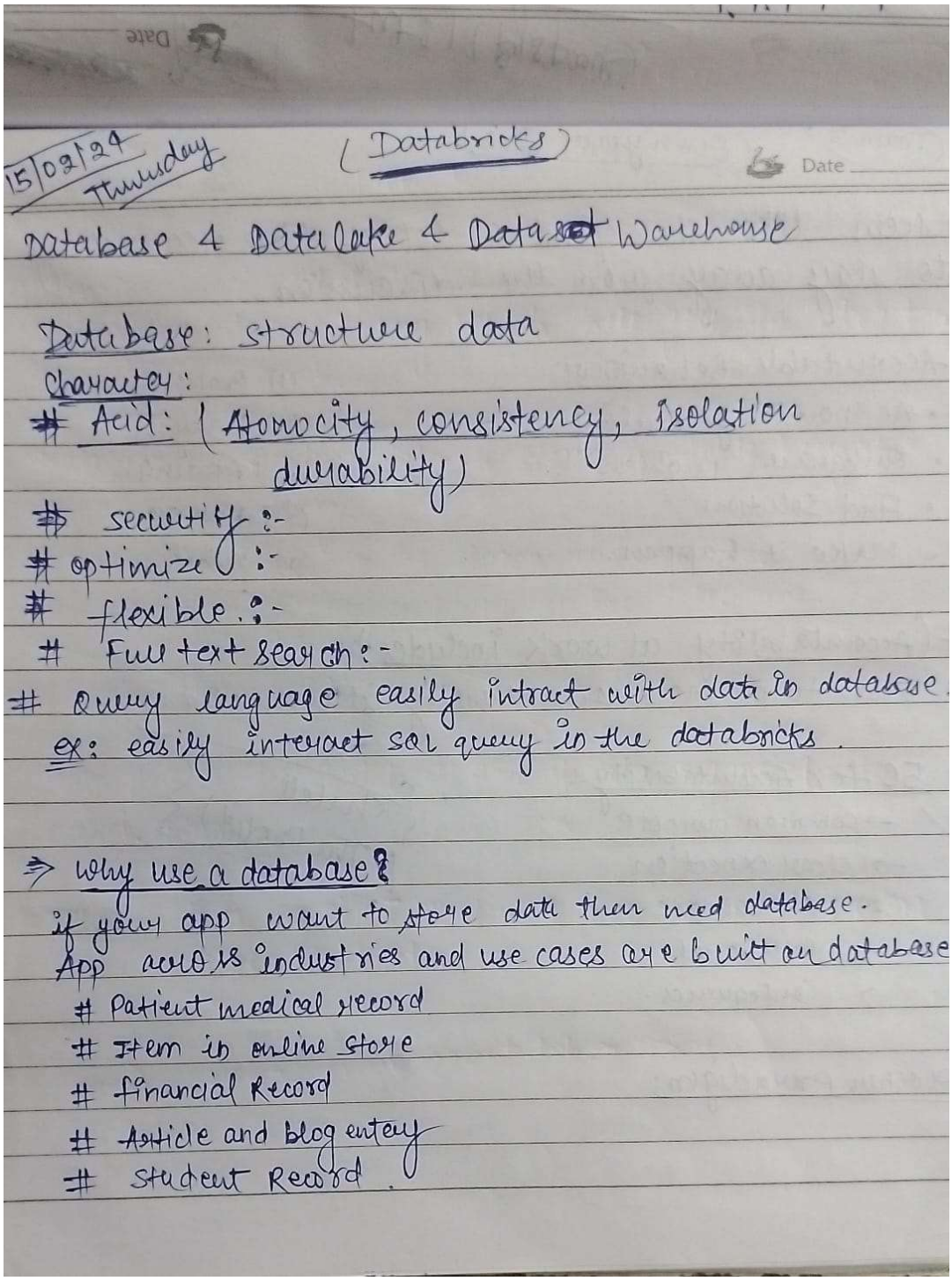
Name: Abhishek Kanoujia

DATA ENGINEERING BATCH 1

DAY 20 ASSIGNMENT

Class hand written notes:-



15/02/24 Thursday (Databricks)

Database & Data lake & Data set Warehouse

Database: structure data

Character:

Acid: (Atomocity, consistency, Isolation durability)

security :-

optimize :-

flexible :-

Full text search :-

Query language easily intract with data in database.

ex: easily interact sql query in the databricks.

⇒ Why use a database?

if your app want to store data then need database.

App and OS is industries and use cases are built on database

Patient medical record

Item in online store

financial Record

Article and blog entery

student Record.

Date

OLTP + data warehouse and data lake:

Both support OLTP. its system are typically used to collect data from variety of source. its data is used to power a range of analytical use cases.

ex: Quaterly sales report. to forcasting (predicts smart sales according to historical trends)

ex: Relation: oracal, mysql, micro, SQL, Postgre SQL
Doc: Mongo, couch, Key-value:- Redix, Dynmo,
wide-column: cassandra, HBase
Graph: Neo4j, Amazon Neptune.

database examples

datawarehouse:

its a system that store structured information from various source. Data warehouse typically store current and historical data from one or more system (its giant database that is optimized for analystics)

characteristics:

(i) store large amount of data
(ii) ETL (extract, transform, load)

iii) it typically pre-define and fixed relational schem
iv) data in the warehouse, business analysts can connect data warehouse with BI.

~~Use of data warehouse?~~ etl vs elt

Extract, transform, load (ETL) * (sequence is import

Exact → transform → load (Data)

Exact: Read data from source

transform: (highly curated cleaned and aggregated data)

load: Data loading in DW, DBC, Datalake, DB

→ DW
DBC (optional)
DL
DB

What is data lake?

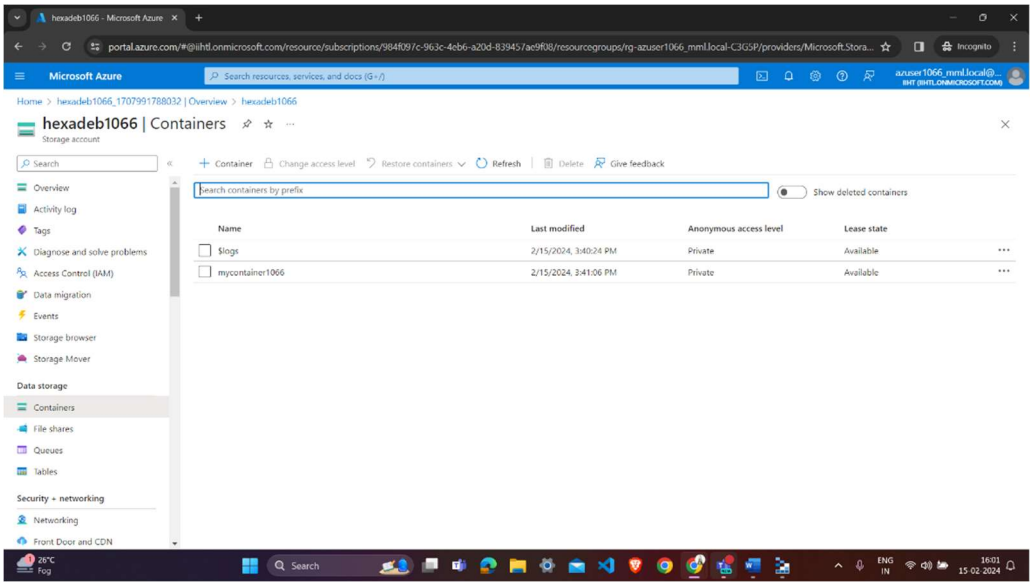
It is a repository of data from disparate source that is stored in its original raw formate.

purpose of data lake is to analyze the data to gain insight. (gain knowledge for particular solution) (by analyzing)

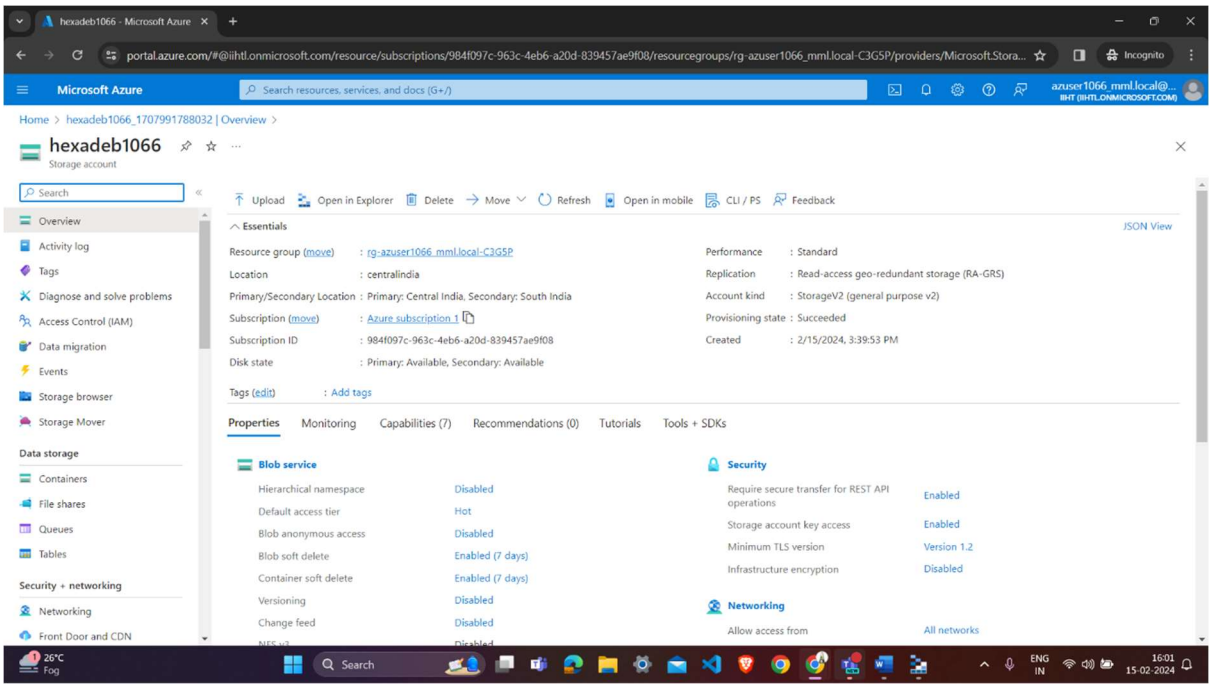
(Key difference btw. them)

	Database	Data lake	Data Warehouse
Workload	Operational & transactional	Analytical	Analytical
Schema			
Pros			
Cons			
datatyp	structured or semi-structu	structured, semi-structu unstructured ..	structured or semi-structured.

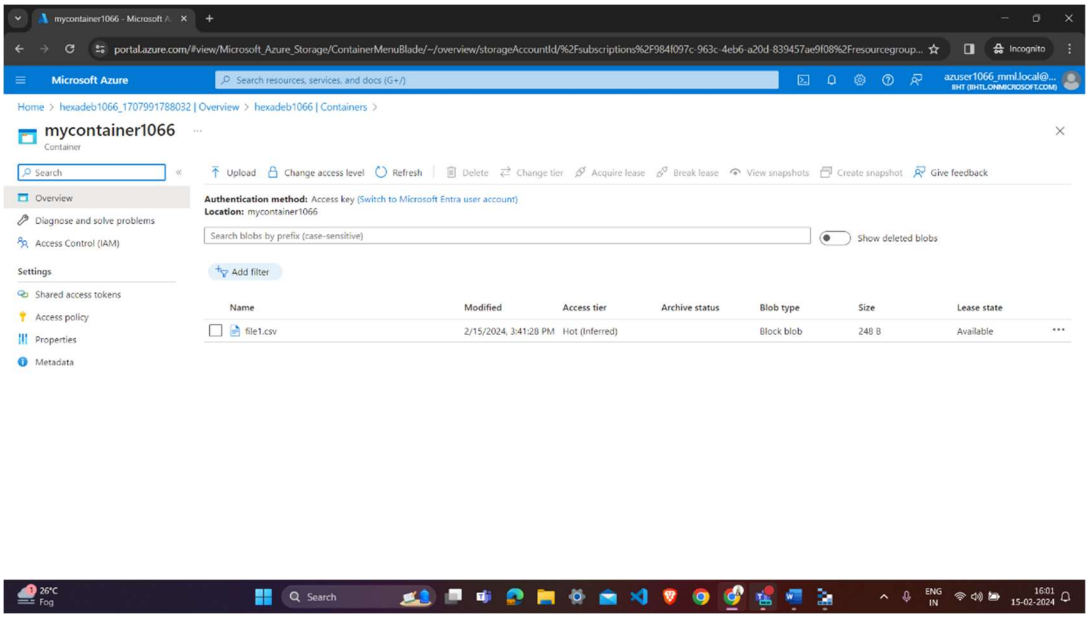
Create data storage :- all the credential provided for creation



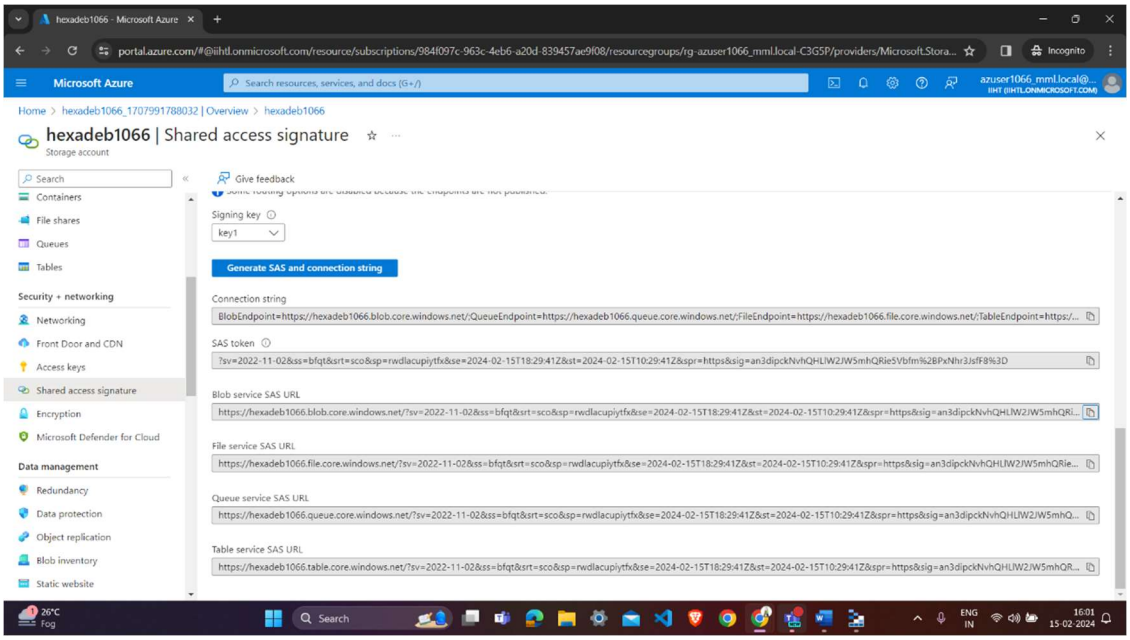
Create container :-



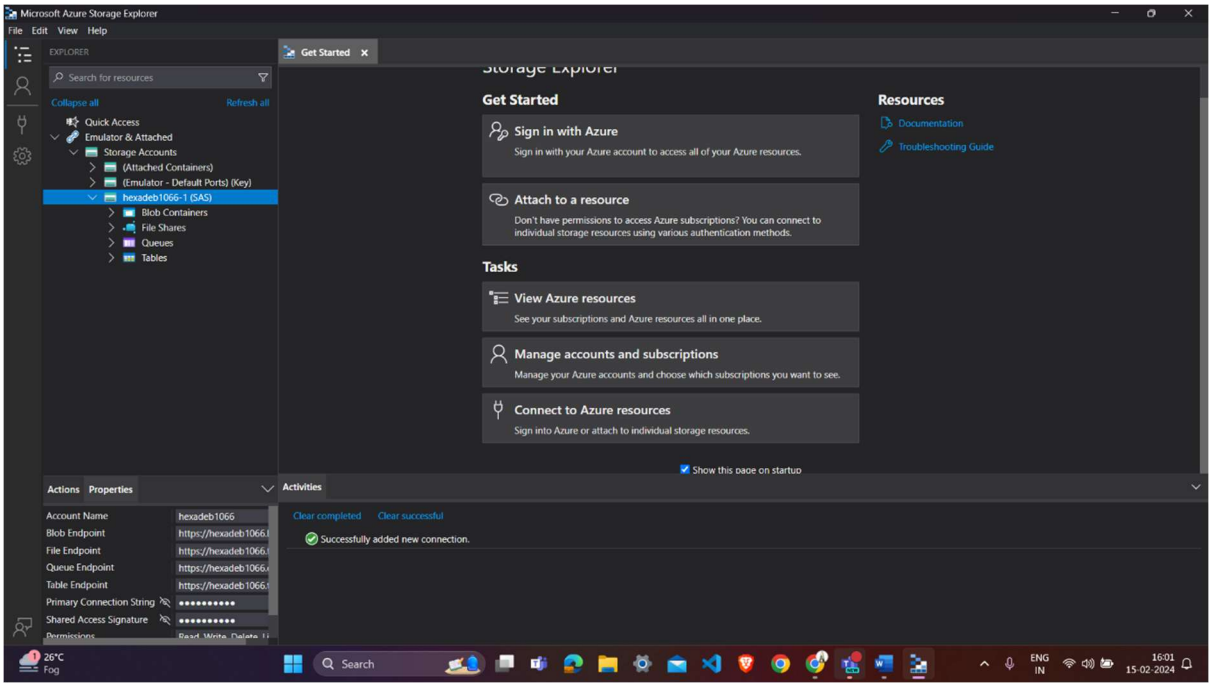
Upload file inside the container:-



Copy path of blob service sas URL:-



Connect to azure resources and paste the sas url in it:



Succesfully added new connection uploaded file is shown here:-