Start with Azure Machine Learning Studio & get on with Azure Storage

Introduction:

This writing is a continuation from my previous one by which we will be sketching across how to get started with Azure Machine Learning and an introduction for azure storage.

Keypoints in short for this article:

- Kickstart for Azure Machine Learning.
- Overview of Azure Storage.
- Creating an experiment in Azure ML studio.

Get started with Azure Machine Learning:

Follow the below steps to get started with Azure Machine Learning Studio..

Requirements –

01. Microsoft account – outlook or Hotmail email account.

Try signing up from http://studio.azureml.net/ for Azure Machine Learning Studio, once after signing in you will be landing up at a page as below:
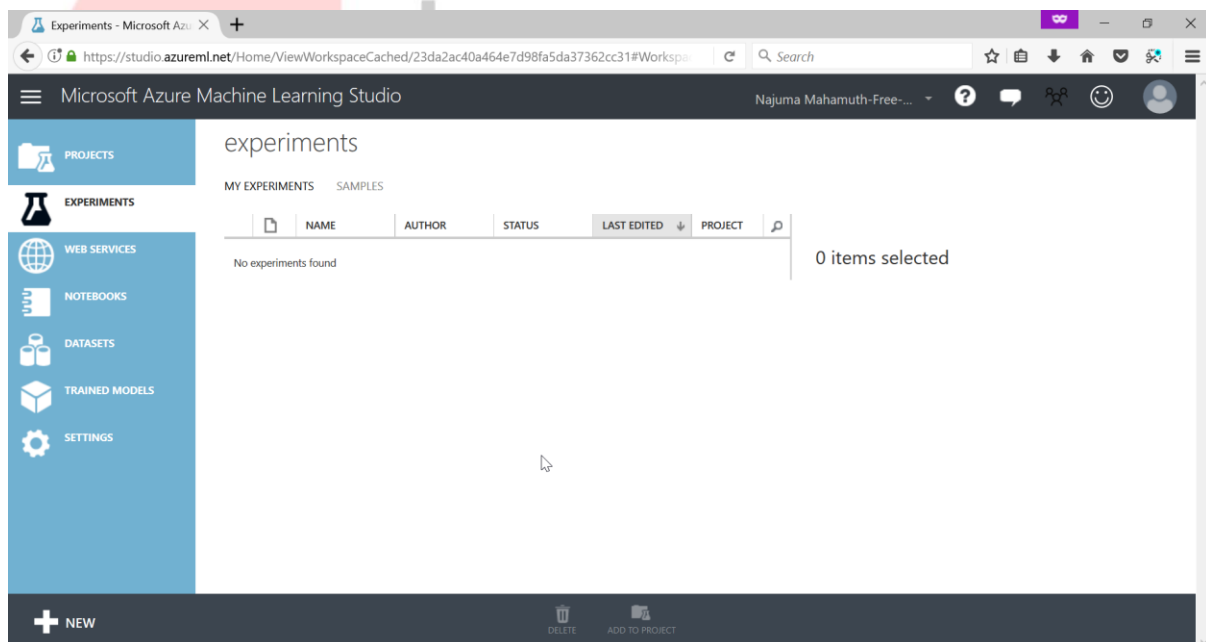
The home page of the Machine Learning Studio has several menus and options on it, "Microsoft Azure Machine Learning Studio" will help you to take forward the browser for home page, "Projects" deals with collection of experiments, datasets, notebooks and other resources representing a single project, "Experiments" helps in experimenting the experiment which we have created, "Web Services" helps in deploying the experiments as web services namely web apps, "Notebooks" helps in moving with jupyter notebooks, "Datasets" deals with the datasets that we have uploaded in Azure Machine Learning Studio, "Trainer Models" helps in training the models and save the experiments on the Azure ML Studio, "Settings" goes with a collection of settings that we can use to configure our account.

Experiment: An experiment normally holds one dataset atleast which provides data for analytical modules which will help us to connect together to construct a predictive model. An experiment will be having datasets connected towards modules with work flows and required parameters for each modules. There are lot many experiments available in Azure ML studio.

Datasets: Datasets are used in modelling process, we have lot many datasets available in Azure ML studio. Datasets can be of any ways like calculation of diagnostics data for cancer, credit card fraud transaction details, etc.,

Modules: Modules are algorithms that can perform on the data provided, this holds training, scoring and validation process.

Once after completing all these things we can deploy the Azure ML experiment as web service.

Brief intro for Azure Storage:

Azure Storage is of high availability, multi-tenant services designed for flexible storage options and it consists of four services blobs, files, tables and queues. Azure storage supports with scalability, durability, high availability and cost effective. Azure storage is multi-tenant which refers to an architecture in which a single instance of software runs on a server and serves for multiple tenants (group of users who share the same resources).

Each storage over here has its own http or https endpoint, for example a storage account created as najumamahamuth as name will be as follows:

Blob – https://najumamahamuth.blob.core.windows.net

Table – https://najumamahamuth.table.core.windows.net

Queue – https://najumamahamuth.queue.core.windows.net

Files - https://najumamahamuth.files.core.windows.net

This storage service can support upto 500TB's of capacity with 20,000 IOPS and upto 20Gbps in and 30 out for LRS - upto 10Gbps in and 20 out for GRS, helps in security through management certificate, RBAC, or name and authentication key.

Replications:

Azure Storage comes up with four types of replications, they are:

- RA-GRS – Read access – Geo-redundant storage, this enables the replication target for secondary locations as same as for the primary location.
- GRS – Geo redundant Storage.
- ZRS – Zone redundant storage which is available only for Block blobs.
- LRS – Locally Redundant Storage.

Blob:

Blob Storage – used for storage of files, VHD's, mp4s, pngs, .txt, etc.,

Blob Storage URL can be illustrated simply as below:

Blob Storage Concepts
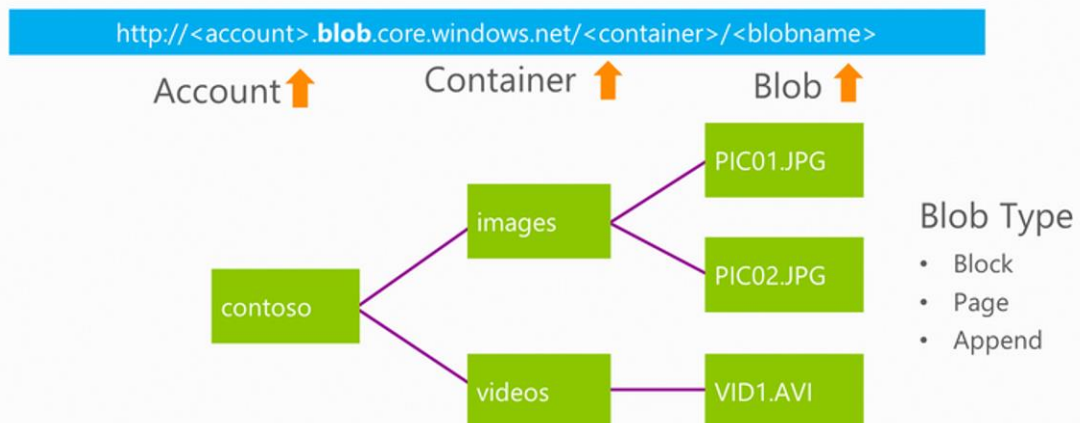
Blob comes with two types as Block blobs and Page Blobs:

Block Blob – this supports upto 200 GB which is designed for uploading and downloading in blocks, recommended for storage of movies, images, text files, etc., which is commonly used for downloads.

Page Blob – this supports upto 1TB and it is designed for reading and writing in 512 byte pages, this is recommended for applications that support seek and random/read write like virtual hard disks.

Tables:

Tables – this is another highly scalable service for non-relational structured data.

Tables are used for massive auto-scaling with NoSQL store. Helps for user, device and service metadata, structured data. Its schema less entities which comes with strong consistency, we have no limits on number of table rows or table size, helps us with dynamic load balancing of table regions, its best for key/value lookups on partition key and row key.

Queues:

Queues – this is for low latency message processing.

Queues are reliable messaging system at scale for cloud services, it's for decoupling components and scale them independently, scheduling of asynchronous tasks, building processes and work flows, no limits on number of queues or messages, message visibility timeout to protect from component issues, update the messages to checkpoint progress part way through.

Files:

Files – SMB (Server Message Block) file sharing which is a common internet file sharing system, this is commonly used for sharing files, printers and ports or communications between nodes of a network. Helps us in Lift and Shift on-premises applications, its natively supported by OS API's, libraries and tools, its built on SMB 3 which can work with Windows and Linux followed by no limit on shares.

Surf more about all these tenant services over here - https://azure.microsoft.com/en-in/services/storage/

Working with sample datasets:

To work with sample datasets which we have in Azure Machine Learning Studio we should first create an experiment on it, so lets start with creating an experiment as start:

Step 1: Goto Azure Machine Learning Studio page using - http://studio.azureml.net/

Signup or sign in with your Microsoft account here, note – its free of cost.
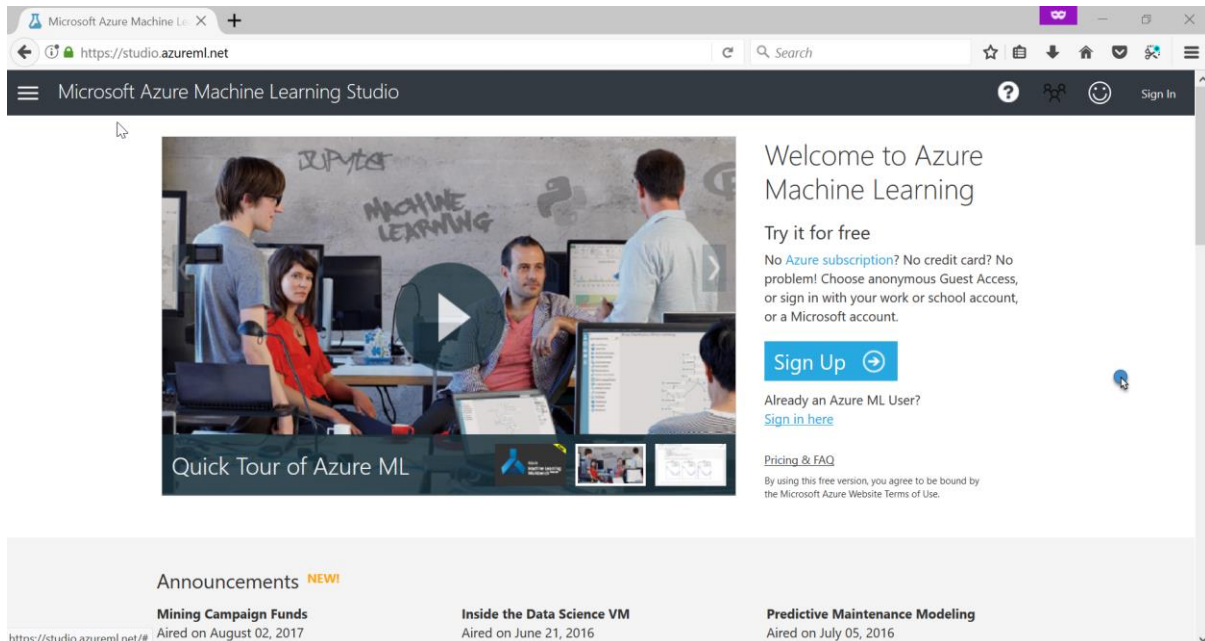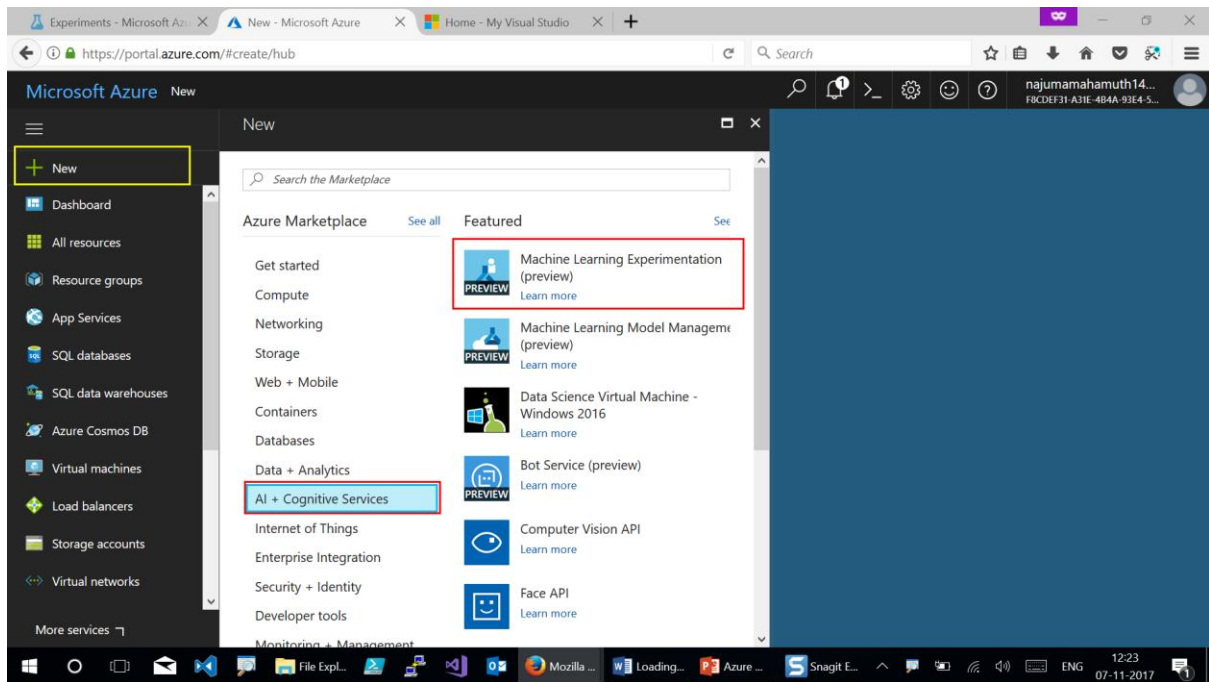

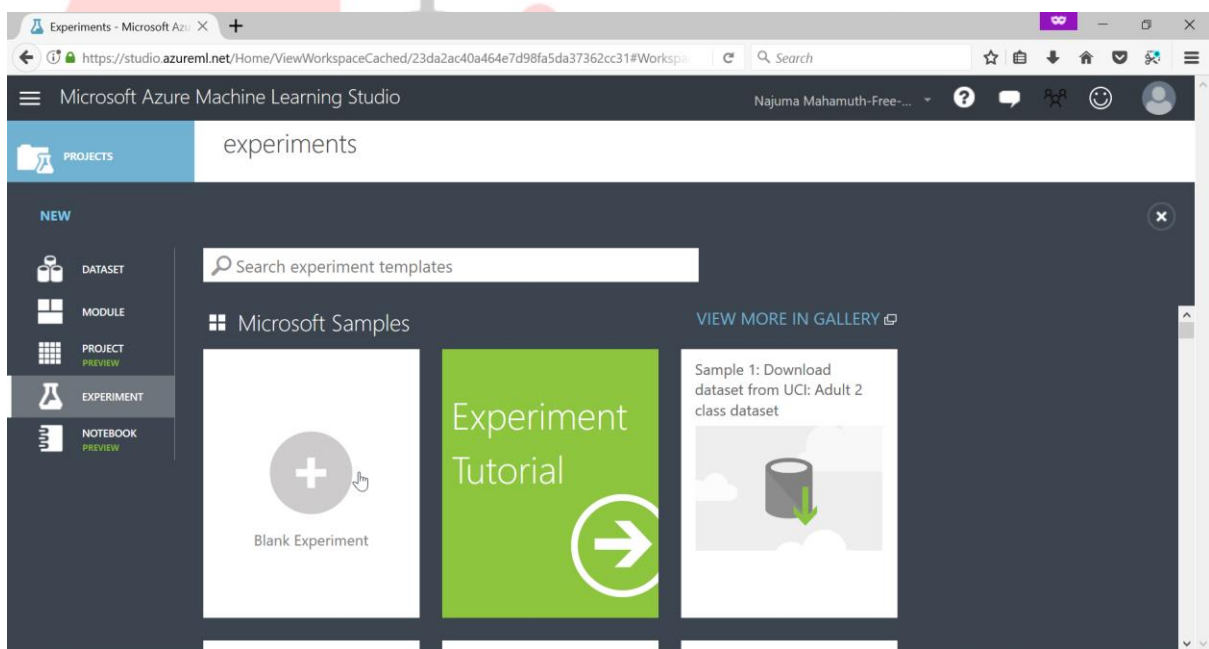Click on Sign in and sign in with Microsoft account.
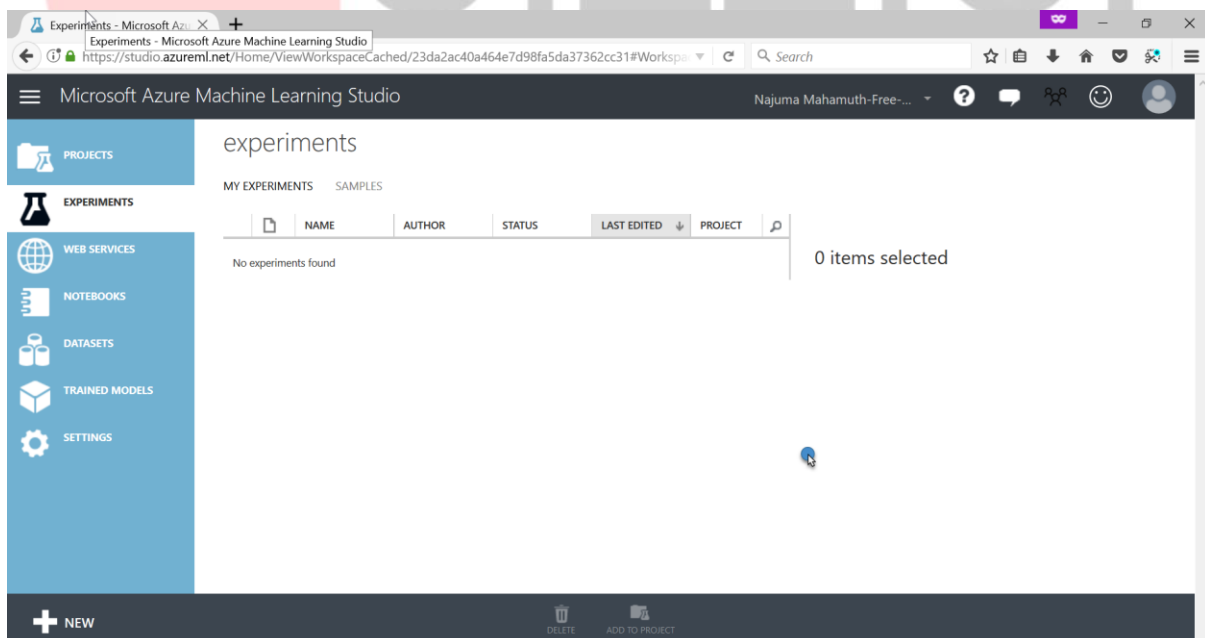
Either you can go for Machine Learning Studio by this way or by using azure portal - https://portal.azure.com
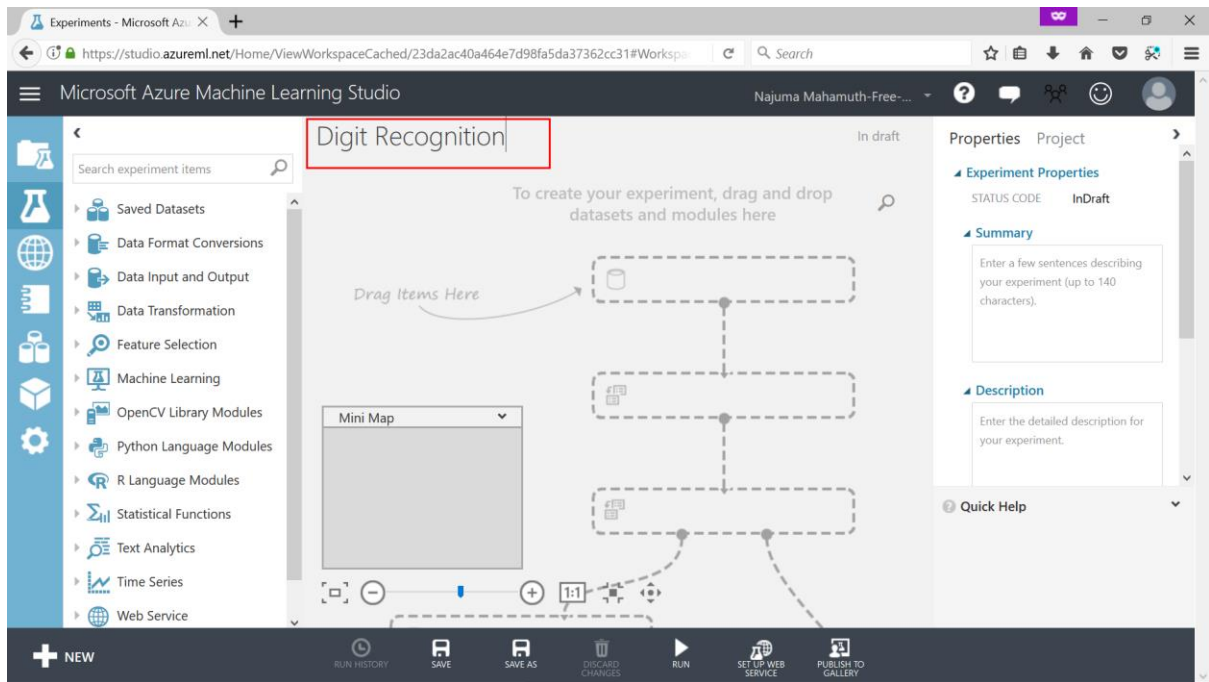
New → AI + Cognitive Services → Machine Learning Experimentation (preview)

Now move back to Azure Machine Learning Studio and click on New → Experiments → Blank Experiments.

Rename the experiment over here at workspace as shown below, here I have renamed as Digit Recognition.

Now the experiment has been created, and the next step is to upload the datasets.

Follow my next article to upload datasets on Azure ML Studio and to work with a scenario for this experiment.