

# Journey Scrapbook

Tanya Singh  
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

# Day-1... continued

- Waterfall Model
- Agile Method
- Scrum Events
- Kanban Board
  
- Role play to demonstrate agile methodology
- Assignment: Requirement Traceability

1	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
2										Test Env	UAT Env	Prod Env				
3	1	Req01	Search for a product	As a user enter item name in the search bar and products displayed 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
4	2				TC02	User searched a product and got a valid dropdown list	Completed	ABC	No	Passed	No Run	No Run	None	No	NA	Partial
5	3				TC03	User searched a product got the partial keywords matched	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial
6	4				TC04	User doesn't type anything and pressed enter	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial
7	1				TC01	User searched a product but its not available	Completed	XYZ	No	Passed	No Run	No Run	None	No	NA	Partial
8	2				TC02	User searched a product and got a valid dropdown list	Completed	ABC	No	Passed	No Run	No Run	None	No	NA	Partial
9	3				TC03	User searched a product got the partial keywords matched	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial
10	4				TC04	User doesn't type anything and pressed enter	Completed	DEF	Yes	Passed	Passed	No Run	Yes	DFCT001	Test OK	Partial

Requirement Traceability for setting up search option for an application

# Day-2

## Business Analysis and Agile

- ❑ Introduction to Dev Ops
- ❑ Creating EPIC, User Story, Tasks and Sub-tasks
- ❑ Story point estimation
- ❑ Scrum fundamentals
- ❑ Sprint backlog
- ❑ Burnup, Burndown, Velocity Charts
- ❑ Assignment: Hands-on practice on Azure

S Search +

Overview

Boards

Work items

Boards

Backlogs

Sprints

Queries

Delivery Plans

Analytics views

Repos

Pipelines

Test Plans

Artifacts

Project settings <

Try the New Boards Hub for improved performance, accessibility, and new features. [Click here](#) to learn more.

### Work items

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

ID	Title	Assigned To	Type	Assigned to	States	Area	Tags	X
31	Review	Tanya Singh	Task	New	New	Search		
30	System testing	Tanya Singh	Task	New	New	Search		
29	Test Case Creation	Tanya Singh	Task	New	New	Search		
28	Unit Testing	Tanya Singh	Task	New	New	Search		
27	Coding	Tanya Singh	Task	New	New	Search		
26	Requirement Analysis	Tanya Singh	Task	New	New	Search		
13	As a user searches by giving voice command.	Unassigned	Task	New	New	Search		
12	As a user searches by uploading an image.	Tanya Singh	Task	Active	Active	Search		
25	Review	Tanya Singh	Task	New	New	Search		
24	System Testing	Tanya Singh	Task	New	New	Search		
10	Search Functionality	Unassigned	Task	New	New	Search		

- Search +
- Overview
- Boards
- Work items
- Boards
- Backlogs
- Sprints
- Queries
- Delivery Plans
- Analytics views
- Repos
- Pipelines
- Test Plans
- Artifacts

Try the New Boards Hub for improved performance, accessibility, and new features. [Click here](#) to learn more.

### Search Team

[Board](#)[Analytics](#)[View as Backlog](#)[Stories](#)[New](#)[Active](#)[1/5](#)[Resolved](#)[0/5](#)[Closed](#)[New item](#)

13 As a user searches by giving voice command.

Unassigned

11

State

New

Tanya Singh

8

State

Active

1/6

Add Task

- Requirement Analysis
- Coding
- Unit Testing
- Test Case Creation
- System Testing
- Review

12 As a user searches by uploading an image.

Tanya Singh

8

State

Active

1/6

Add Task

- Requirement Analysis
- Coding
- Unit Testing
- Test Case Creation
- System Testing
- Review

11 As a user enters a product name in the search bar.

Tanya Singh

5

State

Closed

6/6

Add Task

- Requirement Analysis
- Coding
- Unit Testing
- Test Case Creation
- System Testing
- Review

- [Search](#)
- [Overview](#)
- [Boards](#)
- [Work items](#)
- [Boards](#)
- [Backlogs](#)
- [Sprints](#)
- [Queries](#)
- [Delivery Plans](#)
- [Analytics views](#)
- [Repos](#)
- [Pipelines](#)
- [Test Plans](#)
- [Artifacts](#)

Project settings

Try the New Boards Hub for improved performance, accessibility, and new features. [Click here to learn more.](#)

Search Team

Board [Analytics](#)

Last 3 iterations

Count of work items  
Average Velocity

1.1  
1  
0.9  
0.8  
0.7  
0.6  
0.5  
0.4  
0.3  
0.2  
0

Iteration 1 Iteration 2 Iteration 3

Iteration 1

Planned: 0
Completed: 0
Completed Late: 1
Incomplete: 0

Completed Late Incomplete

# Day-3

## Database Management System

- Brief about data and information
- Introduction to database and DBMS
- Types of databases
- Discussion about ACID property
- Introduction to SQL
- Introduction to Normalization and it's need
- Discussion about functions – group by, order by

# Day-3...continued

- Joins and its types
- Types of aggregate functions
- Sub queries overview
- Fetching data from multiple tables
- Assignment – Hands on experience with Postgre SQL



Applications root@ip-10-0-22-238: /h...

root@ip-10-0-22-238: /home/ubuntu

File Edit View Search Terminal Help

```
Main PID: 955 (code=exited, status=0/SUCCESS)
CPU: 1ms
```

```
Aug 22 02:49:10 ip-10-0-22-238 systemd[1]: Starting PostgreSQL RDBMS...
Aug 22 02:49:10 ip-10-0-22-238 systemd[1]: Finished PostgreSQL RDBMS.
```

```
root@ip-10-0-22-238:/home/ubuntu# systemctl start postgresql
root@ip-10-0-22-238:/home/ubuntu# systemctl enable postgresql
Synchronizing state of postgresql.service with SysV service script with /lib/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable postgresql
root@ip-10-0-22-238:/home/ubuntu# postgres psql
Command 'postgres' not found, did you mean:
  command 'postgrey' from deb postgrey (1.36-5.2)
Try: apt install <deb name>
root@ip-10-0-22-238:/home/ubuntu# sduo -u postgres psql
Command 'sduo' not found, did you mean:
  command 'sudo' from deb sudo (1.9.9-1ubuntu2.4)
  command 'sudo' from deb sudo-ldap (1.9.9-1ubuntu2.4)
Try: apt install <deb name>
root@ip-10-0-22-238:/home/ubuntu# sudo -u postgres psql
could not change directory to "/home/ubuntu": Permission denied
psql (14.8 (Ubuntu 14.8-0ubuntu0.22.04.1))
Type "help" for help.
```

```
postgres=# exit
root@ip-10-0-22-238:/home/ubuntu# sudo -u postgres psql
could not change directory to "/home/ubuntu": Permission denied
psql (14.8 (Ubuntu 14.8-0ubuntu0.22.04.1))
Type "help" for help.
```

```
postgres=# \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
(3 rows)
```

```
postgres=# createdatabase DemoDB
```

Applications root@ip-10-0-22-238: /h...

root@ip-10-0-22-238: /home/ubuntu

File Edit View Search Terminal Help

```
postgres=# create database DemoDB;
ERROR:  syntax error at or near "createdatabase"
LINE 1: createdatabase DemoDB
          ^
postgres=# create database DemoDB;
CREATE DATABASE
postgres=# \c DemoDB
connection to server on socket "/var/run/postgresql/.s.PGSQL.5432" failed: FATAL:  database "DemoDB" does not exist
Previous connection kept
postgres=# \l
```

List of databases					
Name	Owner	Encoding	Collate	Ctype	Access privileges
demodb	postgres	UTF8	C.UTF-8	C.UTF-8	
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

(4 rows)

```
postgres=# \c demodb
You are now connected to database "demodb" as user "postgres".
demodb=# create table Student(Sname varchar(20), rollNo int);
CREATE TABLE
demodb=# insert into student values ('XYZ', 2);
INSERT 0 1
demodb=# select * from student
demodb-# ;
sname | rollno
-----+-----
XYZ   |    2
(1 row)
```

```
demodb=# insert into student values ('ABC', 1);
INSERT 0 1
demodb=# select * from student;
sname | rollno
-----+-----
XYZ   |    2
ABC   |    1
(2 rows)
```

Applications root@ip-10-0-22-238: /h...

root@ip-10-0-22-238: /home/ubuntu

```
File Edit View Search Terminal Help
demodb=# select s.rollno, s.sname, sp.sportsid, sp.sportname
rollno | sname | sportid | sportname
-----+-----+-----+
 1 | Abc   | fb123   | football
 2 | Xyz   | fb123   | football
 1 | Abc   | ckt12   | cricket
 2 | Xyz   | ckt12   | cricket
(4 rows)
```

```
demodb=# select s.rollno, s.sname, sp.sportsid, sp.sportname
demodb-# from student s
demodb-# inner join sports sp on
demodb-# s.rollno = sp.studentid;
ERROR: column sp.sportsid does not exist
LINE 1: select s.rollno, s.sname, sp.sportsid, sp.sportname
^
```

HINT: Perhaps you meant to reference the column "sp.sportid".

```
demodb=# select s.rollno, s.sname, sp.sportid, sp.sportname
from student s
inner join sports sp on
s.rollno = sp.studentid;
rollno | sname | sportid | sportname
-----+-----+-----+
 1 | Abc   | fb123   | football
 2 | Xyz   | ckt12   | cricket
(2 rows)
```

```
demodb=# select s.rollno, s.sname, sp.sportid, sp.sportname
from student s
left join sports sp on
s.rollno = sp.studentid;
rollno | sname | sportid | sportname
-----+-----+-----+
 1 | Abc   | fb123   | football
 2 | Xyz   | ckt12   | cricket
(2 rows)
```

```
demodb=# insert into student values(3,"Efg");
ERROR: column "Efg" does not exist
LINE 1: insert into student values(3,"Efg")
^
```

```
demodb=# insert into student values(3,'Efg');
INSERT 0 1
```

```
demodb=# select * from student s1, student s2 where s1.rollno = s2.rollno;
```

rollno	sname	rollno	sname
1	Abc	1	Abc
2	Xyz	2	Xyz
3	Efg	3	Efg

(3 rows)

```
demodb=# select s.rollno, s.sname, sp.sportid, sp.sportname
```

```
from student s
right join sports sp on
s.rollno = sp.studentid;
```

rollno	sname	sportid	sportname
--------	-------	---------	-----------

1	Abc	fb123	football
2	Xyz	ckt12	cricket

(2 rows)

```
demodb=# select s.rollno, s.sname, sp.sportid, sp.sportname
```

```
from student s
full outer join sports sp on
s.rollno = sp.studentid;
```

rollno	sname	sportid	sportname
--------	-------	---------	-----------

1	Abc	fb123	football
2	Xyz	ckt12	cricket
3	Efg		

(3 rows)

```
demodb=# select rollno, sname, sportid, sportname
```

```
from student
```

```
cross join sports;
```

rollno	sname	sportid	sportname
--------	-------	---------	-----------

1	Abc	fb123	football
2	Xyz	fb123	football
3	Efg	fb123	football
1	Abc	ckt12	cricket
2	Xyz	ckt12	cricket
3	Efg	ckt12	cricket

(6 rows)

```
demodb=#
```

```
ABORT
```

```
ALTER
```

```
COMMENT
```

```
COMMIT
```

```
DO
```

```
DROP
```

```
INSERT INTO
```

```
LISTEN
```

```
REFRESH MATERIALIZED VIEW
```

```
REINDEX
```

```
SELECT
```

```
SET
```

```
V
```

```
V
```

# Day-4

## Testing

- Introduction to V-model
- Discussion about various types of environments
- Discussion on types of testing:
  1. Static
  2. Dynamic
    - i. Functional
      - White Box
      - Black Box
    - ii. Non- Functional

# Day-4... continued

- Introduction to STLC and it's various phases
- Deliverables in the various phases of STLC
- Discussion about Testing triangle
- Various approaches for development: TDD, BDD, DDD, ADD
- Discussion about Test case and test case writing
- **ASSIGNMENT** – To write testcases for testing the login functionality

Test case ID	Test Case Description	Pre-requisites/ Pre-conditions	Test Step	Test data/ Test Input	Expected Results	Actual Results	Test Result (Pass/Fail)
TC01	Test case for the valid username and password and successful login	The user id john@gmail.com with password j123 should reside in the system	1. Invoke google chrome and type in the url name.	<a href="http://www.gmail.com">http://www.gmail.com</a>	The gmail homepage should displayed	The gmail homepage was diaplayed	Pass
			2. In the username, enter a valid value.	<a href="mailto:john@gmail.com">john@gmail.com</a>	The field should accept the username	The field accepted the username	Pass
			3. In the password field, enter a valid value.	j123	The field should accept the password	The field accepted the password	Pass
			4. Click on sign in button	Select sign in button	logged in and the gmail page should be displayed	The user was able to logged in and the gmail page was displayed	Pass
TC02	Test case for the valid username and invalid password	The user id john@gmail.com with password j123 should reside in the system	1. Invoke google chrome and type in the url name.	<a href="http://www.gmail.com">http://www.gmail.com</a>	The gmail homepage should displayed	The gmail homepage was diaplayed	Pass
			2. In the username, enter a valid value.	<a href="mailto:john@gmail.com">john@gmail.com</a>	The field should accept the username	The field accepted the username	Pass
			3. In the password field, enter a invalid value.	j12345	The field should not accept the password	The field doesn't accepted the password	Pass
			4. Click on sign in button	Select sign in button	The user should not be able to log in.	The user was not able to logged in	Pass

## Test case creation

	The user id john@gmail.com with password j123 should reside in the system and user id						
TC03	Test case for the invalid username and valid password	johny@gmail.com shouldn't reside in the system	1. Invoke google chrome and type in the url name.  2. In the username, enter a invalid value.  3. In the password field, enter a valid value.  4. Click on sign in button	<a href="http://www.gmail.com">http://www.gmail.com</a>  <a href="mailto:johny@gmail.com">johny@gmail.com</a>  j123  Select sign in button	The gmail homepage should displayed  The field shouldn't accept the username  The field should accept the password  The user should not be able to log in	The gmail homepage was diaplyed  The field doesn't accepted the username  The field accepted the password  The user was not able to logged in	Pass
TC04	Test case for the invalid username and invalid password	The user id johny@gmail.com with password j12345 shouldn't reside in the system	1. Invoke google chrome and type in the url name.  2. In the username, enter a invalid value.  3. In the password field, enter a invalid value.  4. Click on sign in button	<a href="http://www.gmail.com">http://www.gmail.com</a>  <a href="mailto:johny@gmail.com">johny@gmail.com</a>  j12345  Click on sign in button	The gmail homepage should displayed  The field shouldn't accept the username  The field shouldn't accept the password  The user should not be able to log in	The gmail homepage was diaplyed  The field doesn't accepted the username  The field doesn't accepted the password  The user was not able to logged in	Pass
TC05	Test case to verify forget passowrd is giving reset url	The user id john@gmail.com should reside in the system	1. Invoke google chrome and type in the url name.  2. In the username, enter a valid value.  3. Click on forget password button	<a href="http://www.gmail.com">http://www.gmail.com</a>  <a href="mailto:john@gmail.com">john@gmail.com</a>  Click on forget password button	The gmail homepage should displayed  The field should accept the username  The user should be able to reset password through url	The gmail homepage was displayed  The field accepted the username  The user was able to reset the password through url	Pass

# Day-5

DevOps

- Introduction to Dev Ops
- Software Development
- Dev Ops life cycle
- Dev Ops principles
- Dev Ops stages
- Dev Ops tools

# Day-5... continued

- CI and 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

[Files](#)[main](#)[Go to file](#)[.github/workflows](#)[blank.yml](#)[githubActionsDemo / .github / workflows / blank.yml](#)[View Runs](#)

ad722aa · 11 hours ago

[History](#)[Code](#)[Blame](#) 13 lines (13 loc) · 328 Bytes[Raw](#)

```
1 name: learn-github-actions
2 run-name: ${{ github.actor }} is learning GitHub Actions
3 on: [push]
4 jobs:
5   check-bats-version:
6     runs-on: ubuntu-latest
7     steps:
8       - uses: actions/checkout@v3
9       - uses: actions/setup-node@v3
10      with:
11        node-version: '14'
12      - run: npm install -g bats
13      - run: bats -v
```

← learn-github-actions

Re-run all jobs

Latest #3

...

Summary

Jobs

check-bats-version

Run details

Usage

Workflow file

### check-bats-version

succeeded 1 minute ago in 5s

- > ✓ Set up job 1s
- > ✓ Run actions/checkout@v3 1s
- > ✓ Run actions/setup-node@v3 0s
- > ✓ Run npm install -g bats 1s
- > ✓ Run bats -v 0s
- > ✓ Post Run actions/setup-node@v3 0s
- > ✓ Post Run actions/checkout@v3 0s
- > ✓ Complete job 0s

Search logs



# Day-6

Cloud

- 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

# Day-6... continued

- 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

# Day-7

## 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



Azure DevOps tanyasingh15121

/ Case\_study / Boards / Work items

Search

≡ ⌂ ⓘ ⚙ TS

C Case\_study +

Overview

Boards

Work items

Boards

Backlogs

Sprints

Queries

Delivery Plans

Analytics views

Repos

Pipelines

Test Plans

Project settings <

Try the New Boards Hub for improved performance, accessibility, and new features. [Click here](#) to learn more.

## Work items

Recently updated

+ New Work Item

Open in Queries

Column Options

Import Work Items

... ⚙ ⌂ ↗

Filter by keyword Types Assigned to States Area Tags X

ID	Title	Assigned To	State	Area Path
34	Deployment of Java Spring Boot Application	Unassigned	New	Case_study
32	Java Spring Boot Application	Unassigned	New	Case_study
33	Implementing CI/CD workflow	Unassigned	New	Case_study
43	Output Testing	Unassigned	New	Case_study
42	GitHub Actions Workflow	Unassigned	New	Case_study
41	Pull Request Collaboration	Unassigned	New	Case_study
38	Prometheus Monitoring Activity	Unassigned	New	Case_study
39	Docker Creation	Unassigned	New	Case_study
40	Azure Web App Configuration	Unassigned	New	Case_study

[Home >](#)

# CreateVm-canonical.0001-com-ubuntu-server-focal-2-20230828104450 | Overview

Deployment

[Search](#)

&lt;&lt;

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

## Overview

## Inputs

## Outputs

## Template



### Your deployment is complete



Deployment name: CreateVm-canonical.0001-com-ubuntu...

Start time: 8/28/2023, 12:11:25 PM

Subscription: RDBG

Correlation ID: 0406e192-d368-440c-bcf2

Resource group: RDBG

#### Deployment details

#### Next steps

[Setup auto-shutdown](#) Recommended[Monitor VM health, performance and network dependencies](#) Recommended[Run a script inside the virtual machine](#) Recommended[Go to resource](#)[Create another VM](#)

#### 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.

[HOME](#)[FIND OWNERS](#)[VETERINARIANS](#)[ERROR](#)

## Welcome



 **spring**<sup>®</sup> by VMware Tanzu

Connected to tanya15@51.141.42.98

```
(26 kB at 1.0 MB/s)
Downloading from central: https://repo.maven.apache.org/maven2/org/tukaani/xz/1.9/xz-1.9.jar
Downloaded from central: https://repo.maven.apache.org/maven2/org/apache/maven/shared/file-management/3.1.0/file-management-3.1.0.jar (36 kB at 1.1 MB/s)
Downloaded from central: https://repo.maven.apache.org/maven2/org/tukaani/xz/1.9/xz-1.9.jar (116 kB at 1.8 MB/s)
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-io/3.4.0/plexus-io-3.4.0.jar (79 kB at 1.1 MB/s)
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexus/plexus-archiver/4.4.0/plexus-archiver-4.0.jar (211 kB at 3.0 MB/s)
Downloaded from central: https://repo.maven.apache.org/maven2/org/iq80/snappy/snappy/0.4/snappy-0.4.jar (58 kB at 805 kB/s)
[INFO] Building jar: /home/tanya15/shell_7_foundation_Java_Spring/target/spring-petclinic-3.1.0-SNAPSHOT.jar
[INFO]
[INFO] --- spring-boot-maven-plugin:3.1.1:repackage (repackage) @ spring-petclinic ---
[INFO] Replacing main artifact /home/tanya15/shell_7_foundation_Java_Spring/target/spring-petclinic-3.1.0-SNAPSHOT.jar with repackaged archive, adding nested dependencies in BOOT-INF/.
[INFO] The original artifact has been renamed to /home/tanya15/shell_7_foundation_Java_Spring/target/spring-petclinic-3.1.0-SNAPSHOT.jar.original
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 03:23 min
[INFO] Finished at: 2023-08-28T09:03:17Z
[INFO] -----
```

tanya15@CaseStudy-tanya:~/shell\_7\_foundation\_Java\_Spring\$

```
tanya15@CaseStudy-tanya:~/shell_7_foundation_Java_Spring$ sudo nano /etc/systemd/system/prometheus.service
tanya15@CaseStudy-tanya:~/shell_7_foundation_Java_Spring$ sudo systemctl daemon-reload
tanya15@CaseStudy-tanya:~/shell_7_foundation_Java_Spring$ sudo systemctl status prometheus
● prometheus.service - Prometheus
   Loaded: loaded (/etc/systemd/system/prometheus.service; enabled; vendor preset: enabled)
   Active: active (running) since Mon 2023-08-28 08:57:52 UTC; 15min ago
     Main PID: 7498 (prometheus)
        Tasks: 8 (limit: 4685)
       Memory: 38.3M
          CPU: 1.478s
        CGroup: /system.slice/prometheus.service
                  └─7498 /usr/bin/prometheus

Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.895Z caller=head.go:515 level=info component=tsdb
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.895Z caller=head.go:521 level=info component=tsdb
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.895Z caller=head.go:592 level=info component=tsdb
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.895Z caller=head.go:598 level=info component=tsdb
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.896Z caller=main.go:850 level=info fs_type=EXT4_S
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.896Z caller=main.go:853 level=info msg="TSDB star
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.896Z caller=main.go:980 level=info msg="Loading c
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.905Z caller=main.go:1017 level=info msg="Comple
Aug 28 08:57:52 CaseStudy-tanya prometheus[7498]: ts=2023-08-28T08:57:52.905Z caller=main.go:795 level=info msg="Server is
Aug 28 09:12:33 CaseStudy-tanya systemd[1]: prometheus.service: Current command vanished from the unit file, execution of
lines 1-20/20 (END)
```

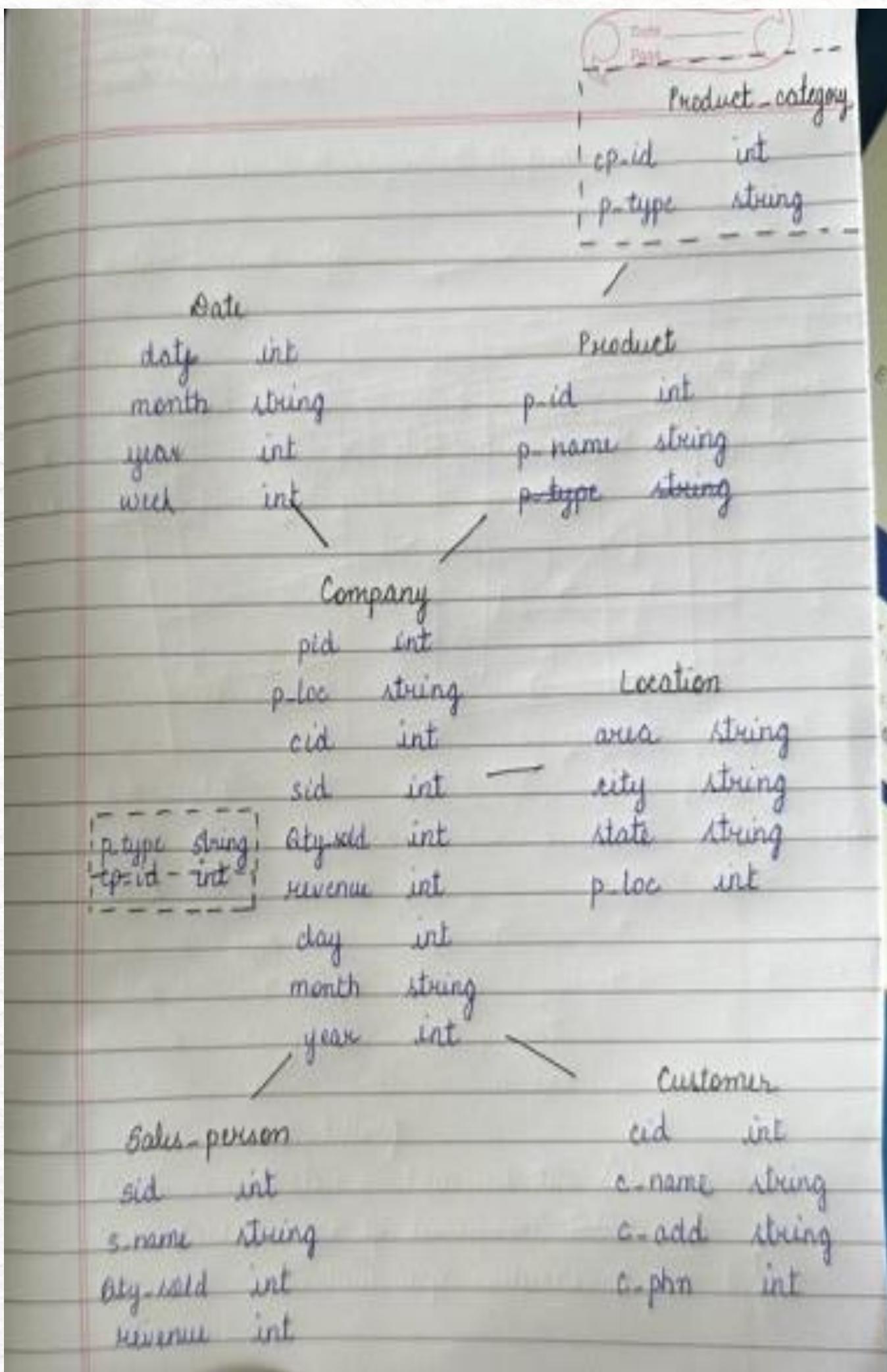
# Day-8

## Database Management System

- ❖ 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)

# Day-8... continued

- ❖ 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
- ❖ SCD (Slowly Changing Dimension)



## Activity

Implement The Concepts Of Star And Snowflake Schema By Drawing Respective Diagrams

# Day-9

## Big Data

- Big Data
- 4 V's Of Big Data ~ Volume, Variety, Velocity, Veracity
- Batch And Stream Data Processing.
- Extract, Load, Transform(ETL) Pipelines
- Processing Big Data
- Parallel Processing And Distributive Computing
- Hadoop Used For Batch Processing & Pyspark Used For Stream Processing.
- Data Warehouse
- ACID Property In DBMS

# Day-9

## Big Data

- Data lake and lakehouse
- Creating A VM Using Microsoft Azure.
- Cloud Computing.
- Types Of Cloud
- Shared Responsibility Model
- Serverless Computing
- Azure Architecture
- Azure Resource Manager

# Day-10

## Azure SQL

- Setting Up Azure SQL Services By Logging In On Azure Portal.
- Introduction To SQL
- DDL (Data Definition Language): create, Alter, Drop, Truncate
- DML(data Manipulation Language): Select, Insert, Update, Delete
- DCL(data Control Language):grant, REVOKE
- Constraints
- Schema In Sql
- Difference Between DELETE, DROP, TRUNCATE.
- Group by, having and where clause
- Grant and Revoke
- Activity : Create Tables For Online Book Store

Microsoft Azure Search resources, services, and docs (G+/-) Home Shellunext\_1693422225...  
UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home > tanyaserver (idatanyaserver/tanyaserver)

## tanyaserver (idatanyaserver/tanyaserver) | Query editor (preview)

SQL database

Search Login New Query Open query Feedback Getting started

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

tanyaserver (tanya15) Refresh

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

> Tables Views Stored Procedures

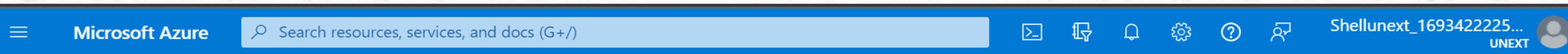
Query 1

Run  Cancel query Save query Export data as Show only Editor

```
1 create table emp (emp_id int, emp_name char)
2 create schema ven
3 create table ven.emp(emp_id int, emp_name char)
```

Results Messages

Query succeeded: Affected rows: 0



Home > tanyaserver (idatanyaserver/tanyaserver)

## tanyaserver (idatanyaserver/tanyaserver) | 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

tanyaserver (tanya15)

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

Tables

> dbo.Author

...

> dbo.Book

...

> dbo.customer

...

> dbo.emp

...

> dbo.Orders

...

> dbo.Publisher

...

> ven.emp

...

Query 1 X

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),Pid int foreign key references Publisher(Pid),B_name varchar(255),B_desc varchar(255));
4  create table Orders (Oid int primary key,odate datetime,bid int foreign key references Book(bid),cid int foreign key references customer(Cid));
5  create table customer(Cid int primary key,c_name varchar(255),oid int foreign key references Orders(Oid));
```

Results

Messages

Query succeeded: Affected rows: 0



# Day-11

Azure SQL

- Create A Database And Elastic Pool
- Joins
- Stored Procedure
- Views
- Functions
- Sub-queries
- UNION And UNION ALL
- Intersection

## 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

# Day-12

## Azure Storage Types

- Resource Group
- Storage Account
- Storage Services
- Redundancy
- RAGRS (Read Access Geo Redundant Storage)
- Security
- Access Protocols
- Access Tiers
- Purging & Archiving
- Containers - Folder For Storing Files
- Azure Tables
- Azure Queues
- File Shares
- Snapshots
- Ingress & Egress
- Static Website

**tanya15\_1693806985733 | Overview**

Deployment

Search Delete Cancel Redeploy Download Refresh

**Your deployment is complete**

Deployment name: tanya15\_169380698... Start time: 9/4/2023, 11:26:31 AM  
Subscription: npunext-1680261340336 Correlation ID: 9732a8be-8a9e-4f7e-8501-74f35040d129  
Resource group: IDA\_Tanya

**Deployment details**

Resource	Type	Status	Operation details
tanya15/default	Microsoft.Storage/stor...	OK	<a href="#">Operation details</a>
tanya15/default	Microsoft.Storage/stor...	OK	<a href="#">Operation details</a>
tanya15	Microsoft.Storage/stor...	OK	<a href="#">Operation details</a>

**Next steps**

[Go to resource](#)

**tanya16 | Containers**

Storage account

Search Container Change access level Restore containers Refresh Delete Give feedback

Events Storage browser

Search containers by prefix Show deleted containers

Name	Last modified	Anonymous access level	Lease state
\$logs	9/4/2023, 12:00:11 PM	Private	Available
gen2	9/4/2023, 1:43:43 PM	Private	Available

**idashell**

Container

Search Upload Change access level ...

Overview Diagnose and solve problems Access Control (IAM) Shared access tokens Access policy Properties Metadata

Authentication method: Access key (Switch to Azure AD User Account)  
Location: idashell

Search blobs by prefix (case-...) Show deleted blobs Add filter

**CASE STUDY.docx**

URL: https://tanya15.blob.cor...  
LAST MODIFIED: 9/4/2023, 11:45:14 AM  
CREATION TIME: 9/4/2023, 11:45:14 AM  
VERSION ID: 2023-09-04T06:15:14.7074289Z  
TYPE: Block blob  
SIZE: 3.18 MiB  
ACCESS TIER: Hot (Inferred)  
ACCESS TIER LAST MODIFIED: N/A  
ARCHIVE STATUS: -  
REHYDRATE PRIORITY: -  
SERVER ENCRYPTED: true  
ETAG: 0x8DBAD0E4DA836F1  
VERSION-LEVEL IMMUTABILITY POLICY: Disabled  
CACHE-CONTROL:  
CONTENT-TYPE: application/vnd.openxmlfor...  
CONTENT MD5: z1z1OkOh/ZavWDnIYW7iA

**CASE STUDY.docx**

Blob Save Discard Download Refresh Delete Change tier ...

Overview Versions Snapshots Edit Generate SAS

Properties URL: https://tanya15.blob.cor...  
LAST MODIFIED: 9/4/2023, 11:45:14 AM  
CREATION TIME: 9/4/2023, 11:45:14 AM  
VERSION ID: 2023-09-04T06:15:14.7074289Z  
TYPE: Block blob  
SIZE: 3.18 MiB  
ACCESS TIER: Hot (Inferred)  
ACCESS TIER LAST MODIFIED: N/A  
ARCHIVE STATUS: -  
REHYDRATE PRIORITY: -  
SERVER ENCRYPTED: true  
ETAG: 0x8DBAD0E4DA836F1  
VERSION-LEVEL IMMUTABILITY POLICY: Disabled  
CACHE-CONTROL:  
CONTENT-TYPE: application/vnd.openxmlfor...  
CONTENT MD5: z1z1OkOh/ZavWDnIYW7iA

Home > Storage accounts > tanya15 | File shares > hello

### hello | Snapshots

SMB File share

Name	Date created	Initiator
2023-09-04T09:15:36.0000000Z	9/4/2023, 2:45:36 PM	Manual
2023-09-04T09:25:25.0000000Z	9/4/2023, 2:55:25 PM	Manual
2023-09-04T09:26:15.0000000Z	9/4/2023, 2:56:15 PM	Manual
2023-09-04T09:27:12.0000000Z	9/4/2023, 2:57:12 PM	Manual

**Operations**

- 
- 
- 
- 
-

**CreateVm-MicrosoftWindowsServer.WindowsServer-201-20230904140526 | Overview**

Deployment

**Overview**

**Your deployment is complete**

Deployment name: CreateVm-MicrosoftWindowsServer.Wi... Start time: 9/4/2023, 2:10:23 PM  
Subscription: npunext-1680261340336 Correlation ID: 6e2f46f0-2be5-49f6-8484-  
Resource group: IDA\_Tanya15

**Deployment details**

**Next steps**

- Recommended
- Recommended
- Recommended

[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

**Administrator: Windows PowerShell**

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

PS C:\Users\tanya15> $connectTestResult = Test-NetConnection -ComputerName tanya15.file.core.windows.net -Port 445
>>> if ($connectTestResult.TcpTestSucceeded) {
>>>     # Save the password so the drive will persist on reboot
>>>     cmd.exe /C "cmdkey /add:``tanya15.file.core.windows.net`` /user:``localhost\tanya15`` /pass:``4IsWVBTQc8SPlhLKpoXmREkGoIP9XnQpgkG5BZ0Pk1T01FGCVAScKJAW2RJ2jE2QdImiXmQKF+Eh+AStVAxeTw==``"
>>>     # Mount the drive
>>>     New-PSDrive -Name Z -PSProvider FileSystem -Root "\\tanya15.file.core.windows.net\hello" -Persist
>>> } else {
>>>     Write-Error -Message "Unable to reach the Azure storage account via port 445. Check to make sure your organization or ISP is not blocking port 445, or use Azure P2S VPN, Azure S2S VPN, or Express Route to tunnel SMB traffic over a different port."
>>> }

CMDKEY: Credential added successfully.

Name          Used (GB)    Free (GB) Provider      Root           CurrentLocation
----          -----        -----       Filesystem   \\tanya15.file.core.windows.net\...
Z              0.00         102400.00

PS C:\Users\tanya15>
```

20.169.198.101 - Remote Desktop Connection

Manage This PC

This PC

Folders (7)

- 3D Objects
- Desktop
- Downloads
- Documents
- Pictures
- Music
- Desktop

Devices and drives (3)

- Windows (C): 116 GB free of 126 GB
- Temporary Storage (D): 14.0 GB free of 15.9 GB
- DVD Drive (E):

Network locations (1)

- hello (\\\tanya15.file.core.windows.net\...)

# tanya16 | Static website

Storage account

Search Save Discard Give feedback

Enabling static websites on the data lake service allows you to host static content. Webpages may include static content and client-side scripts. Server-side scripting is not supported. If a static website is enabled on a data lake service, ACLs will not be honored. As data is replicated asynchronously from primary to secondary regions, files at the secondary endpoint may not be immediately available or in sync with files at the primary endpoint. [Learn more](#)

Static website

Disabled Enabled

An Azure Storage container has been created to host your static website.  
\$web

Primary endpoint <https://tanya16.z13.web.core.windows.net/>

Index document name index.html ✓

Error document path

Back Home https://tanya16.z13.web.core.windows.net

## Team activity

# Day-13

## Azure Storage Types

- Access Keys : Lifetime
- SAS (Shared Access Signature) : Time-bounded
- Azure Key Vaults
- Access Control (IAM) : Check Access Here
- Azure Data Factory
  - a. Data Orchestration
  - b. 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
  - c. Parameterization
  - d. Pipelines

# tanya15 | Shared access signature

Storage account

[Give feedback](#)

[Generate SAS and connection string](#)

## Connection string

```
BlobEndpoint=https://tanya15.blob.core.windows.net/QueueEndpoint=https://tanya15.queue.core.windows.net/FileEndpoint=https://tanya15.file.core.windows.net/
```

## SAS token

```
?sv=2022-11-02&ss=bfqt&srt=c&sp=rwdlacupiytfx&se=2023-09-05T11:51:01Z&st=2023-09-05T03:51:01Z&spr=https&sig=zziiehNC3DBD4W6Cxuo2%2BrezxJ...
```

## Blob service SAS URL

```
https://tanya15.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=c&sp=rwdlacupiytfx&se=2023-09-05T11:51:01Z&st=2023-09-05T03:51:01Z&spr=https&si...
```

## File service SAS URL

```
https://tanya15.file.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=c&sp=rwdlacupiytfx&se=2023-09-05T11:51:01Z&st=2023-09-05T03:51:01Z&spr=https&si...
```

## Queue service SAS URL

```
https://tanya15.queue.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=c&sp=rwdlacupiytfx&se=2023-09-05T11:51:01Z&st=2023-09-05T03:51:01Z&spr=https&si...
```

## Table service SAS URL

```
https://tanya15.table.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=c&sp=rwdlacupiytfx&se=2023-09-05T11:51:01Z&st=2023-09-05T03:51:01Z&spr=https&si...
```

# tanyakey

Key vault

[Delete](#) [Move](#) [Refresh](#) [Open in mobile](#)

⚠ Upcoming TLS 1.0, 1.1 deprecation: Please enable support for TLS 1.2 on clients (applications/platform) to avoid any service impact. Learn more here. →

[JSON View](#)

## Essentials

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

[idatanya](#)

### Vault URI

<https://tanyakey.vault.azure.net/>

### Location

East US

### Sku (Pricing tier)

Standard

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

[npunext-1680261340336](#)

### Directory ID

dce87315-8ffa-4a01-ab40-0de5a7214b2f

### Subscription ID

779629bd-924c-4afb-ba8e-f99be154aa3c

### Directory Name

Unext

### Keys

### Soft-delete

[Enabled](#)

### Secrets

### Purge protection

[Disabled](#)

### Certificates

# tanyakey2 | Secrets

Key vault

[Generate/Import](#)[Refresh](#)[Restore Backup](#)[View sample code](#)[Manage deleted secrets](#)[Overview](#)[Activity log](#)[Access control \(IAM\)](#)[Tags](#)[Diagnose and solve problems](#)

**i** The secret 'tanyasecret' has been successfully created.

Name	Type	Status	Expiration date
tanyasecret		✓ Enabled	

✓ Creating the secret 'tanyasecret'.

The secret 'tanyasecret' has been successfully created.

Microsoft Azure

Data Factory &gt; adftanya

Search factory and documentation

Shellunext\_169342225890@npunext.onmicrosoft.com  
UNEXT

**i** 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!

» [Data Factory](#) [Validate all](#) [Publish all 1](#) [Preview experience](#) [Off](#) [Refresh](#) [Edit](#)

[General](#) [Factory settings](#) [Connections](#) [Linked services](#) [Integration runtimes](#) [Microsoft Purview](#) [Source control](#) [Git configuration](#)

## Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across different network environment. [Learn more](#)

[+ New](#) [Refresh](#)

[Filter by name](#)

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationR...	Azure	Public	✓ Running	0	Auto Resolve	---
integrationRuntime1	Self-Hosted	---	✗ Unavail... <a href="#">More</a>	0	---	---

Microsoft Azure | Data Factory ▶ adftanya

Search factory and documentation

Validate all Publish all 3

Preview experience Off

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name Annotations : Any

Showing 1 - 2 of 2 items

Name	Type	Related	Annotations
AzureBlobStorage1	Azure Blob Storage	1	
Lsidaleshell	Azure Data Lake Storage Gen2	0	

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

Microsoft Azure | Data Factory ▶ adftanya

Search factory and documentation

Validate all Publish all 2

Preview experience Off

Factory Resources

Filter resources by name

Pipelines 0

Change Data Capture (preview) 0

Datasets 1

DS\_Source

Data flows 0

Power Query 0

DS\_Source

Properties

General Related

Name \* DS\_Source

Description

Annotations

+ New

DS\_Source

DelimitedText

Connection Schema Parameters

Linked service \* AzureBlobStorage1

Test connection Edit New Learn more

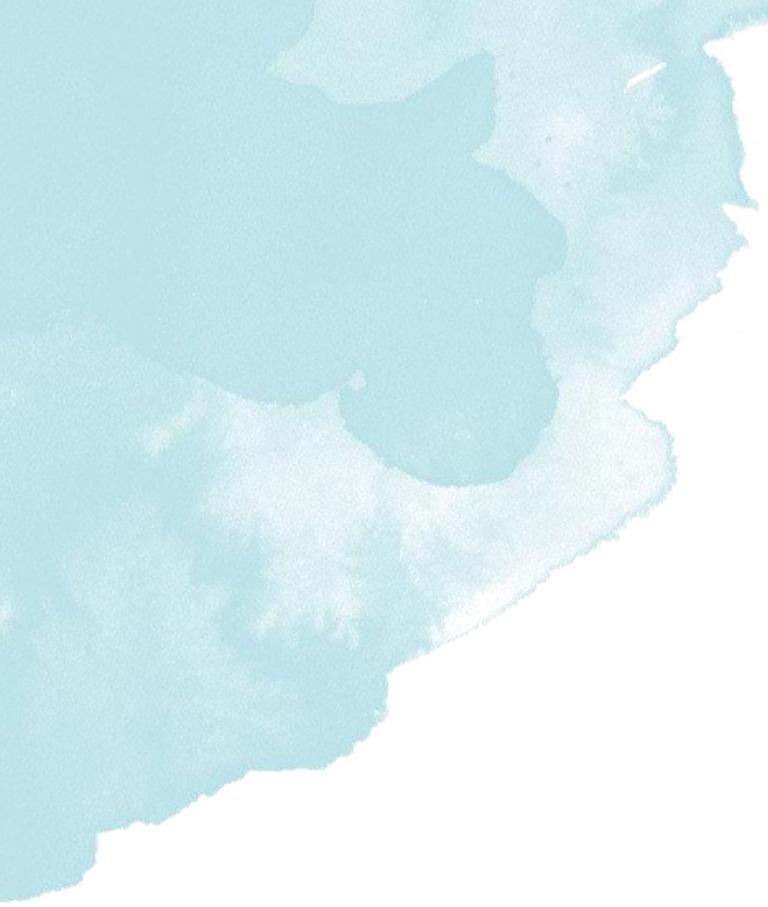
File path \* input / Directory / iris\_d

Compression type Select...

Column delimiter ⓘ Comma (,)

Row delimiter ⓘ Default (\r,\n, or \r\n)

Encoding ⓘ Default(UTF-8)



# Day-14

## Azure Data Factory

- Create Data Factory
- 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
- Copy Data Activity Implementation
- Parameter Concept Implementation
- Triggers
- Get Metadata Activity
- If-else Condition For Get Metadata In Pipelines.
- For Each Loop

Microsoft Azure Data Factory > adftanya

Search factory and documentation

Shellunext\_169342225890@npunext.onmicrosoft.com UNEXT

Microsoft Azure Data Factory > adftanya15

Search factory and documentation

Shellunext\_169342225890@npunext.onmicrosoft.com UNEXT

**Microsoft Fabric** 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!

**Data Factory** Validate all Publish all 2

Preview experience Off

**Activities**

- Validate
- Debug
- Add trigger

**Append variable**

**Copy data**

**Execute Pipeline**

**Append variable**

**PL1\_tanya** pipeline1

**Factory Resources**

DS\_Source PL1\_tanya Output

**Activities**

Pipelines 1

- PL1\_tanya

Validate Validate copy runtime Debug Add trigger

**Copy data**

**Copy\_Customer**

**General Source Sink Mapping Settings User properties**

Source dataset \* DS\_Source

Open New Preview data Learn more

Dataset properties

Name	Value	Type
DS_input	@pipeline().parameters.PL_input	string

File path type

File path in dataset Wildcard file path List of files

```

graph LR
    A[Copy data] --> B[Append variable]
    B --> C[Append variable]
    C --> D[Execute Pipeline]
    D --> E[Append variable]
    E --> F[Append variable]
  
```

Home > satanyagen2 | Containers >

**output** Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

**Overview**

**Authentication method:** Access key ([Switch to Azure AD User Account](#))

**Location:** output

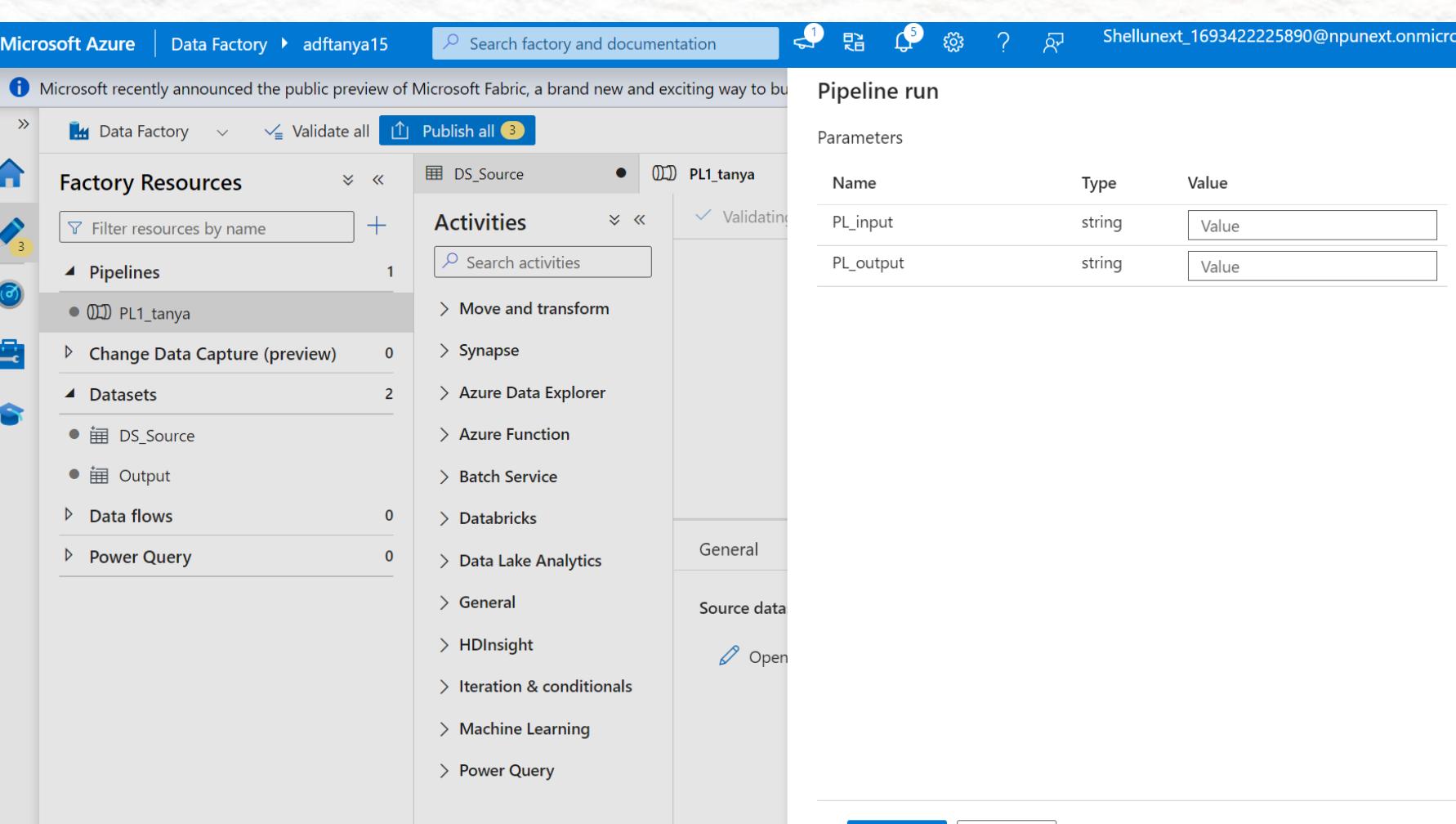
Search blobs by prefix (case-sensitive)

Show deleted objects

**Settings**

Shared access tokens Manage ACL Access policy

Name	Modified	Access tier	Archive status	Blob type	Size
ida.csv	9/6/2023, 11:57:22 AM	Hot (Inferred)		Block blob	2.96
zipcodes.csv	9/6/2023, 11:53:27 AM	Hot (Inferred)		Block blob	2.96



The screenshot shows the Microsoft Azure Data Factory interface. The top navigation bar includes 'Microsoft Azure' and 'Data Factory adftanya15'. A search bar says 'Search factory and documentation'. The top right has icons for notifications, settings, and help.

A banner at the top left says 'Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to bu...'.

The main left sidebar is titled 'Factory Resources' and lists:

- Pipelines (1 item: PL1\_tanya)
- Datasets (2 items: DS\_Source, Output)
- Data flows (0 items)
- Power Query (0 items)

The central workspace shows a pipeline named 'PL1\_tanya'. The 'Activities' section is expanded, showing a list of available activities:

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

To the right of the pipeline, a 'Pipeline run' dialog is open. It shows the following parameters:

Name	Type	Value
PL_input	string	<input type="text"/>
PL_output	string	<input type="text"/>

At the bottom of the dialog are 'File path type' dropdowns, and 'OK' and 'Cancel' buttons.

The screenshot shows the Microsoft Azure Data Factory Pipeline Editor. The top navigation bar includes 'Microsoft Azure' (with a profile icon), 'Data Factory' (selected), 'adftanya15', a search bar ('Search factory and documentation'), and various status indicators (1 notifications, 6 scheduled runs, gear settings, help, and a user icon).

A banner at the top states: "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!"

The main workspace displays a pipeline named 'PL1\_tanya'. The pipeline structure is as follows:

```
graph LR; Start(( )) --> Copy[Copy data: Copy_Customer]; Copy --> GetMetadata[Get Metadata: Get_Metadata1];
```

The 'Activities' pane on the left lists various data integration tasks:

- General: Append variable, Delete, Execute Pipeline, Execute SSIS package, Fail, Get Metadata, Lookup, Stored procedure, Script, Set variable.

The 'Output' pane shows the pipeline's run history:

type	Run start	Duration	Log
Get Metadata1	9/6/2023, 2:22:12 PM	4s	
Copy Customer	9/6/2023, 2:22:02 PM	9s	

The 'Pipeline status' is listed as 'Succeeded'. There are links to 'View debug run consumption', 'Monitor in Azure Metrics', and 'Export to CSV'.

Microsoft Azure | Data Factory > adftanya15 Search factory and documentation 1 ? Shellunext\_1693422225890@npunext.onmicrosoft.com

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!

Data Factory > DS\_Source > PL1\_tanya > Output

Preview experience

Activities

- Execute SSIS package
- Fail
- Get Metadata
- Lookup
- Stored procedure
- Script
- (X) Set variable
- Validation
- Web
- WebHook
- Wait
- HDInsight

PL1\_tanya

Output

Validate all Publish all 1

Validate Debug Add trigger

Copy data Get Metadata Set variable

Copy\_Customer Get Metadata1 Set variable1

Output

Copy to clipboard

{  
    "name": "var1",  
    "value": 3035  
}

Pipeline status Succeeded View debug run

Monitor in Azure Metrics Exp

type	Run start	Duration	Log
Set variable1	Succeeded	9/6/2023, 2:33:47 PM	Less than 1s
Get Metadata1	Succeeded	9/6/2023, 2:33:43 PM	3s
Get Metadata			

Microsoft Azure | Data Factory > adftanya15 | Search factory and documentation | 1 2 3 7 ? 1 UNEXT Shellunext\_169342225890@npunext.onmicrosoft.com

Data Factory Validate all Publish all 2 Preview experience Off

DS\_Source PL1\_tanya Output child\_PL

Activities Validate Debug Add trigger { }

Search activities

Execute pipeline

Execute SSIS package

Fail

Get Metadata

Lookup

Stored procedure

Script

(x) Set variable

Validation

Web

WebHook

Wait

HDInsight

Get Metadata1

If Condition1

True

Copy data1

False

Parameters Variables Settings Output

Pipeline run ID: a96dbdc2-7da6-4b9e-96ea-4b453d27c3ad

Pipeline status Succeeded View debug run consumption

All status List

Showing 1 - 3 of 3 items

Activity name	Activity status	Activity type	Run start	Duration	Log
Copy data1	Succeeded	Copy data	9/6/2023, 3:00:42 PM	9s	



# Day-15

Azure Data Factory

- ❑ Data(resource)transformation Using Data Flow.
- ❑ For Each Activity
- ❑ Data Flows
- ❑ Dataflow Expression Builder And Derived Columns
- ❑ Surrogate Key
- ❑ For Each Container And Get Metadata
- ❑ API

Microsoft Azure | Data Factory > adftanya

Search factory and documentation

Validate all Publish all 4

Preview experience Off

PL\_tanya DS\_Output DS\_S\_Cust DS\_S\_Order DS\_Orders DS\_Odetails dataf ...

Activities Validate Debug Add trigger Data flow debug

Data flow dataflow1

Parameters Variables Settings Output

Pipeline run ID: b03bb5d0-379d-477e-bc6f-055f96904c5f

Pipeline status Succeeded View debug run consumption

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 1 of 1 items

Activity name Activity status Run start Duration Integration runtime User properties Activi

Microsoft Azure | Data Factory > adftanya

Search factory and documentation

Validate all Publish all 1

Preview experience Off

DS\_Output DS\_S\_Cust DS\_S\_Order DS\_Orders DS\_Odetails dataflow1

Validate Data flow debug Debug Settings

source1 Import data from DS\_Orders + join1 Inner join on 'source1' and 'source2' + select1 Renaming join1 to select1 with columns 'Amount, Quantity' + filter1 Filtering rows using expressions on columns 'Amount' + derivedColumn1 Creating/updating the columns 'Amount, Quantity, Updated\_Amt' + sort1 Sorting rows on columns 'Updated\_Amt' + sink1 Columns: 3 total

source2 Import data from DS\_Odetails +

Sink Settings Errors Mapping Optimize Inspect Data preview

Number of rows + INSERT N/A \* UPDATE N/A × DELETE N/A + UPSERT N/A 🔎 LOOKUP N/A ✖ ERROR N/A TOTAL 354

Refresh Statistics Export to CSV

Amount	Quantity	Updated_Amt
3873.00	6	38730
3151.00	7	31510
2830.00	13	28300
2617.00	4	26170

Microsoft Azure | Data Factory > adftanya

Search factory and documentation

Validate all Publish all 1

Preview experience Off

PL\_tanya DS\_Output DS\_S\_Cust DS\_S\_Order DS\_Orders DS\_Odetails dataflow1

Validate Data flow debug Debug Settings

source1 Import data from DS\_Orders + join1 Inner join on 'source1' and 'source2' + select1 Renaming join1 to select1 with columns 'Amount, Quantity' + filter1 Filtering rows using expressions on columns 'Amount' + derivedColumn1 Creating/updating the columns 'Amount, Quantity, Updated\_Amt' + sort1 Sorting rows on columns 'Updated\_Amt' + sink1 Columns: 3 total

source2 Import data from DS\_Odetails +

Sink Settings Errors Mapping Optimize Inspect Data preview

Number of rows + INSERT N/A \* UPDATE N/A × DELETE N/A + UPSERT N/A 🔎 LOOKUP N/A ✖ ERROR N/A TOTAL 354

Refresh Statistics Export to CSV

Amount	Quantity	Updated_Amt
3873.00	6	38730
3151.00	7	31510
2830.00	13	28300
2617.00	4	26170

The screenshot shows the Microsoft Azure Storage Explorer interface. On the left, the 'output1' container is selected under 'Containers'. The 'Overview' tab is active, displaying the container's details: 'Authentication method: Access key (Switch to Azure AD User Account)' and 'Location: output1'. A search bar and a 'Show deleted objects' toggle are also present. In the center, a blob named 'todos' is displayed. The 'Edit' tab is selected, showing the blob's content as a JSON file:

```
1 {"todos": [{"id": 1, "todo": "Do something nice for someone I care about", "completed": false}]} 
```

Microsoft Azure | Data Factory ▶ adftanya Search factory and documentation 1 8 ? ? Shellunext\_1693422225890@npunext.onmicrosoft.com UNEXT

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!

Data Factory Validate all Publish all (6) Preview experience Off

DS\_new PL\_tanya PL\_new DS\_newfile DS\_new\_source DS\_new\_output

Activities Validate Debug Add trigger

Get Metadata Get Metadata1

ForEach ForEach1

Activities Copy data1

Copy data1 Succeeded 9/7/2023, 2:53:02 PM 8s AutoResolveIntegration

Copy data1 Succeeded 9/7/2023, 2:53:02 PM 9s AutoResolveIntegration

Copy data1 Succeeded 9/7/2023, 2:53:02 PM 8s AutoResolveIntegration

ForEach1 Succeeded 9/7/2023, 2:53:02 PM 11s

Get Metadata1 Succeeded 9/7/2023, 2:52:58 PM 2s AutoResolveIntegration



# Day-16

Azure Data Factory

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

Microsoft Azure | Data Factory > adftanya Search factory and documentation 1 Help ? Feedback Shellunext\_169342225890@npunext.onmicrosoft.com UNEXT

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!

Data Factory Validate all Publish all Preview experience Off

PL1

Activities

Search activities

> Move and transform

> Synapse

> Azure Data Explorer

> Azure Function

> Batch Service

> Databricks

> Data Lake Analytics

> General

> HDInsight

> Iteration & conditionals

> Machine Learning

> Power Query

Lookup

ForEach

Copy data1

Validate Debug Add trigger

Parameters Variables Settings Output

Pipeline run ID: 273adc81-ede3-4d0e-b1f3-ca429f482fbf View debug run consumption

Pipeline status Succeeded

All status List Monitor in Azure Metrics Export to CSV

Showing 1 - 18 of 18 items

Activity name	Activity status	Run start	Duration	Integration runtime	User properties	Act
Copy data1	<span>Succeeded</span>	9/8/2023, 1:51:20 PM	11s	AutoResolveIntegration	6f9	
Copy data1	<span>Succeeded</span>	9/8/2023, 1:51:20 PM	10s	AutoResolveIntegration	4e5	

# Day-17

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

We use optional cookies to provide a better experience. [Learn more](#)

Accept

Reject

More options



Synapse live

Validate all

Publish all 1



input

SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Run Undo Publish Query plan

Connect to Built-in

Use database master



```
2 with cte_ida AS(
3     SELECT
4         *
5     FROM
6         OPENROWSET(
7             BULK 'https://tanyasingh.dfs.core.windows.net/input/zipcodes.csv',
8             FORMAT = 'CSV',
9             PARSER_VERSION = '2.0',
10            HEADER_ROW = TRUE
11        ) AS [result]
12    )
13    select * from cte_ida
14
```

### Properties

General Related (0)

Name \*

SQL script 1

Description

Type

.sql script

Size

223 bytes

Results settings per query

First 5000 rows (default)

All rows

Results Messages

View Table

Chart

Export results

Search

RecordNumber

Zipcode

ZipCodeType

City

State

LocationType

Lat

1

704

STANDARD

PARC PARQUE

PR

NOT ACCEPTAB...

17.96

Synapse Analytics ▶ tanya15

cookies to provide a better experience. [Learn more](#)

Accept Reject More options ×

live ▾ Validate all Publish all 1

Apache Spark pool

Apache Spark pools can be tuned to run different kinds of Apache Spark workloads using specific configuration libraries, permissions, etc. [Learn more](#)

+ New Refresh

Filter by name

Showing 1-1 of 1 item

Name	Node size family	Size
tanya15	Memory Optimized	Medium (8 vCores / 64 GB) - 3 to 10 nodes

The screenshot shows the Microsoft Azure Synapse Analytics workspace interface. The top navigation bar includes 'Microsoft Azure' and 'Synapse Analytics > tanya15'. A search bar and various workspace management icons are also present. A cookie consent banner at the top right provides options to 'Accept', 'Reject', or 'More options'. The main workspace area has a sidebar on the left with icons for Home, Databases, Tables, Scripts, Notebooks, and Pipelines. The 'Develop' section is selected, showing a list of resources: 'SQL scripts' (1 item: 'SQL script 1') and 'Notebooks' (1 item: 'Notebook 1'). The 'Notebook 1' tab is active, displaying the following code and output:

```
1 print("Hello")
```

[1] ✓ 3 min 54 sec 3 min 53 sec 649 ms. Command executed in 222 ms by Shellunext\_1693

... Hello

The notebook toolbar includes 'Run all', 'Undo', 'Publish', 'Outline', and an 'Attach to' dropdown set to 'tanya15'. A message box on the right informs the user that other users in the workspace may have access to modify the item. The 'Properties' panel on the right shows the 'General' tab selected, with the name set to 'Notebook 1'.

# Day-18

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

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Manage Manage Columns Reduce Rows Sort Split Column Group By Replace Values Data Type: Any Use First Row as Headers Combine Text Analytics Vision Azure Machine Learning Close New Query Data Sources Parameters Query Transform AI Insights

Queries [2]

= Table.AddColumn(#"Added Custom1", "Short\_names", each if [Country] = "Germany" then "GER")

	Year	Total_Sales	SegCountry	Short_names
1	2014	32390	Government_Canada Changed@shell.com	CAN
2	2014	26440	Government_Germany@shell.com	GER
3	2014	32685	Midmarket_France@shell.com	France
4	2014	13335	Midmarket_Germany@shell.com	GER
5	2014	37065	Midmarket_Mexico@shell.com	Mexico
6	2014	529900	Government_Germany@shell.com	GER
7	2014	13830	Midmarket_Germany@shell.com	GER
8	2014	30228	Channel Partners_Canada Changed@shell.com	CAN
9	2014	38000	Government_France@shell.com	France
10	2014	18552	Channel Partners_Germany@shell.com	GER
11	2014	37065	Midmarket_Mexico@shell.com	Mexico
12	2014	333313	Enterprise_Canada Changed@shell.com	CAN
13	2014	287700	Small Business_Mexico@shell.com	Mexico
14	2014	15029	Government_Germany@shell.com	GER
15	2013	43250	Enterprise_Canada Changed@shell.com	CAN
16	2014	9240	Midmarket_United States of America@shell.c...	United States of America
17	2014	5860	Government_Canada Changed@shell.com	CAN
18	2014	14625	Midmarket_Mexico@shell.com	Mexico
19	2014	30228	Channel Partners_Canada Changed@shell.com	CAN

Query Settings

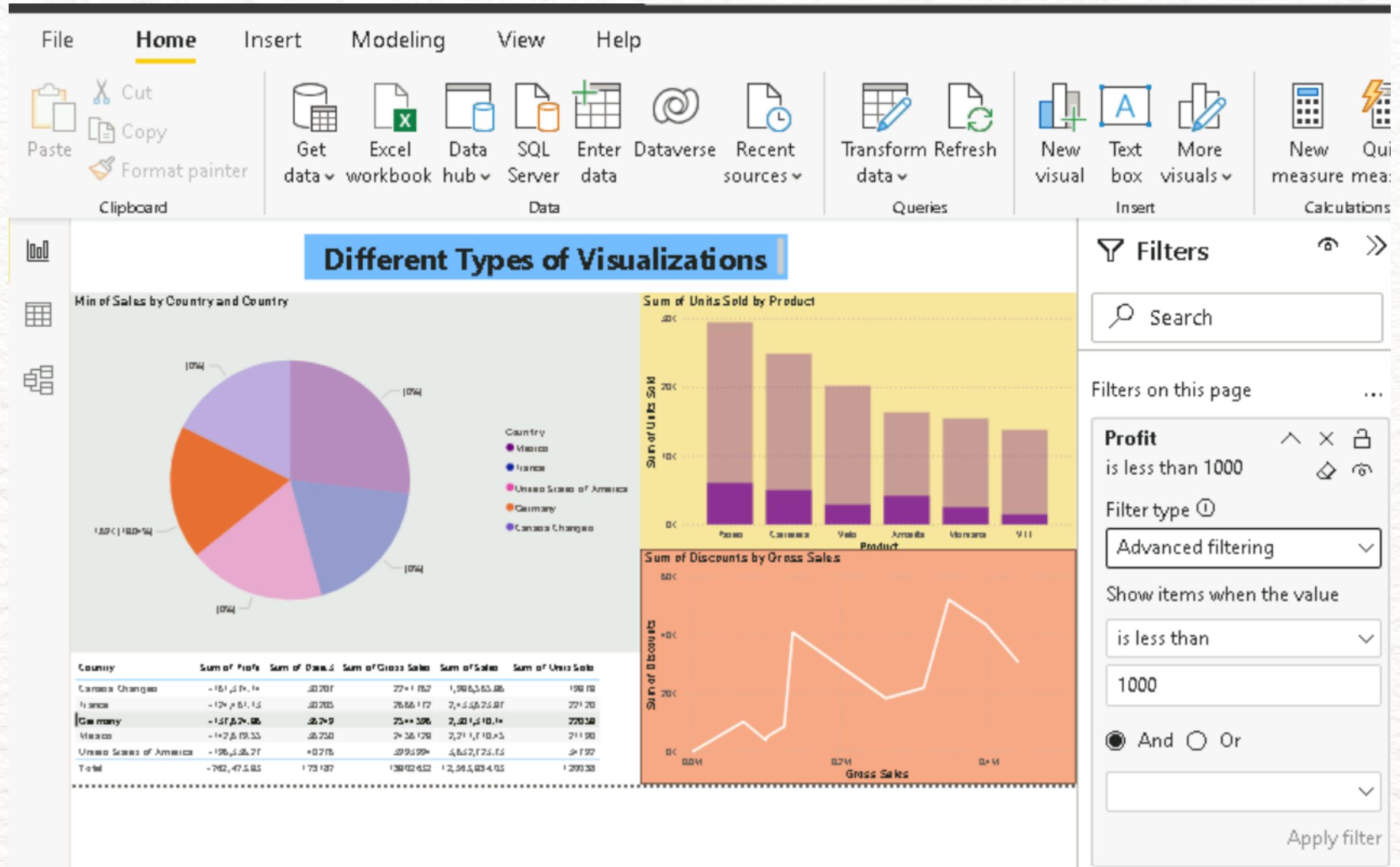
PROPERTIES

Name: financials

All Properties

APPLIED STEPS

Source, Navigation, Changed Type, Split Column by Delimiter, Changed Type1, Replaced Value, Changed Type2, Rounded Up, Added Custom, Added Custom1, Added Conditional Column



First Report - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Manage Properties Advanced Editor Manage Columns Reduce Rows Sort

Close New Query Data Sources Parameters Query

Queries [3]

financials

Sheet1

source\_param (C:\Users\...)

Current Value C:\Users\Administrator\Desktop\Financial.xlsx

Manage Parameter

First Report - Power BI Desktop

File Home Insert Modeling View Help Format Data / Drill

Cut Copy Paste Format painter Clipboard

Get data workbook hub Data SQL Server Enter data Dataverse Recent sources Transform Refresh data New visual Text box More visuals New measure Quick measure Sensitivity Publish Share

Min of Sales by Country and Country

Country United States of America

Canada Changed Mexico

France United States of America

Germany United States of America

1.72 < (100%)

Visualizations

Format visual Filters

Search

Visual General

Slicer settings

Options Orientation Horizontal Selection Reset to default Slicer header On

Page 1 +

This screenshot shows the Power BI Desktop interface. The main area displays a matrix visual titled "Min of Sales by Country and Country". The matrix has "Country" on both the rows and columns. The visible data points are Canada (Changed), Mexico, France, United States of America, and Germany. A large pink circle is overlaid on the matrix. To the right, the "Slicer settings" pane is open, showing options for orientation (set to horizontal) and selection. The "Visual" tab is selected. The ribbon at the top includes tabs like File, Home, Insert, Modeling, View, Help, Format, and Data / Drill.

First Report - Power BI Desktop

File Home

Manage relationships Relationships

New measure

Manage roles

Roles

CEO Manager New role

Create Delete

Tables

financials Sheet1

Table filter DAX expression

```
[Country] = "Value" && [Product] = "Value" && [Segment] = "Value"
```

S >

search

financials Sheet1

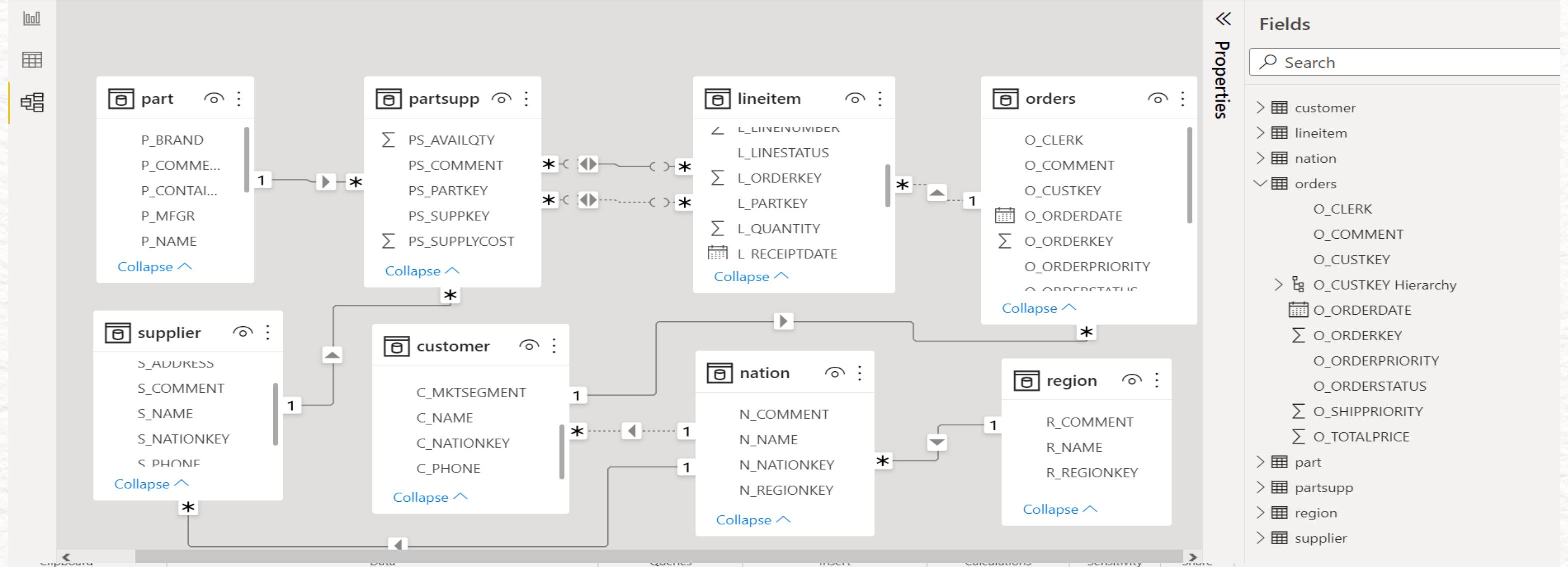
This screenshot shows the "Manage roles" dialog in Power BI Desktop. It lists roles such as CEO, Manager, and New role. Below this is a "Tables" section showing a table named "financials" with a single row "Sheet1". To the right, a "Table filter DAX expression" editor is open, displaying the formula: "[Country] = "Value" && [Product] = "Value" && [Segment] = "Value"". The ribbon at the top includes tabs like File, Home, Insert, Modeling, View, Help, Format, and Data / Drill.



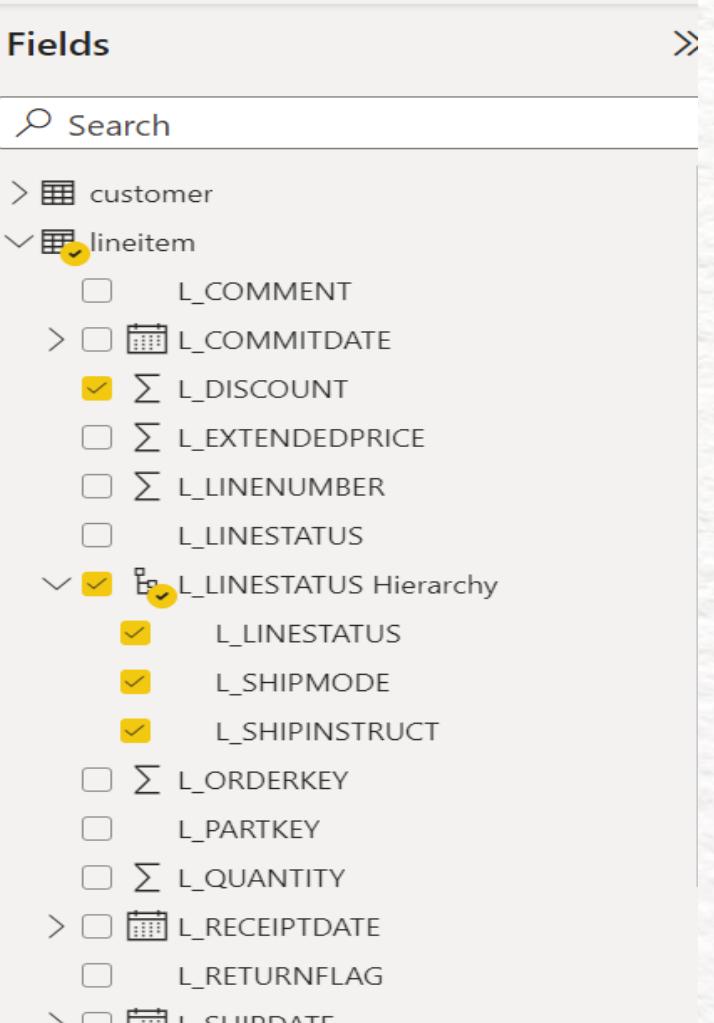
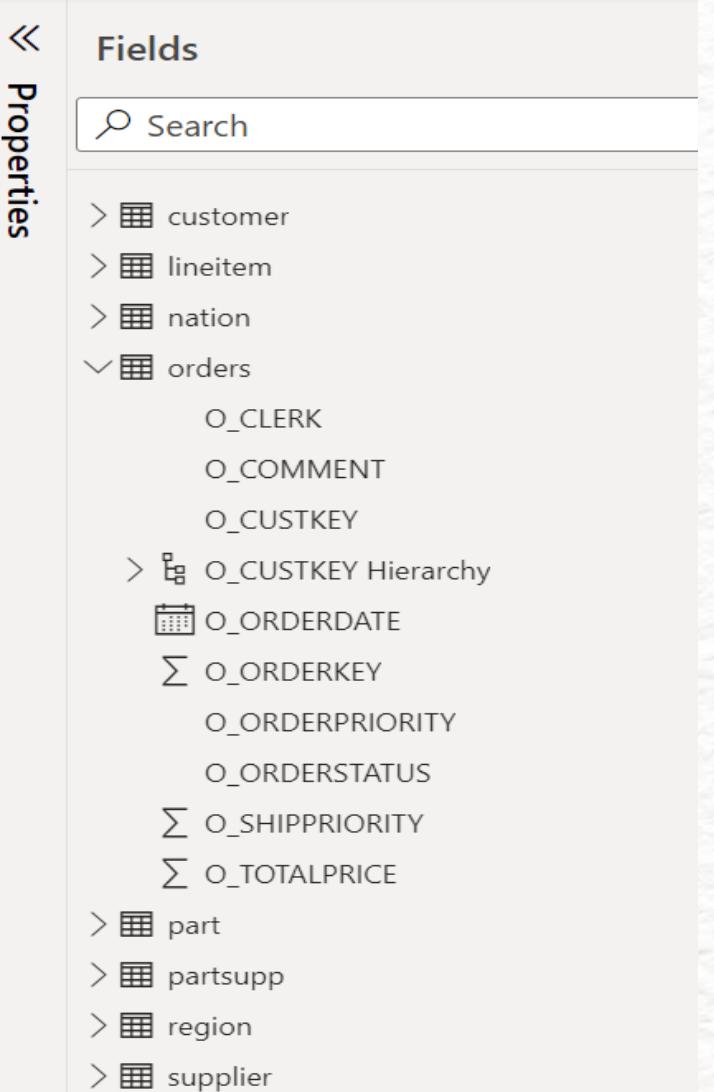
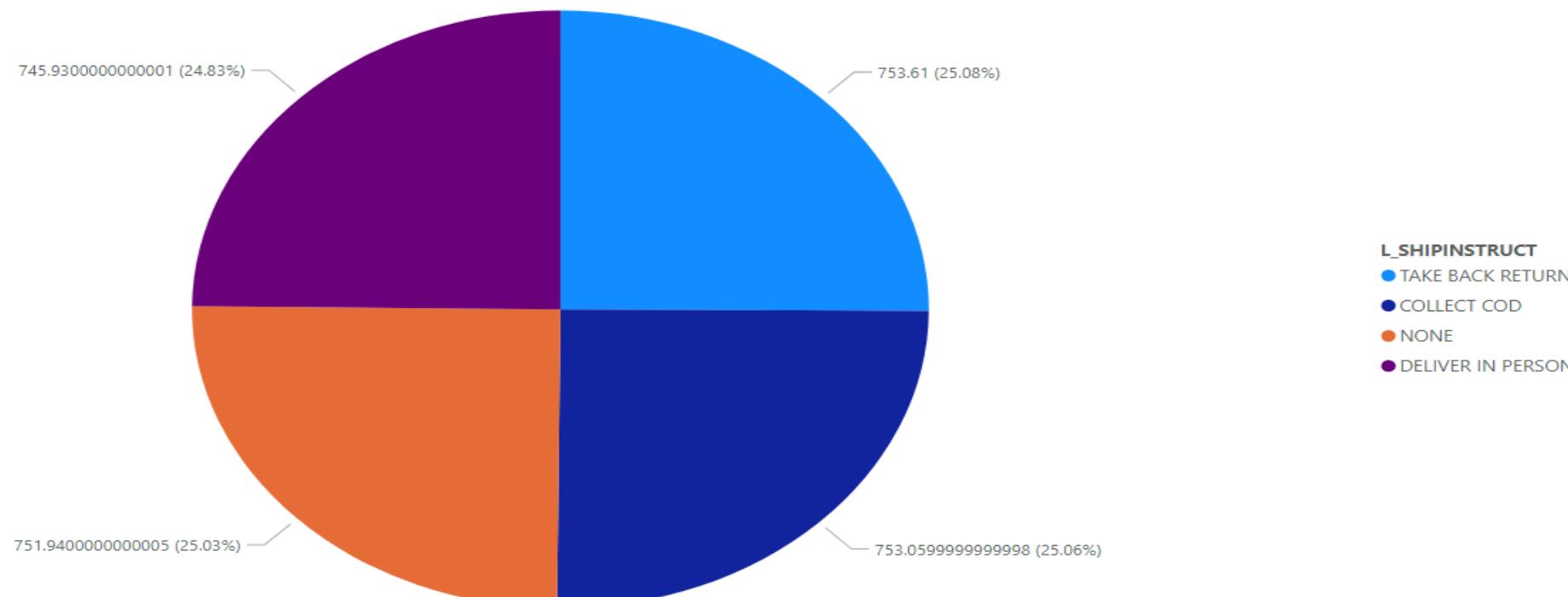
# Day-19

Power BI

- 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
- 8.Case Studies



## Drill through graph



**Table tools**

Name: DateTable

Mark as date table ▾ Manage relationships

New measure New measure column New table

Structure Calendars Relationships Calculations

```

1 DateTable =
2 ADDCOLUMNS(
3   CALENDAR(DATE(2023,1,1), DATE(2023,12,31)),
4   "Year", YEAR([Date]),
5   "Month Number", MONTH([Date]),
6   "Month Name", FORMAT([Date], "MMMM"),
7   "Quarter", QUARTER([Date]),
8   "Day of Week", WEEKDAY([Date]),
9   "Day Name", FORMAT([Date], "dddd"),
10  "Year-Month", FORMAT([Date], "YYYY-MM")
11 )
12

```

Date	Year	Month Number	Month Name	Quarter	Day of Week	Day Name	Year-Month
7/1/2023 12:00:00 AM	2023	7	July	3	7	Saturday	2023-07
7/2/2023 12:00:00 AM	2023	7	July	3	1	Sunday	2023-07
7/3/2023 12:00:00 AM	2023	7	July	3	2	Monday	2023-07
7/4/2023 12:00:00 AM	2023	7	July	3	3	Tuesday	2023-07
7/5/2023 12:00:00 AM	2023	7	July	3	4	Wednesday	2023-07
7/6/2023 12:00:00 AM	2023	7	July	3	5	Thursday	2023-07
7/7/2023 12:00:00 AM	2023	7	July	3	6	Friday	2023-07
7/8/2023 12:00:00 AM	2023	7	July	3	7	Saturday	2023-07
7/9/2023 12:00:00 AM	2023	7	July	3	1	Sunday	2023-07
7/10/2023 12:00:00 AM	2023	7	July	3	2	Monday	2023-07
7/11/2023 12:00:00 AM	2023	7	July	3	3	Tuesday	2023-07
7/12/2023 12:00:00 AM	2023	7	July	3	4	Wednesday	2023-07
7/13/2023 12:00:00 AM	2023	7	July	3	5	Thursday	2023-07
7/14/2023 12:00:00 AM	2023	7	July	3	6	Friday	2023-07
7/15/2023 12:00:00 AM	2023	7	July	3	7	Saturday	2023-07

**Home**

Cut Copy Paste Get data from workbook hub Data SQL Server Enter data Recent sources Transform data Refresh data Insert Queries Calculations Sensitivity Share

Clipboard

### Drill Bit Changes

Date Friday, September 01, 2023 Saturday, September 02, 2023 Sunday, September 03, 2023 Monday, September 04, 2023 Tuesday, September 05, 2023

The chart displays the distribution of Drill Bit Changes across four categories (1, 2, 3, 4). Rig A is represented by blue segments, and Rig B by purple segments. Category 1 is the largest, followed by Category 2, then Category 3, and Category 4 is the smallest.

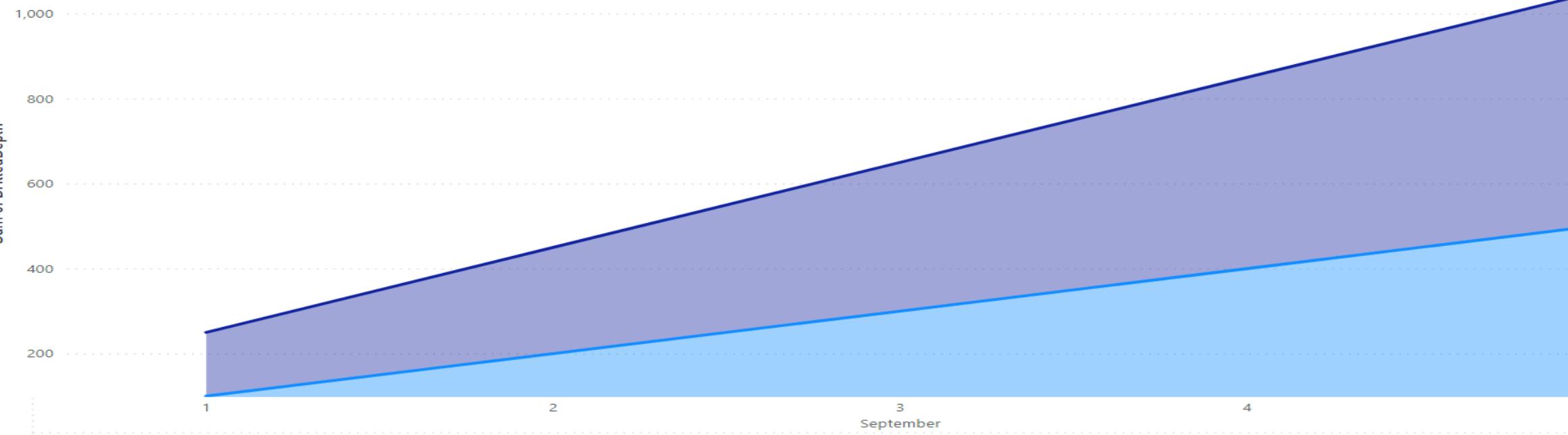
Category	RigName	Value
1	Rig A	~45%
1	Rig B	~35%
2	Rig A	~25%
2	Rig B	~25%
3	Rig A	~15%
3	Rig B	~15%
4	Rig A	~5%
4	Rig B	~5%

Fields

- customer
  - C\_ACCTBAL
  - C\_ADDRESS
  - C\_COMMENT
  - C\_CUSTKEY
  - C\_MKTSEGMENT
  - C\_NAME.1
  - C\_NAME.2
  - C\_NATIONKEY
  - C\_PHONE
- DateTable
- lineitem
- nation
- orders
- part
- partsupp
- region
- supplier

## Drilled Depth of each Rig

RigName ● Rig A ● Rig B



Date  
9/1/2023 9/5/2023

September  
2023  
Day

## Summary Report

RigName	Sum of DrilledDepth	Sum of DrillingSpeed	Sum of OperatingHours	Sum of DrillBitChanges	Year	Quarter	Month	Day
Rig A	100.00	20.00	10.00	1	2023	Qtr 3	September	1
Rig A	200.00	18.00	12.00	2	2023	Qtr 3	September	2
Rig A	300.00	22.00	9.00	0	2023	Qtr 3	September	3
Rig A	400.00	19.00	11.00	1	2023	Qtr 3	September	4
Rig A	500.00	21.00	8.00	0	2023	Qtr 3	September	5
Rig B	150.00	25.00	9.00	0	2023	Qtr 3	September	1
Rig B	250.00	24.00	10.00	1	2023	Qtr 3	September	2
Rig B	350.00	26.00	8.00	0	2023	Qtr 3	September	3
Rig B	450.00	23.00	11.00	2	2023	Qtr 3	September	4
Rig B	550.00	27.00	7.00	0	2023	Qtr 3	September	5
<b>Total</b>	<b>3,250.00</b>	<b>225.00</b>	<b>95.00</b>	<b>7</b>				

RigName  
 Rig A  
 Rig B

9/1/2023 9/5/2023



# Day-20

Python

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

SQLQuery1.sql - EC2AMAZ-OIIOMA\SQLEXPRESS.Sales (EC2AMAZ-OIIOMA\Administrator (55))\* - Microsoft SQL Server Management Studio (Administrator)

File Edit View Project Debug Tools Window Help

New Query Execute Debug

Sales

Object Explorer

SQLQuery1.sql - EC...Administrator (55)\*

```
CREATE database Sales
Use Sales
CREATE TABLE SalesTransactions (
    TransactionID INT PRIMARY KEY,
    ProductID INT,
    StoreID INT,
    TransactionDate DATE,
    QuantitySold INT,
    Revenue DECIMAL(10, 2)
);

-- Insert sample data into Sales Transactions Data
INSERT INTO SalesTransactions (TransactionID, ProductID, StoreID, TransactionDate, QuantitySold, Revenue)
VALUES
    (1, 1, 1, '2023-09-15', 10, 100),
    (2, 2, 1, '2023-09-15', 15, 150),
    (3, 3, 1, '2023-09-15', 20, 200),
    (4, 4, 1, '2023-09-15', 25, 250),
    (5, 5, 1, '2023-09-15', 30, 300),
    (6, 6, 1, '2023-09-15', 35, 350),
    (7, 7, 1, '2023-09-15', 40, 400)
```

Properties

Current connection parameters

Aggregate Status

- Connection failures
- Elapsed time 00:00:00.032
- Finish time 9/15/2023 9:21:09 AM
- Name EC2AMAZ-OIIOMA\SQLEXPRESS
- Rows returned 0
- Start time 9/15/2023 9:21:09 AM
- State Open

Connection

- Connection name EC2AMAZ-OIIOMA\SQLEXPRESS

Connection Details

- Connection elapsed 00:00:00.032
- Connection finish time 9/15/2023 9:21:09 AM
- Connection rows returned 0
- Connection start time 9/15/2023 9:21:09 AM
- Connection state Open
- Display name EC2AMAZ-OIIOMA\SQLEXPRESS
- Login name EC2AMAZ-OIIOMA\Administrator
- Server name EC2AMAZ-OIIOMA\SQLEXPRESS
- Server version 12.0.2000
- Session Tracing ID
- SPID 55

Messages

(7 row(s) affected)

(3 row(s) affected)

(7 row(s) affected)

100 % < >

Query executed successfully.

EC2AMAZ-OIIOMA\SQLEXPRESS ... | EC2AMAZ-OIIOMA\Administrator ... | Sales | 00:00:00 | 0 rows

File Home Insert Modeling View Help

Cut Copy Format painter Paste

Get data from workbook hub Data Server Recent sources

Transform Refresh data New visual Text box More visuals

New measure Quick Sensitivity Publish

Sensitivity Calculations Share

View recovered files

Auto recovery contains some recovered files that haven't been opened.

Load

- Customers 7 rows loaded.
- Products 7 rows loaded.
- SalesTransactions 10 rows loaded.
- Stores 3 rows loaded.

Import data from Excel Import data from

Build visual

Filters

Search

Visualizations Fields

Build visual

Values Add data fields here

# Case Study

Power BI Data Editor																																																																																							
File		Home		Transform		Add Column		View		Tools																																																																													
	New Source		Enter Data		Manage Parameters			Properties	Choose Columns		Keep Rows		A-Z	Z-A	Sort	Data Type: Whole Number	Merge Queries	Text Analytics																																																																					
Close	New Query	Data Sources	Parameters	Query	Manage Columns	Reduce Rows	Sort	Transform	Split Column	Group By	1 2 Replace Values	Combine	AI Insights																																																																										
Queries [4]												Query Settings																																																																											
<table border="1"> <thead> <tr> <th colspan="6">= Sales{[Schema="dbo",Item="SalesTransactions"]}[Data]</th> </tr> <tr> <th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th> </tr> </thead> <tbody> <tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr> <tr><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr> <tr><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr> <tr><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td></tr> <tr><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr> <tr><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td></td></tr> <tr><td>7</td><td>8</td><td>9</td><td>10</td><td></td><td></td></tr> <tr><td>8</td><td>9</td><td>10</td><td></td><td></td><td></td></tr> <tr><td>9</td><td>10</td><td></td><td></td><td></td><td></td></tr> <tr><td>10</td><td></td><td></td><td></td><td></td><td></td></tr> </tbody> </table>												= Sales{[Schema="dbo",Item="SalesTransactions"]}[Data]						1	2	3	4	5	6	1	2	3	4	5	6	2	3	4	5	6	7	3	4	5	6	7	8	4	5	6	7	8	9	5	6	7	8	9	10	6	7	8	9	10		7	8	9	10			8	9	10				9	10					10						<b>PROPERTIES</b> Name: SalesTransactions <a href="#">All Properties</a>			
= Sales{[Schema="dbo",Item="SalesTransactions"]}[Data]																																																																																							
1	2	3	4	5	6																																																																																		
1	2	3	4	5	6																																																																																		
2	3	4	5	6	7																																																																																		
3	4	5	6	7	8																																																																																		
4	5	6	7	8	9																																																																																		
5	6	7	8	9	10																																																																																		
6	7	8	9	10																																																																																			
7	8	9	10																																																																																				
8	9	10																																																																																					
9	10																																																																																						
10																																																																																							
<b>APPLIED STEPS</b> Source: Navigation																																																																																							

Power BI Data Editor																																																																																							
File		Home		Transform		Add Column		View		Tools																																																																													
	New Source		Enter Data		Manage Parameters			Properties	Choose Columns		Keep Rows		A-Z	Z-A	Sort	Data Type: Any	Merge Queries	Text Analytics																																																																					
Close	New Query	Data Sources	Parameters	Query	Manage Columns	Reduce Rows	Sort	Transform	Split Column	Group By	1 2 Replace Values	Combine	AI Insights																																																																										
Queries [5]												Query Settings																																																																											
<table border="1"> <thead> <tr> <th colspan="6">= Table.ExpandTableColumn(source, "SalesTransactions", {"TransactionID", "ProductID", "Revenue"}, true)</th> </tr> <tr> <th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th> </tr> </thead> <tbody> <tr><td>1</td><td>customerA@email.com</td><td>(123) 456-7890</td><td></td><td>1</td><td>101</td></tr> <tr><td>2</td><td>customerB@email.com</td><td>(234) 567-8901</td><td></td><td>2</td><td>102</td></tr> <tr><td>3</td><td>customerC@email.com</td><td>(345) 678-9012</td><td></td><td>3</td><td>103</td></tr> <tr><td>4</td><td>customerD@email.com</td><td>(456) 789-0123</td><td></td><td>4</td><td>104</td></tr> <tr><td>5</td><td>customerA@email.com</td><td>(123) 456-7890</td><td></td><td>5</td><td>101</td></tr> <tr><td>6</td><td>customerB@email.com</td><td>(234) 567-8901</td><td></td><td>6</td><td>102</td></tr> <tr><td>7</td><td>customerE@email.com</td><td>(567) 890-1234</td><td></td><td>7</td><td>105</td></tr> <tr><td>8</td><td>customerF@email.com</td><td>(678) 901-2345</td><td></td><td>8</td><td>106</td></tr> <tr><td>9</td><td>customerG@email.com</td><td>(789) 012-3456</td><td></td><td>9</td><td>107</td></tr> <tr><td>10</td><td>customerE@email.com</td><td>(567) 890-1234</td><td></td><td>10</td><td>105</td></tr> </tbody> </table>												= Table.ExpandTableColumn(source, "SalesTransactions", {"TransactionID", "ProductID", "Revenue"}, true)						1	2	3	4	5	6	1	customerA@email.com	(123) 456-7890		1	101	2	customerB@email.com	(234) 567-8901		2	102	3	customerC@email.com	(345) 678-9012		3	103	4	customerD@email.com	(456) 789-0123		4	104	5	customerA@email.com	(123) 456-7890		5	101	6	customerB@email.com	(234) 567-8901		6	102	7	customerE@email.com	(567) 890-1234		7	105	8	customerF@email.com	(678) 901-2345		8	106	9	customerG@email.com	(789) 012-3456		9	107	10	customerE@email.com	(567) 890-1234		10	105	<b>PROPERTIES</b> Name: Customer Sales Merge <a href="#">All Properties</a>			
= Table.ExpandTableColumn(source, "SalesTransactions", {"TransactionID", "ProductID", "Revenue"}, true)																																																																																							
1	2	3	4	5	6																																																																																		
1	customerA@email.com	(123) 456-7890		1	101																																																																																		
2	customerB@email.com	(234) 567-8901		2	102																																																																																		
3	customerC@email.com	(345) 678-9012		3	103																																																																																		
4	customerD@email.com	(456) 789-0123		4	104																																																																																		
5	customerA@email.com	(123) 456-7890		5	101																																																																																		
6	customerB@email.com	(234) 567-8901		6	102																																																																																		
7	customerE@email.com	(567) 890-1234		7	105																																																																																		
8	customerF@email.com	(678) 901-2345		8	106																																																																																		
9	customerG@email.com	(789) 012-3456		9	107																																																																																		
10	customerE@email.com	(567) 890-1234		10	105																																																																																		
<b>APPLIED STEPS</b> Source: Expanded SalesTransactions																																																																																							

Power BI Data Editor																																																																											
File		Home		Transform		Add Column		View		Tools																																																																	
	New Source		Enter Data		Manage Parameters			Properties	Choose Columns		Keep Rows		A-Z	Z-A	Sort	Data Type: Whole Number	Merge Queries	Text Analytics																																																									
Close	New Query	Data Sources	Parameters	Query	Manage Columns	Reduce Rows	Sort	Transform	Split Column	Group By	1 2 Replace Values	Combine	AI Insights																																																														
Queries [4]												Query Settings																																																															
<table border="1"> <thead> <tr> <th colspan="6">= sales{[Schema="dbo",Item="Products"]}[Data]</th> </tr> <tr> <th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th> </tr> </thead> <tbody> <tr><td>1</td><td>0</td><td>xxx</td><td>xxx</td><td>0</td><td></td></tr> <tr><td>2</td><td>101</td><td>Laptop XYZ</td><td>Electronics</td><td>800</td><td></td></tr> <tr><td>3</td><td>102</td><td>T-Shirt Blue</td><td>Clothing</td><td>15</td><td></td></tr> <tr><td>4</td><td>103</td><td>Smartphone ABC</td><td>Electronics</td><td>400</td><td></td></tr> <tr><td>5</td><td>104</td><td>Sofa Set</td><td>Furniture</td><td>1200</td><td></td></tr> <tr><td>6</td><td>105</td><td>Tablet PQR</td><td>Electronics</td><td>300</td><td></td></tr> <tr><td>7</td><td>106</td><td>Jeans Black</td><td>Clothing</td><td>40</td><td></td></tr> <tr><td>8</td><td>107</td><td>Chair</td><td>Furniture</td><td>80</td><td></td></tr> </tbody> </table>												= sales{[Schema="dbo",Item="Products"]}[Data]						1	2	3	4	5	6	1	0	xxx	xxx	0		2	101	Laptop XYZ	Electronics	800		3	102	T-Shirt Blue	Clothing	15		4	103	Smartphone ABC	Electronics	400		5	104	Sofa Set	Furniture	1200		6	105	Tablet PQR	Electronics	300		7	106	Jeans Black	Clothing	40		8	107	Chair	Furniture	80		<b>PROPERTIES</b> Name: Products <a href="#">All Properties</a>			
= sales{[Schema="dbo",Item="Products"]}[Data]																																																																											
1	2	3	4	5	6																																																																						
1	0	xxx	xxx	0																																																																							
2	101	Laptop XYZ	Electronics	800																																																																							
3	102	T-Shirt Blue	Clothing	15																																																																							
4	103	Smartphone ABC	Electronics	400																																																																							
5	104	Sofa Set	Furniture	1200																																																																							
6	105	Tablet PQR	Electronics	300																																																																							
7	106	Jeans Black	Clothing	40																																																																							
8	107	Chair	Furniture	80																																																																							
<b>APPLIED STEPS</b> Source: Navigation																																																																											

The screenshot displays the Microsoft Power BI desktop interface with several visualizations and the Visualizations pane.

**Home Tab:**

- Data:** Get data, Excel, Data, SQL Server, Enter data, Dataverse, Recent sources.
- Queries:** Transform data, Refresh data, New visual, Text box, More visuals, Insert, New measure, Quick measure, Calculations, Sensitivity, Share.

**Visualizations:**

- Top Spending Customers:** Bar chart showing Sum of Revenue by CustomerName (Customer B, E, A, D, G, F, C).
- Average Price by Category:** Line chart showing Average Price by Category (Furniture, Electronics, Clothing).
- Sum of Revenue by StoreManager:** Pie chart showing revenue distribution by StoreManager (Mike Johnson, John Smith, Jane Doe).
- Sum of Revenue by Day:** Line chart showing Sum of Revenue by Day (Day 5, 10, 15, 20, 25, 30).

**Sales Trends:**

- Sum of QuantitySold by Month:** Line chart showing Sum of QuantitySold by Month (January, May, July, January, February, April, June, September, March, August).
- Sum of Revenue by ProductName:** Donut chart showing revenue distribution by ProductName (T-Shirt Blue, Tablet PQR, Laptop XYZ, Sofa Set, Chair, Jeans Black, Smartphone ABC).

**Count of CustomerID by Price:**

- Sum of QuantitySold by Location and Category:** Stacked bar chart showing Sum of QuantitySold by Location (New York, Los Angeles, Chicago) and Category (Clothing, Electronics, Furniture).

**Visualizations Pane:**

- Format visual:** Tools for styling and layout.
- Filters:** Tools for filtering data.
- Search:** Search bar.
- General:** General properties for the selected visual.
- Title:** Title settings (On/Off).
- Text:** Text input for title.
- Heading:** Heading input for title.
- Font:** Font selection.
- Text color:** Text color selection.
- Background color:** Background color selection.

The screenshot shows the Microsoft Power BI desktop application. The ribbon at the top has tabs: File, Home (selected), Insert, Modeling, View, Help, Format, and Data / Drill. The Home tab contains icons for Cut, Copy, Paste, Format painter, Get data (with sub-options for Excel, Data, SQL Server, Enter data, Dataverse, and Recent sources), Transform Refresh data (with sub-options for New visual, Text box, More visuals, New measure, Quick measure, Sensitivity, and Publish), and Insert (with sub-options for Calculations, Sensitivity, and Share). On the left, there's a vertical ribbon bar with icons for List, Grid, and Matrix. The main area displays a funnel chart titled "Sum of QuantitySold by Category". The Y-axis is labeled "Sum of QuantitySold" and ranges from 100 to 250. The X-axis is labeled "Category" and shows three categories: Electronics, Clothing, and Furniture. The funnel is blue, starting at approximately 260 for Electronics and narrowing down to about 100 for Furniture. To the right of the chart is the "Visualizations" pane, which lists various chart types like Bar, Line, and Map. Below it is the "Fields" pane, which shows a search bar and a hierarchical list of fields from different tables. The "Products" table is expanded, showing fields for Category, Price, ProductID, and ProductName. The "SalesTransactions" table is also expanded, showing fields for CustomerID, ProductID, Revenue, StoreID, and TransactionID. The "QuantitySold" field under SalesTransactions is selected, indicated by a yellow checkmark.

File Home Insert Modeling View Help Format Data / Drill

Cut Copy Paste Format painter Get data (Excel, Data, SQL Server, Enter data, Dataverse, Recent sources) Transform Refresh data (New visual, Text box, More visuals, New measure, Quick measure, Sensitivity, Publish) Insert (Calculations, Sensitivity, Share)

Clipboard Sum of QuantitySold by Category

Sum of QuantitySold

Electronics

Clothing

Furniture

Category

X-axis

Y-axis

Visualizations

Build visual

Filters

Fields

Search

> Customer Sales Merge

> Customers

Products

- Category
- $\sum$  Price
- ProductID
- ProductName

SalesTransactions

- CustomerID
- ProductID
- $\sum$  QuantitySold
- $\sum$  Revenue
- StoreID

TransactionDate

$\sum$  TransactionID



File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel) |



```
In [1]: a=1  
       type(a)
```

```
Out[1]: int
```

```
In [2]: b="Hello"  
       type(b)
```

```
Out[2]: str
```

```
In [6]: f=[1,2,3,4,5]  
       print(f[:-1])
```

```
[1, 2, 3, 4]
```

```
In [7]: type(f)
```

```
Out[7]: list
```

```
In [8]: e=(1,2,3,4)  
       type(e)
```

```
Out[8]: tuple
```

```
In [16]: f[4]=6  
        print(f)
```

```
[1, 2, 3, 4, 6]
```

```
In [20]: price = 19.99  
       type(price)
```

```
In [23]: is_python_fun = True  
       is_learning = False  
       type(is_learning)
```

```
Out[23]: bool
```

```
In [26]: # Accessing elements in lists and dictionaries  
first_number = numbers[0]  
print(first_number)  
last_fruit = fruits[-1]  
print(last_fruit)  
name = person["name"]  
print(name)  
math_grade = grades["math"]  
print(math_grade)
```

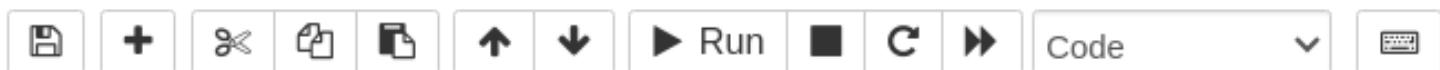
# Day-21

- 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

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel) O



```
In [8]: gc = 8.314
def volume(p, temp):
    """
    09/10/22 - Tanya
    19/09/23 - Singh - 21334 based on this I updated
    Calculate the volume of gas using the ideal gas law.
    Args:
        pressure(float): Pressure in pascals (Pa).
        temperature (float): Temperature in Kelvin (K).
        gas_constant (float): Gas constant for the specific gas.
    Returns:
        float: Volume in cubic meters (m^3).
    """
    return (p*1.0)/(gc * temp)
result = volume(4,5)
print (f"Gas Volume : {result} m^3")
```

```
Gas Volume : 0.09622323791195574 m^3
```

```
In [11]: def mass (p,temp,m):
    """
    Calculate the mass of gas using gas law
    Args:
        pressure(float): Pressure in pascals (Pa).
        temperature (float): Temperature in Kelvin (K).
        gas_constant (float): Gas constant for the specific gas.
        mass (float): mass of gas (kg).
    Returns:
        float: Mass in kilograms (kg).
    """
    return volume(p,temp)* m
result1 = mass (8,6,9)
print(f"Mass of gas : {result1} kg")
```

```
Mass of gas : 1.443348568679336 kg
```

```
In [14]: def factorial(n):
    if n==0: #base_condition
        return 1
    else:
        return n*factorial(n-1) #recursive_call
print(factorial(5))
```

120

```
In [17]: def calculate_total_depth(segments):
    if not segments:
        return 0
    else:
        curr_seg_depth = segments[0]
        remaining_seg = segments[1:]
        return curr_seg_depth + calculate_total_depth(remaining_seg)
calculate_total_depth([1,2,3])
```

Out[17]: 6

```
In [19]: def calculate_total_depth(segments):
    if not segments:
        return 0
    else:
        curr_seg_depth = segments[0]
        remaining_seg = segments[1:]
        return curr_seg_depth + calculate_total_depth(remaining_seg)
calculate_total_depth([1,2,3])
```

Out[19]: 6

```
In [26]: def generate_squares(n):
    for i in range(1,n+1):
        yield i ** 2
```

```
In [28]: generate_squares(10)
```

Out[28]: <generator object generate\_squares at 0x7f97b4306ce0>

```
In [27]: for i in generate_squares(5):
    print(i)
```

1  
4  
9  
16  
25

```
In [35]: def oil_production_m(early_value):
    months = ["Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
    monthly_oil_prod = yearly_value/12

    for month in months:
        yield month, monthly_oil_prod
```

```
In [36]: for month, production in oil_production_m(120000):
    print(f'{month}: {production}')
```

```
Jan: 10000.0
Feb: 10000.0
Mar: 10000.0
Apr: 10000.0
May: 10000.0
Jun: 10000.0
Jul: 10000.0
Aug: 10000.0
Sep: 10000.0
Oct: 10000.0
Nov: 10000.0
Dec: 10000.0
```

```
In [18]: class Person:
    def __init__(self, name, age):
        self.name = name
        self.age = age

    def greet(self):
        return f"Hello, my name is {self.name} and I am {self.age} years old."

    def bdy(self):
        self.age += 1
        return f"I am now {self.age} years old."
```

```
In [19]: abc = Person("john", 20)
```

```
In [20]: abc.name
```

```
Out[20]: 'john'
```

```
In [21]: abc.age
```

```
Out[21]: 20
```

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

```
<__main__.Person object at 0x7f09dd6dd290>
```

```
In [42]: def my_dec(fun):
    def wrapper():
        print("I am starting")
        fun()
        print("I am completed")
    return wrapper
```

```
@my_dec
def ida_hello():
    print("Hello")
```

```
In [43]: ida_hello()
```

```
I am starting
Hello
I am completed
```

```
In [46]: import logging
def my_dec01(fun):
    def wrapper(*args, **kwargs):
        logging.info(f"Calling the function:{fun.__name__}")
        result = fun(*args, **kwargs)
        logging.info(f"{fun.__name__} Completed")
        return result
    return wrapper
```

```
@my_dec01
def calculate_total_depth(segments):
    if not segments:
        return 0
    else:
        curr_seg_depth = segments[0]
        remaining_seg = segments[1:]
        return curr_seg_depth + calculate_total_depth(remaining_seg)
```

```
In [47]: calculate_total_depth([100, 200, 200])
```

```
Out[47]: 500
```

```
In [23]: from datetime import datetime
curr_datetime = datetime.now()
print(curr_datetime)

2023-09-19 08:47:12.206799
```

```
In [26]: a=10
b=0
try:
    result=a/b
    print(result)
except:
    print("Error: Divided by zero")
```

```
Error: Divided by zero
```

# Day-22

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

```
In [1]: import pandas as pd
```

```
In [2]: data={'orange':["kashmir","ooty","blr"], 'apples':["chennai","delhi","mumbai"]}  
print(data)
```

```
{'orange': ['kashmir', 'ooty', 'blr'], 'apples': ['chennai', 'delhi', 'mumbai']}
```

```
In [4]: fruit = pd.DataFrame(data)  
fruit.head()
```

```
Out[4]:
```

	orange	apples
0	kashmir	chennai
1	ooty	delhi
2	blr	mumbai

```
In [7]: fruit = pd.DataFrame(data, index = ['jan','feb','mar'])  
fruit
```

```
Out[7]:
```

	orange	apples
jan	kashmir	chennai
feb	ooty	delhi
mar	blr	mumbai

```
In [8]: test = fruit.loc['feb']  
test
```

```
Out[8]: orange    ooty  
apples    delhi  
Name: feb, dtype: object
```

```
In [10]: movie = pd.read_csv ("~/home/labuser/Downloads/IMDB-Movie-Data.csv", index_col=0)  
movie.head()
```

```
Out[10]:
```

Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	333.13	76.0
2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael F...	2012	124	7.0	485820	126.46	65.0
3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	7.3	157606	138.12	62.0
4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32	59.0
5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016	123	6.2	393727	325.02	40.0

```
In [11]: movie.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 1000 entries, 1 to 1000  
Data columns (total 11 columns):  
 #   Column           Non-Null Count  Dtype     
---  --  
 0   Title            1000 non-null    object    
 1   Genre             1000 non-null    object    
 2   Description       1000 non-null    object    
 3   Director          1000 non-null    object    
 4   Actors            1000 non-null    object    
 5   Year              1000 non-null    int64     
 6   Runtime (Minutes) 1000 non-null    int64     
 7   Rating            1000 non-null    float64  
 8   Votes              1000 non-null    int64     
 9   Revenue (Millions) 972 non-null    float64
```

```
In [47]: final = temp.drop_duplicates(keep="first")
final.shape
```

```
Out[47]: (1000, 11)
```

```
In [48]: final.columns
```

```
Out[48]: Index(['Title', 'Genre', 'Description', 'Director', 'Actors', 'Year',
       'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
       'Metascore'],
      dtype='object')
```

```
In [49]: final.rename(columns={'Runtime (Minutes)': "Runtime", 'Revenue (Millions)': 'Revenue'},inplace = True)
''
```

```
/tmp/ipykernel_1751/3244259771.py:1: 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#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

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

```
Out[49]: ''
```

```
In [50]: final.columns = [i.upper() for i in final.columns]
final.columns
```

```
Out[50]: Index(['TITLE', 'GENRE', 'DESCRIPTION', 'DIRECTOR', 'ACTORS', 'YEAR',
       'RUNTIME', 'RATING', 'VOTES', 'REVENUE', 'METASCORE'],
      dtype='object')
```

```
In [51]: final.head()
```

```
Out[51]:    TITLE      GENRE DESCRIPTION DIRECTOR      ACTORS YEAR  RUNTIME RATING VOTES REVENUE METASCORE
Rank
1   Guardians of the Galaxy  Action,Adventure,Sci-Fi  A group of intergalactic criminals are forced ...
     Action,Adventure,Sci-Fi  James Gunn  Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...
          2014        121       8.1  757074    333.13      76.0
Following clues  Nomi Rapace, Logan
```

```
In [57]: revenue.head()
```

```
Out[57]: Rank
1    333.13
2    126.46
3    138.12
4    270.32
5    325.02
Name: REVENUE, dtype: float64
```

```
In [58]: rev_mean = revenue.mean()
rev_mean
```

```
Out[58]: 82.95637614678898
```

```
In [59]: revenue.fillna(rev_mean,inplace=True)
```

```
/tmp/ipykernel_1751/3062699824.py:1: 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#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
revenue.fillna(rev_mean,inplace=True)
```

```
In [60]: final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 1 to 1000
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   TITLE       1000 non-null   object 
 1   GENRE      1000 non-null   object 
 2   DESCRIPTION 1000 non-null   object 
 3   DIRECTOR    1000 non-null   object 
 4   ACTORS      1000 non-null   object 
 5   YEAR        1000 non-null   int64  
 6   RUNTIME     1000 non-null   int64  
 7   RATING      1000 non-null   float64
 8   VOTES       1000 non-null   int64  
 9   REVENUE     1000 non-null   float64
 10  METASCORE   1000 non-null   float64
```

# Day-23

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

```
In [5]: import findspark
```

```
In [6]: findspark.init()
```

```
In [7]: from pyspark.sql import SparkSession  
spark=SparkSession.builder.appName("WordCount").getOrCreate()
```

```
In [15]: sc = spark.sparkContext  
rdd = sc.parallelize([1,23,4,5])  
rdd.collect()
```

```
Out[15]: [1, 23, 4, 5]
```

```
In [26]: rdd1 = rdd.map(lambda x: x*2)  
rdd1.collect()
```

```
Out[26]: [2, 46, 8, 10]
```

```
In [21]: resultrdd = rdd.flatMap(lambda x:(x,x*2))  
resultrdd.collect()
```

```
Out[21]: [1, 2, 23, 46, 4, 8, 5, 10]
```

```
In [24]: rdd3 = rdd.filter(lambda x: x%2==0)  
rdd3.collect()
```

```
Out[24]: [4]
```

```
In [32]: rdd4 = sc.parallelize([(1,2),(3,4),(1,7),(7,8)])  
res4 = rdd4.reduceByKey(lambda x,y: x+y)  
res4.collect()
```

```
Out[32]: [(1, 9), (3, 4), (7, 8)]
```

```
In [70]: rdd5=sc.parallelize([(1,2),(3,4),(1,6),(2,3)])  
rdd4=rdd5.groupByKey()  
rdd4.collect()
```

```
Out[70]: [(2, <pyspark.resultiterable.ResultIterable at 0x7f36642f73d0>),  
(1, <pyspark.resultiterable.ResultIterable at 0x7f36644c6a90>),  
(3, <pyspark.resultiterable.ResultIterable at 0x7f36644c7490>)]
```

```
In [71]: for key, values in rdd4.collect():  
    print(f"Key : {key} , values : {list(values)}")
```

```
Key : 2 , values : [3]  
Key : 1 , values : [2, 6]  
Key : 3 , values : [4]
```

```
In [74]: wordlist=['this','is','a','sample','text','document','for','word','count','example']  
rdd=sc.parallelize(wordlist)  
wordcount=rdd.flatMap(lambda x:x.split(','))  
wordcount.collect()
```

```
Out[74]: ['this',  
         'is',  
         'a',  
         'sample',  
         'text',  
         'document',  
         'for',  
         'word',  
         'count',  
         'example']
```

```
In [78]: wordlist=['this','is','a','sample','text','document','for','word','count','example']
rdd=sc.parallelize(wordlist)
count = rdd.map(lambda word:(word,1)).groupByKey().mapValues(sum)
count.collect()
```

```
Out[78]: [('this', 1),
('sample', 1),
('text', 1),
('for', 1),
('word', 1),
('is', 1),
('a', 1),
('document', 1),
('count', 1),
('example', 1)]
```

```
In [82]: purchase =sc.textFile('/home/labuser/Downloads/purchases.csv')
purchase.collect()
```

```
Out[82]: [,apples,oranges', 'June,3,0', 'Robert,2,3', 'Lily,0,7', 'David,1,2']
```

```
In [83]: purchasedf=spark.read.option("inferSchema",True).option("header",True).csv('/home/labuser/Downloads/purchases.csv')
```

```
In [84]: purchasedf.show()
```

```
+-----+-----+-----+
| _c0|apples|oranges|
+-----+-----+-----+
| June|     3|      0|
| Robert|    2|      3|
| Lily|     0|      7|
| David|     1|      2|
+-----+-----+-----+
```

```
23/09/21 09:13:47 WARN CSVHeaderChecker: CSV header does not conform to the schema.
Header: , apples, oranges
Schema: _c0, apples, oranges
Expected: _c0 but found:
CSV file: file:///home/labuser/Downloads/purchases.csv
```

```
In [85]: purchasedf.printSchema()
```

```
root
 |-- _c0: string (nullable = true)
 |-- apples: integer (nullable = true)
 |-- oranges: integer (nullable = true)
```

```
In [110]: purchasedf=spark.read.schema(udfschema).option("header",True).csv('~/home/labuser/Downloads/IMDB-Movie-Data.csv')
```

```
In [111]: purchasedf.show()
```

Rank	Title	Genre
1	Guardians of the ...	Action, Adventure, ...
2	Prometheus	Adventure, Mystery...
3	Split	Horror, Thriller
4	Sing	Animation, Comedy, ...
5	Suicide Squad	Action, Adventure, ...
6	The Great Wall	Action, Adventure, ...
7	La La Land	Comedy, Drama, Music
8	Mindhorn	Comedy
9	The Lost City of Z	Action, Adventure, ...
10	Passengers	Adventure, Drama, R...
11	Fantastic Beasts ...	Adventure, Family, ...
12	Hidden Figures	Biography, Drama, H...
13	Rogue One	Action, Adventure, ...
14	Moana	Animation, Adventure
15	Colossal	Action, Comedy, Drama
16	The Secret Life o...	Animation, Adventure
17	Hacksaw Ridge	Biography, Drama, H...
18	Jason Bourne	Action, Thriller
19	Lion	Biography, Drama

```
In [112]: purchasedf.printSchema()
```

```
root
|-- Rank: integer (nullable = true)
|-- Title: string (nullable = true)
|-- Genre: string (nullable = true)
```

```
In [113]: from pyspark.sql.types import *
udfschema = StructType([StructField("Rank", IntegerType(), True),
                       StructField("Title", StringType(), True),
                       StructField("Genre", StringType(), True)])
```

# Day-24

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

```
In [22]: purchasedf.printSchema()
```

```
root
|-- Rank: integer (nullable = true)
|-- Title: string (nullable = true)
|-- Genre: string (nullable = true)
```

```
In [19]: from pyspark.sql.types import *
udfschema = StructType([StructField("Rank",IntegerType(),True),
                       StructField("Title",StringType(),True),
                       StructField("Genre",StringType(),True)])
```

```
In [23]: purchasedf.rdd.getNumPartitions()
```

```
Out[23]: 1
```

```
In [29]: newdf = purchasedf.repartition(10)
```

```
In [30]: newdf.rdd.getNumPartitions()
```

```
23/09/22 04:17:52 WARN CSVHeaderChecker: Number of column in CSV header is not equal to number of fields in the schema:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Downloads/IMDB-Movie-Data.csv
```

```
Out[30]: 10
```

```
In [33]: newdf.write.csv("/home/labuser/Desktop/output/test")
```

```
23/09/22 04:18:42 WARN CSVHeaderChecker: Number of column in CSV header is not equal to number of fields in the schema:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Downloads/IMDB-Movie-Data.csv
```

```
In [40]: spark_ui_url = f'{spark._jsc.sc().uiWebUrl().get()}/'
print("Spark UI URL:",spark_ui_url)
```

```
Spark UI URL: http://ip-172-31-9-31.ap-south-1.compute.internal:4040/
```

```
In [41]: sc.uiWebUrl
```

```
Out[41]: 'http://ip-172-31-9-31.ap-south-1.compute.internal:4040'
```

```
In [43]: purchasedf.createOrReplaceTempView("purchase")
```

```
In [45]: result = spark.sql ("select * from purchase")
type(result)
```

```
Out[45]: pyspark.sql.dataframe.DataFrame
```

In [48]: result.show()

Rank	Title	Genre
1	Guardians of the Galaxy	Action, Adventure, Sci-Fi
2	Prometheus	Adventure, Mystery, Sci-Fi
3	Split	Horror, Thriller
4	Sing	Animation, Comedy, Family
5	Suicide Squad	Action, Adventure, Crime
6	The Great Wall	Action, Adventure, History
7	La La Land	Comedy, Drama, Music
8	Mindhorn	Comedy
9	The Lost City of Z	Action, Adventure, History
10	Passengers	Adventure, Drama, Romance
11	Fantastic Beasts and Where to Find Them	Adventure, Family, Fantasy
12	Hidden Figures	Biography, Drama, History
13	Rogue One: A Star Wars Story	Action, Adventure, Sci-Fi
14	Moana	Animation, Adventure, Family
15	Colossal	Action, Comedy, Drama
16	The Secret Life of Walter Mitty	Animation, Adventure, Family
17	Hacksaw Ridge	Biography, Drama, History
18	Jason Bourne	Action, Thriller
19	Lion	Biography, Drama
20	Arrival	Drama, Mystery, Sci-Fi

3.7.70.65 - Remote Desktop Connection

Applications : [out - File Manager] WordCount - Spark Jobs ...

Desktop/ PySpark - Jupyter Notebook WordCount - Spark Jobs +

Not secure | ip-172-31-10-64.ap-south-1.compute.internal:4040/jobs/

Spark 3.4.1 Jobs Stages Storage Environment Executors SQL / DataFrame WordCount application

## Spark Jobs (?)

User: labuser  
Total Uptime: 55 min  
Scheduling Mode: FIFO  
Completed Jobs: 19

Event Timeline

### Completed Jobs (19)

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page.

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
18	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/09/22 04:21:32	1 s	1/1 (1 skipped)	10/10 (1 skipped)
17	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/09/22 04:21:32	0.2 s	1/1	1/1
16	javaToPython at NativeMethodAccessorImpl.java:0 javaToPython at NativeMethodAccessorImpl.java:0	2023/09/22 04:21:19	0.8 s	1/1	1/1
15	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/09/22 03:44:38	0.5 s	1/1	1/1
14	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/09/22 03:44:38	0.1 s	1/1	1/1
13	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/09/22 03:44:37	0.2 s	1/1	1/1
12	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/09/22 03:44:37	37 ms	1/1	1/1
11	showString at NativeMethodAccessorImpl.java:0 showString at NativeMethodAccessorImpl.java:0	2023/09/22 03:44:37	0.1 s	1/1	1/1

13.235.48.253 - Remote Desktop Connection

```
+-----+-----+-----+-----+
|-----+-----+-----+-----+
only showing top 20 rows
```

In [52]:

```
1 customerDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/
2 lineitemDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/
3 nationDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/Do
4 ordersDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/Do
5 partDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/Down
6 partsuppDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/
7 regionDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/Do
8 supplierDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/
```

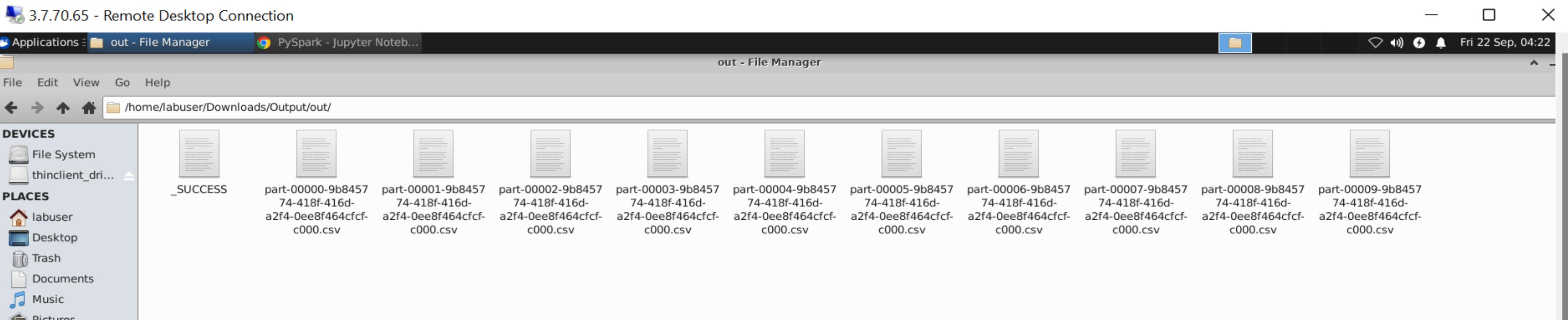
In [55]:

```
1 joinDf=customerDf.join(ordersDf, customerDf.C_CUSTKEY==ordersDf.O_CUSTKEY,how='inner')
2 joinDf.show()
```

CUSTOMER#	NAME#	ADDRESS#	TERRITORY#	PHONE#	FAX#	QUANTITY#	ITEM#	SHIPMODE#	SHIPINSTRUCT#	SHIPVIA#	SHIPDATE#	ACCTBAL#	CREDITLIM#	DISCOUNT#	TERMS#	CONTACT#	COMMENTS#
101 Customer#000000781 PARKAY TWO RESTAURANT 102-470-388-5801 0403.02 1-URGENT Clerk#000000880	ly blit...  2  781  0  38426.09  1996-12-01  1-URGENT Clerk#000000880	0  foxes. pending a...    1234 Customer#000001234 B30hbH0MRJE,F0Lc7...  1 11-742-434-6436  -982.32  FURNITURE y ironic i	nstruct...  3  1234  F  205654.3  1993-10-14  5-LOW Clerk#000000955	0 sly final account...    1369 Customer#000001369 rXTw0zU0a2ak4Nj5l...  10 20-232-617-7418  498.77  AUTOMOBILE ong the ir	onic id...  4  1369  0  56000.91  1995-10-11  5-LOW Clerk#000000124	0 sits. slyly regul...    445 Customer#000000445 MX1UA0KUJzIGyWM p...  20 30-849-846-6070  8018.81  FURNITURE e ironic,	special...  5  445  F  105367.67  1994-07-30  5-LOW Clerk#000000925	0 quickly. bold dep...    557 Customer#000000557  Nt6FUuDR7v  15 25-390-153-6699  9559.04  BUILDING furiously	pending...  6  557  F  45523.1  1992-02-21 4-NOT SPECIFIED Clerk#000000058	0 ggle. special, fi...    392 Customer#000000392  H7M6J0bnd0  17 27-601-793-2507  8492.33  BUILDING efully bol	d ideas...  7  392  0  271885.66  1996-01-10  2-HIGH Clerk#000000470	0 ly special requests     1301 Customer#000001301 oR0kHfL6GWhF VPD,...  10 20-339-347-9046  8966.63  MACHINERY egular asy	mptotes...  32  1301  0  198665.57  1995-07-16  2-HIGH Clerk#000000616				

In [ ]:

```
1
```



13.235.48.253 - Remote Desktop Connection

```
1 | 2016| 123| 6.7|150823| 162.16| 58.0| Good| Garth Davis|Dev Patel, Nicole...
| 19| Lion| Biography,Drama|A five-year-old I...| ...
2016| 118| 8.1|102061| 51.69| 69.0| Good| ...
| 20| Arrival|Drama,Mystery,Sci-Fi|When twelve myste...| Denis Villeneuve|Amy Adams, Jeremy...
2016| 116| 8|340798| 100.5| 81.0| Good| ...
+---+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+
only showing top 20 rows
```

```
In [42]: 1 def concat_shell(column):
2     return column+_shell
```

```
In [43]: 1 from pyspark.sql.functions import *
2 my_uds=udf(concat_shell,StringType())
```

```
In [44]: 1 test=movieDF.withColumn("new_col",my_uds(col("Title")))
```

```
In [45]: 1 test.show()
```

[Stage 38:> (0 + 1) / 1]

Rank	Title	Genre	Description	Director	Actors
Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the ...	Action,Adventure,...	A group of interg...	James Gunn	Chris Pratt, Vin ...
2014	121	8.1 757074	333.13	76.0 Guardians of the ...	
2	Prometheus	Adventure,Mystery...	Following clues t...	Ridley Scott	Noomi Rapace, Log...
2012	124	7 485820	126.46	65.0  Prometheus shell	
3	Split	Horror,Thriller	Three girls are k...	M. Night Shyamalan	James McAvoy, Any...
2016	117	7.3 157606	138.12	62.0  Split shell	
4	Sing	Animation,Comedy,...	In a city of huma...	Christophe Lourdelet	Matthew McConaugh...
2016	108	7.2  60545	270.32	59.0  Sing shell	
5	Suicide Squad	Action,Adventure,...	A secret governme...	David Ayer	Will Smith, Jared...
2016	123	6.2 393727	325.02	40.0  Suicide Squad shell	
6	The Great Wall	Action,Adventure,...	European mercenar...	Yimou Zhang	Matt Damon, Tian ...
2016	103	6.1  560361	45.131	42.0 The Great Wall shell	

3.7.70.65 - Remote Desktop Connection

```
23/09/22 04:21:52 WARN CSVHeaderChecker: NUMBER OF COLUMN IN CSV HEADER IS NOT EQUAL TO NUMBER OF FIELDS IN THE SCHEMA:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Desktop/Pandas_datasets/IMDB-Movie-Data.csv
```

```
In [29]: spark.ui_url = f'{spark.jsc.sc().uiWebUrl().get()}/'
print("Spark UI URL:",spark.ui_url)
```

Spark UI URL: http://ip-172-31-10-64.ap-south-1.compute.internal:4040/

```
In [31]: dsc=spark.sparkContext
```

```
In [32]: dsc.uiWebUrl
```

```
Out[32]: 'http://ip-172-31-10-64.ap-south-1.compute.internal:4040'
```

```
In [37]: purchasedf_01.createOrReplaceTempView("purchase")
```

```
In [38]: result=spark.sql("select * from purchase")
```

```
In [39]: type(result)
Out[39]: pyspark.sql.DataFrame
```

```
In [40]: result.show()
```

_c0	apples	oranges
June	3	0
Robert	2	3
Lily	0	7
David	1	2

```
23/09/22 04:59:33 WARN CSVHeaderChecker: CSV header does not conform to the schema.
Header: , apples, oranges
Schema: _c0, apples, oranges
Expected: _c0 but found:
CSV file: file:///home/labuser/Desktop/Pandas_datasets/purchases.csv
```

```
In [ ]:
```

# Day-25

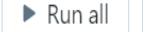
- 1.Caching
- 2.Persist and different storage levels
- 3.Aggregation
- 4.Managed Tables
- 5.External Tables
- 6.Partitioning
- 7.For-Each

# Day-26

1. 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

Microsoft Azure |  databricks | Search data, notebooks, recents, and more... | CTRL + P | tanya15 | shellunext\_1693422225890@npunext.onmicrosoft.com

Untitled Notebook 2023-09-26 09:45:02 Python 

File Edit View Run Help Last edit was 1 minute ago Provide feedback    Share

Cmd 2

```
1 dbutils.fs.mount(source = "wasbs://input@tanyagen2.blob.core.windows.net",mount_point = "/mnt/input/", extra_configs = {"fs.azure.account.key.tanyagen2.blob.core.windows.net":dbutils.secrets.get(scope = "tanyascope1", key = "keygen3")})
```

True

Command took 11.69 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 12:19:01 PM on Shellunext's Cluster

Cmd 3

```
1 dbutils.fs.mounts()
```

```
[MountInfo(mountPoint='/databricks-datasets', source='databricks-datasets', encryptionType=''), MountInfo(mountPoint='/Volumes', source='UnityCatalogVolumes', encryptionType=''), MountInfo(mountPoint='/databricks/mlflow-tracking', source='databricks/mlflow-tracking', encryptionType=''), MountInfo(mountPoint='/databricks-results', source='databricks-results', encryptionType=''), MountInfo(mountPoint='/databricks/mlflow-registry', source='databricks/mlflow-registry', encryptionType=''), MountInfo(mountPoint='/mnt/input/', source='wasbs://input@tanyagen2.blob.core.windows.net', encryptionType=''), MountInfo(mountPoint='/Volume', source='DbfsReserved', encryptionType=''), MountInfo(mountPoint='/volumes', source='DbfsReserved', encryptionType=''), MountInfo(mountPoint='/', source='DatabricksRoot', encryptionType=''), MountInfo(mountPoint='/volume', source='DbfsReserved', encryptionType='')]
```

Command took 0.27 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 12:19:56 PM on Shellunext's Cluster

Cmd 4

```
1
```

Python    

+ New  Workspace  Recents  Catalog  Workflows  Compute SQL  SQL Editor  Queries  Dashboards  Alerts  Query History  SQL Warehouses Data Engineering  Job Runs  Data Ingestion  Delta Live Tables Machine Learning  Experiments

+ New  Workspace  Recents  Catalog  Workflows  Compute SQL  SQL Editor  Queries  Dashboards  Alerts  Query History  SQL Warehouses Data Engineering  Job Runs  Data Ingestion  Delta Live Tables Machine Learning  Experiments

Untitled Notebook 2023-09-26 09:45:02 Python 

File Edit View Run Help Last edit was now Provide feedback   

choose\_colors combo\_ida test

white test ida

Cmd 2

```
1 dbutils.widgets.text("test","ida")
```

Command took 0.07 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 12:31:18 PM on Shellunext's Cluster

Cmd 5

```
1 a=dbutils.widgets.get("test")  
2 print(a)
```

ida

Command took 0.16 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 2:09:57 PM on Shellunext's Cluster

Cmd 6

```
1 dbutils.widgets.dropdown("choose_colors","white",["white","red","black"])
```

Command took 0.10 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 2:13:16 PM on Shellunext's Cluster

Cmd 7

```
1 a=dbutils.widgets.get("choose_colors")  
2 print(a)
```

white

Cmd 9

```
1 dbutils.widgets.combo("combo_ida","test",["1","2","3"])
```

Command took 0.09 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 2:13:23 PM on Shellunext's Cluster

```
1 a=dbutils.widgets.get("combo_ida")  
2 print(a)
```

test

Command took 0.12 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/26/2023, 2:13:27 PM on Shellunext's Cluster

# Day-27

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

```
1 dbutils.fs.ls("dbfs:/user/hive/warehouse/zipcodes_1_csv")
```

```
[FileInfo(path='dbfs:/user/hive/warehouse/zipcodes_1_csv/_delta_log/', name='_delta_log', size=0, modificationTime=1695795043000),  
 FileInfo(path='dbfs:/user/hive/warehouse/zipcodes_1_csv/part-00000-793ac015-2c9b-4e64-a781-2a2c04ead96a.c000.snappy.parquet', name='part-00000-793ac015-2c9b-4e64-a781-2a2c04ead96a.c000.snappy.parquet', size=6781, modificationTime=1695795054000)]
```

Command took 0.19 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 11:42:03 AM on Shellunext's Cluster

Cmd 5

```
1 %sql  
2 describe history zipcodes_1_csv
```

▶ (1) Spark Jobs

▶ `_sqldf: pyspark.sql.DataFrame = [version: long, timestamp: timestamp ... 13 more fields]`

Table +

	version	timestamp	userId	userName	op
1	1	2023-09-27T06:10:55.000+0000	2250536170854016	shellunext_1693422225890@npunext.onmicrosoft.com	COP
2	0	2023-09-27T06:10:43.000+0000	2250536170854016	shellunext_1693422225890@npunext.onmicrosoft.com	CRE

↓ 2 rows | 1.86 seconds runtime

Refreshed 5 minutes ago

```
1 df.write.parquet("/mnt/input/destination")
```

AnalysisException: [PATH\_ALREADY\_EXISTS] Path dbfs:/mnt/input/destination already exists. Set mode as "overwrite" to overwrite the existing path.

Command took 0.86 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 11:43:04 AM on Shellunext's Cluster

Cmd 12

```
1 df.write.saveAsTable("e_zip_tables")
```

▶ (4) Spark Jobs

Command took 3.90 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 11:43:14 AM on Shellunext's Cluster

Cmd 13

```
1 df.write.option("path","/mnt/input/tables").saveAsTable("e_zip_tables12")
```

▶ (4) Spark Jobs

```
1 %sql  
2 select * from zipcodes_1_csv;  
3 delete from zipcodes_1_csv where `_c0` = 61391
```

▶ (6) Spark Jobs

▶ `_sqldf: pyspark.sql.DataFrame = [num_affected_rows: long]`

Table +

	num_affected_rows
1	1

↓ 1 row | 6.00 seconds runtime

Refreshed 5 minutes ago

ⓘ This result is stored as PySpark data frame `_sqldf` and in the IPython output cache as `Out[15]`. [Learn more](#)

Command took 6.00 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 11:42:27 AM on Shellunext's Cluster

Cmd 7

```
1 %sql  
2 describe history zipcodes_1_csv
```

▶ (1) Spark Jobs

▶ `_sqldf: pyspark.sql.DataFrame = [version: long, timestamp: timestamp ... 13 more fields]`

Table +

	version	timestamp	userId	userName	op
1	2	2023-09-27T06:12:32.000+0000	2250536170854016	shellunext_1693422225890@npunext.onmicrosoft.com	DEL
2	1	2023-09-27T06:10:55.000+0000	2250536170854016	shellunext_1693422225890@npunext.onmicrosoft.com	CO
3	0	2023-09-27T06:10:43.000+0000	2250536170854016	shellunext_1693422225890@npunext.onmicrosoft.com	CRE

```
1 %sql  
2 VACUUM e_zip_tables12 RETAIN 169 HOURS
```

▶ (18) Spark Jobs

▶ `_sqldf: pyspark.sql.DataFrame = [path: string]`

Table +

	path
1	dbfs:/mnt/input/tables

↓ 1 row | 48.27 seconds runtime

Refreshed 5 minutes ago

```

1 %sql
2 OPTIMIZE e_zip_tables12 ZORDER BY (RecordNumber)

```

▶ (4) Spark Jobs

[\_sqldf: pyspark.sql.dataframe.DataFrame = [path: string, metrics: struct]

Table +

	path	metrics
1	dbfs:/mnt/input/tables	▶ {"numFilesAdded": 0, "numFilesRemoved": 0, "filesAdded": {"min": null, "max": null, "avg": 0, "totalFiles": 0, "totalSize": 0}, "filesRemoved": {"min": null, "max": null, "avg": 0, "totalFiles": 0, "totalSize": 0}, "partitionsOptimized": 0, "zOrderStats": {"strategyName": "minCubeSize(107374182400)", "inputCubeFiles": {"num": 0, "size": 0}, "inputOtherFiles": {"num": 1, "size": 0}, "inputNumCubes": 0, "mergedFiles": {"num": 0, "size": 0}, "numOutputCubes": 0, "mergedNumCubes": null}, "numBatches": 0, "totalConsideredFiles": 1, "totalFilesSkipped": 1, "preserveInsertionOrder": false, "numFilesSkippedToReduceWriteAmplification": 0, "numBytesSkippedToReduceWriteAmplification": 0, "startTimeMs": 1695795281884, "endTimeMs": 1695795282676, "totalClusterParallelism": 4, "totalScheduledTasks": 0, "autoCompactParallelismStats": null, "deletionVectorStats": {"numDeletionVectorsRemoved": 0, "numDeletionVectorRowsRemoved": 0}, "numTableColumns": 20, "numTableColumnsWithStats": 20, "totalTaskExecutionTimeMs": 0, "skippedArchivedFiles": 0, "clusteringMetrics": null}

↓ 1 row | 2.10 seconds runtime

Refreshed 5 minutes ago

i This result is stored as PySpark data frame \_sqldf and in the IPython output cache as Out[24]. Learn more

Microsoft Azure Search resources, services, and docs (G+) Shellunext\_1693422225... UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home > Storage accounts > tanyagen2 | Containers >

**input** Container

Search Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Azure AD User Account)  
Location: input / tablesnew

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
[..]					
_delta_log					
State=AL					
State=AZ					
State=FL					
State=NC					
State=PR					
State=TX					

Shared access tokens  
Manage ACL  
Access policy  
Properties  
Metadata

```

1 df.display()
2 df.write.partitionBy("State").option("path","/mnt/input/tablesnew").saveAsTable("e_zip_tablesnew80")

```

▶ (5) Spark Jobs

Table +

	RecordNumber	Zipcode	ZipCodeType	City	State	LocationType	Lat	Lon
1	1	704	STANDARD	PARC PARQUE	PR	NOT ACCEPTABLE	17.96	-66.54
2	2	704	STANDARD	PASEO COSTA DEL SUR	PR	NOT ACCEPTABLE	17.96	-66.54
3	10	709	STANDARD	BDA SAN LUIS	PR	NOT ACCEPTABLE	18.14	-66.54
4	61391	76166	UNIQUE	CINGULAR WIRELESS	TX	NOT ACCEPTABLE	32.72	-97.18
5	61392	76177	STANDARD	FORT WORTH	TX	PRIMARY	32.75	-97.18
6	61393	76177	STANDARD	FT WORTH	TX	ACCEPTABLE	32.75	-97.18
7	4	704	STANDARD	IIRR FLIGNFN RICF	PR	NOT ACCEPTABLE	17.96	-66.54

↓ 21 rows | 8.25 seconds runtime

Refreshed 1 minute ago

Command took 8.25 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 11:45:15 AM on Shellunext's Cluster

## Workflows

Jobs Job runs Delta Live Tables

Only my job runs Run status 2023-09-25 14:30:00 → 2023-09-27 14:30:00 Create job



Start time	Job	Run as	Launched	Duration	Status	Run parameters	⋮
Sep 27, 2023, 01:49 PM	idajob1	Shellunext	By scheduler	18s	Pending		⋮
Sep 27, 2023, 01:47 PM	idajob1	Shellunext	By scheduler	2s	Skipped		⋮
Sep 27, 2023, 01:45 PM	idajob1	Shellunext	By scheduler	3m 50s	Succeeded		⋮

Cmd 1

```
1 jdbc_url = "jdbc:sqlserver://idaservernew.database.windows.net:1433;databaseName=idadb02"
2 connection_properties = { "user":"sqladmin", "password":"IDAshell@123", "driver":"com.microsoft.sqlserver.
  jdbc.SQLServerDriver"}
```

Command took 0.93 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 2:28:22 PM on Shellunext's Cluster

Cmd 2

```
1 query="select * from SalesLT.Customer"
```

Command took 0.12 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 2:29:28 PM on Shellunext's Cluster

Cmd 3

```
1 df=spark.read.jdbc(url=jdbc_url,table="SalesLT.Customer",properties=connection_properties)
```

▶ df: pyspark.sql.dataframe.DataFrame = [CustomerID: integer, NameStyle: boolean ... 13 more fields]

Command took 2.17 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 2:29:29 PM on Shellunext's Cluster

```
1 df.display()
```

▶ (1) Spark Jobs

Table ▾ +

	CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix
1	1	false	Mr.	Orlando	N.	Gee	null
2	2	false	Mr.	Keith	null	Harris	null
3	3	false	Ms.	Donna	F.	Carreras	null
4	4	false	Ms.	Janet	M.	Gates	null
5	5	false	Mr.	Lucy	null	Harrington	null
6	6	false	Ms.	Rosmarie	J.	Carroll	null
7	7	false	Mr	Dominic	P	Gash	null

↓ 847 rows | 12.12 seconds runtime

Refreshed 7 minutes ago

Command took 12.12 seconds -- by shellunext\_1693422225890@npunext.onmicrosoft.com at 9/27/2023, 2:29:30 PM on Shellunext's Cluster

# Day-28

- 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

```
i-0ad71c6ac54629843
Thu 28 Sep, 04:45 labuser@ip-172-31-4-200:~/Desktop/docker_project
File Edit View Search Terminal Help
Downloading itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting Jinja2>=3.0
  Downloading Jinja2-3.1.2-py3-none-any.whl (133 kB)
    133.1/133.1 kB 15.6 MB/s eta 0:00:00
Collecting Werkzeug>=2.0
  Downloading werkzeug-2.3.7-py3-none-any.whl (242 kB)
    242.2/242.2 kB 18.7 MB/s eta 0:00:00
Collecting MarkupSafe>=2.0
  Downloading MarkupSafe-2.1.3-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (25 kB)
Installing collected packages: MarkupSafe, itsdangerous, click, Werkzeug, Jinja2, Flask
Successfully installed Flask-2.0.1 Jinja2-3.1.2 MarkupSafe-2.1.3 Werkzeug-2.3.7 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 to use a virtual environment instead: 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 bde919f88970
--> 9f881faef0c4
Step 5/7 : EXPOSE 80
--> Running in d4bf3b268e46
Removing intermediate container d4bf3b268e46
--> 216cb0c1cfbb
Step 6/7 : ENV NAME World
--> Running in b44470d334e3
Removing intermediate container b44470d334e3
--> e6d3ff81fd03
Step 7/7 : CMD ["python", "app.py"]
--> Running in 27ea499d830f
Removing intermediate container 27ea499d830f
--> 020f67ebe140
Successfully built 020f67ebe140
Successfully tagged my-python-app:latest
labuser@ip-172-31-4-200:~/Desktop/docker_project$ sudo docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
labuser@ip-172-31-4-200:~/Desktop/docker_projects$
```

10:15 AM  
9/28/2023

```
i-0ad71c6ac54629843
Thu 28 Sep, 04:46 labuser@ip-172-31-4-200:~/Desktop/docker_project
File Edit View Search Terminal Help
docke_project - File Manager
File Edit View Go Help
/home/labuser/Desktop/docker_project/
DEVICES
  File System
  thinclient_dri...
PLACES
  labuser
  Desktop
  Trash
  Documents
  Music
  Pictures
  Videos
  Downloads
NETWORK
  Browse Network
3 items: 621 bytes, Free space: 10.5 GiB
File Edit View Search Terminal Help
Downloaded itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting Jinja2>=3.0
  Downloading Jinja2-3.1.2-py3-none-any.whl (133 kB)
    133.1/133.1 kB 15.6 MB/s eta 0:00:00
Collecting Werkzeug>=2.0
  Downloading werkzeug-2.3.7-py3-none-any.whl (242 kB)
    242.2/242.2 kB 18.7 MB/s eta 0:00:00
Collecting MarkupSafe>=2.0
  Downloading MarkupSafe-2.1.3-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (25 kB)
Installing collected packages: MarkupSafe, itsdangerous, click, Werkzeug, Jinja2, Flask
Successfully installed Flask-2.0.1 Jinja2-3.1.2 MarkupSafe-2.1.3 Werkzeug-2.3.7 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 to use a virtual environment instead: 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 bde919f88970
--> 9f881faef0c4
Step 5/7 : EXPOSE 80
--> Running in d4bf3b268e46
Removing intermediate container d4bf3b268e46
--> 216cb0c1cfbb
Step 6/7 : ENV NAME World
--> Running in b44470d334e3
Removing intermediate container b44470d334e3
--> e6d3ff81fd03
Step 7/7 : CMD ["python", "app.py"]
--> Running in 27ea499d830f
Removing intermediate container 27ea499d830f
--> 020f67ebe140
Successfully built 020f67ebe140
Successfully tagged my-python-app:latest
labuser@ip-172-31-4-200:~/Desktop/docker_project$ sudo docker ps
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
labuser@ip-172-31-4-200:~/Desktop/docker_projects$
```

10:15 AM  
9/28/2023

```
i-0ad71c6ac54629843
Thu 28 Sep, 05:39 labuser@ip-172-31-4-200:/home/labuser/Desktop/docker_project
File Edit View Search Terminal Help
172.17.0.2:81 - Google ...
root@ip-172-31-4-200: /home/labuser/Desktop/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: 23.0.1 -> 23.2.1
[notice] To update, run: pip install --upgrade pip
Removing intermediate container 73dd2043b6d8
--> 818160232fc
Step 5/7 : EXPOSE 80
--> Running in cdd4638646bc
Removing intermediate container cdd4638646bc
--> 23c3b897c926
Step 6/7 : ENV NAME World
--> Running in df5516edc1ef
Removing intermediate container df5516edc1ef
--> 162babf8ae5e
Step 7/7 : CMD ["python", "app.py"]
--> Running in ebd904206342
Removing intermediate container ebd904206342
--> 5eab6c0d1d21
Successfully built 5eab6c0d1d21
Successfully tagged my-python-app:latest
root@ip-172-31-4-200:/home/labuser/Desktop/docker_project# docker run my-python-app
* Serving 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.
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on all addresses (0.0.0.0)
* Running on http://127.0.0.1:81
* Running on http://172.17.0.2:81
Press CTRL+C to quit
172.17.0.1 - [28/Sep/2023 05:39:01] "GET / HTTP/1.1" 200 -
172.17.0.1 - [28/Sep/2023 05:39:01] "GET /favicon.ico HTTP/1.1" 404 -
labuser@ip-172-31-4-200:~/Desktop/docker_project$
```

11:09 AM  
9/28/2023

root@ip-172-31-4-200: / ...

File Edit View Search Terminal Help

```
},
  "supportPlan": "KubernetesOfficial",
  "systemData": null,
  "tags": null,
  "type": "Microsoft.ContainerService/ManagedClusters",
  "upgradeSettings": null,
  "windowsProfile": {
    "adminPassword": null,
    "adminUsername": "azureuser",
    "enableCsiproxy": true,
    "gmsaProfile": null,
    "licenseType": null
  },
  "workloadAutoScalerProfile": {
    "keda": null,
    "verticalPodAutoscaler": null
  }
}
labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ kubectl apply -f deployment.yaml deployment.apps/my-app unchanged
labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ kubectl apply -f service.yaml
error: the path "service.yaml" does not exist
labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ kubectl apply -f services.yaml
service/my-python-app unchanged
labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ kubectl get pods
NAME READY STATUS RESTARTS AGE
my-app-5dbff5c5df-9s4rd 1/1 Running 0 39m
my-app-5dbff5c5df-vfrbc 1/1 Running 0 39m
labuser@ip-172-31-4-200:~/Desktop/docker_project/kubernetes$ kubectl get svc
NAME TYPE CLUSTER-IP EXTERNAL-IP PORT(S) AGE
kubernetes ClusterIP <none> 443/TCP 88m
my-python-app LoadBalancer 10.0.37.174 20.242.178.81 80:31957/TCP 38m
```

Thu 28 Sep, 05:39

Thu 28 Sep, 10:50 labuser

Thu 28 Sep, 10:51 labuser

3:29 PM  
9/28/2023

Microsoft Azure Portal | Mic ... Deploy a quickstart applic... Azure Voting App | notepad.pw / idashell | The ...

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

Home > AnuKuberClust > Create a starter application >

### Deploy a quickstart application

Resource	Type	Status
azure-vote	Namespace	Success
azure-vote-back	Deployment	Success
azure-vote-back	Service	Success
azure-vote-front	Deployment	Success
azure-vote-front	Service	Success

**2. Next steps**

Here are some actions you can take to further explore your new application and learn

Previous Close

Microsoft Azure Portal | Mic ... Deploy a quickstart applic... Azure Voting App | notepad.pw / idashell | The ...

Thu 28 Sep, 09:59 labuser

File Edit View Search Terminal Help

labuser@ip-172-31-4-200: ~/Desktop/docker\_project/kubernetes\$

**Azure Voting App**

Cats

Dogs

Reset

Cats - 0 | Dogs - 0

4:20 PM  
9/28/2023

4:21 PM  
9/28/2023

# Day-29

## 1. 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

## Pull requests

[New pull request](#)[Mine](#) [Active](#) [Completed](#) [Abandoned](#)

Pull Request ID	Created by	Assigned to	Target branch
SU updated test Shellunext unext!DA90 request l3 into ↗ main	Shellunext	Q 0	Completed 52m ago
SU Updated test.json Shellunext unext!DA90 request l2 into ↗ main	Shellunext	Q 0	Completed 1h ago
SU updated test.json Shellunext unext!DA90 request l1 into ↗ main	Shellunext	Q 0	Completed 1h ago

[Errors 4](#)[Warnings 2](#)

- x 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
- x 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
- x 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
- x 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

## Branches

[New branch](#)

Mine	All	Stale				
Branch	Com...	Author	Authored ...	Behind   Ahead	Status	Pull ...
feature	21c4f62	Shellunext unex...	52m ago	1 0	<span>●</span>	
main	b03970f	Shellunext unex...	51m ago		<span>★</span>	

## Jobs

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

# Day-30

- 1.What is Machine learning
- 2.Its uses in daily life
- 3.Types of machine learning algorithms :
  - Supervised
  - Unsupervised
  - Semi-supervised
  - Reinforcement
- 4.Overview of Azure ML Studio

Azure AI | Machine Learning Studio

Unext > tanya15 > Notebooks

## Notebooks

Files Samples

Assets

- Data
- Jobs
- Components
- Pipelines
- Environments
- Models
- Endpoints

Manage

- Compute
- Monitoring PREVIEW
- Data Labeling
- Linked Services

File Explorer

- Users
- Shellunext\_1693422225890
- azureml-getting-started

  - media

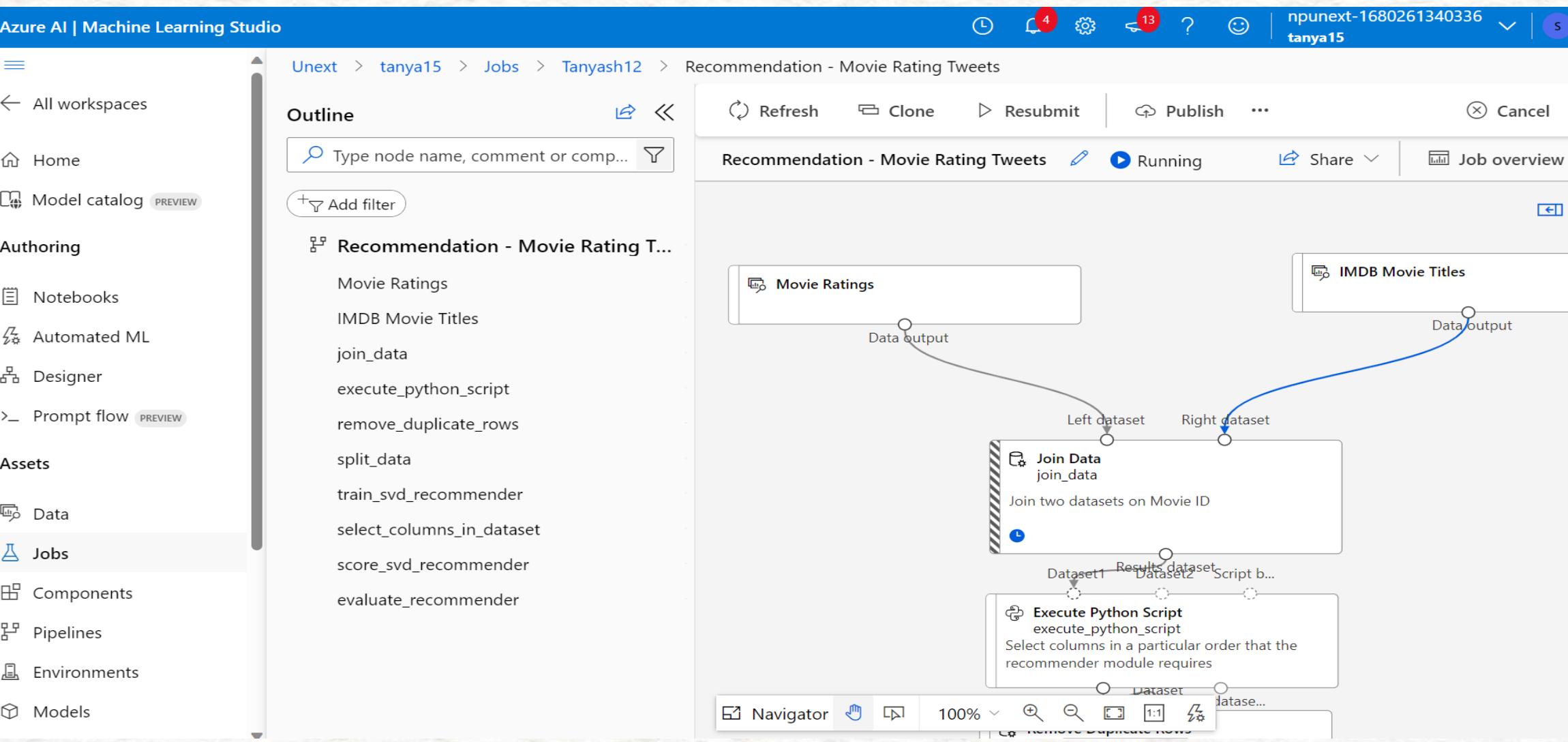
- azureml-getting-started-studio.ipynb

Code Editor

```
1 # Handle to the workspace
2 from azure.ai.ml import MLClient
3
4 # Authentication package
5 from azure.identity import DefaultAzureCredential
6
7 credential = DefaultAzureCredential()
8
9 # Get a handle to the workspace. You can find the info on the workspace
10 ml_client = MLClient(
11     credential=credential,
12     subscription_id="779629bd-924c-4afb-ba8e-f99be154aa3c", # this will
13     resource_group_name="tanya_ida",
14     workspace_name="tanya15",
15 )
16 )
```

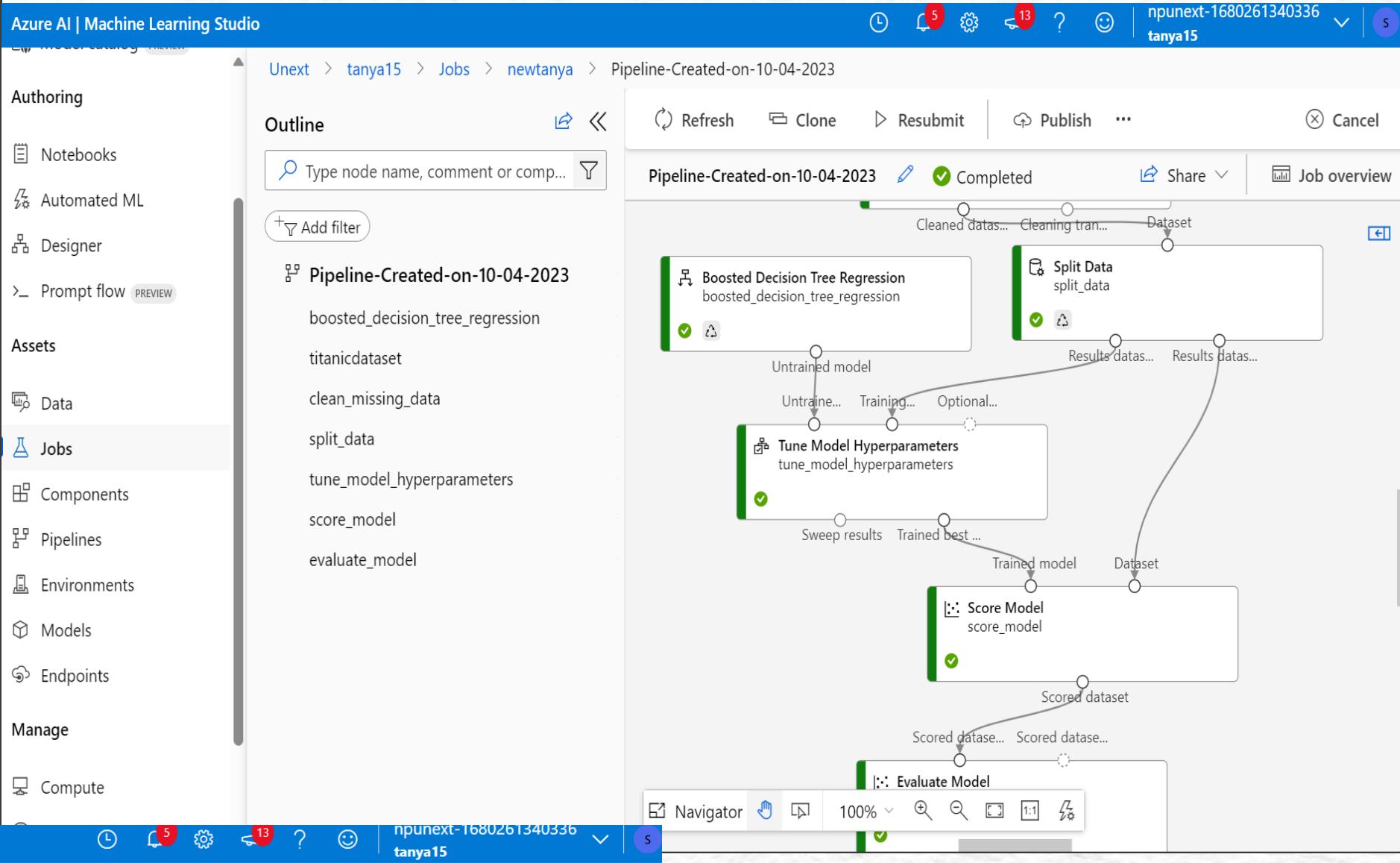
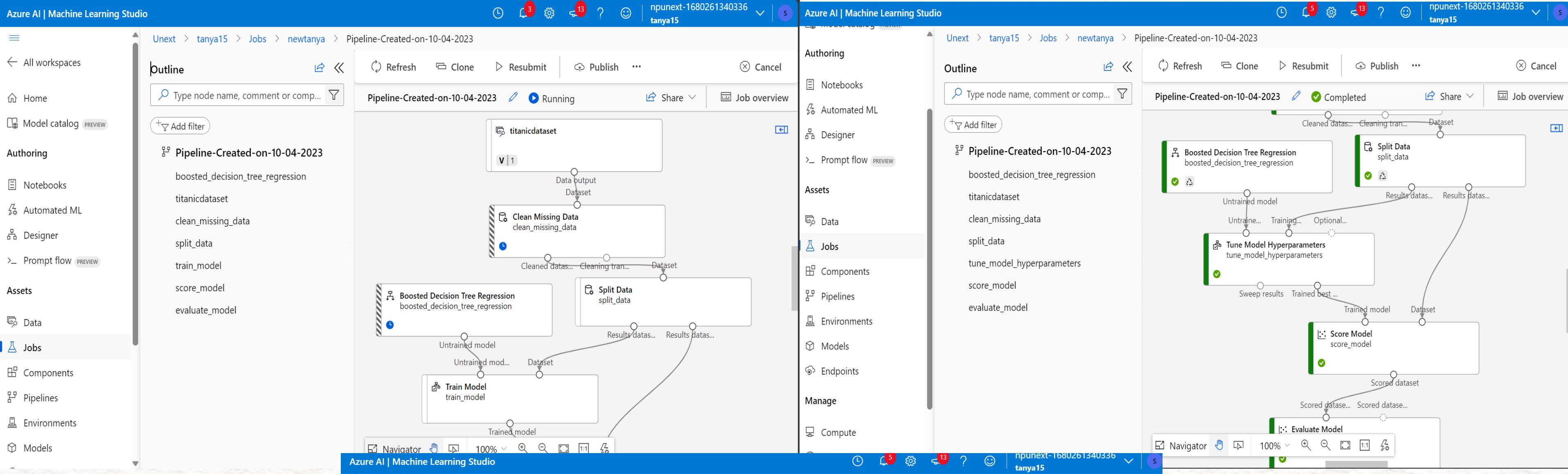
[2] ✓ 26 sec- Command executed in 17 sec 110 ms by Shellunext\_1693422225890 on 11:13:43 AM, 10/03/23

Class WorkspaceHubOperations: This is an experimental class, and may change at any time. Please see <https://aka.ms/azuremlexperimental> for more information.



# Day-31

1. Azure ML Studio Pipeline
2. Tune Model Hyperparameters
3. Automated ML



Unext > tanya15 > Jobs > newtanya > Pipeline-Created-on-10-04-2023

Scored\_dataset

Rows 267 Columns 13

PassengerId	Survived	Pclass	Name
756	1	2	Hamalainen, Master. Viljo
204	0	3	Youseff, Mr. Gerious
563	0	2	Norman, Mr. Robert Douglas
54	1	2	Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkinson)

To view, select a column in the table



Thank You