



JOURNEY SCRAPBOOK

Foundational Training Week (18th August- 28th August)

Anushka Dhir

Batch-6

DAY-1

Business Analysis and Agile

- ❖ Introduction to Business Analysis
- ❖ Functional and Non-functional requirements
- ❖ Requirements Elicitation
- ❖ Requirements Modelling
- ❖ Requirement Management Techniques
- ❖ Requirements Traceability
- ❖ SDLC Model
- ❖ Waterfall Model
- ❖ Agile Method
- ❖ Scrum Events
- ❖ Kanban Board
- ❖ ACTIVITY – To roleplay the various phases involved in SCRUM
- ❖ ASSIGNMENT – Requirement Traceability for given scenario

Requirement Traceability for setting up add to cart option for an application

Sno.	Req ID	Req Desc	User Story	TC ID	TC Desc	Test Design	Test Designer	UAT Test Req?	Test Execution			Defects?	Defect ID	Defect Status	Req Coverage Status
									Test Env	UAT Env	Prod Env				
1	Req01	Search for a product	Enter item name in the search bar and display the products in a dropdown format according to the given input.	TC01	User searched a product but its not available	Completed	XYZ	No	Passed	No Run	No Run	None	No	NA	Partial
2				TC02	User searched a product and got a valid dropdown list	Completed	ABC	No	Passed	No Run	No Run	None	No	NA	Partial
3				TC03	User searched a product got the partial keywords matched	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial
4				TC04	User doesn't type anything and pressed enter	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial
1			User gives the item name in the search bar and hit the enter button.	TC01	User searched a product but its not available	Completed	XYZ	No	Passed	No Run	No Run	None	No	NA	Partial
2				TC02	User searched a product and got a valid dropdown list	Completed	ABC	No	Passed	No Run	No Run	None	No	NA	Partial
3				TC03	User searched a product got the partial keywords matched	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial
4				TC04	User doesn't type anything and pressed enter	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial

DAY-2

Business Analysis and Agile

- ❖ Discussion On Agile Methodology
 - Scrum
 - Kanban
- ❖ Comparison Between KANBAN And SCRUM
- ❖ Values And Principles Of Agile Methodology
- ❖ Discussion Of Document Of Ready (Dor) And Document Of Done (Dod)
- ❖ Estimation In Agile:
 - Scope
 - Time
 - Cost
- ❖ Various Charts Used:
 - Burndown Chart
 - Burnup Chart
 - Velocity Chart
- ❖ ACTIVITY – To Roleplay The Various Phases Involved In SCRUM
 - Sprint Review
 - Sprint Retrospection
- ❖ ASSIGNMENT – Introduction To Azure Devops

AA Assignment - Agile +

Overview

Boards

Work items

- Boards
- Backlogs
- Sprints
- Queries
- Delivery Plans
- Analytics views

Repos

Pipelines

Test Plans

Artifacts

Work items

Recently updated | + New Work Item | Open in Queries | Column Options | Import Work Items | Recycle Bin | ⚙️ ⚙️ ⚙️

ID	Title	Assigned To	State	Area Path	Tags
30	Coding	👤 Unassigned	● New	Assignment - Agile	
29	Review	👤 Unassigned	● New	Assignment - Agile	
28	System Testing	👤 Unassigned	● New	Assignment - Agile	
27	Test Case Creating	👤 Unassigned	● New	Assignment - Agile	
26	Unit Testiong	👤 Unassigned	● New	Assignment - Agile	
25	Coding	👤 Unassigned	● New	Assignment - Agile	
24	Requirement Analysis	👤 Unassigned	● New	Assignment - Agile	
23	Review	👤 Unassigned	● New	Assignment - Agile	
22	System Testing	👤 Unassigned	● New	Assignment - Agile	
21	Test Case Creating	👤 Unassigned	● New	Assignment - Agile	
20	Unit Testing	👤 Unassigned	● New	Assignment - Agile	
10	Search Functionality	👤 Unassigned	● New	Assignment - Agile	

Assignment - Agile +

Overview

Boards

Work items

Boards

Backlogs

Sprints

Queries

Delivery Plans

Analytics views

Repos

Pipelines

Test Plans

Artifacts

Assignment - Agile Team ▾ ⭐ ⓘ

Board Analytics

View as Backlog

Stories

Filter

Settings

More

New

Active

1/5

Resolved

0/5

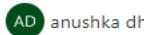
Closed

<

+ New item



13 Search Via Image



AD

anushka dhir



State

New



1/6

11 Search Via Speech



Unassigned

11

State

Active



0/6

12 Search Via Text



Unassigned

5

State

Closed



6/6

DAY-3

Database Management System

- ❖ Brief About Data And Information
- ❖ Introduction To Database And Database Management System
- ❖ Types Of Databases
- ❖ Discussion About ACID Property Of DBMS
- ❖ Introduction To The Various Operations Of SQL – DDL, DQL, DML, DCL And TCL
- ❖ Introduction To Normalization And It's Need
- ❖ Discussion About:
 - Functions – Group By, Order By
 - Joins And Its Types
 - Aggregate Functions
 - Sub Queries
- ❖ ASSIGNMENT – Hands On Experience With Postgre SQL

```
+-----+-----+-----+-----+-----+
| postgres | postgres | UTF8   | C.UTF-8 | C.UTF-8 |
| template0 | postgres | UTF8   | C.UTF-8 | C.UTF-8 | =c/postgres      +
|          |          |          |          |          | postgres=CTc/postgres
| template1 | postgres | UTF8   | C.UTF-8 | C.UTF-8 | =c/postgres      +
|          |          |          |          |          | postgres=CTc/postgres
| training  | postgres | UTF8   | C.UTF-8 | C.UTF-8 |
+-----+
(4 rows)
```

```
training=# create table guest (gid int, mob bigint, email varchar(50), aadhaar bigint);
CREATE TABLE
```

```
training=# insert into guest values (1, 1234567890, 'abx@gmail.com', 23456123);
INSERT 0 1
```

```
training=# insert into guest values (2, 9876554322, 'abc@gmail.com', 23456345);
INSERT 0 1
```

```
training=# create table room (gid int, room_id int, days int);
CREATE TABLE
```

```
training=# insert into room values (1, 10, 4);
INSERT 0 1
```

```
training=# insert into room values (2, 15, 2);
INSERT 0 1
```

```
training=# select * from guest
```

```
training-# ;
+-----+-----+-----+
| gid | mob    | email   | aadhaar |
+-----+-----+-----+
| 1  | 1234567890 | abx@gmail.com | 23456123 |
| 2  | 9876554322 | abc@gmail.com | 23456345 |
+-----+
(2 rows)
```

```
training=# select * from guest cross join room;
```

```
training-# ;
+-----+-----+-----+-----+-----+
| gid | mob    | email   | aadhaar | gid | room_id | days |
+-----+-----+-----+-----+-----+
| 1  | 1234567890 | abx@gmail.com | 23456123 | 1  | 10       | 4    |
| 2  | 9876554322 | abc@gmail.com | 23456345 | 1  | 10       | 4    |
| 1  | 1234567890 | abx@gmail.com | 23456123 | 2  | 15       | 2    |
| 2  | 9876554322 | abc@gmail.com | 23456345 | 2  | 15       | 2    |
+-----+
(4 rows)
```

```
training=# select * from guest inner join room on guest.gid = room.gid;
```

```
training-# ;
+-----+-----+-----+-----+-----+
| gid | mob    | email   | aadhaar | gid | room_id | days |
+-----+-----+-----+-----+-----+
| 1  | 1234567890 | abx@gmail.com | 23456123 | 1  | 10       | 4    |
| 2  | 9876554322 | abc@gmail.com | 23456345 | 2  | 15       | 2    |
+-----+
(2 rows)
```

```
training=#
8
```

```

training=# create table room (gid int, room_id int, days int);
CREATE TABLE
training=# insert into room values (1, 10, 4);
INSERT 0 1
training=# insert into room values (2, 15, 2);
INSERT 0 1
training=# select * from guest
training-# ;
+-----+-----+-----+
| gid | mob | email | aadhaar |
+-----+-----+-----+
| 1 | 1234567890 | abx@gmail.com | 23456123 |
| 2 | 9876554322 | abc@gmail.com | 23456345 |
+-----+
(2 rows)

training=# select * from guest cross join room;
+-----+-----+-----+-----+-----+-----+
| gid | mob | email | aadhaar | gid | room_id | days |
+-----+-----+-----+-----+-----+-----+
| 1 | 1234567890 | abx@gmail.com | 23456123 | 1 | 10 | 4 |
| 2 | 9876554322 | abc@gmail.com | 23456345 | 1 | 10 | 4 |
| 1 | 1234567890 | abx@gmail.com | 23456123 | 2 | 15 | 2 |
| 2 | 9876554322 | abc@gmail.com | 23456345 | 2 | 15 | 2 |
+-----+
(4 rows)

training=# select * from guest inner join room on guest.gid = room.gid;
+-----+-----+-----+-----+-----+-----+
| gid | mob | email | aadhaar | gid | room_id | days |
+-----+-----+-----+-----+-----+-----+
| 1 | 1234567890 | abx@gmail.com | 23456123 | 1 | 10 | 4 |
| 2 | 9876554322 | abc@gmail.com | 23456345 | 2 | 15 | 2 |
+-----+
(2 rows)

training=# insert into room values (1, 14, 7);
INSERT 0 1
training=# insert into room values (1, 4, 1);
INSERT 0 1
training=# insert into room values (2, 18, 1);
INSERT 0 1
training=# select count(gid), gid, room_id, days from room group by gid;
ERROR: column "room.room_id" must appear in the GROUP BY clause or be used in an aggregate function
LINE 1: select count(gid), gid, room_id, days from room group by gid...
^

training=# select count(room_id), gid from room group by gid;
+-----+
| count | gid |
+-----+
| 2 | 2 |
| 3 | 1 |
+-----+
(2 rows)

training=#

```

```
training=# \l
      List of databases
   Name | Owner | Encoding | Collate | Ctype | Access privileges
-----+-----+-----+-----+-----+
postgres | postgres | UTF8 | C.UTF-8 | C.UTF-8 |
template0 | postgres | UTF8 | C.UTF-8 | C.UTF-8 | =c/postgres      +
          |           |       |           |       | postgres=CTc/postgres
template1 | postgres | UTF8 | C.UTF-8 | C.UTF-8 | =c/postgres      +
          |           |       |           |       | postgres=CTc/postgres
training  | postgres | UTF8 | C.UTF-8 | C.UTF-8 |
(4 rows)
```

```
training=# create table guest (gid int, mob bigint, email varchar(50), aadhaar bigint);
```

```
CREATE TABLE
```

```
training=# insert into guest values (1, 1234567890, 'abx@gmail.com', 23456123);
```

```
INSERT 0 1
```

```
training=# insert into guest values (2, 9876554322, 'abc@gmail.com', 23456345);
```

```
INSERT 0 1
```

```
training=# create table room (gid int, room_id int, days int);
```

```
CREATE TABLE
```

```
training=# insert into room values (1, 10, 4);
```

```
INSERT 0 1
```

```
training=# insert into room values (2, 15, 2);
```

```
INSERT 0 1
```

```
training=# select * from guest
```

```
training-# ;
```

gid	mob	email	aadhaar
1	1234567890	abx@gmail.com	23456123
2	9876554322	abc@gmail.com	23456345

```
(2 rows)
```

```
training=#
```

DAY-4

Testing

- ❖ Introduction To V-model
- ❖ Discussion About Various Types Of Environments
- ❖ Discussion On Types Of Testing:
 - Static
 - Dynamic
 - Functional
 - White Box
 - Black Box
 - Non- Functional
- ❖ Introduction To Software Test Life Cycle(STLC) And It's Various Phases
- ❖ Deliverables In The Various Phases Of STLC
- ❖ Discussion About Testing Triangle
- ❖ Various Approaches For Development
 - Test Driven Development (TDD)
 - Behavior Driven Development (BDD)
 - Domain Driven Development (DDD)
 - Acceptance Driven Development (ADD)
- ❖ Discussion About Test Case And Test Case Writing
- ❖ ASSIGNMENT – To Write Testcases For Testing The Login Functionality

Test Case Description	Pre-requisites	Test Case Id	Test Step	Test data	Expected results	Actual Results	Test result(pass/fail)
Test for valid user name and password	The user and password need to be existed in the system.i.e abc@yahoo.com,xyz	TC-01	User name and password are valid	<u>abc@yahoo.com</u> , <u>xyz</u>	User will be logged in	User logged in successfully	Pass
Test for invalid user name and password	The user and password need to be existed in the system.i.e abc@yahoo.com,xyz	TC-02	Username and password are invalid	<u>uzq@yahoo.com,ys</u> <u>x</u>	User is not able to login	User was not able to log in	Pass
Test for valid user name and invalid password	The user and password need to be existed in the system.i.e abc@yahoo.com,xyz	TC-03	Username is valid password is invalid	<u>abc@yahoo.com</u> , <u>xqq</u>	User is not able to login because password is wrong	User is not able to login	Pass
Test for invalid user name and valid password	The user and password need to be existed in the system.i.e abc@yahoo.com,xyz	Tc-04	Username is invalid password is valid	<u>xyx@yahoo.com</u> , <u>xyz</u>	User is not able to login because username is wrong	User is not able to login	Pass

DAY-5

DevOps & GitHub Actions

- ❖ Introduction To Devops
- ❖ Discussion On Principles Of Devops
- ❖ Discussion On Devops Lifecycle And It's Phases
- ❖ Introduction To The Various Devops Tools Depending On The Phase
- ❖ Introduction To Continuous Integration And Continuous Deployment (Ci/Cd)
- ❖ Introduction To Github Actions
- ❖ Discussion On The Various Components Of Github Action
 - Workflow
 - Event
 - Job
 - Action
 - Runner
- ❖ Assignment – To Setup A Github Action Workflow



Anushka622001 / GA_Session_AD

 Type / to search

Issues



Actions



Wiki



Insights

[← GA_Session_AD](#)

✓ Anushka622001 is learning GitHub Actions #1

[Re-run all jobs](#)[Latest #2](#)

Summary

Jobs

[check-bats-version](#)

Run details

[Usage](#)[Workflow file](#)

Re-run triggered 6 hours ago



Anushka622001

o- 63ac449

main

Status

Success

Total duration

17s

Artifacts

-

GA_Session_AD.yml

on: push

[check-bats-version](#)

6s



DAY-6

Cloud Computing

- ❖ Introduction To Cloud Computing
- ❖ Discussion On Various Types Of Cloud Service Models
 - Software As A Service (SaaS)
 - Platform As A Service (PaaS)
 - Infrastructure As A Service (IaaS)
- ❖ Introduction Service Layers Of Cloud Platform
- ❖ Discussion On Cloud Sourcing And The Advantages Of Cloud Services
- ❖ Introduction Cloud Taxonomy And It's Divisions
- ❖ Introduction To Microsoft Azure And It's Various Offerings
 - Analytics, Computing, Networking, Storage, Etc.
- ❖ Discussion About The Drawbacks And Challenges In Cloud Computing
- ❖ Discussion About The Future Of Cloud Computing
- ❖ Assignment – Setting Up An Azure Virtual Machine



Home >

NewVMAnu



Virtual machine

 Search[Connect](#) [Start](#) [Restart](#) [Stop](#) [Capture](#) [Delete](#) [Refresh](#) [Open in mobile](#) [Feedback](#) [CLI / PS](#)[JSON View](#)[Overview](#)[Activity log](#)[Access control \(IAM\)](#)[Tags](#)[Diagnose and solve problems](#)

Settings

[Networking](#)[Connect](#)[Disks](#)[Size](#)[Microsoft Defender for Cloud](#)[Advisor recommendations](#)[Extensions + applications](#)[Availability + scaling](#)[Configuration](#)[Identity](#)

^ Essentials

Resource group ([move](#))[RDBG](#)

Status

Stopped (deallocated)

Location

East US

Subscription ([move](#))[RDBG](#)

Subscription ID

23d99a1f-8507-41a7-b40b-64402a51e95e

Health state

-

Tags ([edit](#))[Add tags](#)

Operating system

Linux

VM availability status

Available

Size

Standard B2s (2 vcpus, 4 GiB memory)

Public IP address

[NewVMAnu-ip](#)

Virtual network/subnet

[ManiVM-vnet/default](#)

DNS name

[Not configured](#)[Properties](#)[Monitoring](#)[Capabilities \(7\)](#)[Recommendations](#)[Tutorials](#)

Virtual machine

Computer name

NewVMAnu

Operating system

Linux

Networking

Public IP address

NewVMAnu- (Network
ip)
interface

newvmanu107)

DAY-7

Case Study Evaluation and Assessment

- ❖ CASE STUDY – Deployment Of Java Spring Boot App
 - Azure Devops
 - Azure VM
 - Azure Web App Services
 - GitHub – Fork, Feature Branch, Pull Request, Merge
 - GitHub Actions – Automate Workflow
 - Docker – Make Docker Image And Dockerfile
 - Maven – Run Maven Package
 - Prometheus – To Monitor Local Environment

Anushka_Dhir_Assessment Case Study • Last Modified: Mon at 4:25 PM

Comments Editing Share

Paste
Clipboard
Font
Paragraph
Normal No Spacing Heading 1

Name- Anushka Dhir (Batch-6)
Reg ID- SH7B23188

➤ Azure Board Activities (Project Management Setup):

Azure DevOps Project Overview

Project stats

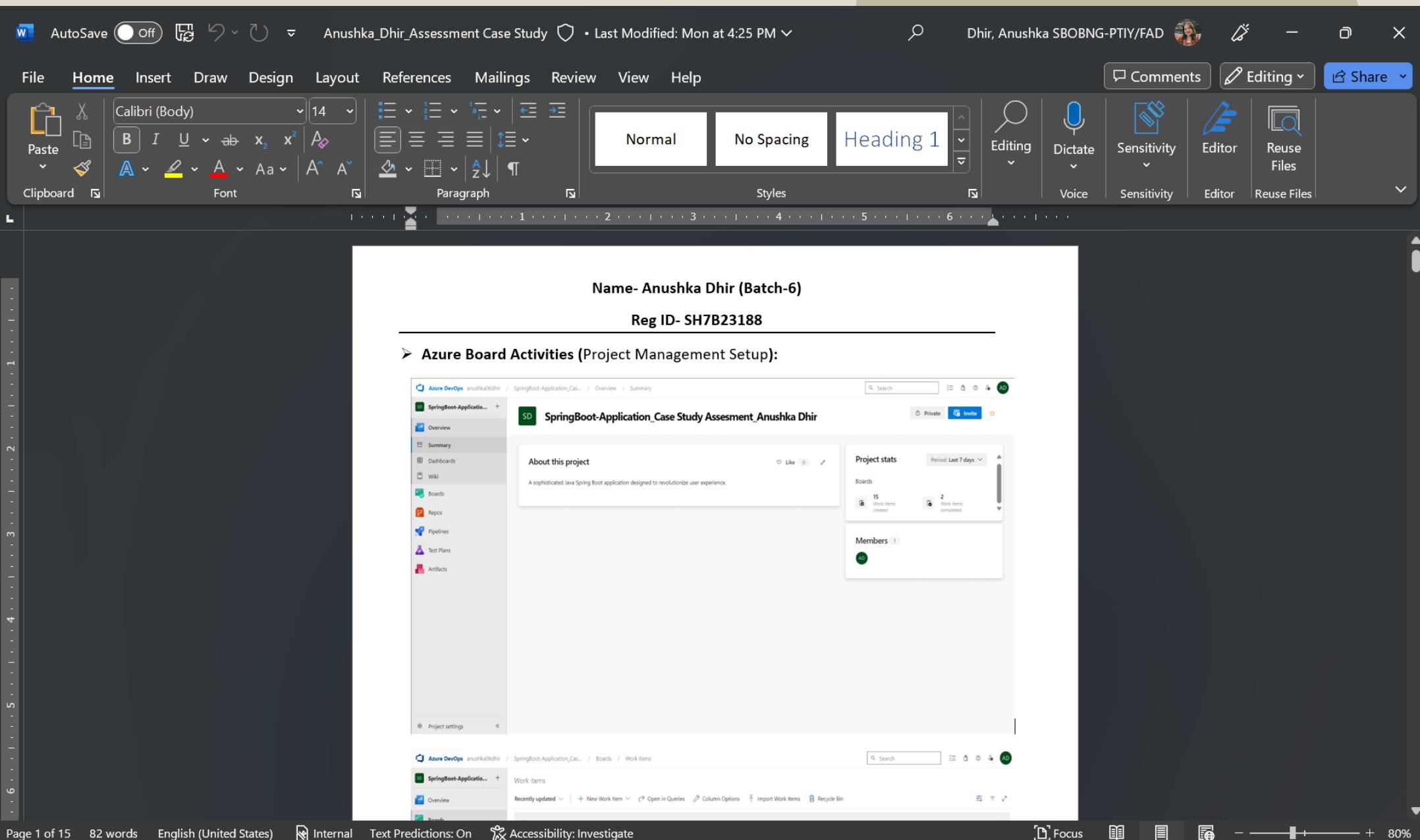
Boards

Members

Azure DevOps Work Items

Recently updated

Page 1 of 15 82 words English (United States) Internal Text Predictions: On Accessibility: Investigate Focus





JOURNEY SCRAPBOOK

Custom Bootcamp (29th August – 6th October)

Week-1

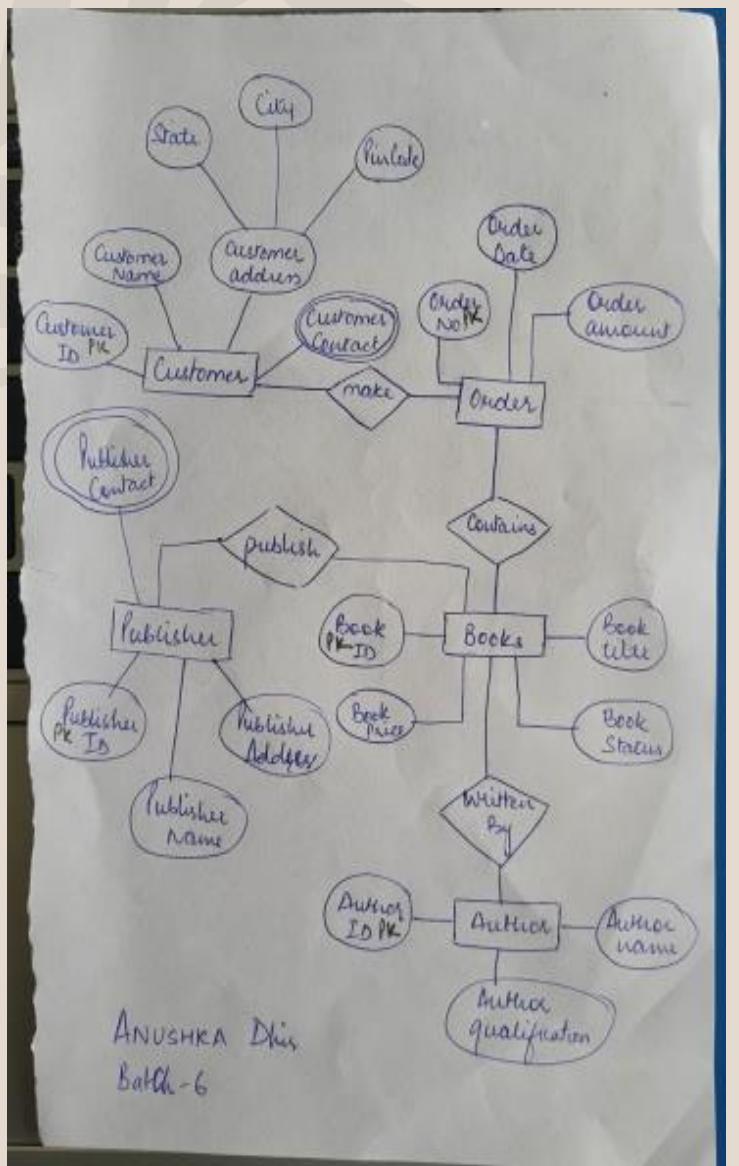
Anushka Dhir

Batch-6

DAY-1

Data Fundamentals

- ❖ What Is Data?
- ❖ Types Of Data
- ❖ What Is Database?
- ❖ What Does Data Engineer Do?
- ❖ Normalization And Denormalization
- ❖ Normalization Forms (1NF,2NF,3NF,BCNF)
- ❖ Dependencies (Partial, Transitive, Functional)
- ❖ Attribute Keys (Candidate, Primary, Foreign, Super)
- ❖ Anomalies- Insert, Update, Delete
- ❖ Data Modelling And ER Diagram
- Activity- To Make An ER Diagram For Ordering A Book From Online Store, Perform Normalization
 - ❖ Dimensional Modelling
 - ❖ Fact Table And Dimension Table
 - ❖ Star Schema & Snowflake Schema
- Activity- Implement The Concepts Of Star And Snowflake Schema By Drawing Respective Diagrams
 - ❖ SCD (Slowly Changing Dimension)



INF

Order ID	Date	Book ID	Customer ID
1	21.08.23	1	30
2	22.08.23	1,5,7,9	31
3	23.08.23	2,5	32
4	24.08.23	5,7	33

Order ID	Date	Book ID	Customer ID
1	21.08.23	1	30
2	22.08.23	1	31
2	22.08.23	5	31
2	22.08.23	7	31
2	22.08.23	9	31
3	23.08.23	2	32
3	23.08.23	5	32
4	24.08.23	5	33
4	24.08.23	7	33

Product-category

cp-id	int
p-type	string

Date

date	int
month	string
year	int
week	int

Product

p-id	int
p-name	string
p-type	string

Company

pid	int
p-loc	string
cid	int
sid	int
p-type	string
p-rid	int
revenue	int

Location

area	string
city	string
state	string
p-loc	int

Customer

cid	int
c-name	string
c-add	string
c-phn	int

Sales-person

sid	int
s-name	string
aty-rid	int
revenue	int

Day

day	int
month	string
year	int

DAY-2

Big Data & Azure Services Fundamentals

Big Data Fundamentals

- 1.What Is Data?
- 2.What Is Big Data?
- 3.4 V's Of Big Data - Volume, Variety, Velocity, Veracity
- 4.Batch And Stream Data Processing.
- 5.Extract, Load, Transform(ETL) Pipelines
- 6.How To Process Big Data?
 - a. Hadoop
 - b. Pyspark
- 7.Parallel Processing And Distributive Computing
- 8.Hadoop Used For Batch Processing & Pyspark Used For Stream Processing.
- 9.Types Of Data Structure
 - a. Structured Data
 - b. Semi-structured Data
 - c. Unstructured Data
- 10.Data Warehouse -
 - a. Save Data As Object
 - i. Tables
 - ii. Views
 - b. Structured Data (Fixed Schema-data Allocated Even If Value Is Null)
 - c. Dimension Modelling
 - d. ACID Property In DBMS -
 - i. Atomicity
 - ii. Consistency
 - iii. Isolation
 - iv. Durability
- 11.Datalake -
 - a. Saves All Types Of Data (Structured, Semi-structured And Unstructured)
 - b. Flexible Schema (No Data Allocated For Null Values)
 - c. No ACID Properties

DAY-2

Big Data & Azure Services Fundamentals

Big Data Fundamentals (Contd.)

1. Data Lakehouse
 - a. Combination Of Data Warehouse And Datalake.
 - b. Took Advantages Of Both Data Warehouse And Datalake.
2. Creating A VM Using Microsoft Azure.
3. What Is Cloud Computing.
4. Types Of Cloud -
 1. Public Cloud
 2. Private Cloud
 3. Hybrid Cloud
5. Cloud Benefits -
 1. High Availability
 2. Scalability
 3. Agility
 4. Fault Tolerance
 5. Elasticity
 6. Security
 7. Disaster Recovery
 8. Customer Latency Capabilities
6. Compare Capex(capital Expenditure) Vs OpEx(operational Expenditure) -
 1. Cloud Is OpEx.
7. Infrastructure As A Service(IAAS)
8. Platform As A Service(PAAS)
9. Software As A Service(SAAS)
10. Shared Responsibility Model
11. Serverless Computing

DAY-2

Big Data & Azure Services Fundamentals

Azure Services Fundamentals

- 1. Azure Architecture
 - a. Regions
 - b. Region Pairs
 - c. Availability Zones
- 2. Azure Resources
- 3. Azure Subscriptions -
 - i. Billing Boundary
 - ii. Access Control Boundary
- 4. Azure Resource Manager
- 5. Resource Groups
- 6. Subscriptions -> Resource Groups

DAY-3

Azure SQL

1. Setting Up Azure SQL Services By Logging In On Azure Portal.
2. SQL Deployment Options For Azure SQL:
 1. SQL Database
 2. SQL Managed Instances
 3. SQL Virtual Machines
3. Select SQL Database -> Single Database -> Create Database.
4. Introduction To SQL
5. Database Created On Microsoft Azure SQL.
6. Running Query Editor On Our Server.
7. DDL (Data Definition Language): create, Alter, Drop, Truncate
8. DML (data Manipulation Language): Select, Insert, Update, Delete
9. DCL (data Control Language): grant, REVOKE
10. Constraints
11. View: View Is A Virtual Table Based On The Result-set Of An SQL Statement.
12. Schema In Sql
13. Difference Between DELETE, DROP, TRUNCATE.
14. Activity : Create Tables For Online Book Store
15. Group By Clause
16. Having Clause : For Filtration
17. Where Clause
18. GRANT: Grant Select Permission On <Table_name> To <User>
19. Revoke

Home >



Microsoft.SQLDatabase.newDatabaseNewServer_d2bf60d3482c4af19dc27 | Overview

Deployment

[Delete](#) [Cancel](#) [Redeploy](#) [Download](#) [Refresh](#)

Overview

Inputs

Outputs

Template

Your deployment is complete

 Deployment name : Microsoft.SQLDatabase.newDatabase... Start time : 8/31/2023, 10:28:52 AM
 Subscription : npunext-1680261348707 Correlation ID : d77af3b5-3d31-4ac7-bd1f-aa286...
 Resource group : RG_IDAShell_Anushka

[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 >](#)



Home > Anushka_IDA (idashellserver-anushka/Anushka_IDA)



Anushka_IDA (idashellserver-anushka/Anushka_IDA) | Query editor (preview)



Search



Login

New Query

Open query

Feedback

Getting started

Overview

Activity log

Tags

Diagnose and solve problems

Query editor (preview)

Settings

Compute + storage

Connection strings

Properties

Locks

Data management

Replicas

Sync to other databases

Integrations

Azure Synapse Link

Stream analytics (preview)

Add Azure Search

Anushka_IDA (Anushka0602)



Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

> Tables

> Views

> Stored Procedures

Query 1

 Run Cancel query Save query Export data as Show only Editor

1 select getdate()

Results

Search to filter items...

2023-08-31T06:18:05.4100000

Query succeeded | 0s

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

Home > Anushka_IDA (idashellserver-anushka/Anushka_IDA)

Anushka_IDA (idashellserver-anushka/Anushka_IDA) | Query editor (preview)

SQL database

Search Login New Query Open query Feedback Getting started

Anushka_IDA (Anushka0602)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables: dbo.emp_tbl, ven.emp_tbl
Views
Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 --create schema ven
2 --create table emp_tbl(emp_id int)
3 create table ven.emp_tbl(emp_id int)
4
```

Results Messages

Query succeeded: Affected rows: 0

Query succeeded | 0s



Anushka_IDA (idashellserver-anushka/Anushka_IDA) | Query editor (preview)



Search

Login New Query Open query Feedback Getting started

Overview

Activity log

Tags

Diagnose and solve problems

Query editor (preview)

Settings

Compute + storage

Connection strings

Properties

Locks

Data management

Replicas

Sync to other databases

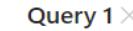
Integrations

Azure Synapse Link

Stream analytics (preview)

Add Azure Search

Anushka_IDA (Anushka0602)



Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

> dbo.Author

...

> dbo.Book

...

> dbo.customer

...

> dbo.Orders

...

> dbo.Publisher

...

Views

> Stored Procedures

Results Messages

Query succeeded: Affected rows: 0

Query succeeded | 0s

Query 1

Run

Cancel query

Save query

Export data as

Show only Editor

```
1 create table Author (Aid int primary key, A_name varchar(255),no_of_books int);
2 create table Publisher(Pid int primary key,p_name varchar(255),p_country varchar(255));
3 create table Book (bid int primary key,Author_id int foreign key references Author(Aid));
4 create table Orders (Oid int primary key,odate datetime,bid int foreign key references Book(bid));
5 create table customer(Cid int primary key,c_name varchar(255),oid int foreign key references Orders(Oid));
```

DAY-4

Azure SQL

1. Create A Database And Elastic Pool.
2. Joins:
 - a. Inner Join
 - b. Left Join
 - c. Right Join
 - d. Self Join
 - e. Outer Join
 - f. Cross Join
3. Stored Procedure
4. Views
5. Functions
 1. Scaler
 2. Table Valued
6. Complex Codes -> Stored Procedures
7. Simple Code -> Views
8. Temporary Table: # Before Table Name Creates A Temporary Table (i.e., No Physical Table Created). It Drops Automatically.
9. Sub-queries
10. UNION And UNION ALL
11. Intersection
12. INDEX
 - a. Clustered Index
 - b. Non-Clustered Index
 - c. Column Stored Index



AnushkaShell (anushkaidashell/AnushkaShell)



Search

Copy Restore Export Set server firewall Delete Connect with... Feedback

Overview

Activity log

Tags

Diagnose and solve problems

Query editor (preview)

Settings

Compute + storage

Connection strings

Properties

Locks

Data management

Replicas

Sync to other databases

Integrations

Azure Synapse Link

Stream analytics (preview)

Add Azure Search

Essentials

Resource group ([move](#)) : [RG_AnushkaIDAShell](#) Server name : [anushkaidashell.database.windows.net](#)

Status : Online Elastic pool : [AnushkaIDAShell](#)

Location : East US Connection strings : [Show database connection strings](#)

Subscription ([move](#)) : [npunext-1680261348707](#) Pricing tier : [Elastic Basic](#)

Subscription ID : 6f5f0ab3-77b8-43f8-9c0e-b251f714e513 Earliest restore point : No restore point available

Tags ([edit](#)) : [Add tags](#)

Getting started

Monitoring

Properties

Features

Notifications (1)

Integrations

Tutorials

Start working with your database

Connect to your database and start working with data with a few simple steps. [Learn more](#)



Configure access

Configure network access to your SQL server. [Learn more](#)

[Configure](#)



Connect to application

Use connection strings to connect to your SQL database from your applications and favorite tools.

[See connection strings](#)



Start developing

Work in your database by using tools to add, modify and query data. [Compare tools](#)

[Open Azure Data Studio](#)

Home > AnushkaIDAShell (anushkaidashell/AnushkaIDAShell)



AnushkaIDAShell (anushkaidashell/AnushkaIDAShell) | Query editor (preview)

X

Search

Login New Query Open query Feedback Getting started

- Overview
- Activity log
- Tags
- Diagnose and solve problems
- Query editor (preview) (selected)

Settings

- Compute + storage
- Connection strings
- Properties
- Locks

Data management

- Replicas
- Sync to other databases

Integrations

- Azure Synapse Link
- Stream analytics (preview)
- Add Azure Search

AnushkaIDAShell (Anushka)

 Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

- Tables
- Views
- Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 Select TOP (1000) *
2 FROM [SalesLT].[Customer]
3 INNER JOIN [SalesLT].[CustomerAddress]
4 ON [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

Results Messages

Search to filter items...

CustomerID	NameStyle	Title	FirstName
29485	False	Ms.	Catherine
29486	False	Ms.	Kim
29489	False	Ms.	Frances
29490	False	Ms.	Margaret

Query succeeded | 5s

Query 1 ×

▶ Run Cancel query ⏪ Save query ⏪ Export data as Show only Editor

```
1  SELECT MAX([SalesLT].[Product].ListPrice)
2  FROM [SalesLT].[Product]
3  WHERE [SalesLT].[Product].ListPrice < (SELECT MAX([SalesLT].[Product].ListPrice) FROM [SalesLT].[Product])
```

Results Messages

🔍 Search to filter items...

3399.9900

Query 1 × Query 2 × Query 3 × Query 4 ×[Run](#) [Cancel query](#) [Save query](#) [Export data as](#) [Show only Editor](#)

```
1  Select TOP (1000) *
2  FROM [SalesLT].[Customer]
3  RIGHT JOIN [SalesLT].[CustomerAddress]
4  ON [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

[Results](#) [Messages](#) Search to filter items...

CustomerID	NameStyle	Title	FirstName	MiddleName
29485	False	Ms.	Catherine	R.
29486	False	Ms.	Kim	
29489	False	Ms.	Frances	B.

[Query succeeded | 1s](#)

Query 1 × Query 2 ×

[Run](#) [Cancel query](#) [Save query](#) [Export data as](#) [Show only Editor](#)

```
1  Select TOP (1000) *
2  FROM [SalesLT].[Customer]
3  LEFT JOIN [SalesLT].[CustomerAddress]
4  ON [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

[Results](#) [Messages](#) Search to filter items...

CustomerID	NameStyle	Title	FirstName	MiddleName
1	False	Mr.	Orlando	N.
2	False	Mr.	Keith	
3	False	Ms.	Donna	F.

Query succeeded | 17s

Query 1 X Query 2 X Query 3 X[Run](#) [Cancel query](#) [Save query](#) [Export data as](#) [Show only Editor](#)

```
1 Select TOP (1000) *
2 FROM [SalesLT].[Customer]
3 CROSS JOIN [SalesLT].[CustomerAddress]
4
```

[Results](#) [Messages](#) Search to filter items...

CustomerID	NameStyle	Title	FirstName	MiddleName
1	False	Mr.	Orlando	N.
1	False	Mr.	Orlando	N.
1	False	Mr.	Orlando	N.

Query succeeded | 2s



JOURNEY SCRAPBOOK

Custom Bootcamp (29th August – 6th October)

Week-2

Anushka Dhir

Batch-6

DAY-1

Azure Storage Types- Day 1

1. Create A Resource -> Azure Storage Account
2. Resource Group
3. Storage Account
4. Storage Services
5. Redundancy
 1. Local Redundancy
 2. Zone Redundancy
 3. Geo Redundancy
 4. Geo Zone Redundancy
6. RAGRS (Read Access Geo Redundant Storage)
 1. Failover
 2. Synchronous Replication
7. Security
8. Access Protocols
9. Access Tiers
 1. Hot
 2. Cold
 3. Cool
 4. Archive Access
10. Purging & Archiving
11. Containers - Folder For Storing Files
12. Azure Tables
13. Azure Queues
14. File Shares
15. Snapshots
16. Ingress & Egress
17. Static Website

DAY-2

Azure Storage Types- Day 2
& Azure Data Factory

1. Access Keys : Lifetime
2. SAS (Shared Access Signature) : Time-bounded
3. Azure Key Vaults
4. Access Control (IAM) : Check Access Here
5. Azure Data Factory :
 - a. Data Engineer Activities:
 - i. ETL (Extract, Transform, Load)
 - ii. ELT
 - iii. Azure Databricks - Transformation
 - b. Data Orchestration
 - c. ADF Components:
 - i. Pipelines : Combination Of One Or More Activities
 - ii. Activities
 - iii. Datasets
 - iv. Linked Services : To Access Particular Directory This Is Used.
 - v. Integration Runtime : Gateway Or Bridge To Connect For Data Extraction.
 1. Azure Auto Resolved (Default)
 2. Self-hosted
 3. Azure SSIS
 - vi. Triggers
 - vii. Data Flow
 - d. Parameterization
 - e. Pipelines

anushkashell | Shared access signature

Storage account

Search

Give feedback

Events

Storage browser

Data storage

Containers

File shares

Queues

Tables

Security + networking

Networking

Access keys

Encryption

Microsoft Defender for Cloud

Data management

Redundancy

Data protection

blob inventory

Generate SAS and connection string

Signing key

Connection string

SAS token

Blob service SAS URL

File service SAS URL

Queue service SAS URL

Table service SAS URL

https://anushkashell.blob.core.windows.net/?sv=2023-09-05t115026z&st=2023-09-05t035020z&sp=https&sr=MsyWAM57WAjw&sig=1a4e16412e0Cw/zB2Zf2...&t=

https://anushkashell.queue.core.windows.net/?sv=2023-09-05t115026z&st=2023-09-05t035020z&sp=https&sr=MsyWAM57WAjw&sig=1a4e16412e0Cw/zB2Zf2...&t=

https://anushkashell.table.core.windows.net/?sv=2023-11-02t0ss+bfqf0rt+clsp+rdlacupy&se=2023-09-05t115026z&st=2023-09-05t035026z&sp=https&sr=MsyWAM57WAjw&sig=1a4e16412e0Cw/zB2Zf2...&t=

https://anushkashell.blob.core.windows.net/?sv=2023-11-02t0ss+bfqf0rt+clsp+rdlacupy&se=2023-09-05t115026z&st=2023-09-05t035026z&sp=https&sr=MsyWAM57WAjw&sig=1a4e16412e0Cw/zB2Zf2...&t=

https://anushkashell.queue.core.windows.net/?sv=2023-11-02t0ss+bfqf0rt+clsp+rdlacupy&se=2023-09-05t115026z&st=2023-09-05t035026z&sp=https&sr=MsyWAM57WAjw&sig=1a4e16412e0Cw/zB2Zf2...&t=

https://anushkashell.table.core.windows.net/?sv=2023-11-02t0ss+bfqf0rt+clsp+rdlacupy&se=2023-09-05t115026z&st=2023-09-05t035026z&sp=https&sr=MsyWAM57WAjw&sig=1a4e16412e0Cw/zB2Zf2...&t=

keyuser | Overview

key vault

Search

Delete Move Refresh Open in mobile

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

Objects

Keys

Secrets

Certificates

Settings

Access configuration

Networking

Microsoft Defender for Cloud

Properties

Locks

Manage keys and secrets used by apps and services

Control access to key vault

Assign access policy and determine whether a given service principal, namely an application or user group, can perform different operations on key vault keys, secrets or certificates.

keyuser | Secrets

key vault

Search

Generate/Import Refresh Restore Backup View sample code Manage deleted secrets

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

Objects

Keys

Secrets

Certificates

Settings

Access configuration

Networking

Microsoft Defender for Cloud

Properties

Locks

Name Type Status Expiration date

TestSecret Enabled

keyuser | Access Control (IAM)

Storage account

Search

Add Download role assignments Edit columns Refresh Remove Feedback

My access

Check access

Role assignments Roles Deny assignments Classic administrators

anushkashell | Access Control (IAM)

Storage account

Search

Add Download role assignments Edit columns Refresh Remove Feedback

My access

Check access

Role assignments Roles Deny assignments Classic administrators

Shellunext unextIDA90 assignments - anushkashell

Current role assignments Eligible assignments

Assignments for the selected user, group, service principal, or managed identity at this scope or inherited to this scope.

Search by assignment name or description

Role assignments (1)

Role Description Scope Group assignment Condition

Owner Grants full access to all resources Subscription (Inherited) None

Deny assignments (0)

Role Description

Classic administrators (0)

Description

User is not a classic administrator.

Microsoft Azure Data Factory

Search factory and documentation

Validate all Publish all

Preview experience Off

Activities

Move and transform

Copy data

File flow

Synapse

Notebook

Spark job definition

Azure Data Explorer

Azure Data Explorer C...

Azure Function

Azure Function C...

Batch Service

General Source Sink Mapping Settings User properties

Name * Copy data1 Learn more

Description

Activity state (preview)

Active Inactive

Timeout 0:00:00

DAY-3

Azure Data Factory - Day 1

1.Create Data Factory

2.Activities In Pipeline :

- a. String, Boolean, Array And Integer Activities In Pipelines
- b. Execute Pipeline Activity
- c. SSIS Package (Azure SSIS IR - Required For This Package)
- d. Get Metadata Activity

1.Copy Data Activity Implementation

2.Parameter Concept Implementation

3.Triggers

4.Get Metadata Activity

7.If-else Condition For Get Metadata In Pipelines.

8.For Each Loop

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

All services > Microsoft.DataFactory-20230906091019 | Overview >

df-anushkashell

Data factory (V2)

Search

Delete

Essentials

Resource group (move) : RG_Anushka
Status : Succeeded
Location : East US
Subscription (move) : npunext-1680261348707
Subscription ID : 6f5f0ab3-77b8-43f8-9c0e-b251f714e513

Type : Data factory (V2)
Getting started : Quick start

Azure Data Factory Studio

Launch studio

Quick Starts

Tutorials

Template Gallery

Training Modules

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Alerts

Metrics

Diagnostic settings

Logs

Monitoring

https://portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/6f5f0ab3-77b8-43f8-9c0e-b251f714e513/resourceGroups/RG_Anushka/providers/Microsoft.DataFactory/factories/df-anush...

Microsoft Azure | Data Factory > df-anushkashell

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience Off

Factory Resources

Pipelines

Activities

DS_Source

Output

Validate all

PL_IDA_01

PL_IDA_01

DS_source

Output

Get Metadata

Set variable

Copy data

Get Metadata1

Set variable1

Activities

Change Data Capture (preview)

Datasets

Data flows

Power Query

Append variable

Delete

Execute Pipeline

Execute SSIS package

Fail

Get Metadata

Lookup

Stored procedure

Script

(X) Set variable

Validation

Web

WebHook

Wait

HDInsight

Iteration & conditionals

Machine Learning

Parameters Variables Settings Output

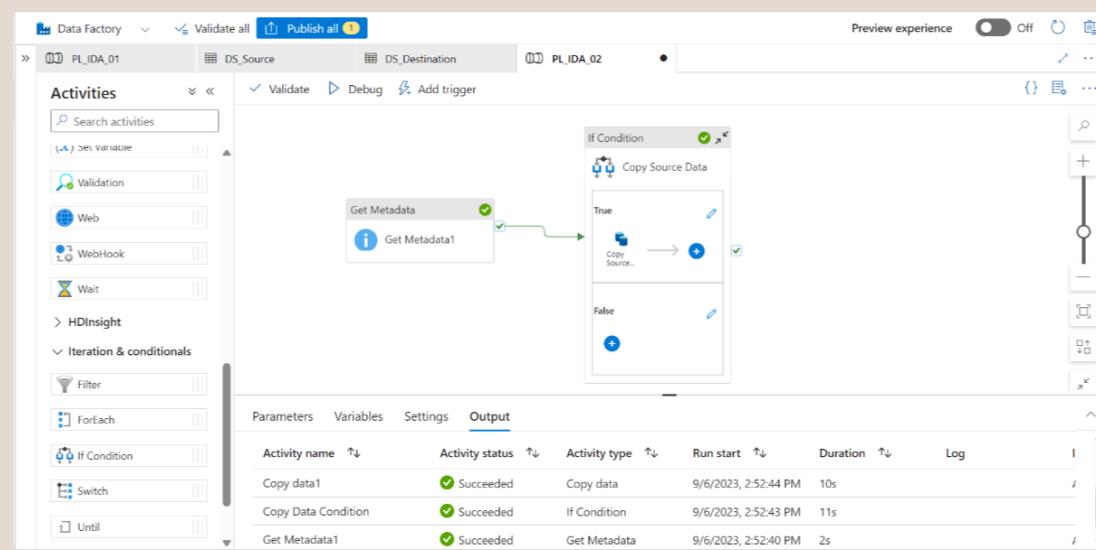
Pipeline run ID: d45d2417-5ff1-426f-bb88-87c17d7123ea

Pipeline status: Succeeded

All status

Showing 1 - 3 of 3 items

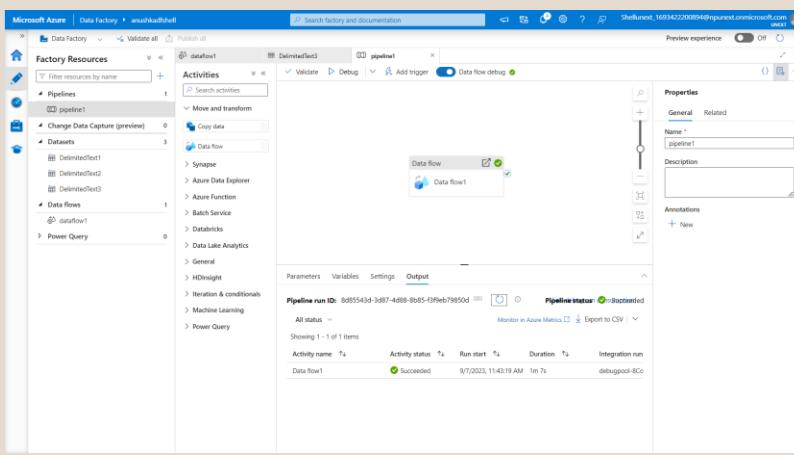
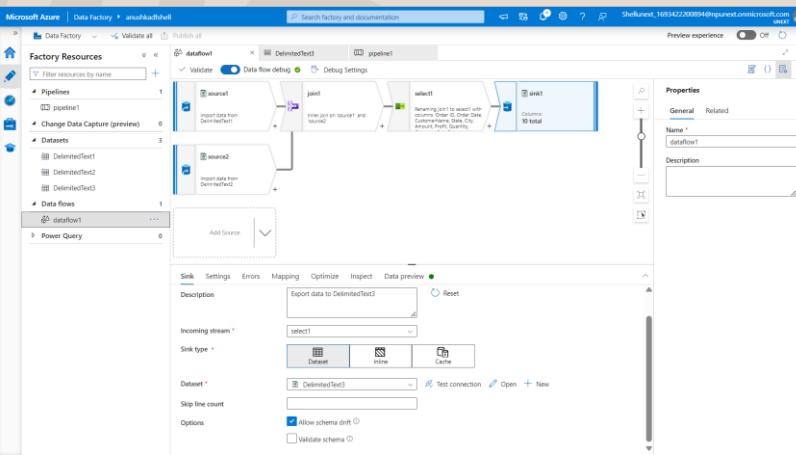
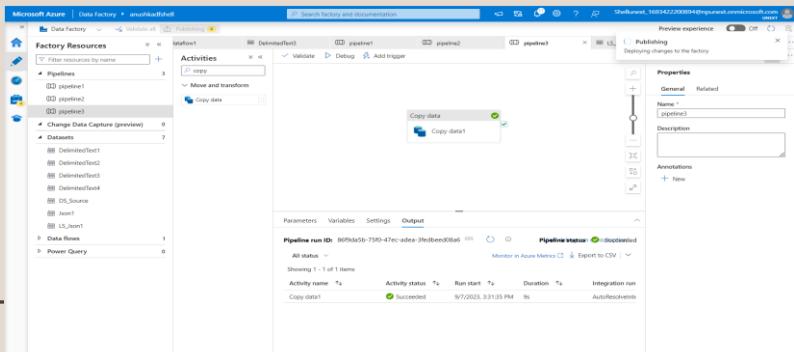
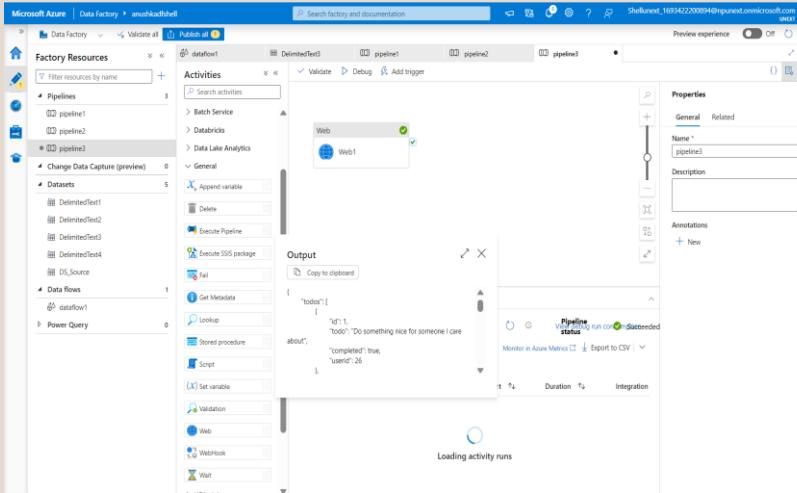
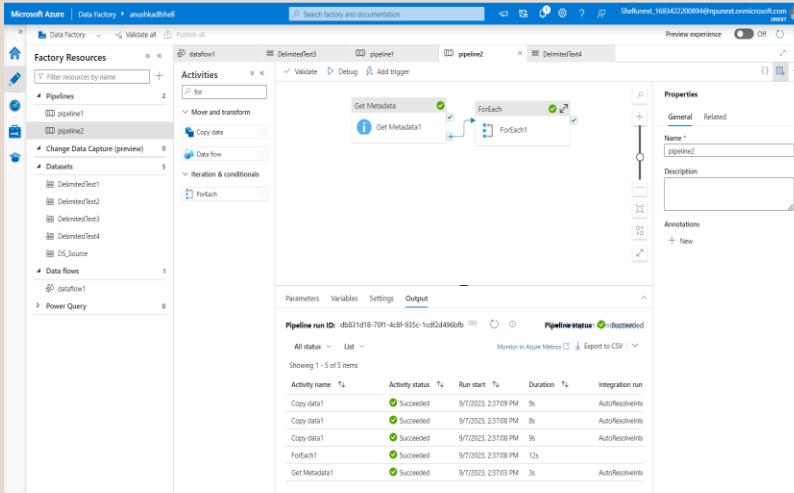
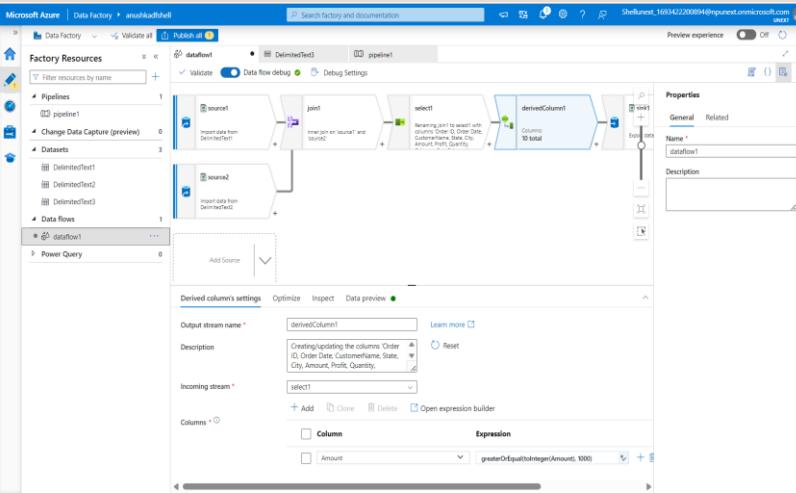
Activity name	Activity status	Activity type	Run start	Duration	Log	Integration runt
Set variable1	Succeeded	Set variable	9/6/2023, 2:45:18 PM	Less than 1s		
Get Metadata1	Succeeded	Get Metadata	9/6/2023, 2:45:13 PM	4s		AutoResolveInteg
Copy data1	Succeeded	Copy data	9/6/2023, 2:45:03 PM	9s		AutoResolveInteg



DAY-4

Azure Data Factory - Day 2

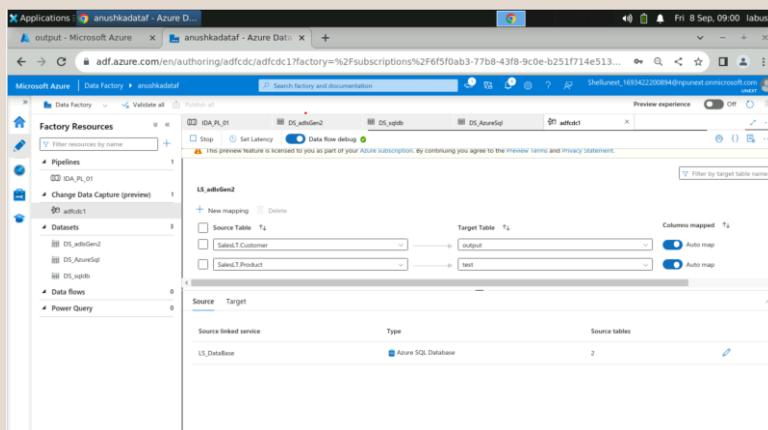
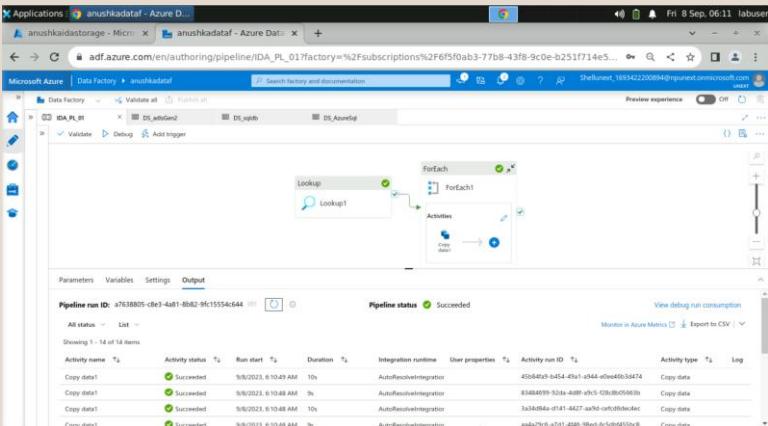
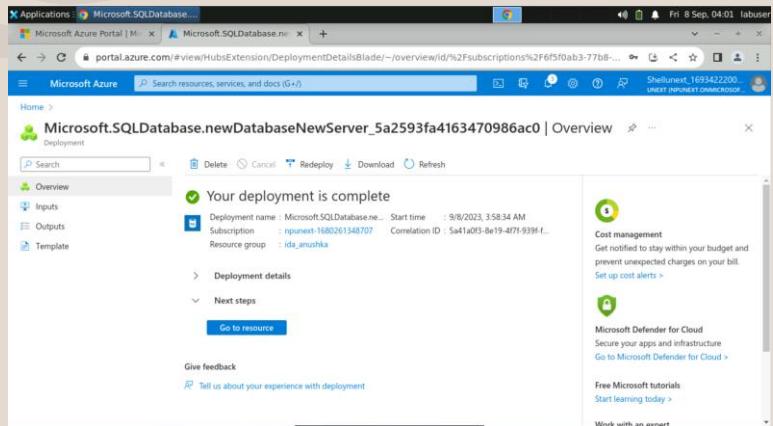
1. Data(resource)transformation Using Data Flow.
2. For Each Activity
3. Data Flows
4. Dataflow Expression Builder And Derived Columns
5. Surrogate Key
6. For Each Container And Get Metadata
7. API

DAY-5

Azure Data Factory - Day 3

- 1.Create SQL Database On VM.
- 2.Lookup Table In Sql Database
- 1.Lookup Table With Foreach Pipeline
- 4.CDC (Change Data Capture)
- 5.Runbook Or Automation





JOURNEY SCRAPBOOK

Custom Bootcamp (29th August – 6th October)

Week-3

Anushka Dhir

Batch-6

DAY-1

Azure Synapse Analytics

1. Azure Synapse Analytics : Used To Create ADF Services, Spark Notebooks, SQL Data Warehouse
2. SQL Pool
 - Serverless
 - Dedicated Pool
3. Spark Pool
4. Data Exploration
5. SQL Database
6. Collate
7. Common Table Expression (CTE)
8. View
9. DWU (Data Warehousing Unit)
10. Storage Node
11. Compute Node
12. Data Movement Service (DMS)
13. Massive Parallel Processing
14. Replicated Tables
15. Hash Distributed Table
16. Implementing Partitions For An SQL Data Warehouse

Microsoft Azure | Search resources, services, and docs (F1)

Home > Microsoft.Azure.SynapseAnalytics-20230912034630 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20230912034630 Start time : 9/12/2023, 3:56:42 AM
Subscription : rpuwest188026134877 Correlation ID : 3eab269-82d0-4cb6-b6fc-bc47b0ffa78
Resource group : RG-Anushka

Deployment details

Next steps

Go to resource group

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 >

Applications : anushkasynapse - Azure... Tue 12 Sep, 04:58 labus

Microsoft Azure Portal | Microsoft Edge | anushkasynapse - Microsoft Edge | anushkasynapse - Azure | +

web.azuressynapse.net/en/authoring/explore/workspace/sqlscripts/SQL%20script%20?workspace=%2Fsubscriptions%2F6... Tue 12 Sep, 04:58 labus

Synapse Analytics - anushkasynapse

Microsoft Azure | Synapse Analytics | anushkasynapse | Search

Synapse live | Validate all | Publish all

Data Workspace Linked

Filter resources by name

SQL database

anushka (SQL)

External tables

External resources

Views

Schemas

Security

math (SQL)

SQL script 1

Run Undo Publish Query plan Connect to Built-in Use database master

alter database anushka collate Latin1_general_100 BIN2 UTF8

Results Messages

No results to show

Your query yielded no displayable results

1000011 Queries executed successfully.

The screenshot displays two separate browser windows for Microsoft Azure Synapse Analytics, both titled "anushkasynapse - Azure...".

Top Window:

- Shows a "Data" workspace.
- Contains a query titled "SQL script 1" using the "anushkafile" database.
- The query uses BULK INSERT to load data from a CSV file into a table named "zipcodes".
- The results show 10 rows of data with columns: RecordNumber, Zipcode, ZipCodeType, City, State, LocationType, Lat, Long, Xaxis, Yaxis, and Zaxis.
- A message at the bottom indicates "00:01:14 Query executed successfully".

Bottom Window:

- Shows a "Data" workspace.
- Contains a query titled "SQL script 2" using the "anushkafile" database.
- The query uses BULK INSERT to load data from a CSV file into a table named "zipcodes".
- The results show 10 rows of data with columns: RecordNumber, Zipcode, ZipCodeType, City, State, LocationType, Lat, Long, Xaxis, Yaxis, and Zaxis.
- A message at the bottom indicates "00:01:14 Query executed successfully".

The screenshot shows a Microsoft Azure Synapse Analytics workspace titled 'anushkasynapse - Azure...'. The left sidebar displays the 'Data' workspace, with a 'Linked' section showing a connection to 'anushka'. The main area contains a code editor with the following SQL script:

```
14 Select * from cte test1/
15
16 CREATE view anushka_view as
17     SELECT
18         TOP 100 *
19     FROM
20     OPENROWSET(
21         BULK 'https://anushkaautomationshell.dfs.core.windows.net/anushkafile/zipcodes.csv',
22         FORMAT = 'CSV',
23         PARSER_VERSION = '2.0',
24         HEADER_ROW = TRUE
25     ) AS [result]
26
27 |
```

The 'Results' tab is selected, showing the message 'No results to show' and 'Your query yielded no displayable results'. To the right, the 'Properties' panel shows the script is named 'SQL script 2' and has a size of 241 bytes. The 'Results settings per query' section is set to 'First 5000 rows (default)'.

DAY-2

Data Visualization - Power BI

1. PBI Desktop
2. Power BI Supports Multiple Data Connectors.
3. Power Query Editor: Used To Transform Data
 - Language Used At Back-end - M Language
 - Different Fields Of This Editor
4. Get Data In PBI Report And Transform It In Power Query Editor
5. Getting Familiar With Different Visuals
6. Parameter
7. Filters :
 - Visual Level
 - Page Level
 - Report Level
8. Slicers
9. DAX Measures
10. DAX Calculated Columns
11. RLS (Manage Roles)

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Column From Custom Invoke Custom Function General

Queries [1] financials

= Table.AddColumn(#"Changed Type3", "Short_Name", each if [Country] = "United States of America" then

#	Year	L2 Total_GrossSales	SngCountry	Short_Name
1	2014	32390	Government Canada	Canada
2	2014	26440	Government Germany	Germany
3	2014	32685	Midmarket France	France
4	2014	13335	Midmarket Germany	Germany
5	2014	37065	Midmarket Mexico	Mexico
6	2014	529600	Government Germany	Germany
7	2014	15830	Midmarket Germany	Germany
8	2014	30228	Channel Partners Canada	Canada
9	2014	38000	Government France	France
10	2014	18552	Channel Partners Germany	Germany
11	2014	37065	Midmarket Mexico	Mexico
12	2014	33312,5	Enterprise Canada	Canada
13	2014	287700	Small Business Mexico	Mexico
14	2014	15029	Government Germany	Germany
15	2013	43250	Enterprise Canada	Canada
16	2014	9240	Midmarket United States of America	USA
17	2014	5860	Government Canada	Canada
18	2014	14625	Midmarket Mexico	Mexico
19	2014	301798	Channel Partners Canada	Canada

19 COLUMNS 700 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 9:56 AM 11:44 AM 9:53/2023

The screenshot shows the Microsoft Power BI Desktop interface. The ribbon menu includes Home, Insert, Modeling, View, Help, Format, Data / Drill, and a context menu for the selected visualization. The Home tab is active. The main workspace displays two visualizations: a pie chart titled 'Sum of Sales by Country' and a stacked bar chart titled 'Sum of Discounts by Country and Discount Band'. The pie chart shows sales distribution across five countries. The bar chart shows the sum of discounts for each country, categorized by discount band (High, Low, Medium, None). The right side of the screen features the 'Visualizations' pane, which lists various financial-related measures and fields, and the 'Fields' pane, which provides detailed information about the selected visual elements.

The screenshot shows the Microsoft Power Query Editor interface. The ribbon at the top includes Home, Transform, Add Column, View, Tools, and Help. A URL bar at the top left shows the address: https://wai.nuvepro.com/guacamole/#/client/a50WYmFKYTRjN2M3NDHZGE3ZBAG51dmVsW5r7ho... . The main area displays a table with columns: Segment, Country, Product, Discount Band, and Units Sold. A formula bar at the top indicates a transformation: #Table.AddColumn(#"Changed Type3", "Short_Name", each if [Country] = "United States of America" then "USA" else [Country]). The right pane contains sections for Properties (with Name set to financials) and Applied Steps (listing Source, Navigation, Changed Type, Replaced Value, Changed Type1, Added Custom, Changed Type2, Added Custom1, Changed Type3, and Added Conditional Column). The bottom status bar shows 19 COLUMNS, 700 ROWS, and a preview download time of 2:19 PM.

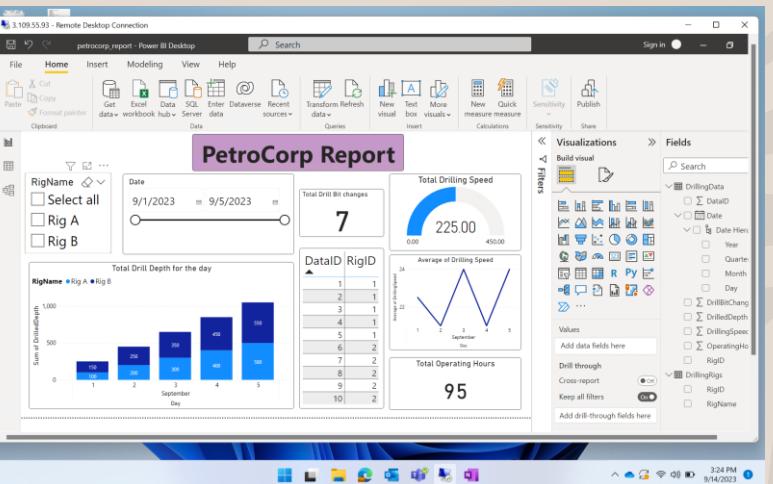
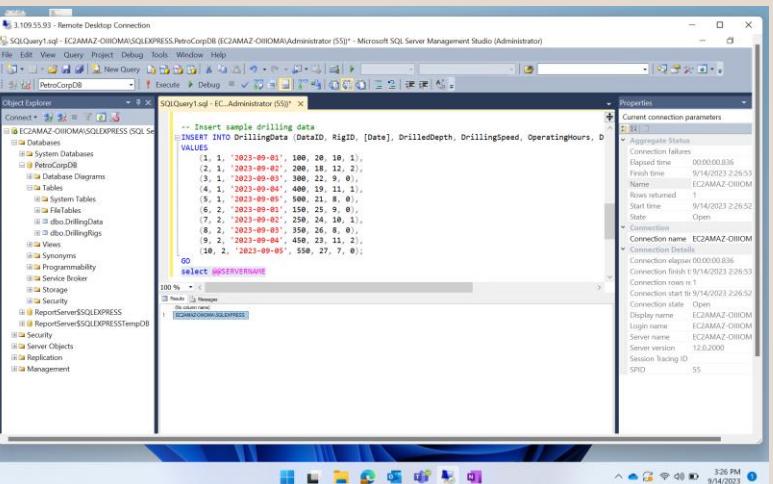
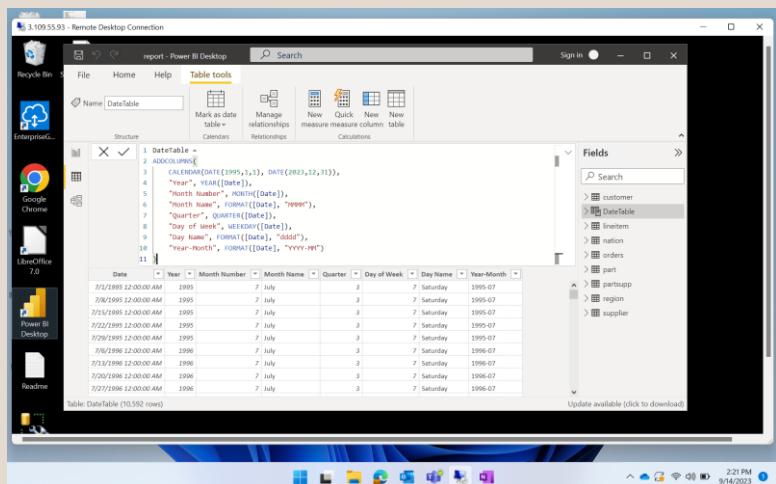
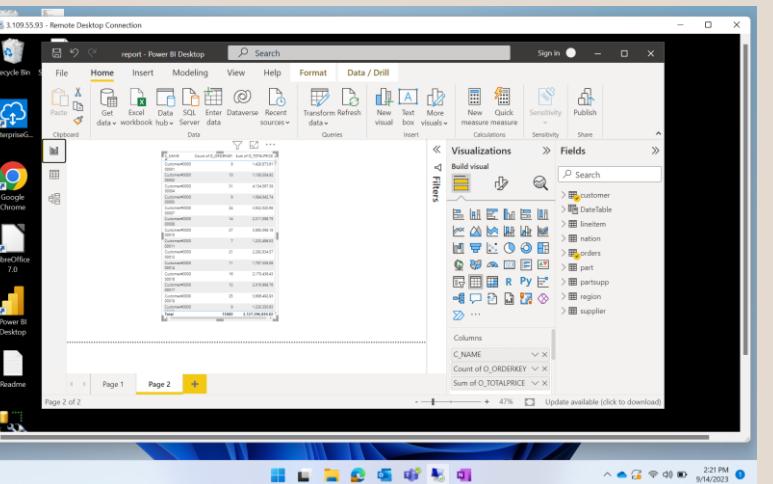
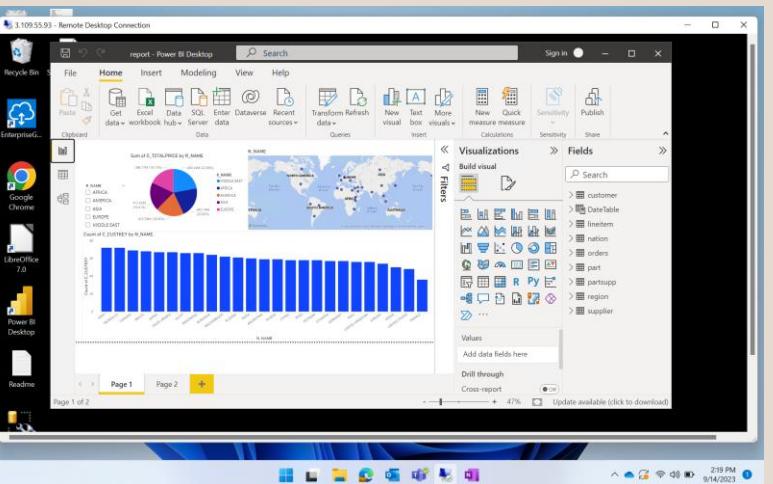
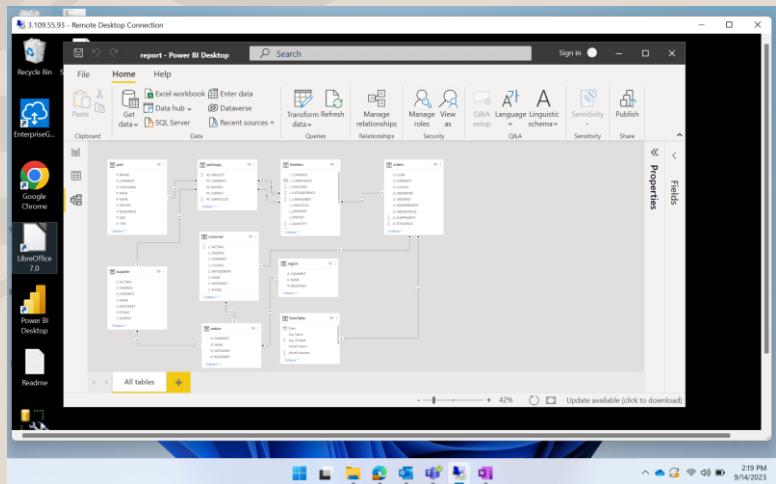
The screenshot shows a Power BI Desktop interface with the following details:

- Title Bar:** Untitled - Power BI Desktop
- Address Bar:** https://wai.nuvepro.com/guacamole/#/client/a50WmFyTRjn2M3NDHZG32ABjG51dmVsAtW5r?ho...
- Header:** File, Home, Insert, Modeling, View, Help, Format, Data / Drill, Table tools
- Table Tools Tab:** Name (Financials), Mark as date table, Manage relationships, New, Quick measure, New measure column, New table, Calculations.
- Structure Tab:** Structure, Mark as date table, Calendars, Relationships.
- Visuals:**
 - Pie Chart:** Sales by Country
 - United States: 25.00M (37.6%)
 - Canada: 21.00M (31.8%)
 - France: 20.50M (30.3%)
 - Mexico: 21.25M (31.9%)
 - Bar Chart:** Sales of Domestic vs. International and Payment Method
 - Legend: Domestic (Orange), International (Blue)
 - Y-axis: Sales (M)
 - X-axis: Country
 - United States of America: Domestic 2.1M, International 1.8M
 - Canada: Domestic 2.1M, International 2.2M
 - Mexico: Domestic 2.0M, International 2.0M
 - France: Domestic 2.1M, International 2.0M
 - Germany: Domestic 2.1M, International 2.0M
- Filters:** Filters on this visual, Filters on this page, Filters on all pages, Add data fields here.
- Visualizations:** Format visual, Slicer settings, Add data fields here.
- Fields:** Fields (Financials), Search, Add data fields here.
- Page Control:** Page 1 of 1, Next page.
- Bottom Bar:** Update available (click to download), 2:57 PM

DAY-3

Power BI - Day 2

- 1.Data Modeling : Establishing Relationships Between Tables
- 2.Univariant
- 3.Bivariant
- 4.Multivariant
- 5.Drill Through
- 6.Date Table
- 7.Connection Of SQL Server And Power BI



PetroCorp Report

Rig Name

Rig A

Rig B

Date

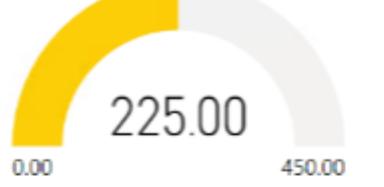
9/1/2023  9/5/2023 



Total Drill Bit changes

7

Total Drilling Speed

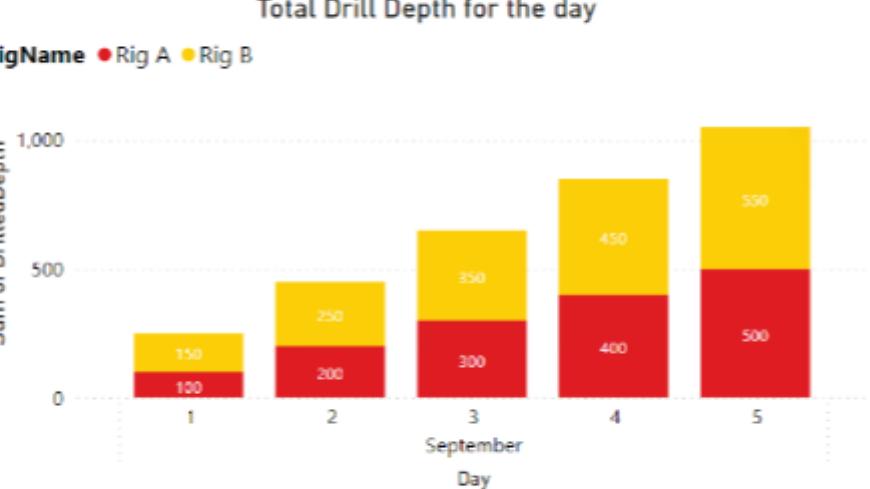


225.00

0.00 450.00

Total Drill Depth for the day

RigName • Rig A • Rig B

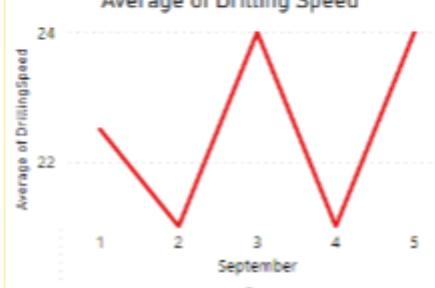


Day	Rig A	Rig B	Total
1	100	150	250
2	200	250	450
3	300	350	650
4	400	450	850
5	500	550	1050

DataID | RigID

DataID	RigID
10	2
9	2
8	2
7	2
6	2
5	1
4	1
3	1
2	1
1	1

Average of Drilling Speed



24

22

1 2 3 4 5

September Day

22.5 23.5 24.5

Average Drilling Speed

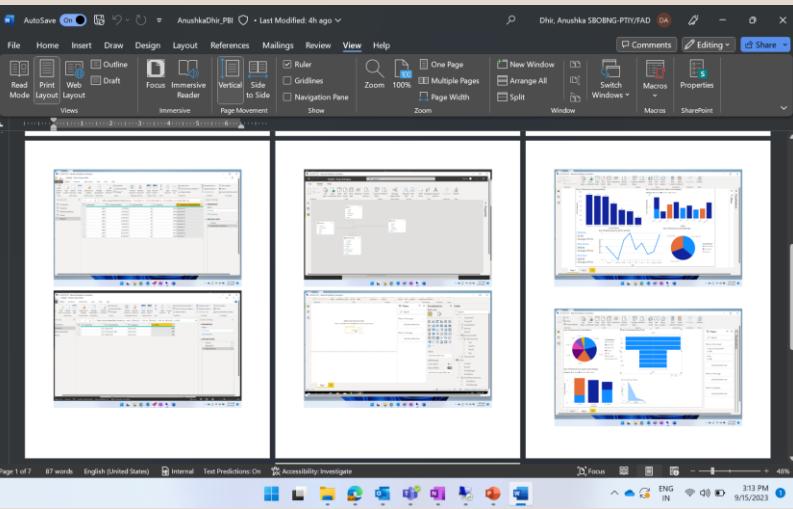
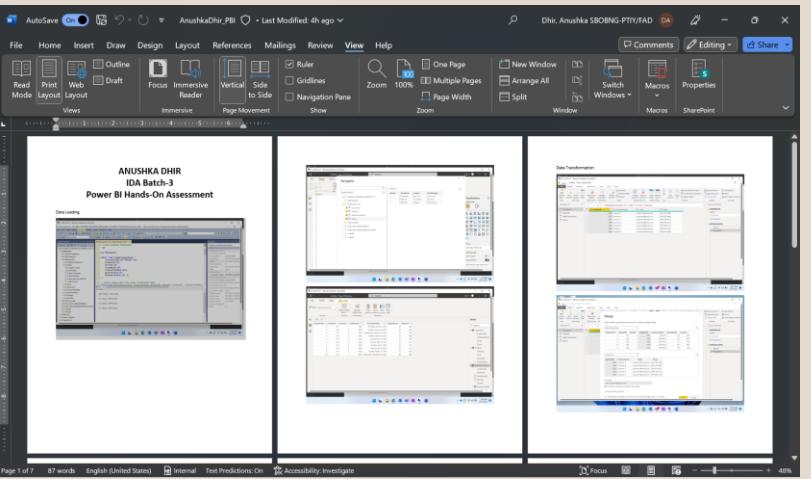
Total Operating Hours

95

DAY-4

Python – Day 1

1. Hands-on Assessment - Power BI
2. Jupyter Notebook Launch
3. Datatypes In Python:
 - Integer
 - Float
 - String
 - List
 - Dictionary
 - Tuple
 - Boolean
4. Arithmetic
5. Logic



```
In [1]: # Integers
age = 20
print("Age is ", age)

# Floating Point Numbers
price = 19.99
print("Price is ", price)

# Strings
name = "Alice"
message = "Hello, World!"
print("My name is ", name)

# Boolean values
is_python_installed = True
is_learning = False

# Lists
numbers = [1, 2, 3, 4, 5]
fruits = ["apple", "banana", "cherry"]

# Tuples
coordinates = (3, 4)

# Dictionaries
person = {"first": "Alice", "age": 20, "city": "New York"}
grades = {"math": 90, "science": 85, "history": 80}

# Accessing elements in lists and dictionaries
first_number = numbers[0]
last_fruit = fruits[-1]
name1 = person["first"]
math_grade = grades["math"]

print(first_number)
print(last_fruit)
print(name1)
print(math_grade)

Age is 20
Price is 19.99
My name is Alice
1
cherry
Alice
92
```

```
In [1]: #!/usr/bin/python
# This script calculates profit based on user satisfaction and price.

class User:
    profit = False
    user_satisfaction = None

    def __init__(self, profit, low_price, user_satisfaction):
        self.profit = profit
        self.low_price = low_price
        self.user_satisfaction = user_satisfaction

    def set_low_price(self):
        self.low_price = profit

    def set_user_satisfaction(self):
        self.user_satisfaction = profit

    def calculate_profit(self, low_price):
        self.low_price = low_price
        self.user_satisfaction = profit

    def get_low_price(self):
        return self.low_price

    def get_user_satisfaction(self):
        return self.user_satisfaction

    def __str__(self):
        return f"User Profit: {self.profit}, Low Price: {self.low_price}, Satisfaction: {self.user_satisfaction}"
```

```
In [2]: #!/usr/bin/python
# This script calculates profit based on user satisfaction and price.

Region = [Region(5, reg, 0, reg, 0) for reg in range(10)]
for i in range(len(Region)):
    if Region[i].low_price < Region[i].user_satisfaction:
        Region[i].profit = True
    else:
        Region[i].profit = False
```

```
In [3]: print([str(region) for region in Region])
```



JOURNEY SCRAPBOOK

Custom Bootcamp (29th August – 6th October)

Week-4

Anushka Dhir

Batch-6

DAY-1

Python – Day 2

1. Function
2. Function Composition
3. Recursive Function
4. Generator Function
5. Decorator (Wrap Function)
6. Types Of Argument
7. Class
8. Date Stripping
9. Date Conversion
10. Exception Handling
11. List Comprehension
12. Lambda Function

11.232.140.15 - Remote Desktop Connection

jupyter Untitled 1 checkpoint [closed]

```
In [1]: print(a)
```

```
In [2]: def f(a):  
    print(a)
```

```
In [3]: f(1)
```

```
In [4]: def f(gas,constant,temperature):  
    vol=pressure*1.0/(gas*constant*temperature)  
    print(vol)
```

```
In [5]: gas.vol_func(10000,0.214,300)
```

```
In [6]: print(gas.vol_func(10000,0.214,300))
```

```
In [7]: Gas.vol_func(10000,0.214,300)
```

```
In [8]: Gas.vol_func(10000,0.214,300,"kg")
```

```
In [9]: 176.499295218 kg
```

```
In [10]: def fact(n):  
    if n==1:  
        return 1  
    else:  
        return n*fact(n-1)
```

```
In [11]: print(fact(5))
```

```
In [12]: def calc_tot_depth(segments):  
    if not segments:  
        return 0  
    else:  
        curr_seg = segments[0]  
        remaining_seg = segments[1:]
```

Remaining: 30 hrs (51 mins) 14:44 PM 9/19/2023

11.232.140.15 - Remote Desktop Connection

jupyter Untitled 1 checkpoint [closed]

```
In [13]: print(a)
```

```
In [14]: def f(a):  
    print(a)
```

```
In [15]: f(1)
```

```
In [16]: def f(gas,constant,temperature):  
    vol=pressure*1.0/(gas*constant*temperature)  
    print(vol)
```

```
In [17]: gas.vol_func(10000,0.214,300,"kg")
```

```
In [18]: 176.499295218 kg
```

```
In [19]: def fact(n):  
    if n==1:  
        return 1  
    else:  
        return n*fact(n-1)
```

```
In [20]: print(fact(5))
```

```
In [21]: def generate_segments():  
    for i in range(1,11):  
        for j in range(i+1,11):  
            print(i,j)
```

Click Notes 14:44 PM 9/19/2023

11.232.140.15 - Remote Desktop Connection

jupyter Untitled 1 checkpoint [closed]

```
In [22]: print(a)
```

```
In [23]: def calculate_yearly_production(yearlyProduction):  
    months = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"]  
    for month in months:  
        x = datetime.datetime(2021, month, 1)  
        print(x.strftime("%Y-%m-%d"), yearlyProduction["yearlyProduction"][month])  
        month += 1
```

```
In [24]: for i in calculate_yearly_production(2000):  
    print(i)
```

```
In [25]: Production in January is 1000.0  
Production in February is 1000.0  
Production in March is 1000.0  
Production in April is 1000.0  
Production in May is 1000.0  
Production in June is 1000.0  
Production in July is 1000.0  
Production in August is 1000.0  
Production in September is 1000.0  
Production in October is 1000.0  
Production in November is 1000.0  
Production in December is 1000.0
```

```
In [26]: def my_decorator(func):  
    def wrapper(*args, **kwargs):  
        print("I am starting")  
        func(*args, **kwargs)  
        print("I am completed")  
    return wrapper
```

```
In [27]: def calculate_yearly_production(yearlyProduction):  
    months = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"]  
    for month in months:  
        x = datetime.datetime(2021, month, 1)  
        print(x.strftime("%Y-%m-%d"), yearlyProduction["yearlyProduction"][month])  
        month += 1
```

In [28]: I am starting
I am completed 14:44 PM 9/19/2023

11.232.140.15 - Remote Desktop Connection

jupyter Untitled 1 checkpoint [closed]

```
In [29]: print(a)
```

```
In [30]: Production in January is 1000.0  
Production in February is 1000.0  
Production in March is 1000.0  
Production in April is 1000.0  
Production in May is 1000.0  
Production in June is 1000.0  
Production in July is 1000.0  
Production in August is 1000.0  
Production in September is 1000.0  
Production in October is 1000.0  
Production in November is 1000.0  
Production in December is 1000.0
```

```
In [31]: def my_decorator(func):  
    def wrapper(*args, **kwargs):  
        print("I am starting")  
        func(*args, **kwargs)  
        print("I am completed")  
    return wrapper
```

```
In [32]: def calculate_yearly_production(yearlyProduction):  
    months = ["January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"]  
    for month in months:  
        x = datetime.datetime(2021, month, 1)  
        print(x.strftime("%Y-%m-%d"), yearlyProduction["yearlyProduction"][month])  
        month += 1
```

Perumanaithel, Sumita D. 14:44 PM 9/19/2023

11.232.140.15 - Remote Desktop Connection

jupyter Untitled 1 checkpoint [closed]

```
In [33]: print(a)
```

```
In [34]: class MyClass:  
    def __init__(self,p,t):  
        self.p=p  
        self.t=t  
    def __str__(self):  
        return self.p,self.t  
    def __add__(self,o):  
        return self.p+o.p, self.t+o.t  
    def __mul__(self,o):  
        return self.p*o.p, self.t*o.t  
    def factorial(self,n):  
        if n==1:  
            return 1  
        else:  
            return n*factorial(n-1)
```

```
In [35]: def my_decorator(func):  
    def wrapper(*args, **kwargs):  
        if not segments:  
            print("No segments")  
        else:  
            curr_seg = segments[0]  
            remaining_seg = segments[1:]
```

```
In [36]: x = MyClass()  
print(x.vol(10000,300))  
print(x.vol(10000,300))  
print(x.vol(10000,300))  
print(x.vol(10000,300))
```

```
In [37]: 4.000917996482  
3.0000000000000004  
120  
4
```

In [38]: In [38]: 14:44 PM 9/19/2023

11.232.140.15 - Remote Desktop Connection

jupyter Untitled 1 checkpoint [closed]

```
In [39]: print(a)
```

```
In [40]: a = MyClass()  
print(a.vol(10000,300))  
print(a.vol(10000,300))  
print(a.vol(10000,300))  
print(a.vol(10000,300))
```

```
In [41]: 4.000917996482  
3.0000000000000004  
120  
4
```

```
In [42]: import time  
curr_time=time.time()  
curr_time=datetime.fromtimestamp(curr_time).strftime('%Y-%m-%d %H:%M:%S')  
print(curr_time)
```

```
In [43]: 2023-09-19 14:44:40
```

```
In [44]: from datetime import datetime  
curr_datetime=datetime.now()  
print(curr_datetime)  
curr_datetime=datetime.now().strftime('%Y')  
print(curr_datetime)
```

```
In [45]: 2023
```

```
In [46]: from datetime import datetime  
curr_datetime=datetime.now()  
print(curr_datetime)  
curr_datetime=datetime.now().strftime('%H')  
print(curr_datetime)
```

```
In [47]: 09
```

```
In [48]: a=MyClass()  
b=MyClass()  
try:  
    result=a+b  
    print(result)  
except:  
    print("Error : Someone divided by zero")
```

```
In [49]: Error : Someone divided by zero
```

In [50]: a = MyClass()
b = MyClass()
try:
 result = a + b
 print(result)
except:
 print("Error : Someone divided by zero")

In [51]: 14:44 PM 9/19/2023

In [52]: a = MyClass()
b = MyClass()
try:
 result = a + b
 print(result)
except:
 print("Error : Someone divided by zero")

In [53]: In [53]: 14:44 PM 9/19/2023

In [54]: a = MyClass()
b = MyClass()
try:
 result = a + b
 print(result)
except:
 print("Error : Someone divided by zero")

In [55]: 14:44 PM 9/19/2023

In [56]: a = MyClass()
b = MyClass()
try:
 result = a + b
 print(result)
except:
 print("Error : Someone divided by zero")

In [57]: a = MyClass()
b = MyClass()
try:
 result = a + b
 print(result)
except:
 print("Error : Someone divided by zero")

In [58]: 14:44 PM 9/19/2023

DAY-2

Python – Day 3

1. Pandas : Panel Datasets
 - a. Dataframe
 - b. Functions on dataframe
 - c. Changing column to lowercase
 - d. Dropping Null Values
 - e. Dropping columns with Null Values
 - f. Filling null values
 - g. Describe columns
 - h. Value counts()
 - i. Selected rows or columns of dataframe
 - j. Conditional selection
2. Matplotlib
3. Arrays In Python

PANDAS

```
In [4]: import pandas as pd
In [7]: data = pd.read_csv('kashmir.csv')
print(data)
#orange : ('kashmir', 'sathy', 'blr')
#apple : ('chennai', 'delta', 'blr')
In [8]: fruit_df=pd.DataFrame(data)
print(fruit_df.head())
Out[8]:
```

orange apples

3 kashmir chennai

1 sathy delta

2 blr blr

```
In [9]: fruit_df.set_index(['jan','feb','mar'])
fruit_df
```

```
Out[9]:
```

orange apples

3 kashmir chennai

1 sathy delta

2 blr blr

```
In [11]: test_fruit=df.loc['feb']
test
```

```
Out[11]:
```

orange apples

3 kashmir chennai

1 sathy delta

2 blr blr

```
In [12]:
```

```
In [14]: movieDF = pd.read_csv('https://www.kaggle.com/mohitkumar1995/IMDb-Movie-Data.csv')
movieDF
```

```
Out[14]:
```

Rank	Title	Genre	Description	Director	Actors	Year	Budget (Millions)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic misfits are recruited as the Earth's last line of defense against an army of powerful alien invaders.	James Gunn	Chris Pratt, Michael Rooker, Dave Bautista, Karen Gillan	2014	121	8.1	757074	203.13	76.0
2	Prometheus	Science Fiction,Mystery,Sci-Fi	Following their discovery of a new planet, a crew of space explorers are forced to return to Earth after a mysterious alien entity attacks them.	Ridley Scott	Norman Reiss, Logan Marshall-Green, Michael Fassbender	2012	124	7.0	469303	128.46	65.0
3	Salt	Horror,Thriller	Three girls are captured by a mysterious alien who has come to Earth to collect salt.	J. August Richards	James McAvoy, Alice Taggart, Lucy Boynton, Michael Fassbender	2010	117	7.3	137906	128.32	62.0
4	Sing	Animation,Comedy,Family	A series of events lead to the rescue of a young orangutan named Bao.	Christopher Landon	Martine McCutcheon, James Corden, Nick Frost, Taron Egerton	2016	106	7.2	69545	270.02	59.0
5	Star Wars: Episode VII - The Force Awakens	Action,Adventure,Sci-Fi	A series of events lead to the rescue of a young orangutan named Bao.	Dave Filoni	Will Smith, Jaden Smith, Diego Luna, Lupita Nyong'o	2015	123	8.2	399777	329.02	40.0

```
In [15]: movieDF.tail(10)
```

```
Out[15]:
```

Rank	Title	Genre	Description	Director	Actors	Year	Budget (Millions)	Rating	Votes	Revenue (Millions)	Metascore
981	Underworld: Blood Wars	Action,Adventure,Horror	A werewolf army is created as the result of a secret experiment.	Parker Posey	Wesley Snipes, Michael Ealy, Shannen Doherty	2016	82	6.0	120706	45.00	44.0
982	Team Zennow	Film	An eight-year-old girl with a rare disease is the key to saving her mother.	Gauri Khan	Diamond Dallas Page, Ginnifer Goodwin, Kristin Chenoweth	2007	105	6.5	100007	1.00	40.0
983	Take Me Home	Comedy,Drama,Romance	A graduation (and wedding) gift for his wife.	Michael Dowd	Tyler Perry, Chandra Wilson, Chandra Wilson, Tyler Perry	2011	87	6.5	40428	5.02	N/A
984	Take This Waltz	Comedy,Drama,Romance	A woman tries to make things right with her ex-husband.	Mike简·Jewell	Alfre Woodard, Joaquin Phoenix, Alfre Woodard, Joaquin Phoenix	2010	100	6.0	27959	1.00	N/A

```
In [16]: movieDF.info()
```

```
Out[16]:
```

class pandas.core.frame.DataFrame:

Introducing the DataFrame constructor.

Data columns (total 13 columns): # Column Type

* rank: int8

* title: object

* genre: object

* description: object

* director: object

* actors: object

* year: int8

* runtime: int8

* rating: float64

* votes: int64

* revenue: float64

* metascore: float64

memory usage: 93.0+ KB

```
In [17]: movieDF.shape
```

```
Out[17]: (2006, 13)
```

```
In [18]: tempDF = movieDF.append(movieDF)
```

```
/usr/local/pyspark/bin/python2.7:2246:559746854:py:1: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.
```

```
tempDF
```

```
Out[18]: (2006, 13)
```

```
In [19]: finalDF=finalDF.drop_duplicates()
```

```
finalDF.shape
```

```
Out[19]: (1900, 13)
```

```
In [20]:
```

```
In [21]: movieDF.drop_duplicates().head(10).T
```

```
Out[21]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [22]: movieDF.drop_duplicates().head(10).T
```

```
Out[22]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [23]: movieDF.drop_duplicates().head(10).T
```

```
Out[23]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [24]: movieDF.drop_duplicates().head(10).T
```

```
Out[24]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [25]: movieDF.drop_duplicates().head(10).T
```

```
Out[25]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [26]: movieDF.drop_duplicates().head(10).T
```

```
Out[26]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [27]: movieDF.drop_duplicates().head(10).T
```

```
Out[27]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [28]: movieDF.drop_duplicates().head(10).T
```

```
Out[28]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [29]: movieDF.drop_duplicates().head(10).T
```

```
Out[29]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [30]: movieDF.drop_duplicates().head(10).T
```

```
Out[30]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [31]: movieDF.drop_duplicates().head(10).T
```

```
Out[31]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [32]: movieDF.drop_duplicates().head(10).T
```

```
Out[32]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [33]: movieDF.drop_duplicates().head(10).T
```

```
Out[33]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [34]: finalDF.rename(columns={'Runtime (Minutes)': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[34]:
```

SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#return-slice-with-a-copy

```
finalDF.rename(columns={'Runtime (Minutes)': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
Out[35]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [36]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[36]:
```

SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#return-slice-with-a-copy

```
finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
Out[37]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [38]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[38]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [39]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[39]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [40]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[40]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [41]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[41]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [42]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[42]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [43]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[43]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [44]: finalDF.rename(columns={'Runtime': 'Runtime', 'Revenue (Millions)': 'RevenueMillion'}, inplace = True)
```

```
finalDF
```

```
Out[44]:
```

ValueError: cannot convert non-finite values (NA or inf) to integer

```
In [45]: finalDF.rename(columns={'RevenueMillion': 'Revenue'}, inplace = True)
```

```
finalDF
```

```
Out[45]:
```

SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#return-slice-with-a-copy

```
finalDF.rename(columns={'RevenueMillion': 'Revenue'}, inplace = True)
```

```
Out[46]:
```

Index(['Title', 'Genre', 'Description', 'Director', 'Actors', 'Year', 'Runtime', 'Rating', 'Votes', 'Revenue', 'RevenueMillion', 'Metascore'], dtype='object')

```
In [47]: finalDF.columns[1].lower()
```

```
finalDF
```

```
Out[47]:
```

Index(['title', 'genre', 'description', 'director', 'actors', 'year', 'runtime', 'rating', 'votes', 'revenue', 'revenue_million', 'metascore'], dtype='object')

```
In [48]: finalDF.isnull().sum()
```

```
finalDF
```

```
Out[48]:
```

title 0

genre 0

description 0

director 0

actors 0

year 0

runtime 0

rating 0

votes 0

revenue 126

revenue_million 126

metascore 64

dtype: int64

```
In [49]: testFinalDF.dropna()
```

```
testFinalDF
```

```
Out[49]:
```

title 0

genre 0

description 0

director 0

actors 0

year 0

runtime 0

rating 0

votes 0

revenue 126

revenue_million 126

metascore 64

dtype: int64

```
In [50]:
```

jupyter

```
In [1]:
```

Untitled

```
In [2]:
```

Last checked: a day ago (Unchanged changes)

```
In [3]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [4]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [5]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [6]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [7]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [8]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [9]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [10]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [11]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [12]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [13]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [14]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [15]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [16]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [17]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [18]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [19]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [20]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [21]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [22]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [23]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [24]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [25]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [26]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [27]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [28]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [29]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [30]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [31]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [32]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [33]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [34]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [35]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [36]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [37]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [38]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [39]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [40]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [41]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [42]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [43]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [44]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [45]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [46]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [47]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [48]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [49]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [50]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [1]:
```

Untitled

```
In [2]:
```

Last checked: a day ago (Unchanged changes)

```
In [3]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [4]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [5]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [6]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [7]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [8]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [9]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [10]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [11]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [12]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [13]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [14]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [15]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [16]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [17]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [18]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [19]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [20]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [21]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [22]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [23]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [24]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [25]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [26]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [27]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [28]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [29]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [30]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [31]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [32]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [33]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [34]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [35]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [36]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [37]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [38]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [39]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [40]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [41]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [42]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [43]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [44]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [45]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [46]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [47]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [48]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [49]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [50]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [1]:
```

Untitled

```
In [2]:
```

Last checked: a day ago (Unchanged changes)

```
In [3]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [4]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [5]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [6]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [7]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [8]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [9]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [10]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [11]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [12]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [13]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [14]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [15]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [16]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [17]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [18]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [19]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [20]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [21]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [22]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [23]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [24]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [25]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [26]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [27]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [28]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [29]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [30]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [31]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [32]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [33]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [34]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [35]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [36]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [37]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [38]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [39]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [40]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [41]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [42]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [43]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [44]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [45]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [46]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [47]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [48]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [49]:
```

File Edit View Insert Cell Kernel Widgets Help

```
In [50]:
```

File Edit View Insert Cell Kernel Widgets Help

In [43]:

```
Int64Index: 1089 entries, 1 to 1089
Data types: float64, int64, object
# Column Non-Null Count Dtype
... # Column Non-Null Count Dtype
0 title    1089 non-null object
1 year     1089 non-null int64
2 description 1089 non-null object
3 director 1089 non-null object
4 actors   1089 non-null object
5 genres   1089 non-null object
6 runtime  1089 non-null int64
7 votes    1089 non-null int64
8 revenue_million 1089 non-null float64
9 metascore 1089 non-null float64
dtype: object, int64, float64, object[10]
memory usage: 92.8+ KB
```

In [44]:

```
finalDF.describe()
```

Out[44]:

	year	revenue_million	votes	revenue	actors	metascore
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	2012.770000	113.770000	4.723203	108900000.00	62.058791	55.599403
std	3.039841	108.000000	1.478700	1.478700e+05	36.420341	19.834006
min	2000.000000	10.000000	1.000000	10.000000	1.000000	1.000000
25%	2010.000000	108.000000	4.300000	1.030000e+05	17.442900	47.700000
50%	2014.000000	111.000000	4.800000	1.107000e+05	62.375000	54.860000
75%	2018.000000	123.000000	7.400000	1.369000e+05	86.175000	71.000000
max	2020.000000	151.000000	9.000000	1.781000e+05	98.000000	100.000000

In [45]:

```
finalDF.describe()
```

Out[45]:

	year	revenue_million	votes	revenue	actors	metascore
count	2000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	2012.770000	113.770000	4.723203	108900000.00	62.058791	55.599403
std	3.039841	108.000000	1.478700	1.478700e+05	36.420341	19.834006
min	2000.000000	10.000000	1.000000	10.000000	1.000000	1.000000
25%	2010.000000	108.000000	4.300000	1.030000e+05	17.442900	47.700000
50%	2014.000000	111.000000	4.800000	1.107000e+05	62.375000	54.860000
75%	2018.000000	123.000000	7.400000	1.369000e+05	86.175000	71.000000
max	2020.000000	151.000000	9.000000	1.781000e+05	98.000000	100.000000

In [46]:

```
finalDF['revenue_million'].value_counts().head()
```

Out[46]:

revenue_million	count
0	7
10	5
20	4
30	1

Name: revenue_million, dtype: int64

In [47]:

```
jupyter Untitled 1 checkpoint, Yesterday at 5:56 AM (unseen changes)
```

File Edit View Insert Cell Kernel Widgets Help

In [48]:

```
finalDF.describe()
```

Out[48]:

	year	revenue_million	votes	revenue	actors	metascore
count	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000
mean	2012.770000	113.770000	4.723203	108900000.00	62.058791	55.599403
std	3.039841	108.000000	1.478700	1.478700e+05	36.420341	19.834006
min	2000.000000	10.000000	1.000000	10.000000	1.000000	1.000000
25%	2010.000000	108.000000	4.300000	1.030000e+05	17.442900	47.700000
50%	2014.000000	111.000000	4.800000	1.107000e+05	62.375000	54.860000
75%	2018.000000	123.000000	7.400000	1.369000e+05	86.175000	71.000000
max	2020.000000	151.000000	9.000000	1.781000e+05	98.000000	100.000000

In [49]:

```
finalDF['revenue_million'].value_counts().head()
```

Out[49]:

revenue_million	count
0	7
10	5
20	4
30	1

Name: revenue_million, dtype: int64

In [45]:

```
finalDF.describe()
```

Out[45]:

	year	name	rating	votes	revenue_million	metascore
count	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000
mean	2012.770000	113.770000	4.723203	108900000.00	62.058791	55.599403
std	3.039841	108.000000	1.478700	1.478700e+05	36.420341	19.834006
min	2000.000000	8	3.8	1000000.00	0.000000	55.599403
25%	2010.000000	108.000000	4.300000	1.030000e+05	62.058791	47.700000
50%	2014.000000	111.000000	4.800000	1.107000e+05	86.175000	54.860000
75%	2018.000000	123.000000	7.400000	1.369000e+05	98.000000	71.000000
max	2020.000000	151.000000	9.000000	1.781000e+05	98.000000	100.000000

In [46]:

```
finalDF['revenue_million'].value_counts().head()
```

Out[46]:

revenue_million	count
0	7
10	5
20	4
30	1

Name: revenue_million, dtype: int64

In [47]:

```
finalDF[['title','revenue_million']]
```

Out[47]:

Rank	title	revenue_million
1	Guardians of the Galaxy	331.13
2	Guardians of the Galaxy	130.46
3	Sing	130.12
4	Sing	270.32
5	Guardians of the Galaxy	329.02

In [48]:

```
df=df.loc[[1,2,3]]
```

Out[48]:

Rank	title	revenue_million
1	Guardians of the Galaxy	331.13
2	Guardians of the Galaxy	130.46
3	Sing	130.12

In [49]:

```
df=df.loc[[0,4]]
```

Out[49]:

Rank	title	revenue_million
1	Guardians of the Galaxy	331.13
2	Guardians of the Galaxy	130.46
3	Sing	130.12
4	Sing	270.32

In [49]:

```
con.selectFinalDF['title']=='Sing'
```

Out[49]:

Rank	title	revenue_million
1	Sing	270.32

In [50]:

```
con.selectFinalDF['title'].value_counts().head()
```

Out[50]:

title	count
Sing	2

In [51]:

```
finalDF[finalDF['revenue_million']>=270.32]
```

Out[51]:

Rank	title	revenue_million	votes	year	rating	actors	director	genres	description
1	Guardians of the Galaxy	331.13	1.1	2014	8.1	757074	331.13	Action, Adventure, Sci-Fi	A group of intergalactic misfits travel through space in a stolen ship to steal one last item before the entire galaxy is destroyed.
4	Sing	270.32	1.2	2014	8.2	60548	270.32	Action, Adventure, Comedy, Family	A team of intergalactic animals, a talking robot, and a tree-lobster must stop an evil alien invasion.
12	Rogue One: A Star Wars Story	269.02	1.2	2016	8.2	390517	269.02	Action, Adventure, Fantasy	A team of间谍 steal the plans to the Death Star from the Empire's secret base on the planet Scarif.
13	The Social Network	269.02	1.9	2010	8.1	32108	269.02	Drama, Thriller, Romance	The social media empire that we know today was built on a lie.

In [44]:

```
finalDF.describe()
```

Out[44]:

	year	revenue_million	votes	revenue	actors	metascore
count	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000
mean	2012.770000	113.770000	4.723203	108900000.00	62.058791	55.599403
std	3.039841	108.000000	1.478700	1.478700e+05	36.420341	19.834006
min	2000.000000	10.000000	1.000000	10.000000	1.000000	1.000000
25%	2010.000000	108.000000	4.300000	1.030000e+05	17.442900	47.700000
50%	2014.000000	111.000000	4.800000	1.107000e+05	62.375000	54.860000
75%	2018.000000	123.000000	7.400000	1.369000e+05	86.175000	71.000000
max	2020.000000	151.000000	9.000000	1.781000e+05	98.000000	100.000000

In [45]:

```
df=df.loc[[1,2,3]]
```

Out[45]:

Rank	title	revenue_million
1	Guardians of the Galaxy	331.13
2	Guardians of the Galaxy	130.46
3	Sing	130.12

In [46]:

```
df=df.loc[[0,4]]
```

Out[46]:

Rank	title	revenue_million
1	Guardians of the Galaxy	331.13
2	Guardians of the Galaxy	130.46
3	Sing	130.12
4	Sing	270.32

In [47]:

```
con.selectFinalDF['title']=='Sing'
```

Out[47]:

Rank	title	revenue_million
1	Sing	270.32

In [48]:

```
finalDF.describe()
```

Out[48]:

	year	name	rating	votes	revenue_million	metascore
count	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000	1089.000000
mean	2012.770000	113.770000	4.723203	108900000.00	62.058791	55.599403
std	3.039841	108.000000	1.478700	1.478700e+05	36.420341	19.834006
min	2000.000000	8	3.8	1000000.00	0.000000	55.599403
25%	2010.000000	108.000000	4.300000	1.030000e+05	17.442900	47.700000
50%	2014.000000	111.000000	4.800000	1.107000e+05	62.375000	54.860000
75%	2018.000000	123.000000	7.400000	1.369000e+05	86.175000	71.000000
max	2020.000000	151.000000	9.000000	1.781000e+05	98.000000	100.000000

In [49]:

```
finalDF[['title','revenue_million']]
```

Out[49]:

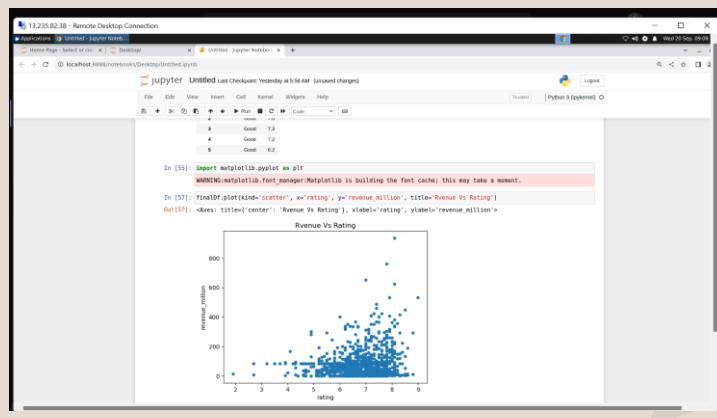
Rank	title	revenue_million
1	Guardians of the Galaxy	331.13
2	Guardians of the Galaxy	130.46
3	Sing	130.12

In [50]:

```
finalDF['revenue_million'].value_counts().head()
```

Out[50]:

revenue_million	count
0	7
10	5
20	4
30	1



DAY-3

Apache PySpark – Day 1

1. Driver process
2. Executors
3. Cluster Manager
4. Partition
5. Job, Stage and Task
6. Lazy Evaluation
7. RDD (Resilient Distributed Dataset)
8. Caching
9. Catalyst and Tungsten
10. RDD vs Dataframe
11. Deployment Types
 1. Client Mode
 2. Cluster Mode

```
15.207.34.155 - Remote Desktop Connection

File Edit View Insert Cell Kernel Help Help [Python (Spark)] 

In [4]: import findspark
In [5]: findspark.init()
In [4]: from pyspark.sql import SparkSession
spark=SparkSession.builder.appname("WordCount").getOrCreate()
In [12]: sc=spark.sparkContext
In [13]: rdd=sc.parallelize([(1,2),(2,3),(3,4)])
In [17]: results=rdd.map(lambda x:x**2)
results.collect()

Out[17]: [(2, 4, 8, 16)]
In [19]: results=rdd.map(lambda x:(x,x**2))
results.collect()

Out[19]: [(1, 2), (2, 4), (3, 8), (4, 16)]
In [21]: results=rdd.flatMap(lambda x:[x,2*x])
results.collect()

Out[21]: [(1, 2), (2, 4)]
In [22]: results=rdd.map(lambda x:(x,x**2))
results.collect()

Out[22]: [(1, 1), (2, 4), (3, 9), (4, 16), (5, 25), (6, 36)]
In [28]: results=rdd.reduceByKey(lambda x,y:[x,y])
results.collect()

Out[28]: [(0, [1, 1]), (1, [2, 4]), (2, [3, 9]), (3, [4, 16]), (4, [5, 25]), (5, [6, 36])]
```

15:20 54.155 - Remote Desktop Connection

From the last few lines of the session, we can see that the user has been updating their Jupyter Notebook. The code they have written so far is as follows:

```
jupyter PySpark Last Checkpoint 3 hours ago [modified changes]
```

```
In [1]:
```

```
from pyspark import SparkContext
```

```
In [2]:
```

```
sc = SparkContext("local[2]", "WordCount")
```

```
In [3]:
```

```
def words(text): return re.findall(r'\w+', text)
```

```
In [4]:
```

```
def count_words(text): return zip(text.split(), map(lambda x: x[1], Counter(text).items()))
```

```
In [5]:
```

```
words_list = sc.parallelize(words(" ".join(open('sample').read().splitlines())))
```

```
In [6]:
```

```
word_count = words_list.map(count_words).reduceByKey(lambda a, b: a + b)
```

```
In [7]:
```

```
word_count.collect()
```

```
Out[7]:
```

```
[('the', 1), ('is', 1), ('a', 1), ('sample', 1), ('text', 1), ('document', 1), ('for', 1), ('word', 1), ('example', 1), ('word', 1), ('count', 1), ('is', 1)]
```

```
In [8]:
```

```
word_count = word_count.collect()
```

```
In [9]:
```

```
word_count
```

```
Out[9]:
```

```
[('the', 1), ('is', 1), ('a', 1), ('sample', 1), ('text', 1), ('document', 1), ('for', 1), ('word', 1), ('example', 1), ('word', 1), ('count', 1), ('is', 1)]
```

```
In [10]:
```

```
wc = word_count[(lambda x: x[1]) > 1]
```

```
wc.collect()
```

```
Out[10]:
```

```
[('the', 1), ('is', 1), ('a', 1), ('sample', 1), ('text', 1), ('document', 1), ('for', 1), ('word', 1), ('example', 1), ('word', 1), ('count', 1), ('is', 1)]
```

13.207.54.155 - Remote Desktop Connection

File Edit View Insert Cell Kernel Widgets Help

Toolbox Python 3 (ipython) □

Start new one

```
example:  
def count(x):  
    count = 0  
  
    In [38]: w = rdd.map(lambda x:(x,1)).reduceByKey(lambda a,b: a+b)  
  
Out[38]: {(1,1): 1,  
           (2,1): 1,  
           (3,1): 1,  
           ('word', 2): 2,  
           ('word', 1): 1,  
           ('a', 1): 1,  
           ('count', 2): 2,  
           ('count', 1): 1}  
  
In [41]: purchases= pd.read_csv("/home/labuser/Desktop/Pandas_datasets/purchases.csv")  
purchases.collect()  
  
Out[41]: [(apple,orange,'June,3.0','Robert,J,T','Lily,B,T','David,I,T')]  
  
In [42]: purchases=spark.read_csv("/home/labuser/Desktop/Pandas_datasets/purchases.csv")  
  
Out[42]: 

| apple | orange | June,3.0 | Robert,J,T | Lily,B,T | David,I,T |
|-------|--------|----------|------------|----------|-----------|

  
In 1 1: □
```

```
15.120.254.155 - Remote Desktop Connection

File Edit View Insert Cell Kernel Widgets Help
Kernel Python (Unknown) 0

[In 1]: purchases = pd.read_csv("C:/Users/Desktop/Pandas_Datasets/purchases.csv")
purchases.collect()

Out[1]: [{'apple': 0.5, 'banana': 1.0, 'orange': 0.5}, {'apple': 1.0, 'banana': 0.5, 'orange': 1.0}, {'apple': 0.5, 'banana': 0.5, 'orange': 0.5}, {'apple': 0.5, 'banana': 0.5, 'orange': 0.5}, {'apple': 0.5, 'banana': 0.5, 'orange': 0.5}]

[In 2]: purchases.read_csv("C:/Users/Desktop/Pandas_Datasets/purchases.csv")

Out[2]: [0, 1, 2, 3, 4]
[apple] [banana] [orange]
[null] [apple] [banana]
[apple] [banana] [apple]
[banana] [apple] [banana]
[apple] [banana] [apple]
[banana] [apple] [banana]

[In 3]: purchases == pd.read_option("inferSchema",True).option("header",True).csv("C:/Users/Desktop/Pandas_Datasets/purchases.csv")

Out[3]: [0, 1, 2, 3, 4]
[apple] [banana] [orange]
[null] [apple] [banana]
[apple] [banana] [apple]
[banana] [apple] [banana]
[apple] [banana] [apple]
[banana] [apple] [banana]

22/09/09 09:05:26 WARN CSVHeaderChecker: CSV header does not conform to the schema.
Records : 50000, Wrongs : 0, Skipped : 0, Truncated : 0
Expected : 3 found : 3
CSV file: file:///C:/Users/Desktop/Pandas_Datasets/purchases.csv
```

DAY-4

Apache PySpark – Day 2

1. Partitioning
2. Spark UI URL
3. SQL in PySpark
4. Drop duplicates
5. When – Otherwise condition
6. Concat and Functions
7. Caching
8. Joins
9. Group By

```
3.77.10.65 - Remote Desktop Connection

[4]: pd.read_csv('IMDb-Movie-Data.csv')
          SingAnimation.Comedy...
          Superhero.Comedy...
          The Great Wall>Action.Adventure...
          Madhouse...Comedy...
          The Long Dumb Walk>Action.Adventure...
          Passengers:Adventure.Drama.H...
          Fantastic Voyage:Science.Fiction.H...
          Hidden Figures:Biography.Drama.H...
          The King's Speech:Biography.Drama.H...
          Human.Calibration.Adventure...
          Goodbye, Darkness:History.War...
          The Secret Life of Walter.Mitty:Animation.Adventu...
          The Martian:Science.Fiction.Thriller...
          Jason Bourne:Action.Thriller...
          Arrival:Science.Fiction.H...
          Arrival:Science.Fiction.H...
only showing top 20 rows

In [10]: testDF = moviesDF.repartition(10)

In [11]: testDF.rdd.getNumPartitions()
Out[11]: 10

In [12]: testDF.write.csv('/home/labour/Desktop/Pandas/datasets/IMDb-Movie-Data.csv')
        (Stage 20)
        10

In [13]: testDF.write.csv('/home/labour/Desktop/Output/out')
        (Stage 21)
        10

In [14]: pd.read_csv('IMDb-Movie-Data.csv')
          SingAnimation.Comedy...
          Superhero.Comedy...
          The Great Wall>Action.Adventure...
          Madhouse...Comedy...
          The Long Dumb Walk>Action.Adventure...
          Passengers:Adventure.Drama.H...
          Fantastic Voyage:Science.Fiction.H...
          Hidden Figures:Biography.Drama.H...
          The King's Speech:Biography.Drama.H...
          Human.Calibration.Adventure...
          Goodbye, Darkness:History.War...
          The Secret Life of Walter.Mitty:Animation.Adventu...
          The Martian:Science.Fiction.Thriller...
          Jason Bourne:Action.Thriller...
          Arrival:Science.Fiction.H...
          Arrival:Science.Fiction.H...
only showing top 20 rows
```

Job ID	Description	Submitted	Duration	Memory (Allocated/Total)	Tasks (0 of 0 assigned) Success/Error
38	car or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	1.4	5/1 (1 assigned)	0/0 (0 assigned)
37	car or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.2 s	1/1	0/0
36	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.8 s	1/1	0/0
35	jewel or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.8 s	1/1	0/0
34	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
33	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
32	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.2 s	1/1	0/0
31	car or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.7 s	1/1	0/0
30	car or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
29	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
28	car or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.2 s	1/1	0/0
27	car or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.2 s	1/1	0/0
26	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
25	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
24	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
23	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
22	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
21	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
20	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
19	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
18	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
17	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
16	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
15	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
14	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
13	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
12	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
11	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
10	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
9	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
8	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
7	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
6	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
5	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
4	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
3	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
2	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
1	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0
0	shelling or fahleitheitfahleitheitcomp/pas 0	2023090023 04/21/22	0.9 s	1/1	0/0



JOURNEY SCRAPBOOK

Custom Bootcamp (29th August – 6th October)

Week-5

Anushka Dhir

Batch-6

DAY-1

Apache PySpark – Day 3

1. Caching
2. Persist and different storage levels
3. Aggregation
4. Managed and External Tables
5. Partitioning
6. For-Each

** Note : Only 4 hours training due to OneIDA event. **

DAY-2

Azure Databricks – Day 1

1. What is databricks?
2. Setting Up azure databricks
3. Overview of Databricks Workspace
4. Cluster configuration
5. Mounting
6. Widgets
7. Parameters
8. Functions
9. %run, %md, %language
10. Structured streaming

AnushkaDA_anushkadbricks | Overview

Your deployment is complete

Deployment name: AnushkaDA_anushkadbricks
Start time: 5/26/2023, 9:42:41 AM
Completion ID: b6c14812-799d-4b19-8f6b-aaf43668...
Resource group: AnushkaDA

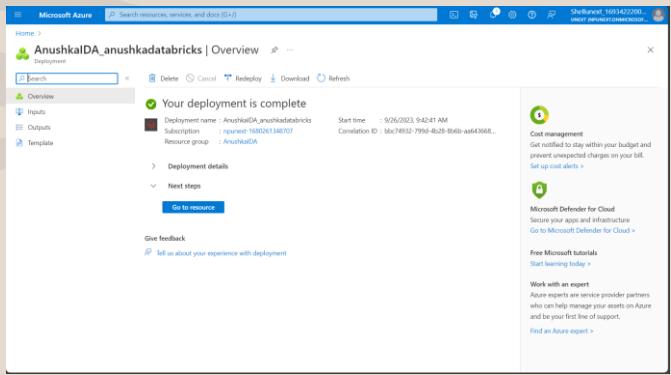
Deployment details

Next steps

Go to resource

Give feedback

Sell us about your experience with deployment



Get started with Databricks

Start the SQL warehouse

A SQL warehouse is a simple compute resource that gives you the power to process data in the cloud. We've created your first warehouse for you. To start it, click the button below.

Start warehouse

Explore sample projects

Don't have your data handy? Learn how to gain insights in just a few steps with these sample projects.

SQL and visualizations

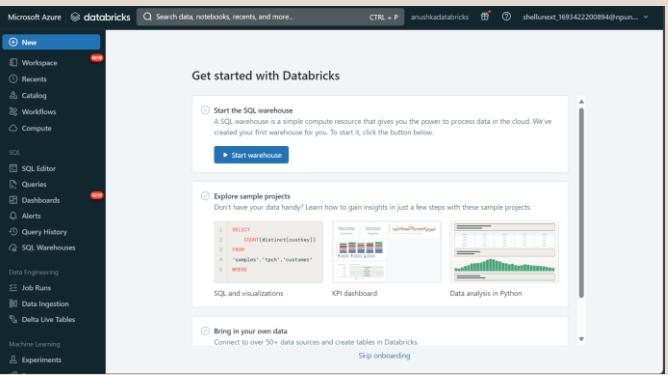
KPI dashboard

Data analysis in Python

Bring in your own data

Connect to over 50+ data sources and create tables in Databricks.

Skip onboarding



Shellnextr's Cluster

Configuration

Compute

Unrestricted

Access mode

Single user access

Single user

Shellnextr

Performance

Databricks Runtime Version

13.3 LTS (includes Apache Spark 3.4.1, Scala 2.12)

Use Photon Acceleration

Worker type

Standard_DS3_v2

Min workers: 2, Max workers: 2, Current: 2

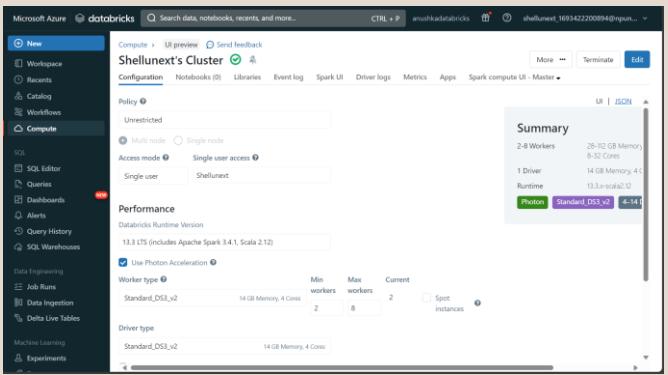
Driver type

Standard_DS3_v2

14 GB Memory, 4 Cores

Summary

2-8 Workers, 28-102 GB Memory, 8-32 Cores
1 Driver, 14 GB Memory, 4 Cores
Runtime: 13.3.x (standard), 14
Photon Standard_DS3_v2, v14



Mount

Mount Python

File Edit View Run Help Last edit was 2 minutes ago Provide feedback

Mount all

shellnextr's Cluster

Mount

Mounting a file system

Code 1

```
1 dbutils.fs.mount(source = "wasbs://input@anushkag2.blob.core.windows.net",
2                   mount_point = "/mnt/input",
3                   extra_configs = [{"fs.azure.account.key": "anushkag2.blob.core.windows.net": "dbutils.secrets.get(scope = "anushkaScope", key = "anushka-secret")"}]
```

True

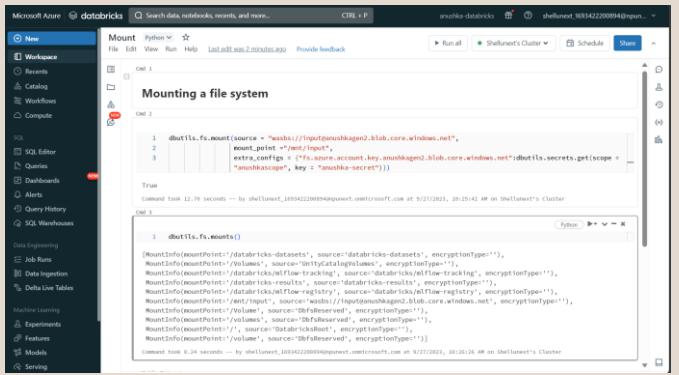
Command took 12.76 seconds -- by shellnextr,189942220894@spark.onmicrosoft.com at 5/27/2023, 09:25:42 AM on Shellnextr's Cluster

Code 2

```
1 dbutils.fs.mount()
```

[MountInfo(name='Pyhive - databricks-datasets', source='databricks-datasets', encryptionType=''), MountInfo(name='Pyhive - /volumes', source='unitycatalog/volumes', encryptionType=''), MountInfo(name='Pyhive - /databricks/mflow-tracking', source='databricks/mflow-tracking', encryptionType=''), MountInfo(name='Pyhive - /databricks/mflow-registry', source='databricks/mflow-registry', encryptionType=''), MountInfo(name='Pyhive - /mflinput', source='wasbs://input@anushkag2.blob.core.windows.net', encryptionType=''), MountInfo(name='Pyhive - /volumes', source='dbfs:/shared', encryptionType=''), MountInfo(name='Pyhive - /volume', source='dbfs:/reserved', encryptionType=''), MountInfo(name='Pyhive - /volume', source='dbfs:/reserved', encryptionType='')

Command took 0.24 seconds -- by shellnextr,189942220894@spark.onmicrosoft.com at 5/27/2023, 09:25:42 AM on Shellnextr's Cluster



DAY-3

Azure Databricks – Day 2

1. Structured streaming
2. DBFS from sample data
3. Delta table & Parquet table
4. Partition Tables
5. Medallion / Multi hop Architecture
6. Workflows → Jobs
7. Accessing data from Azure SQL database
8. Unity catalog in databricks
9. Metastore

DBFS Example

This notebook will show you how to create and query a table or DataFrame that you uploaded to DBFS. DBFS is a Databricks File System that allows you to store data for querying inside of Databricks. This notebook assumes that you have a file already inside of DBFS that you would like to read from.

```

1 #!python
2 #!file location and type
3 file_location = "dbfs:/mnt/datasets/r/zipcodes.csv"
4 file_type = "csv"
5
6 #!CSV options
7 header = "true"
8 delimiter = ","
9 first_row_as_header = "true"
10
11 df = spark.read.format(file_type).options(header=header, delimiter=delimiter).load(file_location)
12 df.show()
13
14 #!spark.read.format("csv").options(header="true", delimiter=",")\n15 df = spark.read.format("csv").options(header="true", delimiter=",")\n16 df = df.withColumnRenamed("lat", "Lat")\n17 display(df)

```

Catalog Explorer

default > default.zipcodes_1.csv

Column	Type	Comment
RecordNumber	int	
ZipCode	int	
ZipCodeType	string	
City	string	
State	string	
LocationType	string	
Lat	double	
Long	double	
Xaxis	double	
Yaxis	double	

SQL Editor

27th September

```

1 VACUUM zipcodes RETAIN 169 HOURS
2
3 !df=spark.read.option("inferSchema",True).option("header","true").csv("dbfs:/mnt/input/zipcodes.csv")
4
5 !df=df.drop("RecordNumber")
6
7 !df=df.withColumn("RecordNumber", monotonically_increasing_id())
8
9 !df.write.option("path","dbfs:/input/tables").saveTable("zipcodes")

```

Workflows

Only my job runs Run status ▾ 2023-09-25 14:00:00 ▾ 2023-09-27 14:00:00 ▾ Create job

Start Time	Job	Run as	Launched	Duration	Status	Run parameters
Sep 27, 2023, 01:48 PM	Anushka-job	Shellenet	By scheduler	2s	Skipped	
Sep 27, 2023, 01:48 PM	Anushka-job	Shellenet	By scheduler	2m 0s	Pending	

Jobs

Workflow > Jobs > Anushka-Job

Run ▾ Tasks

Anushka-Job

Job ID: .0094@sparkpool.com/cicr005/job/Anushka-job

Creator: Shellenet

Run as: Shellenet

Tags: #tag

Git: Not configured

Add Git settings

Schedule: Every 2 minutes

Edit schedule ▾ Pause ▾ Delete

Compute: Available jobs cluster

Driver: Standard (5G) v2 - Workers: Standard (5G) v2 - 0 workers 132 LTS Photon (includes Apache Spark 3.4.1, Scala 2.12)

DAY-4

Docker Kubernetes

1. What is Docker ?
2. Docker initialization in VM
3. Build Docker image
 - Application
 - Requirement.txt
 - Dockerfile
4. Image created and Docker push
5. Azure container registry (ACR)
6. Kubernetes
7. AKS
8. Deploy a single-image application using code and UI

```
 100% |██████████| 0B/0B [00:00<00:00, 0B/s] labuser@ip-172-31-4-700: ~Desktop/docker_project
$ ./run_docker.sh
  File Edit View Search Terminal Help
  Downloading docker-engine-s2i-slim:2.1.2-py3...none-any (15 kB)
  [downloaded]:132 1.1.2-py3-none-any.whl (133 kB)
  Downloading docker-engine-s2i-slim:2.1.2-py3...none-any.whl (133 kB)
  [downloaded]:133 1.1.2-py3-none-any.whl (133 kB) eta 0:00:00
  Downloading docker-engine-s2i-slim:2.1.2-py3...none-any.whl (134 kB)
  [downloaded]:134 1.1.2-py3-none-any.whl (134 kB) eta 0:00:00
  Downloading docker-engine-s2i-slim:2.1.2-py3...none-any.whl (134 kB)
  [downloaded]:134 1.1.2-py3-none-any.whl (134 kB) eta 0:00:00
  Downloading MarkupSafe==2.0
  Downloading MarkupSafe==2.0.1-1.3-cp38-cp38-manylinux2014_27_x86_64-manylinux2014_x86_64.whl (25 kB)
  [downloaded]:135 2.0.1-1.3-cp38-cp38-manylinux2014_27_x86_64-manylinux2014_x86_64.whl (25 kB)
  Successfully installed Flask-2.0.1 Jinja2-3.1.2 MarkupSafe-2.1.3 Werkzeug-2.1.2 click-8.1.7 itsdangerous-2.1.2
  WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended you use a virtual environment instead.
  See https://pip.pypa.io/warnings/venv

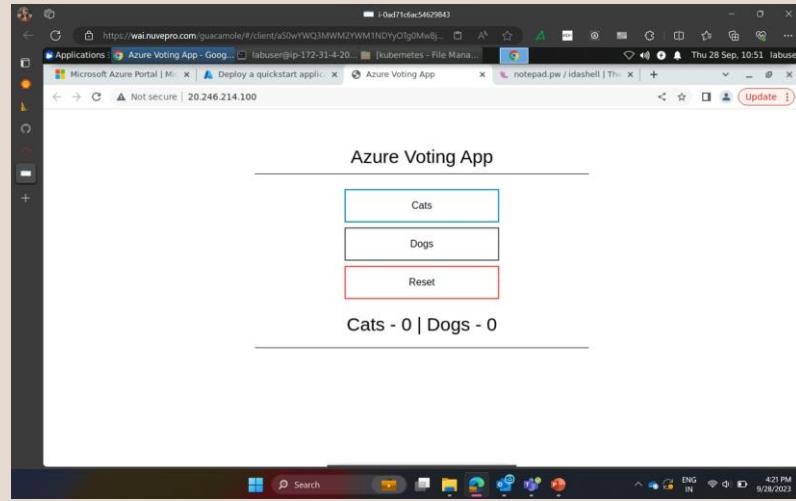
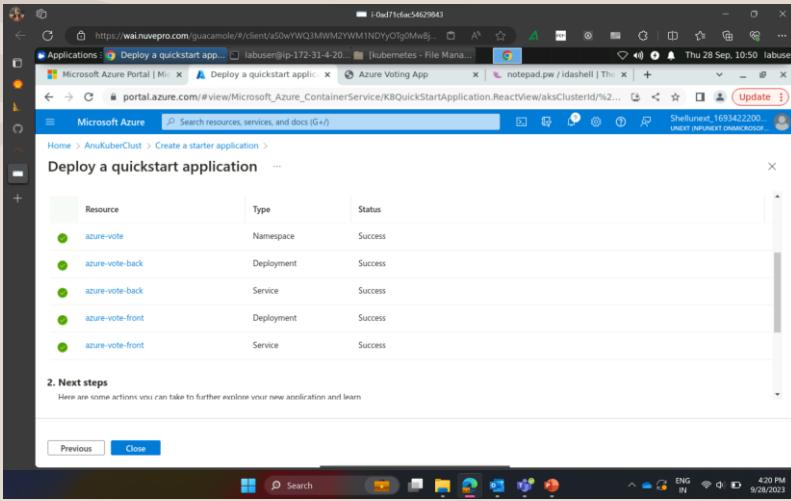
[notice] A new release of pip is available: 23.0.1 => 23.2.1
[notice] To update, run: pip install --upgrade pip
removing intermediate container d4bf3b26de4b
Step 5/7 : EXPOSE 80
  <--> 218cf1c1fb
removing intermediate container d4bf3b26de4b
  <--> 218cf1c1fb
entrypoint: /bin/sh -c
  <--> Running in b44470334e3
  <--> removing intermediate container b44470334e3
  <--> e6d3f781f083
Step 7/7 : CMD ["python", "app.py"]
  <--> Running in feef998b30f
removing intermediate container 27ea4990b30f
  <--> 218cf1c1fb
Successfully built 020976be1a0
Successfully tagged python-app:latest
labuser@ip-172-31-4-700: ~Desktop/docker_project
```

The screenshot shows a Linux desktop environment with several windows open. In the foreground, a terminal window titled 'labouser@ip-172-31-4-200' displays a command to install Python dependencies using pip. The command is: `curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py`. Below this, another command is shown: `python get-pip.py`. A warning message from pip is visible: "WARNING: Running pip as the root user or 'root' is not recommended. This can lead to permission issues when installing packages." The background shows a file manager window titled 'docker_project - File Manager' containing files like app.py, Dockerfile, and requirements.txt. A sidebar on the left lists 'DEVICES' (File System, thincient_dfl...), 'PLACES' (home, Desktop, Trash, Documents, Music, Pictures, Videos, Downloads), and 'NETWORK' (Browse Network). The bottom status bar indicates '3 items: 621 bytes, Free space: 10.5 GB'.

```
[ 0%] [idle] (40GB) 
Applications https://www.python.org/guides/running-on-windows/YWQMAMZQWYHINDYDgjMw/ 
17:21 0.01 Google Root@12-21-4-200 ~ /root/docker_project 
File Edit View Search Terminal Help 
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv 
notice: A new release of pip is available: 23.1.0 >= 23.1.0. 
To update, run: python -m pip install --upgrade pip 
removing intermediate container 738d2043b6d8 
Step 5/7 : EXPAND_R0 
--> Running in cd44539864bc 
  /bin/sh -c curl -sSf https://raw.githubusercontent.com/dotnet/dotnet-docker/2.2.0/containers/dotnet/Dockerfile . 
--> 25c3b097c526 
  /bin/sh -c curl -sSf https://raw.githubusercontent.com/dotnet/dotnet-docker/2.2.0/containers/dotnet/Dockerfile . 
  /bin/sh -c curl -sSf https://raw.githubusercontent.com/dotnet/dotnet-docker/2.2.0/containers/dotnet/Dockerfile . 
removing intermediate container df551dedc1ef 
  /bin/sh -c curl -sSf https://raw.githubusercontent.com/dotnet/dotnet-docker/2.2.0/containers/dotnet/Dockerfile . 
Step 7/7 : CMD ["python", "app.py"] 
  /bin/sh -c curl -sSf https://raw.githubusercontent.com/dotnet/dotnet-docker/2.2.0/containers/dotnet/Dockerfile . 
removing intermediate container ebd94206342 
Success: Image for the Dockerfile was successfully built. 
SealedImage2121 
Successfully tagged my-python-app:latest 
my-python-app:latest > /root/docker_project$ docker run my-python-app 
  * Starting Flask app "app" (lazy loading) 
    Environment: production 
    WARNING: This is a development server. Do not use it in a production deployment. 
    Use a production WSGI server instead. 
    Defaulting to port 5000 
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead. 
  Running on http://127.0.0.1:5000 (Press CTRL+C to quit) 
  Running on http://[::]:5000 (Press CTRL+C to quit) 
172.17.0.1 - - [28/Sep/2023 09:39:01] "GET / HTTP/1.1" 200 - 
172.17.0.1 - - [28/Sep/2023 09:39:01] "GET /favicon.ico HTTP/1.1" 404 - 
17:21 0.01 Google Root@12-21-4-200 ~ /root/docker_project 
[ 0%] [idle] (40GB) 
The 28 Sep, 05:39 lahusen 
11:08 AM 
FIND IN 
9/26/2023
```

A screenshot of a Microsoft Edge browser window. The address bar displays the URL '172.17.0.2:81'. The main content area of the browser shows a simple web application with the title 'Web App with Python Flask!' and a large red 'Run' button. Below the browser window, the Windows taskbar is visible, featuring the Start button, a search bar, and several pinned icons for applications like File Explorer, Microsoft Edge, and the Task View switcher.

```
https://wa.nevpro.com/guacamole/4/cb7ca5aXWQJ3HWDYWM1NDNtyCtgMwfI  
Applications ➔ Run Command Line ➔ labuser@ip-172-31-4-200: ~ labuser@ip-172-31-4-200: ~  
File Edit View Terminal Tabs Help  
root@ip-172-31-4-200:~/home/labuser/Desktop/docker_project  
labuser@ip-172-31-4-200: ~$ docker projects cd  
labuser@ip-172-31-4-200: ~$ az login  
az login: Please sign in to your account. You can do this via "az login" or by opening a browser at https://aka.ms/azelogin. Please continue the login in the web browser.  
Opening in existing browser session...  
libava error: vagetDriverHandleIndex() failed with unknown libava error, driver_name = [null]  
{  
    "cloudName": "AzureCloud",  
    "id": "0f519ab0-7708-4318-90ce-b2517f14e513",  
    "isDefault": true,  
    "name": "AzureCloud",  
    "name": "npunext_1686261348707",  
    "tenantId": "dce87313-8ff4-4ab1-ab40-8de5a7214b2f",  
    "user": {  
        "name": "Shellunnel_169342200894@npunext.onmicrosoft.com",  
        "type": "user"  
    }  
}  
labuser@ip-172-31-4-200: ~
```



DAY-5

Azure Deployment / DevOps

1. What is DevOps ?
 1. Collaborative Branch (Main branch)
 2. Feature branch
2. Azure DevOps
 1. Boards
 2. Repos
 3. Pipelines
 4. Test plans
3. Creating a project Board in Azure DevOps
 - Work items
 - Epic
 - Feature
 - User story
 - Tasks
4. Azure Repos
5. Azure Pipelines

Boards

The Boards hub displays a list of work items. The table shows the following data:

ID	Title	Assigned To	State	Area Path	Tags
24	Implement a search algorithm that queries the database for books	Unassigned	New	IDA_Anushka_Shell	
23	Design a search bar and results page UI	Unassigned	New	IDA_Anushka_Shell	
22	As a User, I want to search for books by title or author, so I can ...	Unassigned	New	IDA_Anushka_Shell	
21	Implement backend logic to fetch books by category	Unassigned	New	IDA_Anushka_Shell	
20	Develop a UI for browsing books by category	Unassigned	New	IDA_Anushka_Shell	
19	Create a database schema for book categories	Unassigned	New	IDA_Anushka_Shell	
18	As a user, I want to browse books by category, so I can find books ...	Unassigned	New	IDA_Anushka_Shell	
17	Search books	Unassigned	New	IDA_Anushka_Shell	
16	Browse books	Unassigned	New	IDA_Anushka_Shell	
15	Browse and Search Books	Unassigned	New	IDA_Anushka_Shell	
14	Handle authentication errors and provide appropriate feedback	Unassigned	New	IDA_Anushka_Shell	
13	Implement backend logic for user authentication	Unassigned	New	IDA_Anushka_Shell	
12	Design the login page UI	Unassigned	New	IDA_Anushka_Shell	
11	As a registered user, I want to log in to my account	Unassigned	New	IDA_Anushka_Shell	
10	User Login	Unassigned	New	IDA_Anushka_Shell	

Branches

The Branches hub shows two branches: main and feature. The main branch is the default.

Branch	Com...	Author	Authored...	Behind/Ahead	Status	Pull ...
feature	2164562	Shellunext unex...	52m ago	1 0		
main	1632708	Shellunext unex...	51m ago			★

Pull requests

The Pull requests hub lists three completed pull requests:

- updated test: Shellunext@20089 request 13 into P_main - Completed 52m ago
- Updated test.json: Shellunext@20089 request 12 into P_main - Completed 1h ago
- updated test.json: Shellunext@20089 request 11 into P_main - Completed 1h ago

Pipelines

The Pipeline details page for Anushka@22001.sample-code shows the following summary:

- Triggered by: Anushka@22001
- Repository and version: Anushka@22001/sample-code, P main, 4ebdb85e
- Time started and elapsed: Just now, 1m 13s
- Related: 0 work items, 0 artifacts
- Tests and coverage: A Get started

Errors and Warnings:

- Failed to download Python from the Github Actions python registry (<https://github.com/actions/python-versions>). Error: Error: Could not find Python matching spec 2.7 (x64) in the python-ve... Job Python27 + Use Python 2.7
- Version spec 2.7 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9.18 (x64) if this ... Job Python27 + Use Python 2.7
- Failed to download Python from the Github Actions python registry (<https://github.com/actions/python-versions>). Error: Error: Could not find Python matching spec 3.6 (x64) in the python-ve... Job Python36 + Use Python 3.6
- Version spec 3.6 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9.18 (x64) if this ... Job Python36 + Use Python 3.6

Jobs:

Name	Status	Duration
Job Python27	Failed	4s
Job Python35	Queued	
Job Python36	Failed	5s
Job Python37	Success	16s



JOURNEY SCRAPBOOK

Custom Bootcamp (29th August – 6th October)

Week-6

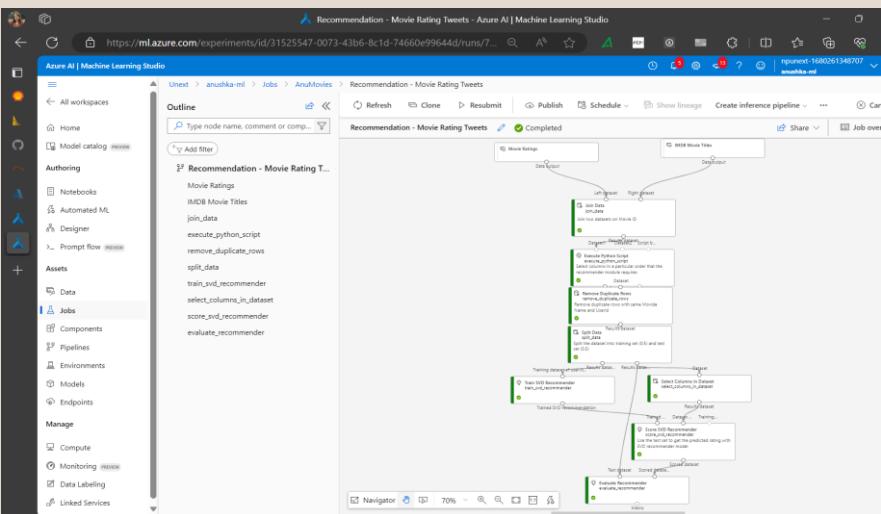
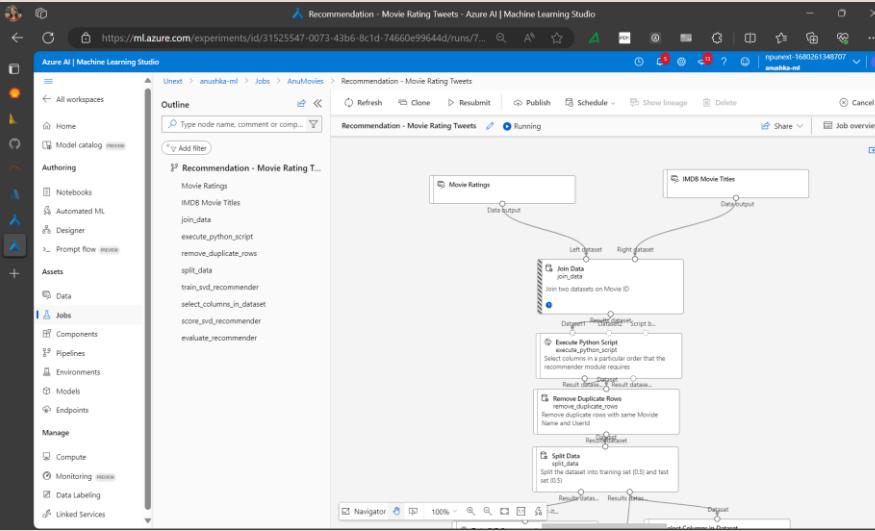
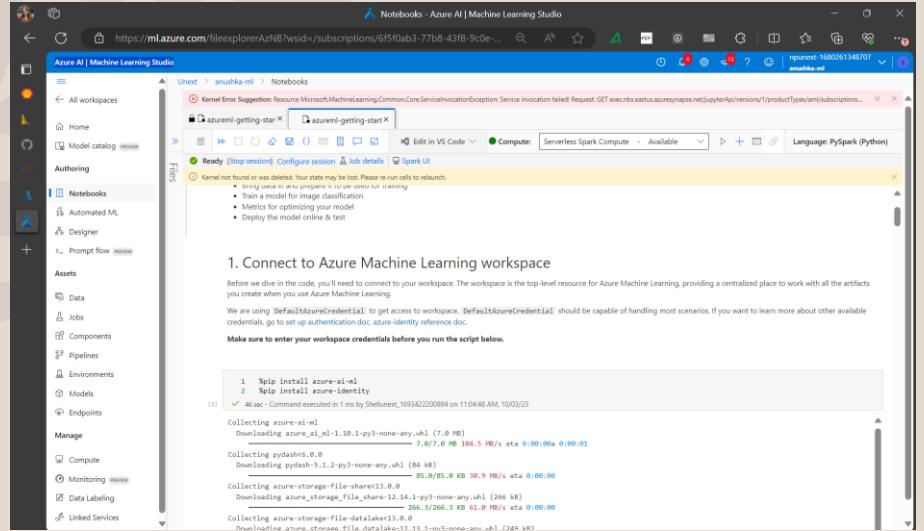
Anushka Dhir

Batch-6

DAY-1

Machine Learning – Day 1

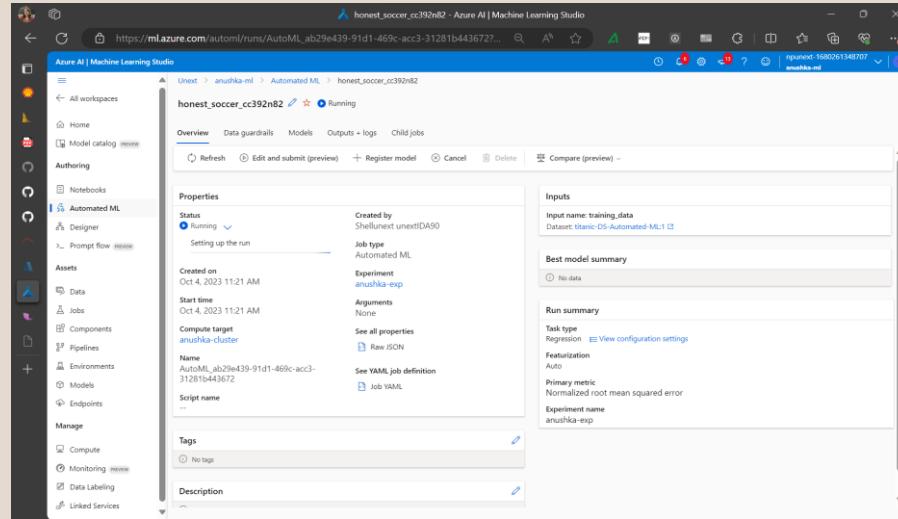
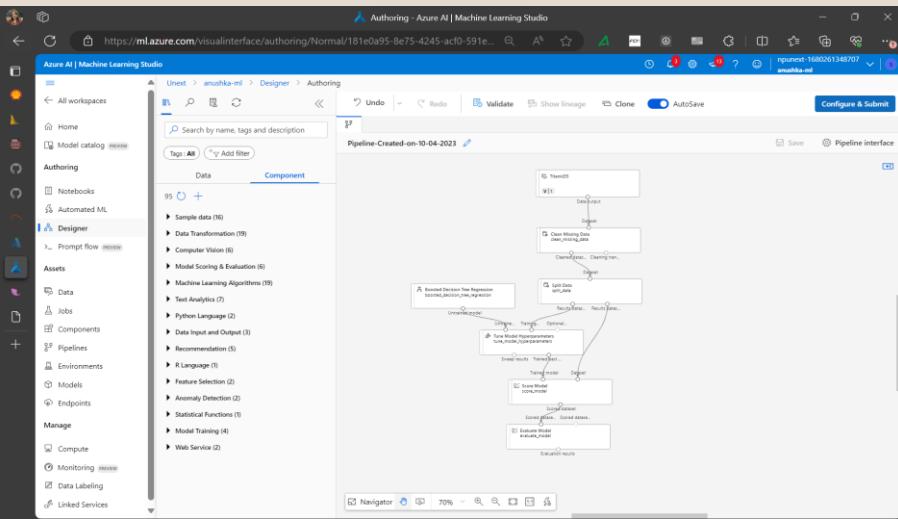
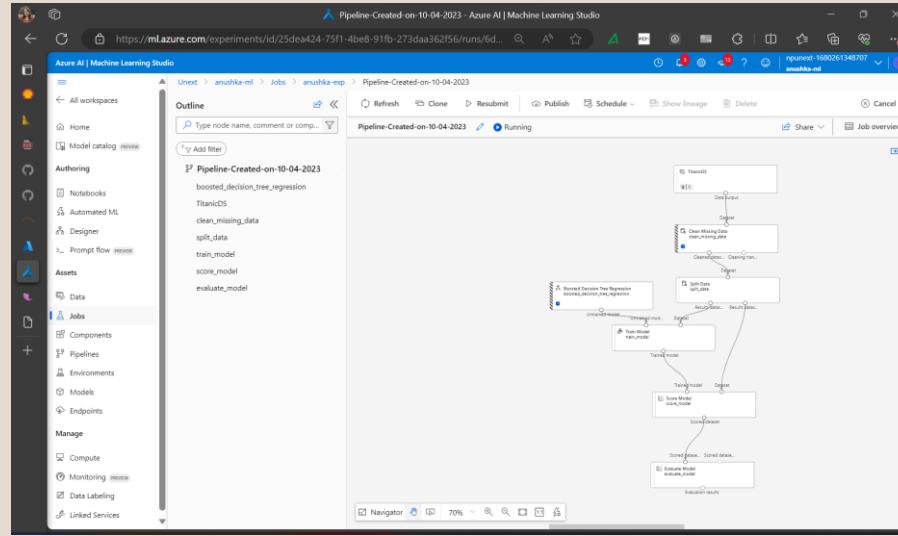
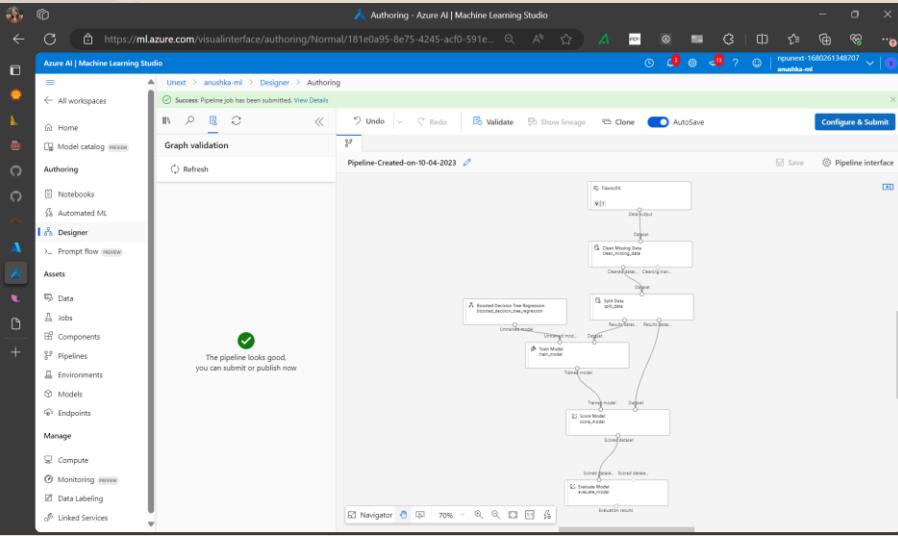
1. What is Machine learning and its uses in daily life ?
2. Types of machine learning algorithms :
 - Supervised
 - Unsupervised
 - Semi-supervised
 - Reinforcement
3. Overview of Azure ML Studio
4. Notebooks and Pipelines
5. Azure ML Designer



DAY-2

Machine Learning – Day 2

1. Azure ML Studio Pipelines
2. Tune model Hyperparameters
3. Automated ML



The background features abstract, organic shapes in muted colors. On the left, there's a large, rounded shape in a reddish-brown hue. Above it and to the right are several thin, light gray lines resembling stylized leaves or petals. To the right of the text, there's a large, rounded shape in a light olive green color, with a thin white line tracing its curved edge.

Thank You