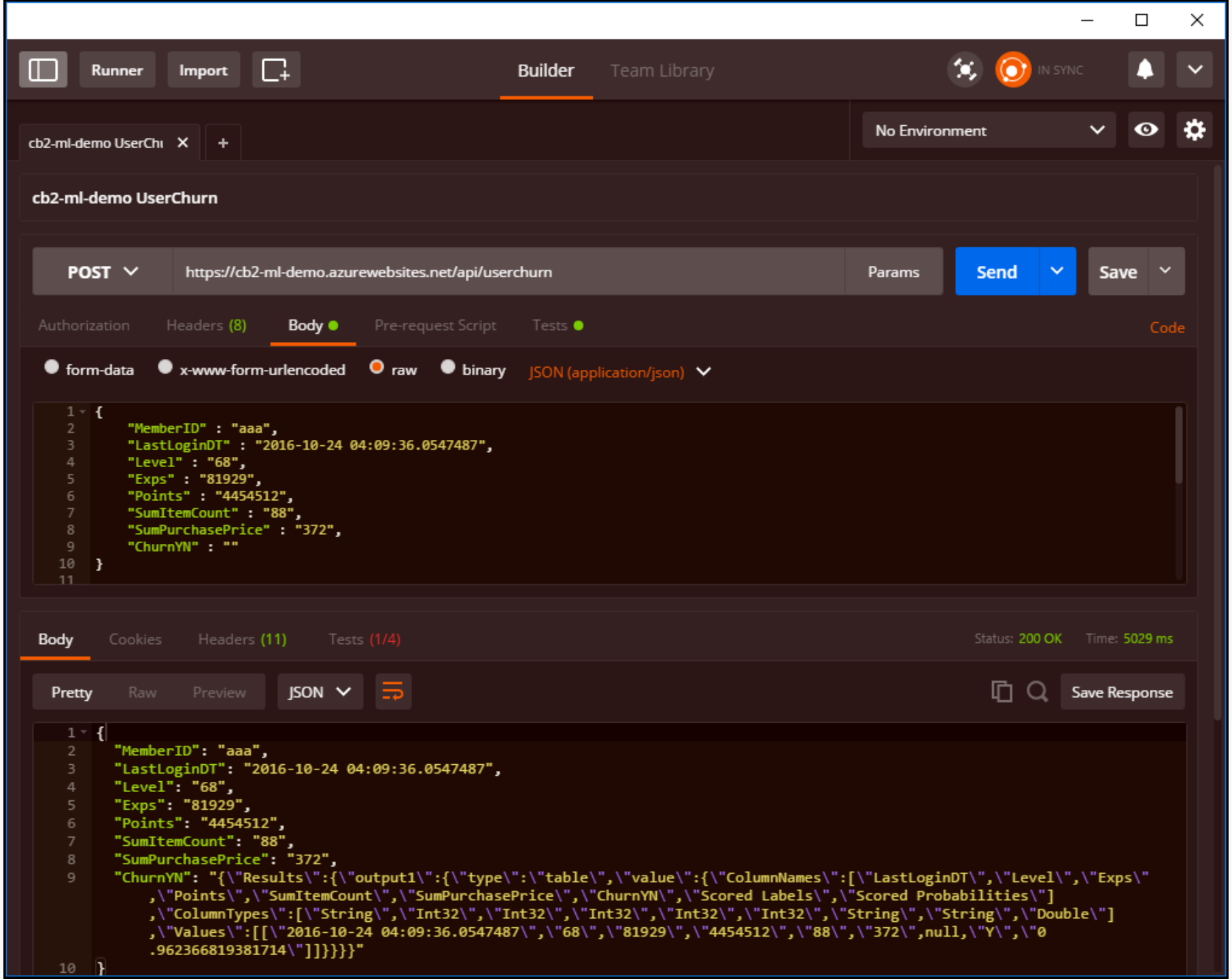
Predict customer churn

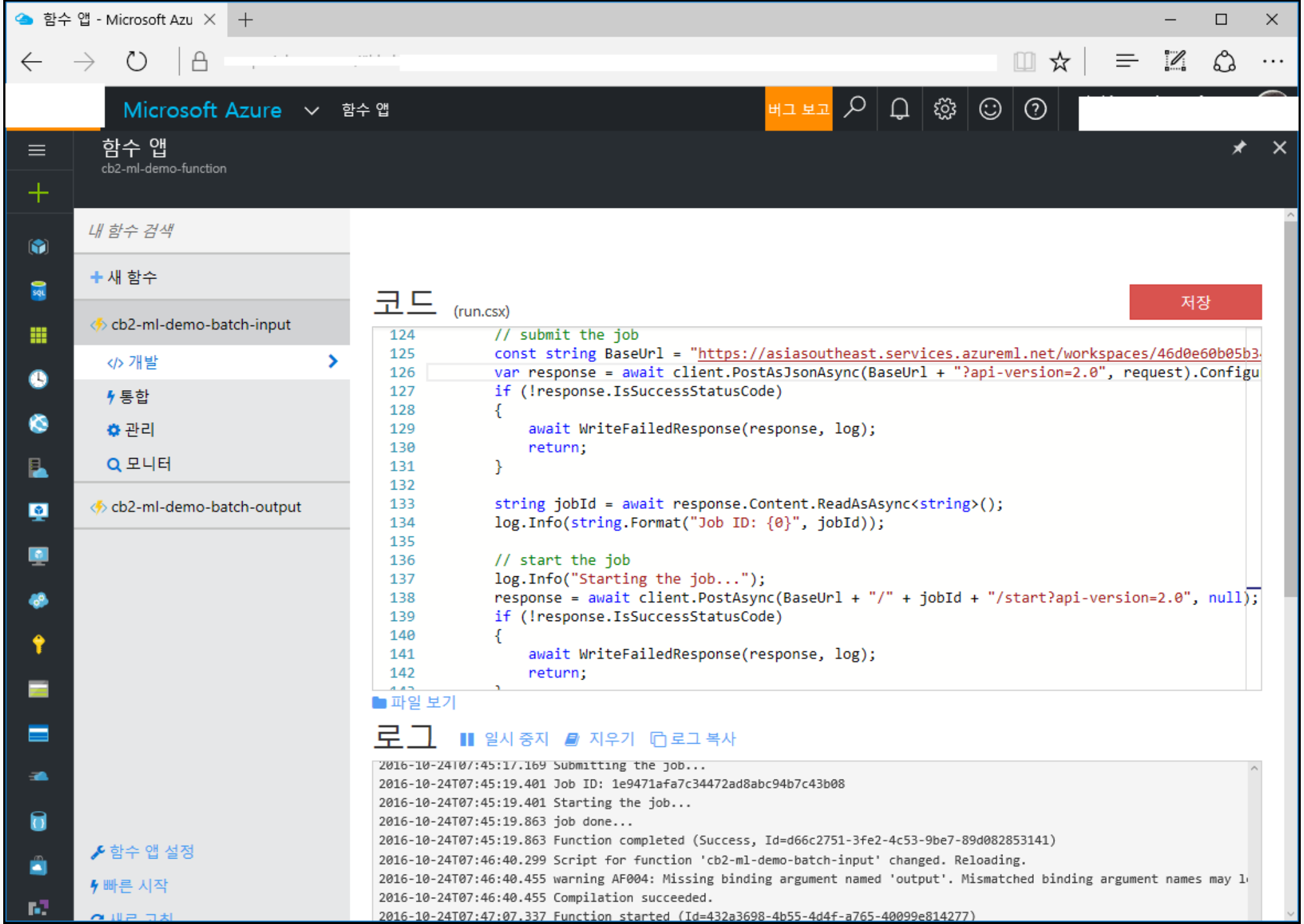
Leave or not

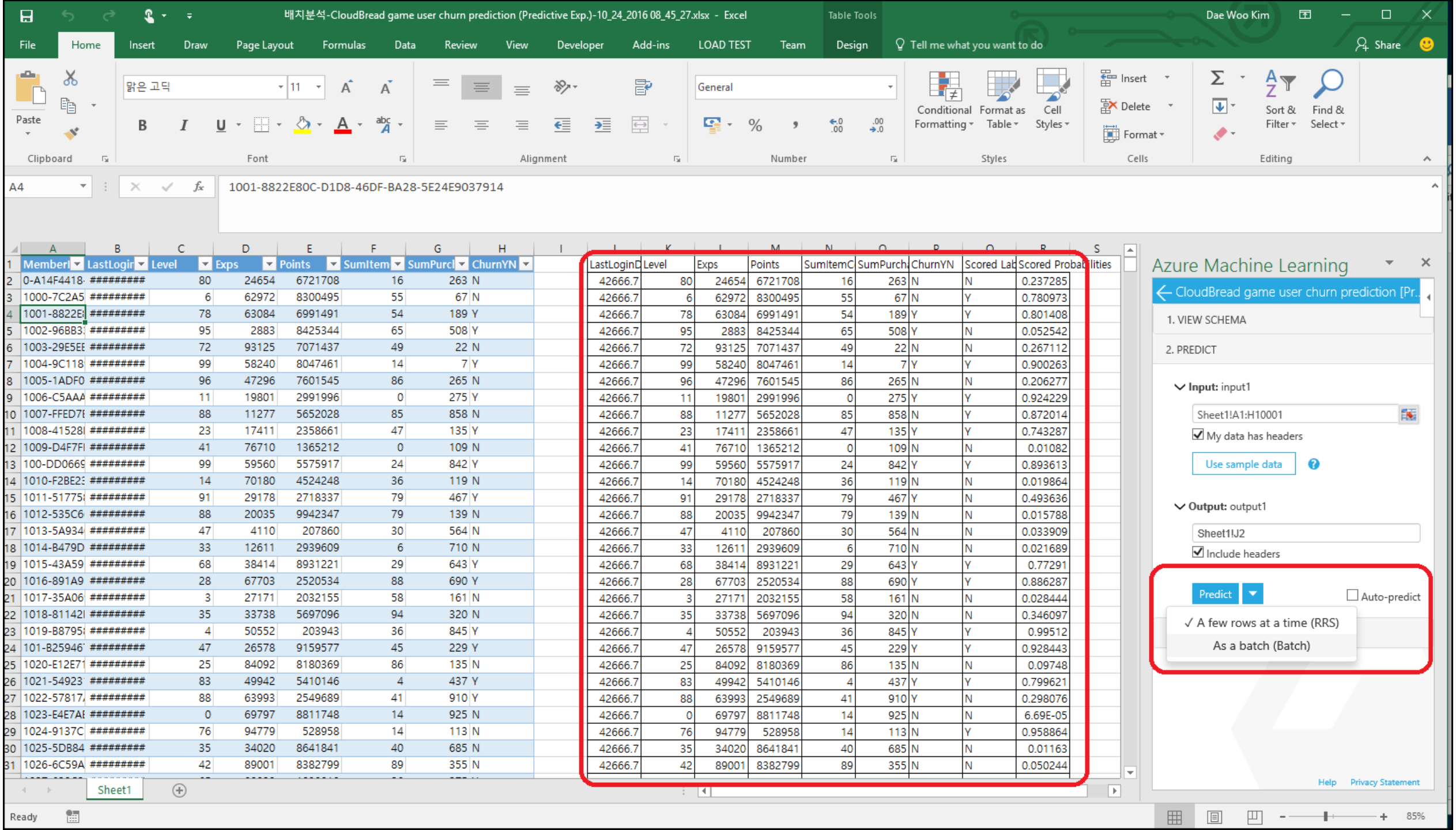
Microsoft

Demo


Microsoft Azure Machine Learning Studio
dw-ml-dev
Search experiment items
Saved Datasets
Trained Models
Data Format Conversions
Data Input and Output
Data Transformation
Feature Selection
Machine Learning
OpenCV Library Modules
Python Language Modules
R Language Modules
Statistical Functions
Text Analytics
Web Service
Deprecated
Training experiment
Predictive experiment
CloudBread game user churn prediction
Finished running
cloudbread-ml-churn-result....
Select Columns in Dataset
Split Data
Two-Class Boosted Decision...
Train Model
Score Model
Evaluate Model
1
Run
Run selected
Properties
Project
Evaluate Model
START TIME 10/24/2016...
END TIME 10/24/2016...
ELAPSED TIME 0:00:00.000
STATUS CODE Finished
STATUS DETAILS Task output was present in output cache
Quick Help
Evaluates a scored classification or regression model with standard metrics
(more help...)
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Runner
Import
Builder
Team Library
IN SYNC
cb2-ml-demo UserChu
No Environment
cb2-ml-demo UserChurn
POST
https://cb2-ml-demo.azurewebsites.net/api/userchurn
Params
Send
Save
Authorization
Headers (8)
Body
Pre-request Script
Tests
Code
form-data
x-www-form-urlencoded
raw
binary
JSON (application/json)
{
"MemberID" : "aaa",
"LastLoginDT" : "2016-10-24 04:09:36.0547487",
"Level" : "68",
"Exps" : "81929",
"Points" : "4454512",
"SumItemCount" : "88",
"SumPurchasePrice" : "372",
"ChurnYN" : ""
}
Body
Cookies
Headers (11)
Tests (1/4)
Status: 200 OK
Time: 5029 ms
Pretty
Raw
Preview
JSON
Save Response
{
"MemberID": "aaa",
"LastLoginDT": "2016-10-24 04:09:36.0547487",
"Level": "68",
"Exps": "81929",
"Points": "4454512",
"SumItemCount": "88",
"SumPurchasePrice": "372",
"ChurnYN": "{\"Results\":{\"output1\":{\"type\":\"table\",\"value\":{\"ColumnNames\":[\"LastLoginDT\",\"Level\",\"Exps\"
,\"Points\",\"SumItemCount\",\"SumPurchasePrice\",\"ChurnYN\",\"Scored Labels\",\"Scored Probabilities\"]
,\"ColumnTypes\":[\"String\",\"Int32\",\"Int32\",\"Int32\",\"Int32\",\"Int32\",\"String\",\"String\",\"Double\"]
,\"Values\":[[\"2016-10-24 04:09:36.0547487\",\"68\",\"81929\",\"4454512\",\"88\",\"372\",null,\"Y\",\"0
.962366819381714\"]]}}}}"
}


함수 앱 - Microsoft Azu
Microsoft Azure
함수 앱
버그 보고
함수 앱
cb2-ml-demo-function
내 함수 검색
새 함수
cb2-ml-demo-batch-input
개발
통합
관리
모니터
cb2-ml-demo-batch-output
함수 앱 설정
빠른 시작
코드
(run.csx)
저장
124 // submit the job
125 const string BaseUrl = "https://asiasoutheast.services.azureml.net/workspaces/46d0e60b05b3
126 var response = await client.PostAsJsonAsync(BaseUrl + "?api-version=2.0", request).Configu
127 if (!response.IsSuccessStatusCode)
128 {
129 await WriteFailedResponse(response, log);
130 return;
131 }
132
133 string jobId = await response.Content.ReadAsAsync<string>();
134 log.Info(string.Format("Job ID: {0}", jobId));
135
136 // start the job
137 log.Info("Starting the job...");
138 response = await client.PostAsync(BaseUrl + "/" + jobId + "/start?api-version=2.0", null);
139 if (!response.IsSuccessStatusCode)
140 {
141 await WriteFailedResponse(response, log);
142 return;
파일 보기
로그
일시 중지
지우기
로그 복사
2016-10-24T07:45:17.169 Submitting the job...
2016-10-24T07:45:19.401 Job ID: 1e9471afa7c34472ad8abc94b7c43b08
2016-10-24T07:45:19.401 Starting the job...
2016-10-24T07:45:19.863 job done...
2016-10-24T07:45:19.863 Function completed (Success, Id=d66c2751-3fe2-4c53-9be7-89d082853141)
2016-10-24T07:46:40.299 Script for function 'cb2-ml-demo-batch-input' changed. Reloading.
2016-10-24T07:46:40.455 warning AF004: Missing binding argument named 'output'. Mismatched binding argument names may l
2016-10-24T07:46:40.455 Compilation succeeded.
2016-10-24T07:47:07.337 Function started (Id=432a3698-4b55-4d4f-a765-40099e814277)


배치분석-CloudBread game user churn prediction (Predictive Exp.)-10_24_2016 08_45_27.xlsx - Excel
Dae Woo Kim
File Home Insert Draw Page Layout Formulas Data Review View Developer Add-ins LOAD TEST Team Design
Table Tools
Tell me what you want to do
Share
Paste
Clipboard
맑은 고딕
11
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Sort & Filter
Find & Select
Editing
A4
1001-8822E80C-D1D8-46DF-BA28-5E24E9037914
Member
LastLogin
Level
Exps
Points
SumItem
SumPurcl
ChurnYN
0-A14F4418 ######### 80 24654 6721708 16 263 N
1000-7C2A5 ######### 6 62972 8300495 55 67 N
1001-8822E ######### 78 63084 6991491 54 189 Y
1002-96BB3 ######### 95 2883 8425344 65 508 Y
1003-29E5E ######### 72 93125 7071437 49 22 N
1004-9C118 ######### 99 58240 8047461 14 7 Y
1005-1ADF0 ######### 96 47296 7601545 86 265 N
1006-C5AAA ######### 11 19801 2991996 0 275 Y
1007-FFED7 ######### 88 11277 5652028 85 858 N
1008-41528 ######### 23 17411 2358661 47 135 Y
1009-D4F7F ######### 41 76710 1365212 0 109 N
100-DD066 ######### 99 59560 5575917 24 842 Y
1010-F2BE2 ######### 14 70180 4524248 36 119 N
1011-51775 ######### 91 29178 2718337 79 467 Y
1012-535C6 ######### 88 20035 9942347 79 139 N
1013-5A934 ######### 47 4110 207860 30 564 N
1014-B479D ######### 33 12611 2939609 6 710 N
1015-43A59 ######### 68 38414 8931221 29 643 Y
1016-891A9 ######### 28 67703 2520534 88 690 Y
1017-35A06 ######### 3 27171 2032155 58 161 N
1018-81142 ######### 35 33738 5697096 94 320 N
1019-B8795 ######### 4 50552 203943 36 845 Y
101-B25946 ######### 47 26578 9159577 45 229 Y
1020-E12E7 ######### 25 84092 8180369 86 135 N
1021-54923 ######### 83 49942 5410146 4 437 Y
1022-57817 ######### 88 63993 2549689 41 910 Y
1023-E4E7AE ######### 0 69797 8811748 14 925 N
1024-9137C ######### 76 94779 528958 14 113 N
1025-5DB84 ######### 35 34020 8641841 40 685 N
1026-6C59A ######### 42 89001 8382799 89 355 N
LastLoginD Level Exps Points SumItemC SumPurch ChurnYN Scored Lab Scored Probabilities
42666.7 80 24654 6721708 16 263 N N 0.237285
42666.7 6 62972 8300495 55 67 N Y 0.780973
42666.7 78 63084 6991491 54 189 Y Y 0.801408
42666.7 95 2883 8425344 65 508 Y N 0.052542
42666.7 72 93125 7071437 49 22 N N 0.267112
42666.7 99 58240 8047461 14 7 Y Y 0.900263
42666.7 96 47296 7601545 86 265 N N 0.206277
42666.7 11 19801 2991996 0 275 Y Y 0.924229
42666.7 88 11277 5652028 85 858 N Y 0.872014
42666.7 23 17411 2358661 47 135 Y Y 0.743287
42666.7 41 76710 1365212 0 109 N N 0.01082
42666.7 99 59560 5575917 24 842 Y Y 0.893613
42666.7 14 70180 4524248 36 119 N N 0.019864
42666.7 91 29178 2718337 79 467 Y N 0.493636
42666.7 88 20035 9942347 79 139 N N 0.015788
42666.7 47 4110 207860 30 564 N N 0.033909
42666.7 33 12611 2939609 6 710 N N 0.021689
42666.7 68 38414 8931221 29 643 Y Y 0.77291
42666.7 28 67703 2520534 88 690 Y Y 0.886287
42666.7 3 27171 2032155 58 161 N N 0.028444
42666.7 35 33738 5697096 94 320 N N 0.346097
42666.7 4 50552 203943 36 845 Y Y 0.99512
42666.7 47 26578 9159577 45 229 Y Y 0.928443
42666.7 25 84092 8180369 86 135 N N 0.09748
42666.7 83 49942 5410146 4 437 Y Y 0.799621
42666.7 88 63993 2549689 41 910 Y N 0.298076
42666.7 0 69797 8811748 14 925 N N 6.69E-05
42666.7 76 94779 528958 14 113 N Y 0.958864
42666.7 35 34020 8641841 40 685 N N 0.01163
42666.7 42 89001 8382799 89 355 N N 0.050244
Sheet1
Ready
Azure Machine Learning
CloudBread game user churn prediction [Pr...
1. VIEW SCHEMA
2. PREDICT
Input: input1
Sheet1!A1:H10001
My data has headers
Use sample data
Output: output1
Sheet1!J2
Include headers
Predict
Auto-predict
✓ A few rows at a time (RRS)
As a batch (Batch)
Help
Privacy Statement
85%

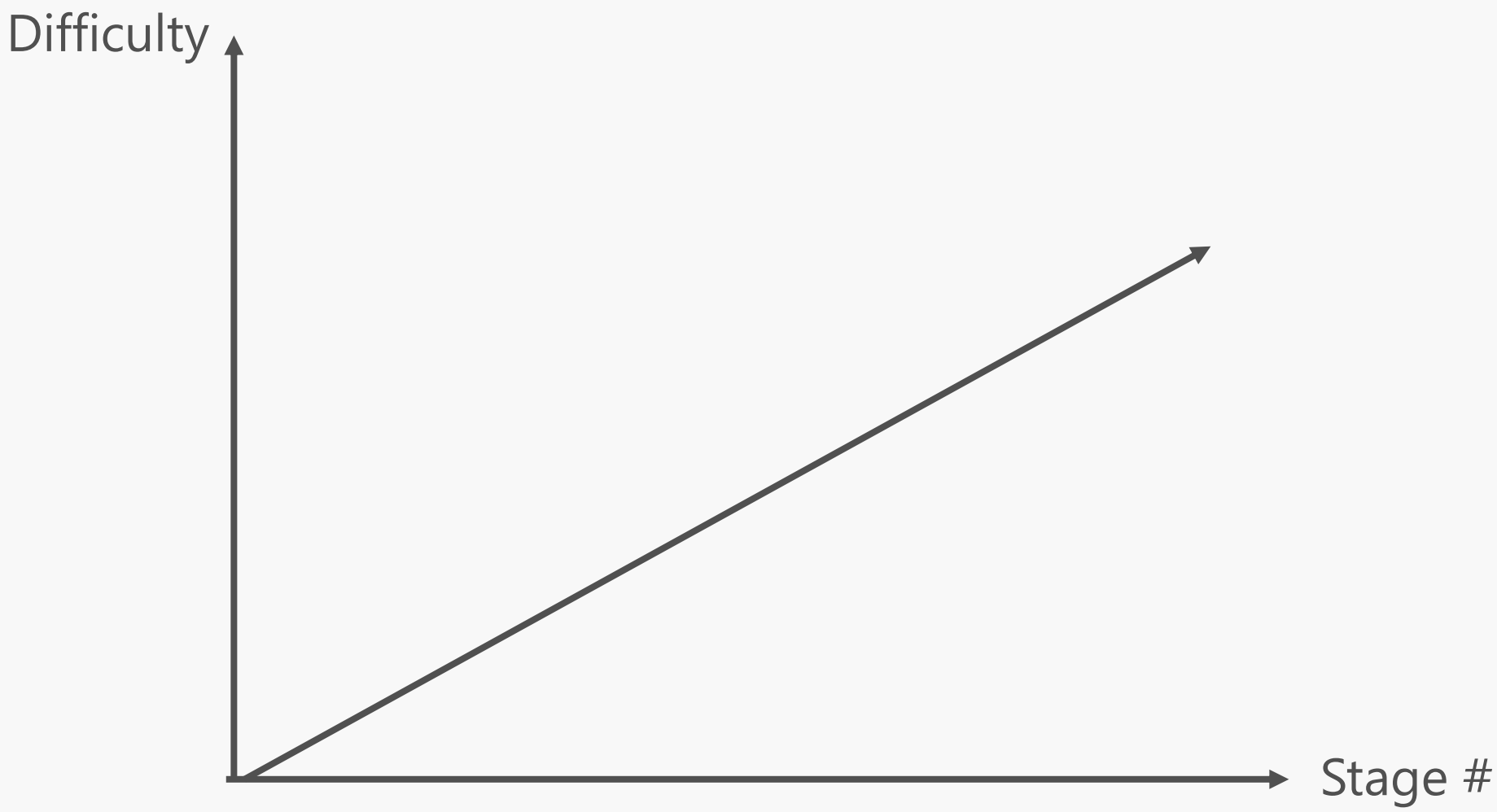
Machine Learning

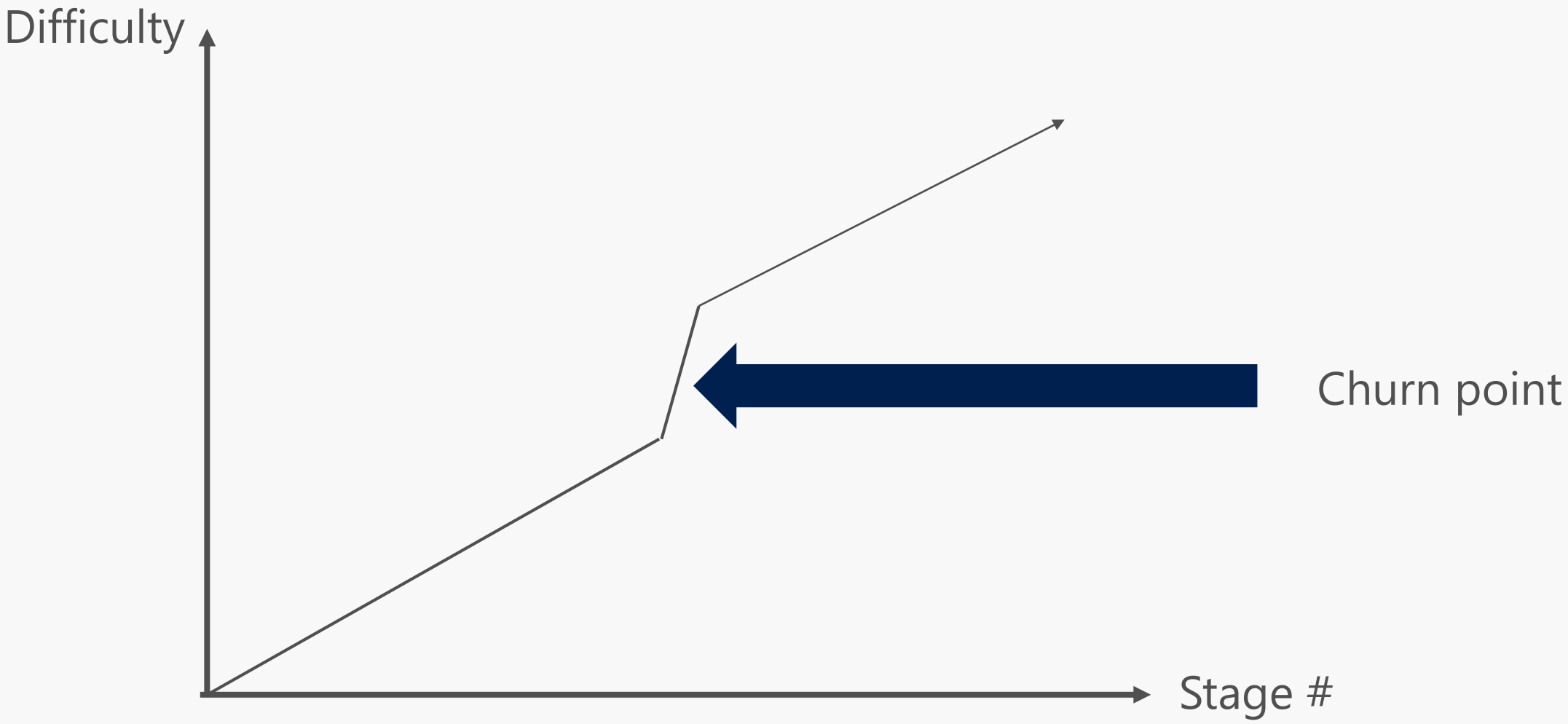
Predict product suggestion

Promote personalized experience

Game level design

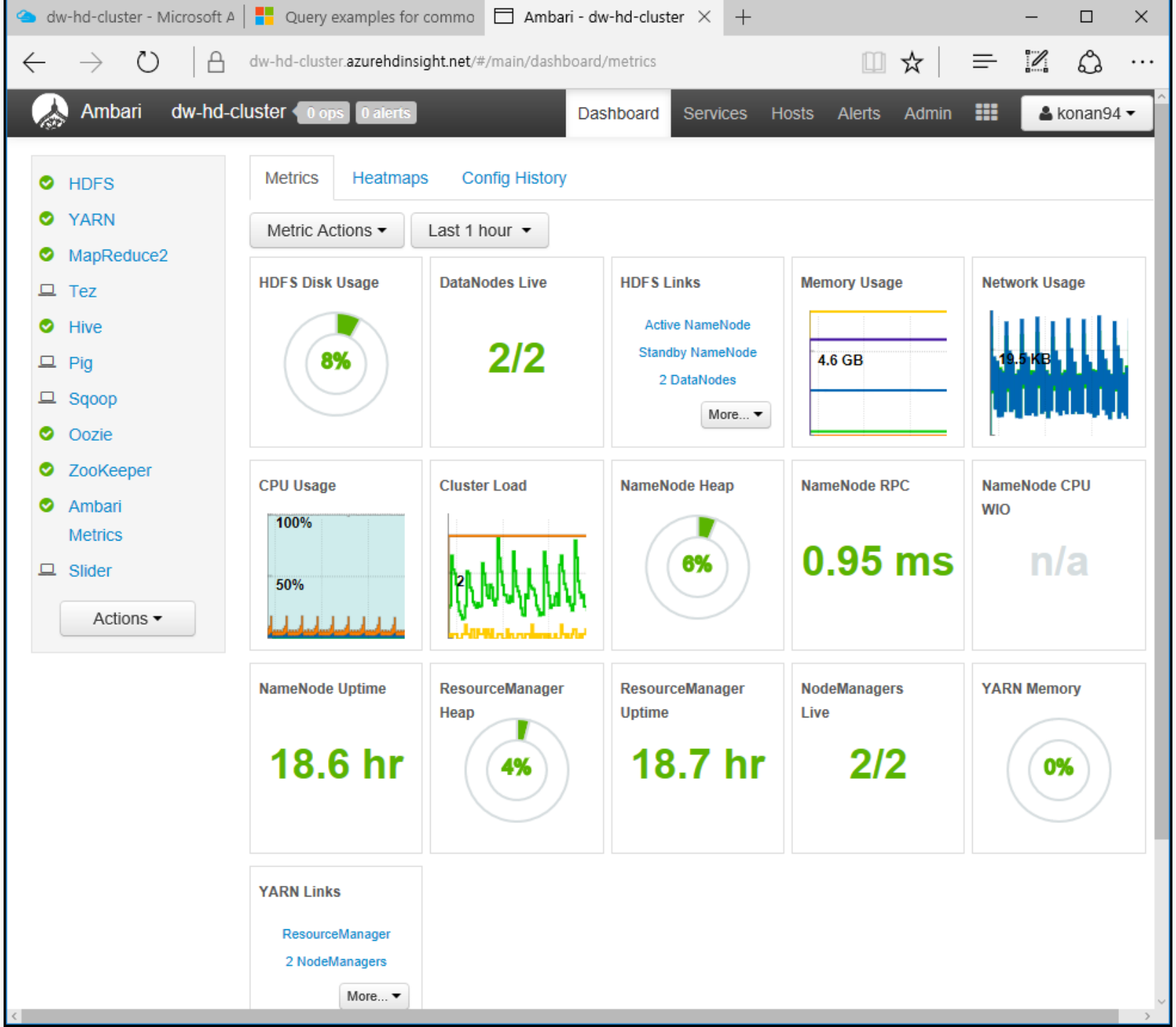
/w big data analytics, Hadoop - HDInsight

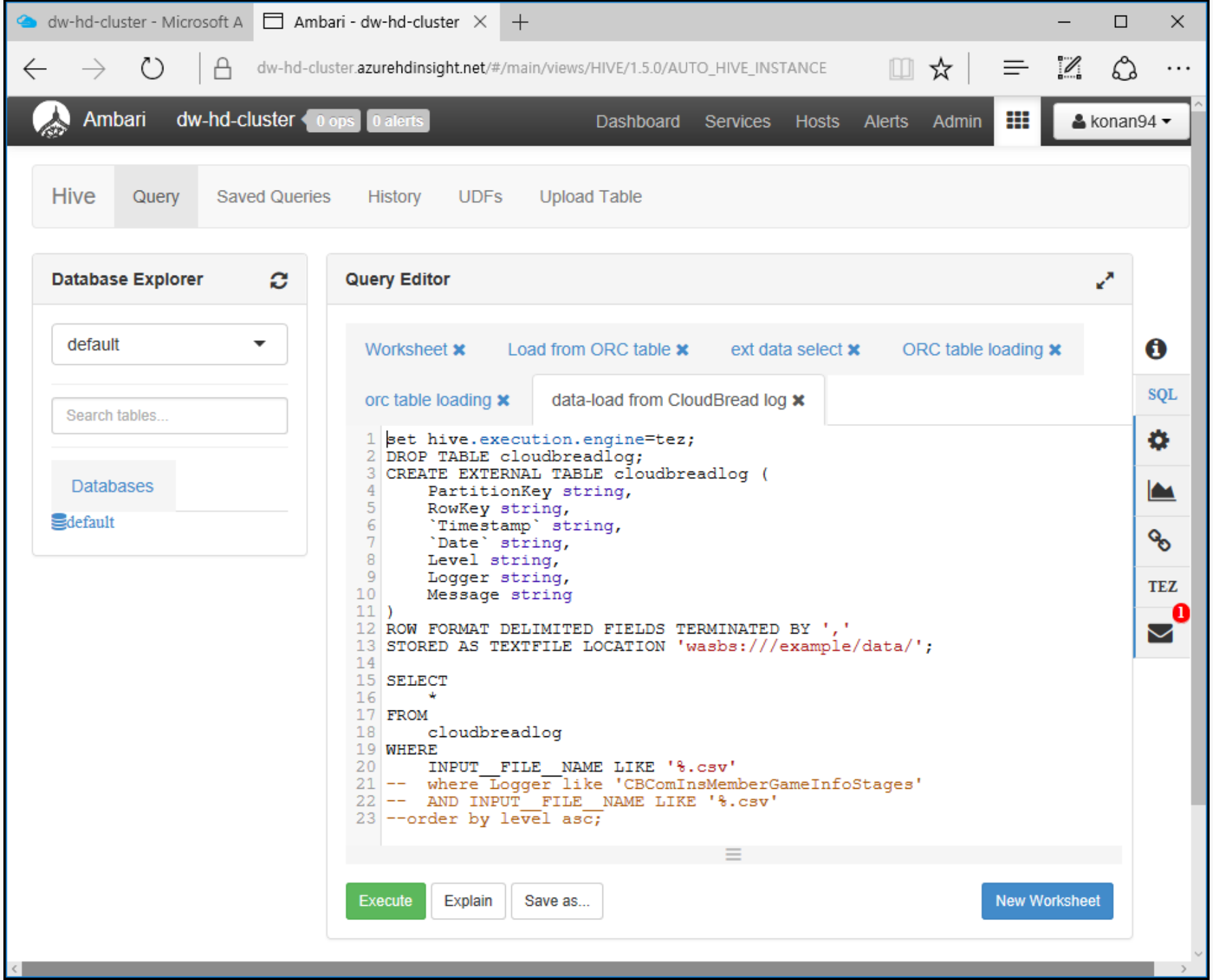

Difficulty
Stage #


Difficulty
Churn point
Stage #

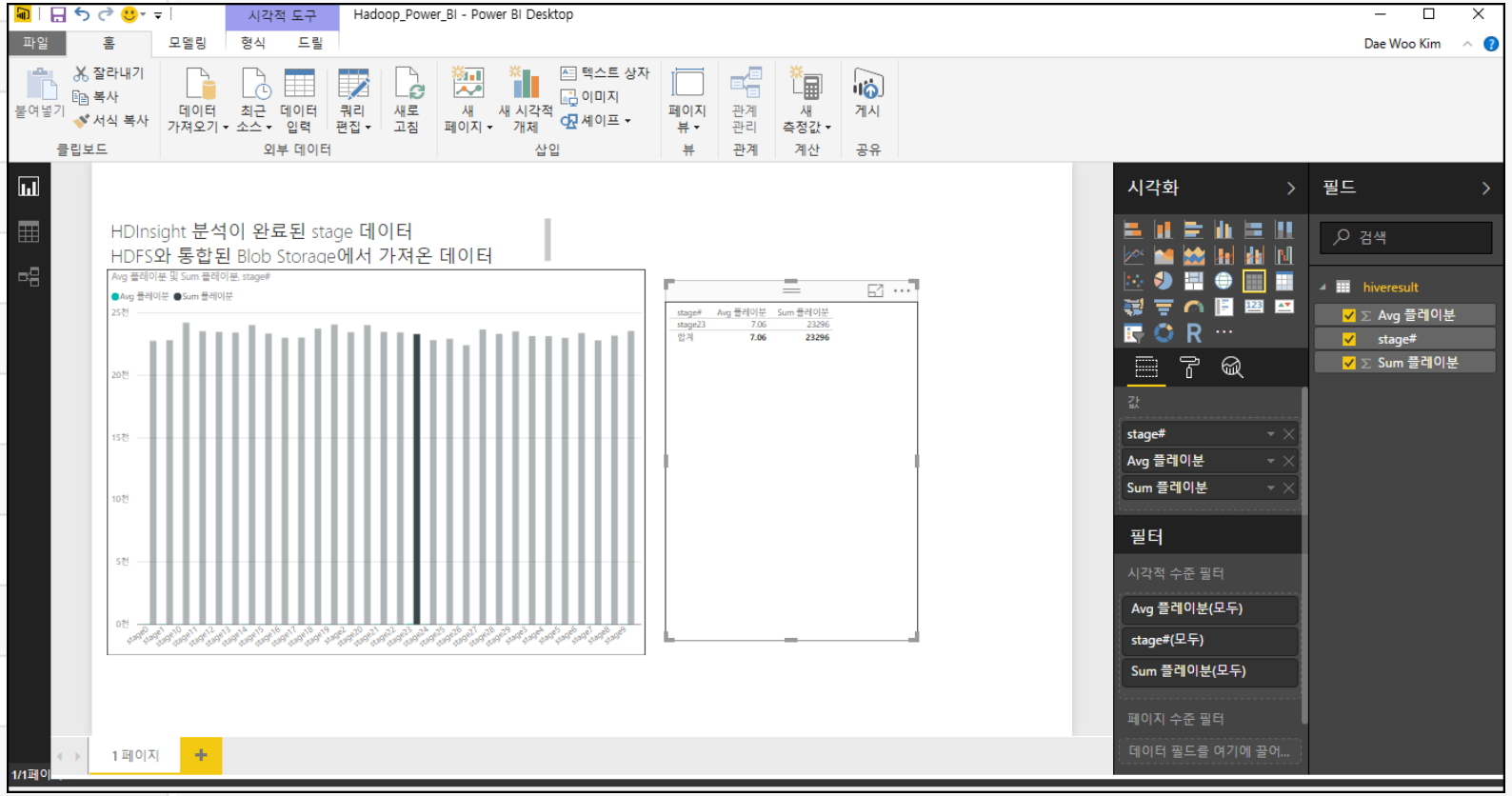
Microsoft

Demo


dw-hd-cluster - Microsoft A
Query examples for commo
Ambari - dw-hd-cluster
dw-hd-cluster.azurehdinsight.net/#/main/dashboard/metrics
Ambari
dw-hd-cluster
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
konan94
HDFS
YARN
MapReduce2
Tez
Hive
Pig
Sqoop
Oozie
ZooKeeper
Ambari Metrics
Slider
Actions
Metrics
Heatmaps
Config History
Metric Actions
Last 1 hour
HDFS Disk Usage
8%
DataNodes Live
2/2
HDFS Links
Active NameNode
Standby NameNode
2 DataNodes
More...
Memory Usage
4.6 GB
Network Usage
19.5 KB
CPU Usage
100%
50%
Cluster Load
2
NameNode Heap
6%
NameNode RPC
0.95 ms
NameNode CPU WIO
n/a
NameNode Uptime
18.6 hr
ResourceManager Heap
4%
ResourceManager Uptime
18.7 hr
NodeManagers Live
2/2
YARN Memory
0%
YARN Links
ResourceManager
2 NodeManagers
More...


dw-hd-cluster - Microsoft A
Ambari - dw-hd-cluster
dw-hd-cluster.azurehdinsight.net/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Ambari
dw-hd-cluster
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
konan94
Hive
Query
Saved Queries
History
UDFs
Upload Table
Database Explorer
default
Search tables...
Databases
default
Query Editor
Worksheet
Load from ORC table
ext data select
ORC table loading
orc table loading
data-load from CloudBread log
set hive.execution.engine=tez;
DROP TABLE cloudbreadlog;
CREATE EXTERNAL TABLE cloudbreadlog (
PartitionKey string,
RowKey string,
`Timestamp` string,
`Date` string,
Level string,
Logger string,
Message string
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE LOCATION 'wasbs:///example/data/';
SELECT
*
FROM
cloudbreadlog
WHERE
INPUT__FILE__NAME LIKE '%.csv'
-- where Logger like 'CBComInsMemberGameInfoStages'
-- AND INPUT__FILE__NAME LIKE '%.csv'
--order by level asc;
SQL
TEZ
Execute
Explain
Save as...
New Worksheet

level	stage_play_avg_min	stage_play_sum_min
stage0	7.032447466	22757
stage1	6.923216995	22812
stage10	6.990762125	24216
stage11	6.962710861	23527
stage12	6.99791294	23471
stage13	7.04329525	23426
stage14	7.060240964	24026
stage15	7.005102041	23341
stage16	6.991793313	23003
stage17	6.954984894	23021
stage18	7.089904421	23737
stage19	7.13037037	24065
stage2	7.110673135	23451



Microsoft

Demo
how it works, w/ code

Q & A

Microsoft

Cloud and Machine Learning in real world business

Dae Woo Kim (daewoo.kim@)
Technology Evangelist
Microsoft