09.2Aug. Streaming, Data Pipeline(Orchestration)

2 August 2023 07:38 PM

- PL
- DB
- ETL
- DWH
- Streaming
- Data Pipeline (Orchestration)
- API Gateway.

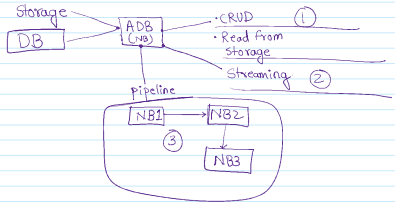
P1 = Serverless (Microservices)

P2- Data Modeling.

P3- Data Pipeline

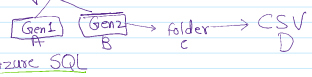
ADB

- BDP — P4
  - Hadoop → Spark → DBrick.
- CI/CD — P5
- BI
- Project → Major Project



Infrastructure (Azure) = RG

1] Storage Account



2] Azure SQL

A. SQL Server

B. SQL DataBase

C. Query (Table)

3] Azure DataBricks

A. WorkSpace

B. Cluster

C. NoteBook

Gen1: (Hierarchical Namespace= FALSE)
Storage Account: spmujahedv1
Container: container1
Data Set File: iris.csv
Key: KzecEGQj+5nekfAHYXaUDalHy7p0Nt+iAvjgoZ2VV3ondMvzx8JJhziTIMnvVDo3ebud3WJf17+H+AStG+n/Kg==

Gen2: (Hierarchical Namespace= TRUE)
Storage Account: spmujahed
Container: container1
Folder Name: input
Key: X3qf6OqlqFIEBESmsoksKfhlm/3X2mvKxh4BbqWu3wIo+7/KpLKUUHdTZhywAtFGeXTewdeLfweh+AStsbk1GQ==

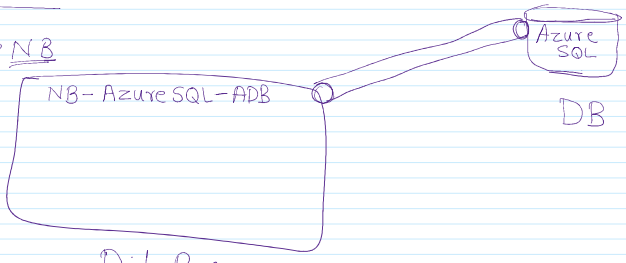
Azure SQL:
Server
ServerName: spmujahed.database.windows.net
User Name: spadmin
Password: spmujahed@123

Database Name: dbspmujahed

Azure Databricks

Azure DataBricks Cluster

→ NB



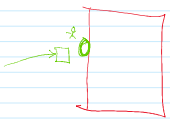
Big Data Platform

Spark → framework

↓

python

```
result = Spark.read.format("sqlserver")
    .option("host", "sql server address")
                    SN. database.windows.net
    .option("port", "1433")
```



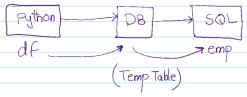
SN. database.windows.net

- .option("port", "1433")
- .option("user", "---")
- .option("password", "---")
- .option("database", "-----")
- .option("dtable", "dbo.iris")
- .load()

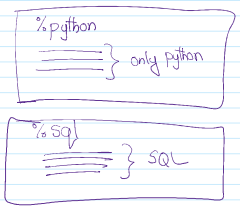
```
DROP TABLE IF EXISTS iris_data;
GO

CREATE TABLE iris_data (
  "Sepal.Length" FLOAT NOT NULL,
  "Sepal.Width" FLOAT NOT NULL,
  "Petal.Length" FLOAT NOT NULL,
  "Petal.Width" FLOAT NOT NULL,
  "Species" VARCHAR(100) NOT NULL
);

INSERT INTO iris_data("Sepal.Length","Sepal.Width","Petal.Length","Petal.Width","Species")
VALUES(5.1,3.5,1.4,.2,'Setosa');
select * from iris_data
```

% ⟶ Magic Command.

- It should in First
- Each Cell / Magic Command



- 1 NB / Multiple Magic Commands



%scala
%r
%sh ⟶ Linux (Command / Script)
%fs ⟶ DBFS dbutils.fs ⟵ Py dbfs
%md ⟶ Documentation

Dataframe = (spark.read.----.load())

WASBS= Window Azure Storage Blob Secure ⟶ Gen 1
ABFSS= Azure Blob File System Secure ⟶ Gen 2

1. Access Key
2. SAS
3. Service Principal (AAD)

WASBS

wasbs://container@storageAccount.blob.core.windows.net