

MARLA Architecture on Azure Services Configuration Manual

MSc Research Project
Cloud Computing

Janit Pathak
Student ID: 20186169

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland
Project Submission Sheet
School of Computing



Student Name:	Janit Pathak
Student ID:	20186169
Programme:	Cloud Computing
Year:	2022
Module:	MSc Research Project
Supervisor:	Vikas Sahni
Submission Due Date:	15/08/2022
Project Title:	Configuration Manual
Word Count:	618
Page Count:	8

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Janit Pathak
Date:	15th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

MARLA Architecture on Azure Services Configuration Manual

Janit Pathak
20186169

1 Introduction

1.1 Purpose of this document

The Configuration Manual is based on NCI Research project requirements guidelines mentioned in Project Handbook. The main objective of the document is to describe the software requirements of the research project and steps to deploy the developed software.

1.2 Document Structure

Section	Purpose
General Information	The module define the required prerequisites of platform setup for the research project and general information regarding implementation
Development Environment prerequisites	The module explains steps to setup the platform for software development
Solution Deployment Procedure	The module explains the steps to deploy the the research project software
Validation	The module explains the steps to validate the success of deployed solution.

2 General Information

2.1 Objective

- Implementing MARLA architecture on Microsoft Azure Cloud.
- Evaluating its performance matrix
- Comparing its makespan with AWS MARLA architecture.

2.2 Solution Summary

The implemented solution consists of 5 stages which perform together to form an architecture.

- The input phase accepts CSV file as input in blob storage and trigger the event to execute mapreduce task.
- The coordinator stage process the input data by dividing the input data into specified chunks and send data to mapper for further processing.
- The mapper phase segregate the functional data from the input data set. The output of the stage is key value pair which get stored in intermediate storage.
- The reducer phase process the intermediate data into two dimensional key value pairs.
- In final stage, the resultant data get store in reducer bucket.

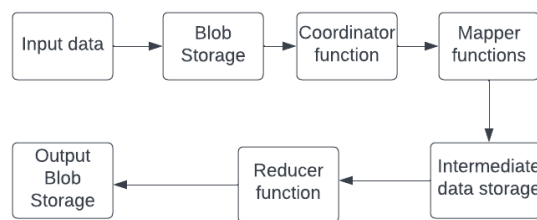


Figure 1: Application Work Flow

2.3 Architecture Requirements

The section describe the required Azure services to deploy research paper solution.

2.3.1 Microsoft Azure Account

The Microsoft Azure account with user details ¹.

2.3.2 Azure function

The Azure function is the computing service provided by Azure which is use to implement mapper, coordinator and reducer. The functions are created by using python 3.8 ².

2.3.3 Azure blob storage

The Azure blob storage stores input and output file and it also trigger the execution of the application³.

¹<https://docs.microsoft.com/en-us/learn/modules/create-an-azure-account/>

²<https://docs.microsoft.com/en-us/azure/azure-functions/functions-get-started>

³<https://docs.microsoft.com/en-us/azure/azure-functions/functions-get-started>

2.3.4 Azure VPC

The Azure VPC(Virtual private Network) connects the different cloud services used in the application ⁴.

2.3.5 Azure Monitor

The Azure monitor provide data insights of the application such as CPU utilization, logs etc⁵.

2.4 Required Skills

The guide assumes that the user has basic knowledge of Microsoft Azure and python.

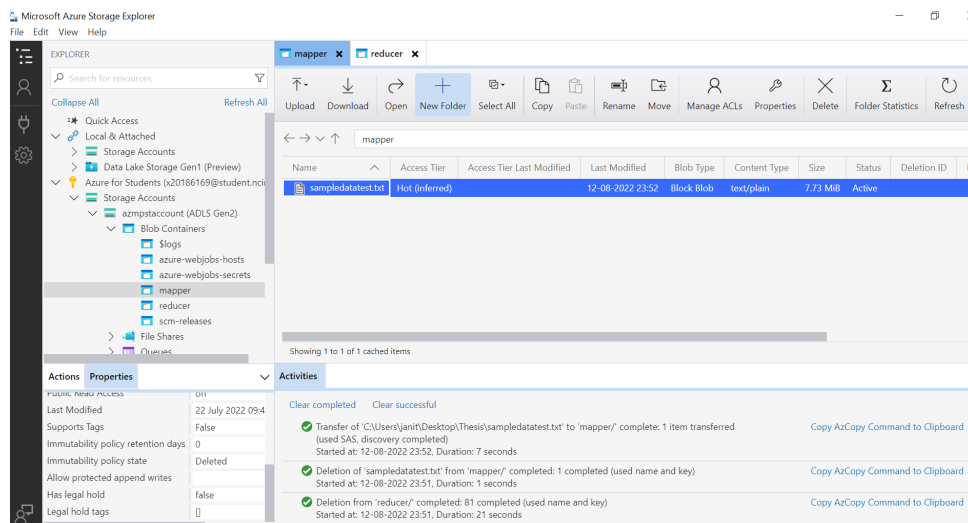


Figure 2: Input in blob storage

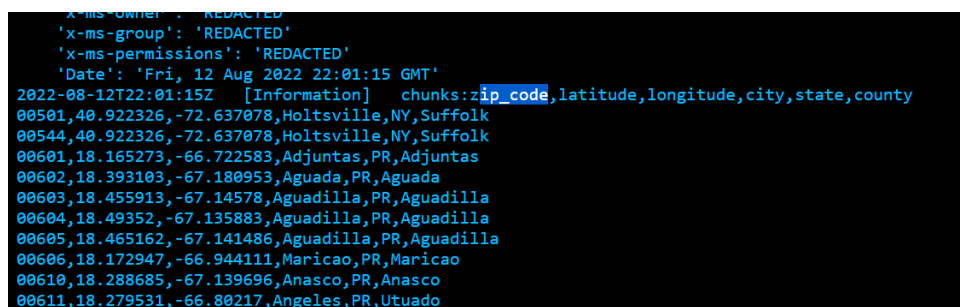


Figure 3: Data chunks in coordinator

⁴<https://docs.microsoft.com/en-us/azure/virtual-network/quick-create-portal>

⁵<https://docs.microsoft.com/en-us/azure/azure-monitor/overview>

```

02216,42.338947,-70.919635,Boston,MA,Suffolk
02217,42.338947,-70.919635,Boston,MA,Suffolk
02222,42.364381,-71.063314,Boston,MA,Suffolk
02228,,East Boston,MA,Suffolk
02238,42.446396,-71.459405,Cambridge,MA,Middlesex
02239,42.446396,-71.45940
2022-08-12T22:01:16Z [Information] pairs:[('county', 1), ('Suffolk', 1), ('Suffolk', 1), ('Adjuntas', 1),
('Aguada', 1), ('Aguadilla', 1), ('Aguadilla', 1), ('Aguadilla', 1), ('Maricao', 1), ('Anasco', 1), ('Utuaedo', 1),
('Arecibo', 1), ('Arecibo', 1), ('Arecibo', 1), ('Arecibo', 1), ('Barceloneta', 1), ('Cabo Rojo', 1), ('Cabo Rojo', 1),
('Penuelas', 1), ('Camuy', 1), ('Lares', 1), ('San German', 1), ('Sabana Grande', 1), ('Ciales', 1), ('Utuaedo', 1),
('Dorado', 1), ('Guanica', 1), ('Florida', 1), ('Arecibo', 1), ('Guanica', 1), ('Guayanilla', 1), ('Hatillo', 1),
('Hormigueros', 1), ('Isabela', 1), ('Jayuya', 1), ('Lajas', 1), ('Lares', 1), ('Las Marias', 1), ('Manati', 1),
('Moca', 1), ('Rincon', 1), ('Quebradillas', 1), ('Mayaguez', 1), ('Mayaguez', 1), ('Mayaguez', 1), ('San German', 1),
('San Sebastian', 1), ('Morovis', 1), ('Arecibo', 1), ('Aguadilla', 1), ('Vega Alta', 1), ('Vega Baja', 1), ('Vega
Baja', 1), ('Yauco', 1), ('Agua Buenas', 1), ('Guayama', 1), ('Aibonito', 1), ('Maunabo', 1), ('Arroyo', 1), ('Ponce',
1), ('Ponce', 1), ('Ponce', 1), ('Naguabo', 1), ('Naranjito', 1), ('Orocovis', 1), ('Rio Grande', 1), ('Patillas', 1),
('Caguas', 1), ('Caguas', 1), ('Caguas', 1), ('Ponce', 1), ('Canovanas', 1), ('Ponce', 1), ('Ponce', 1), ('Ponce', 1),
('Ponce', 1), ('Ponce', 1), ('Ceiba', 1), ('Cayey', 1), ('Cayey', 1), ('Fajardo', 1), ('Cidra', 1), ('Fajardo', 1),
('Humacao', 1), ('Ceiba', 1), ('Naguabo', 1), ('Rio Grande', 1), ('Salinas', 1), ('San Lorenzo', 1), ('Santa Isabel',
1), ('Vieques', 1), ('Villalba', 1), ('Yabucoa', 1), ('Coamo', 1), ('Las Piedras', 1), ('Loiza', 1), ('Luquillo', 1),

```

Figure 4: execution in mapper

```

'x-ms-content-crc64': 'REDACTED'
'x-ms-request-server-encrypted': 'REDACTED'
'Date': 'Fri, 12 Aug 2022 22:01:15 GMT'
2022-08-12T22:01:16Z [Information] Accomack
2022-08-12T22:01:16Z [Information] Adair
2022-08-12T22:01:16Z [Information] Adams
2022-08-12T22:01:16Z [Information] Addison
2022-08-12T22:01:16Z [Information] Adjuntas
2022-08-12T22:01:16Z [Information] Aguada
2022-08-12T22:01:16Z [Information] Aguadilla
2022-08-12T22:01:16Z [Information] Agua Buenas
2022-08-12T22:01:16Z [Information] Aibonito
2022-08-12T22:01:16Z [Information] Alken
2022-08-12T22:01:16Z [Information] Alschua
2022-08-12T22:01:16Z [Information] Alamanca
2022-08-12T22:01:16Z [Information] Albany
2022-08-12T22:01:16Z [Information] Albemarle
2022-08-12T22:01:16Z [Information] Alcona
2022-08-12T22:01:16Z [Information] Alcorn
2022-08-12T22:01:16Z [Information] Alexander

```

Figure 5: execution in reducer

3 Development Environment Requirement

3.1 Code Repository

Refer the zip file which is submitted in the ICT solution.

3.2 Required Programming Languages

Python Version 3.8 The Microsoft Azure provide the platform to write code in python language. The following packages are required for the application.

- csv -13.1.1
- pandas - 0.24.2
- PyPI - 24.0.0
- azure-identity==1.10.0
- azure-storage-blob==12.13.0
- azure-common==1.1.28
- msrestazure==0.6.4

3.3 Creating Azure AD user

These are the permission required by the user to create the infrastructure of the application.

- Azure function
- Azure blob storage
- Azure monitor

3.4 Implementing MARLA on Azure

- **Install Azure CLI** : Follow the link for detail instructions⁶. Run commands under all the steps in powershell with administrator rights
- **Configure Azure CLI**
\$ az login
- **Set Subscription Id**
\$subscription="f111d8d8-5988-4854-9c4a-be4a34bfccc6" # add subscription here
az account set -s \$subscription
- **Initialize Variables**

Note: Please set all the variables according to use

```
$location="eastus"  
$resourceGroup="AzureMapReduce"  
$tag="Azure-Map-Reduce-Tag"  
$storage="azmpstaccount"  
$functionApp="maraf"  
$skuStorage="Standard_LRS"  
$functionsVersion="4"  
$mapperContainer="mapper"  
$reducerContainer="reducer"  
$publishZip="maraf.zip" #This variable denotes the path to source code zip file  
$mapperFunction="mapper"  
$reducerFunction="reducer"
```

- **Create a resource group** echo "Creating \$resourceGroup in "\$location"..."

```
$resgrp = az group create --name $resourceGroup --location $location --tags $tag
```

```
$resgrp = $resgrp -- ConvertFrom-Json
```

⁶<https://docs.microsoft.com/en-us/cli/azure/install-azure-cli-windows?tabs=azure-cli>

- **Create an Azure storage account in the resource group.**

```
echo "Creating $storage"
```

```
az storage account create --name $storage --location "$location" --resource-group $resourceGroup --sku $skuStorage --kind StorageV2 --access-tier=Hot --enable-hierarchical-namespace true
```

```
az storage container create -n $mapperContainer --account-name $storage --resource-group $resourceGroup
```

```
az storage container create -n $reducerContainer --account-name $storage --resource-group $resourceGroup
```

- **Create an Azure storage account in the resource group.** echo "Creating \$functionApp"

```
$funcapp = az functionapp create --name $functionApp --storage-account $storage --consumption-plan-location $location --resource-group $resourceGroup --functions-version $functionsVersion --os-type linux --runtime python -g $resourceGroup --assign-identity [system]
```

```
$funcapp = $funcapp -- ConvertFrom-Json
```

- **Configuring a serverless function app in the resource group.**

```
echo "Configuring $functionApp"
```

```
az functionapp config appsettings set --name $functionApp --resource-group $resourceGroup --settings "BucketIn=$mapperContainer" "BucketOut=$reducerContainer" "ClusterName=MapReduce" "MapperMemory=1536" "MapperNodes=100" "MaxBlockSize=0" "MinBlockSize=1024" "ReducerMemory=1536" "ReducersNumber=10" "TimeOut=180" "StorageAccount=$storage" "FunctionApp=$functionApp" "MapperFunction=$mapperFunction" "ReducerFunction=$reducerFunction"
```

- **Deploying the function app in zip file.**

```
echo "Deploying $functionApp"
```

```
az functionapp deployment source config-zip -g $resourceGroup -n $functionApp --src $publishZip --build-remote true
```

- **Providing Permissions to the function app.** echo "Providing Permissions to \$functionApp"

```
az role assignment create --assignee $funcapp.identity.principalId --role "Storage Blob Data Contributor" --scope $resgrp.id
```



```

##$subscription=<subscriptionId>" # add subscription here
#az account set -s $subscription # ...or use 'az login'
# Function app and storage account names must be unique.
az login

# Variable block
$location="eastus"
$resourceGroup="AzureMapReduceX"
$tag="Azure-Map-Reduce-Tag"
$storage="azmpstaccountX"
$functionApp="marafX"
$skuStorage="Standard_LRS"
$functionsVersion="4.8.0.0"
$mapperContainers="mappeX"
$reducerContainers="reducerX"
$publishZip="maraf.zip"

# Create a resource group
echo "Creating $resourceGroup in \"$location\"..."
az group create --name $resourceGroup --location "$location" --tags $tag

# Create an Azure storage account in the resource group.
echo "Creating $storage"
az storage account create --name $storage --location "$location" --resource-group $resourceGroup --sku $skuStorage --kind StorageV2 --access-tier=Hot

az storage container create -n mapperContainer --resource-group $resourceGroup
az storage container create -n reducerContainer --resource-group $resourceGroup

```

Figure 6: Setup file

3.5 Replicating Existing MARLA architecture on AWS

The existing MARLA architecture Giménez-Alventosa et al. (2019) on AWS can be replicated by running following commands⁷.

- Run Command to clone the repository.
\$ git clone https://github.com/grycap/marla
- Set the configuration⁸.
- To Remove the Lambda Cluster,
\$ sh marla_remove.sh ClusterName

The Solution Summary is same as in Section 2

4 Validation

Create Azure function for Coordinator and reducer function and deploy the code available in section 2. After creating all required services, test the configurations in Config file. There are two types of validation:

- upload the sample file in the mapper storage Figure 2, if the configuration is not appropriate then it throws the error.
- Add input file in mapper storage, the function start executing and input data is divided into chunks by coordinator Figure3. The mapper stores the process data Figure 4 and the output is stored in the reducer bucket Figure5. The output can be observed in Azure logs as well.

⁷<https://github.com/grycap/marla>

⁸<https://github.com/grycap/marla>

References

- Giménez-Alventosa, V., Moltó, G. and Caballer, M. (2019). A framework and a performance assessment for serverless mapreduce on aws lambda, *Future Generation Computer Systems* **97**: 259–274. JCR Impact Factor 2021: 7.187.