ECV-BD-301

Build recommendation engine with Amazon Machine Learning in S3

2016.11.03

Version 1.2

Agenda

Introduction

Overview

In this lab, you will build a smart solution using Amazon Redshift and Amazon Machine Learning that predict rental bikes for Capital bikeshare system.

The dataset contains daily amount of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information.

You will learn how to use Redshift and predict using Machine Learning to create a model that will predict the rental bikes.

Topics covered

By the end of this lab, you will be able to:

- Create a Redshift cluster, build Redshift table and load data from Amazon S3.
- Create a Machine Learning Model
- Train the Machine Learning Model, using historic data about rental bikes.
- Predict the rental amount for the future sharebike system with Redshift and Amazon Machine Learning

S3

What is S3?

Amazon S3 is storage for the Internet. It's a simple storage service that offers software developers a highly-scalable, reliable, and low-latency data storage infrastructure at very low costs.

Amazon Machine Learning

What is Amazon Machine Learning?

Amazon Machine Learning is a machine service that allows you to easily build

predictive applications, including fraud detection, demand forecasting, and click prediction. Amazon Machine Learning uses powerful algorithms that can help you create machine learning models by finding patterns in existing data, and using these patterns to make predictions from new data as it becomes available. The AWS Management Console and API provide data and model visualization tools, as well as wizards to guide you through the process of creating machine learning models, measuring their quality and fine-tuning the predictions to match your application requirements. Once the models are created, you can get predictions for your application by using the simple API, without having to implement custom prediction generation code or manage any infrastructure. Amazon Machine Learning is highly scalable and can generate billions of predictions, and serve those predictions in real-time and at high throughput. With Amazon Machine Learning there is no setup cost and you pay as you go, so you can start small and scale as your application grows.

The workshop's region will be in N. Virginia

About this lab

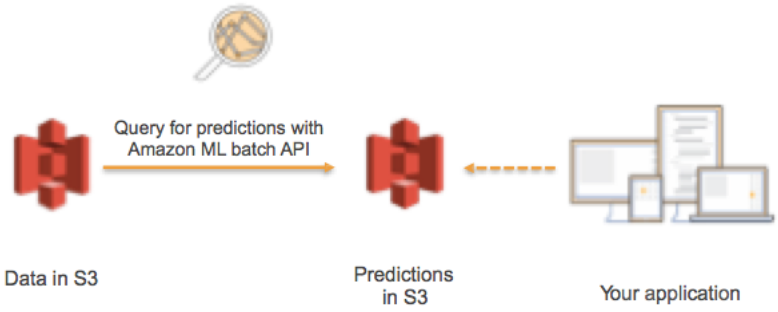
Scenario

The dataset contains daily amount of rental bikes between years 2011 and 2012 in Capital bikeshare system with the corresponding weather and seasonal information.

All we want to know is how much bikes we should prepare for the next week. To avoid the situation when the supply could not meet the demand.

Architecture Diagram

We upload data into S3. Then, we used Amazon Machine Learning for training model and prediction. All of the output will be stored into S3.



Upload file to S3

Upload a file to S3 bucket

On the service menu, click 'S3'

Click 'Create Bucket'

For Bucket Name, type 'Unique Name'

For Region, choose 'US Standard

Click 'Create'

Select the bucket which you created before

Click 'Upload'

Click 'Add files'

Select the csv file which in the share folder, then choose

Click 'Start Upload'

Working with Amazon Machine Learning

Create Model via Amazon Machine Learning

On the service menu, click 'Machine Learning'

Click 'Create a new Datasource and ML model'

For 'where is your data', choose 'S3

For **S3 location**, choose the s3 bucket location/file which you created

For Datasource name, type 'aml-ver2'

Click 'Verify'

For S3 permissions, click 'yes' about Amazon ML needs read permission for this

Amazon S3 location. Do you want to grant permission?

*You will see 'The validation is successful. To go to the next step, choose Continue'

Click 'Continue'

In Schema part

About first line in the column name, click yes when you see the question: Does the first line in your CSV contain the column names?

For Datatype, choose season/mnth/weekday/workingday/weathersit as Catagorical

For Datatype, choose cnt as Numetric

Click 'Continue'

In Target part

For target, choose 'cnt' as target for prediction

Click 'Continue'

In Row ID part

Click 'Review'

In Review part

Click 'Finish

In ML model settings part

Click 'Review

In Review part

Click 'Create ML Model'

For this moment, you will see the message said ' status: Pending', you can test this machine learning until the status go to 'completed'

Testing with Amazon Machine Learning

For Dashboard, click 'ML-model' which AML created

For the left panel, click 'Try real-time predictions'

For season, you can type 1 to 4

For mnth, you can type 1 to 12

For weekday, you can type 1 to 7

For workingday, you can type 1 to 2

For weathersit, you can type 1 to 4

Then, click 'create prediction', you will see the predict value in the right panel

End your lab

Delete S3

On the service menu, click 'S3'.

Select the bucket which you created before.

Select Action, then click 'Delete Bucket'.

Confirm the action, then delete the bucket.

Delete Amazon Machine Learning

On the service menu, click 'Machine Learning'.

Select the items which you created before.

Select 'Actions', then click 'Delete'.

Click 'click'.