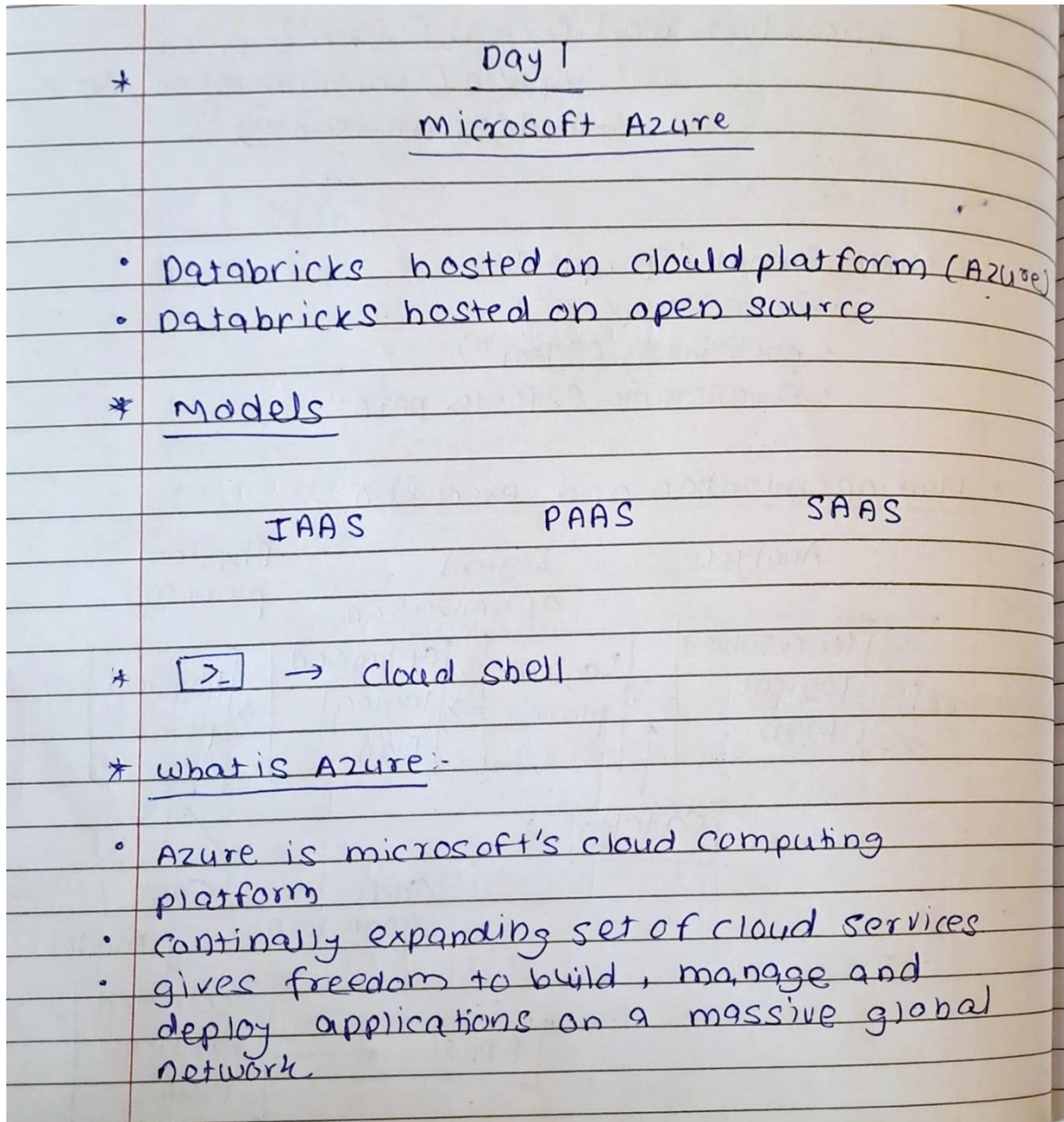


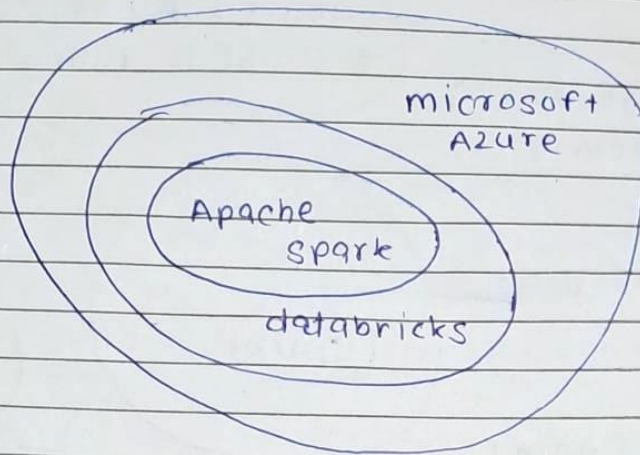
Azure Databricks Assignment 1

Pratik Wani

- Notes



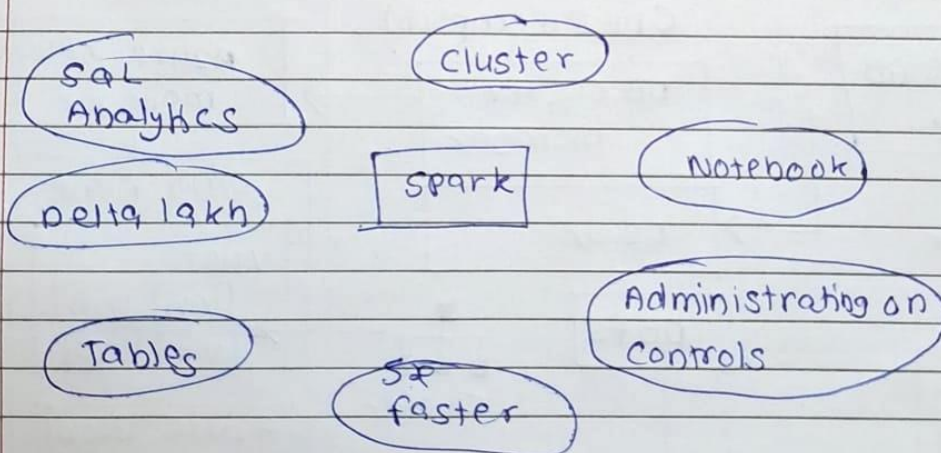
* Azure databricks



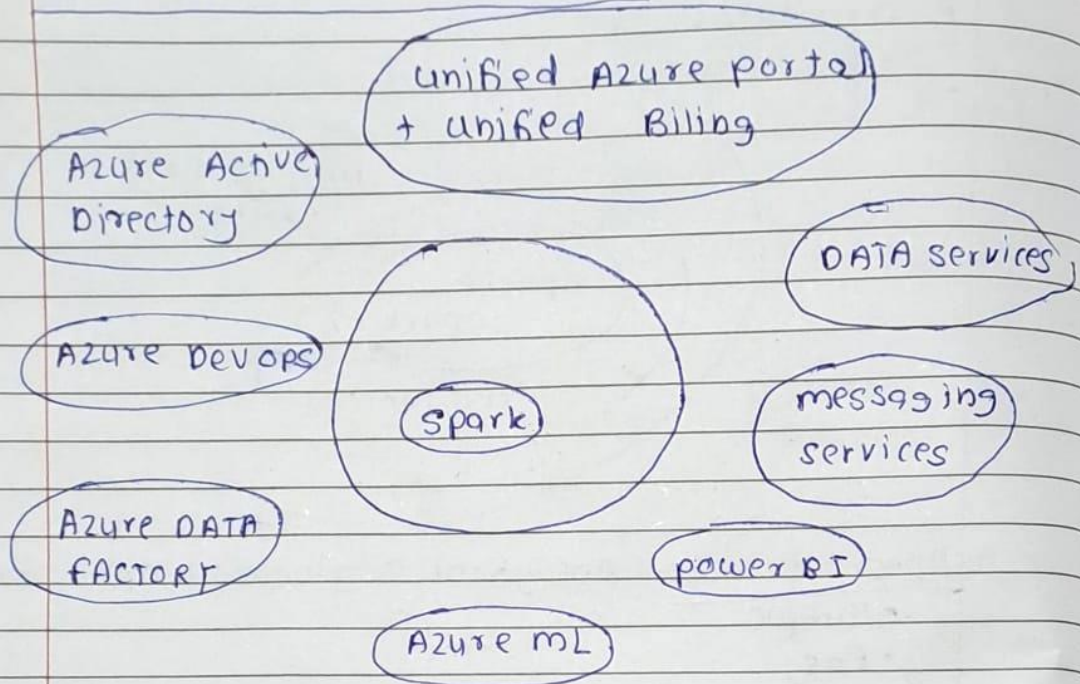
* Apache spark:- Analytical engine and open source

- Simple
- Easy
- Support SQL, Streaming, ML and graph processing
- Distributed computing platform
- in-memory processing engine

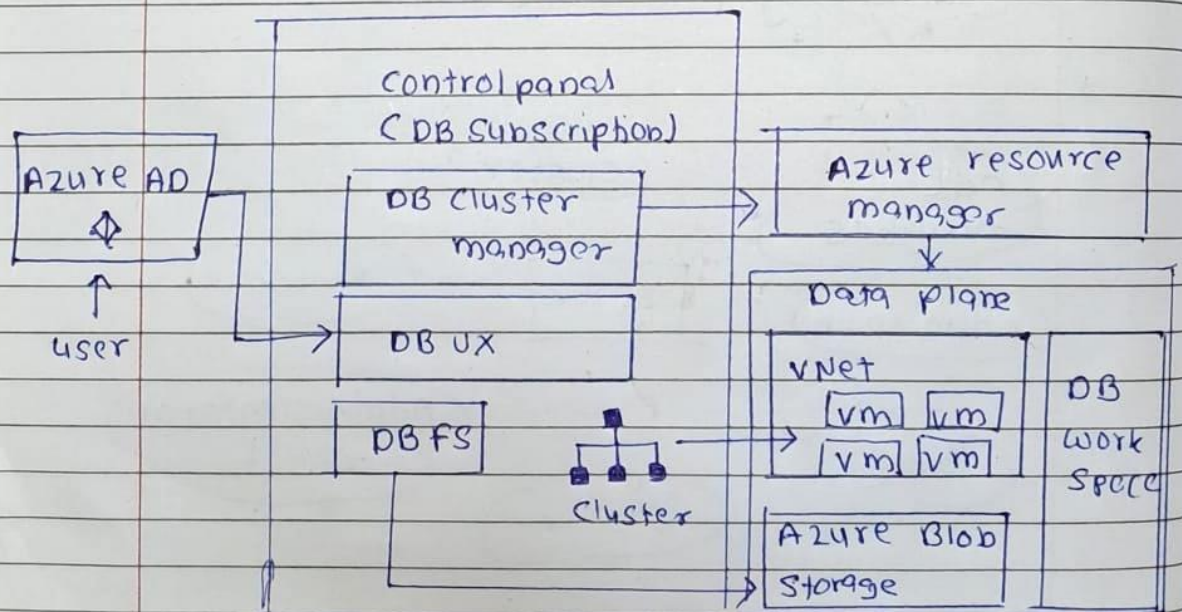
*



* Azure Databricks integration



* Architecture :-



* workspace components

compute
Networking
AI and ML
DATABASE

* cluster types

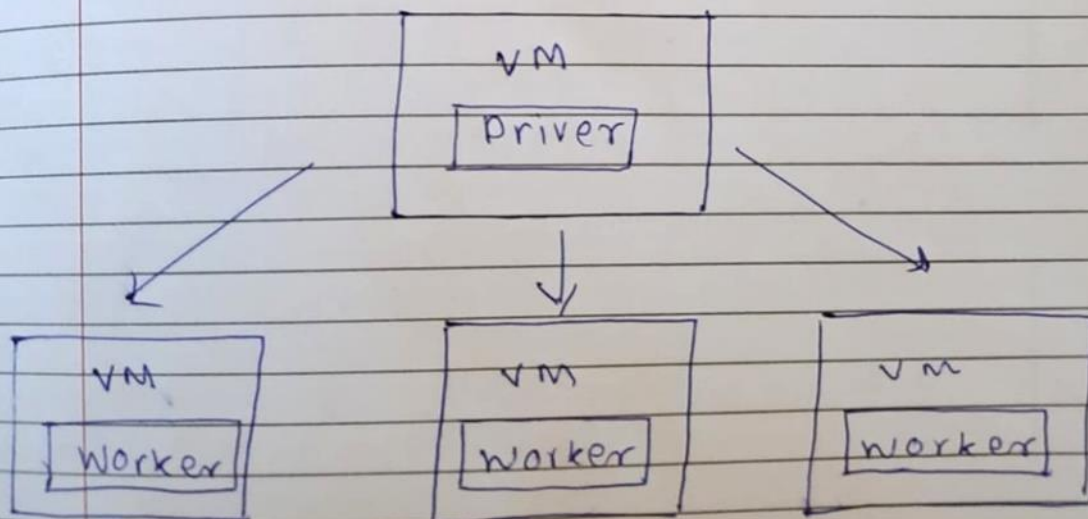
All Purpose

- created manually
- persistent
- shared
- expensive

Job cluster

- created by jobs
- Terminated
- Isolated
- cheaper

* DB cluster



- Azure Databricks Cluster and First Spark Session Creation

The screenshot shows the 'Azure Databricks' workspace list in the Microsoft Azure portal. The page displays a table of 48 records, with the first 10 visible. The table columns are Name, Type, Resource group, Location, and Subscription. The workspaces are listed in descending order of creation date.

Name	Type	Resource group	Location	Subscription
AureDataBricks_1098	Azure Databricks Service	rg-azuser1098_mml.local-1cQ9r	Central India	Azure subscription 1
azure_databricks_stan	Azure Databricks Service	azure-sg	East US	Azure subscription 1
AzureWSTest-AUS	Azure Databricks Service	rg-azuser1092_mml.local-gv7uw	Australia Central	Azure subscription 1
databricks_hexa	Azure Databricks Service	rg-azuser1062_mml.local-xrJDn	Central India	Azure subscription 1
DatabricksTraining-1083	Azure Databricks Service	rg-azuser1083_mml.local-5QutK	East US	Azure subscription 1
DatabricksTraining-1083-EA	Azure Databricks Service	rg-azuser1083_mml.local-5QutK	East Asia	Azure subscription 1
DB_learning	Azure Databricks Service	rg-azuser1089_mml.local-22dNT	Central India	Azure subscription 1
db_ws_siva	Azure Databricks Service	rg_siva	East US	Azure subscription 1
DBOne1099	Azure Databricks Service	rg-azuser1099_mml.local-zlwGt	West Europe	Azure subscription 1
Dbrickslearn1086_AustraliaE	Azure Databricks Service	rg-azuser1086_mml.local-qvNvp	Australia East	Azure subscription 1
hexa-deb-1054	Azure Databricks Service	rg-azuser1062_mml.local-xrJDn	South India	Azure subscription 1
hexa-deb-1062	Azure Databricks Service	rg-azuser1062_mml.local-xrJDn	West Central US	Azure subscription 1

The screenshot shows the 'Create an Azure Databricks workspace' page in the Microsoft Azure portal. The page displays a summary of the workspace configuration, including the workspace name, subscription, resource group, region, pricing tier, and managed resource group name. The 'Review + create' tab is selected, and the 'Create' button is visible at the bottom.

Validation Succeeded

Basics

Property	Value
Workspace name	HexawareDataBricks-1056
Subscription	Azure subscription 1
Resource group	rg-azuser1056_mml.local-SHF8
Region	Central India
Pricing Tier	standard
Managed Resource Group name	

Networking

Property	Value
Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	No

Encryption

Create **< Previous** [Download a template for automation](#)

MakeMyLabs rg-azuser1056_mml.local-SHFsi

portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F984f097c-963c-4eb6-a20d-839457ae9f08%2FresourceGroups%2Frg-azuser10... Incognito

Microsoft Azure Search resources, services, and docs (G+)

Home > rg-azuser1056_mml.local-SHFsi_HexawareDatabricks-1056 | Overview Deployment

Search « Delete Cancel Redeploy Download Refresh

Overview Inputs Outputs Template

✓ Your deployment is complete

Deployment name : rg-azuser1056_mml.local-SHFsi_HexawareDatabricks... Start time : 12/02/2024, 20:39:30
Subscription : Azure subscription 1 Correlation ID : ec8a1958-38f4-447c-a852-d10b4ecef7e
Resource group : rg-azuser1056_mml.local-SHFsiB

> Deployment details
▼ Next steps
Go to resource

Give feedback
Tell us about your experience with deployment

Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >

Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >

Free Microsoft tutorials
Start learning today >

Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

MakeMyLabs HexawareDatabricks-1056 - Mi Databricks

adb-3532112277509880.0.azuredatabricks.net/?o=3532112277509880&l=en-gb Incognito

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P HexawareDatabricks-1056 azuser1056_mml.local@ihitl.onmicr...

New Workspace Recents Catalog Workflows Compute Data Engineering Job Runs Delta Live Tables Machine Learning Experiments Features Models Serving Partner Connect Collapse menu

Get started

Import and transform data
Create a table by uploading local files, or create a pipeline for continuous data ingestion and transformation.
Create table Create pipeline

Notebook
Create a new notebook for data analysis, transformation, and machine learning.
Create notebook

AutoML
Accelerate the training of ML models for efficient discovery and iteration.
Start AutoML

Pick up where you left off Recents Favorites Popular Provide feedback

Recents appear here Popular assets appear here

Microsoft Azure | databricks Search data, notebooks, recents, and more... CTRL + P HexawareDatabricks-1056 azuser1056_mml.local@ihtl.onmicr...

Compute > New compute > Preview Send feedback

azuser1056_mml.local's Cluster

☒ Multi node ☐ Single node

Access mode ☐ Single user access

Single user azuser1056_mml.local

Performance

Databricks runtime version

Runtime: 13.3 LTS (Scala 2.12, Spark 3.4.1)

☒ Use Photon Acceleration

Worker type 16 GB Memory, 4 Cores Workers 1 ☐ Spot instances

Driver type 16 GB Memory, 4 Cores

☐ Enable autoscaling

☒ Terminate after 030 minutes of inactivity

Tags

Add tags

Create compute Cancel

Summary

1 Worker	16 GB Memory 4 Cores
1 Driver	16 GB Memory, 4 Cores
Runtime	13.3.x-scala2.12

Photon Standard_D4ads_v5 4 DBU/h

Microsoft Azure | databricks Search data, notebooks, recents, and more... CTRL + P HexawareDatabricks-1056 azuser1056_mml.local@ihtl.onmicr...

Compute > Preview Send feedback

azuser1056_mml.local's Cluster

Configuration Notebooks (0) Libraries Event log Spark UI Driver logs Metrics Apps Spark compute UI - Master

☐ Multi node ☐ Single node

Access mode ☐ Single user access

Single user azuser1056_mml.local

Performance

Databricks Runtime Version

13.3 LTS (includes Apache Spark 3.4.1, Scala 2.12)

☒ Use Photon Acceleration

Worker type 16 GB Memory, 4 Cores Workers 1 Current 1 ☐ Spot instances

Driver type 16 GB Memory, 4 Cores

☐ Enable autoscaling

☒ Terminate after 30 minutes of inactivity

Tags

No custom tags

> Automatically added tags

Summary

1 Worker	16 GB Memory 4 Cores
1 Driver	16 GB Memory, 4 Cores
Runtime	13.3.x-scala2.12

Photon Standard_D4ads_v5 4 DBU/h

More Terminate Edit

MakeMyLabs

HexawareDatabricks-1056 - Mi...

Untitled Notebook 2024-02-12

adb-3532112277509880.0.azuredatabricks.net/?o=3532112277509880#notebook/1670544436303214/command/1670544436303215

Gmail Maps News YouTube Translate Sign in - Google Ac... Login Home Python Tutorial For... Basic Bank System New Tab

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P HexawareDatabricks-1056 azuser1056_mml.local@iihtl.onmicr...

New

Workspace

Recents

Catalog

Workflows

Compute

Data Engineering

Job Runs

Delta Live Tables

Machine Learning

Experiments

Features

Models

Serving

Partner Connect

Collapse menu

Untitled Notebook 2024-02-12 20:54:20 Python ☆

File Edit View Run Help Last edit was now New cell UI: OFF

Run all azuser1056_mml.local's... Schedule Share

Cmd 1

1 from pyspark.sql import SparkSession

Command took 1.95 seconds -- by azuser1056_mml.local@iihtl.onmicrosoft.com at 2/12/2024, 8:54:41 PM on azuser1056_mml.local's Cluster

Cmd 2

1 spark = SparkSession.builder.appName("FirstSession").getOrCreate()

Command took 0.12 seconds -- by azuser1056_mml.local@iihtl.onmicrosoft.com at 2/12/2024, 8:54:53 PM on azuser1056_mml.local's Cluster

Cmd 3

1 spark

SparkSession - hive

SparkContext

Spark UI

Version

v3.4.1

Master

spark://10.139.64.5:7077

AppName

Databricks Shell

Command took 0.64 seconds -- by azuser1056_mml.local@iihtl.onmicrosoft.com at 2/12/2024, 8:55:00 PM on azuser1056_mml.local's Cluster

[Shift+Enter] to run
[Shift+Ctrl+Enter] to run selected text

