

Name: Pradip Bochare

✚ Notes

The image shows a page from a lined notebook with handwritten notes in black ink. The notes are organized into sections with asterisks. The first section defines a database and its typical use. The second section lists various database characteristics. The third section explains why databases are used, providing several examples. The notebook has a date field at the top right and a page number field at the top left.

* Database
Database is a collection of data or information. Databases are typically accessed electronically and are used to support OLTP.

* Database Characteristics-

- Security
- ACID (Atomicity, Consistency, Isolation, Durability) transactions to ensure data integrity.
- Query languages and API.
- Indexes to optimize query performance.
- Full text search.
- Flexible deployment topologies
 - on premises, private cloud, hybrid cloud, and multi-cloud hosting options

* Why Use Database

- Patient medical records.
- Items in an online store
- Financial records.
- Articles & blog entries.
- Sports scores & statistics.
- Online gaming information.
- Student grades and scores.
- IoT devices readings.
- Mobile application information.

* Database Examples

- Relational Databases: Oracle, MySQL, MS SQL
- Document databases: Mongo DB and Cloud DB.
- Key value databases: Redis & Dynamo DB
- Wide-column stores: Cassandra & HBase
- Graph databases: Neo4j & Amazon Neptune

* Data Warehouse

System that stores highly structured information from various sources. Data warehouses typically store current and historical data from one or more systems.

* Data warehouse characteristics.

- Store large amounts of current and historical data from sources.
- Extract, transform, Load (ETL) processes move data from its original source to data warehouse.
- Have pre-defined and fixed relational schema.

* Data Warehouse examples.

- Amazon Redshift
- Google BigQuery
- IBM Db2 Warehouse
- Microsoft Azure Synapse
- Oracle Autonomous Data Warehouse
- Snowflake
- Teradata Vantage

* Data Lake

- Data Lake is repository of data from disparate sources that is stored in its original, raw format.
- What sets data lakes apart is their ability to store data in variety of formats including JSON, BSON, CSV, TSV, Avro, ORC and Parquet.
- Primary purpose of data lake is to analyze data to gain insights.
- Tools like Starburst, Presto, Dremio and Atlas Data Lake can give a database like view into the data stored in your data lake.

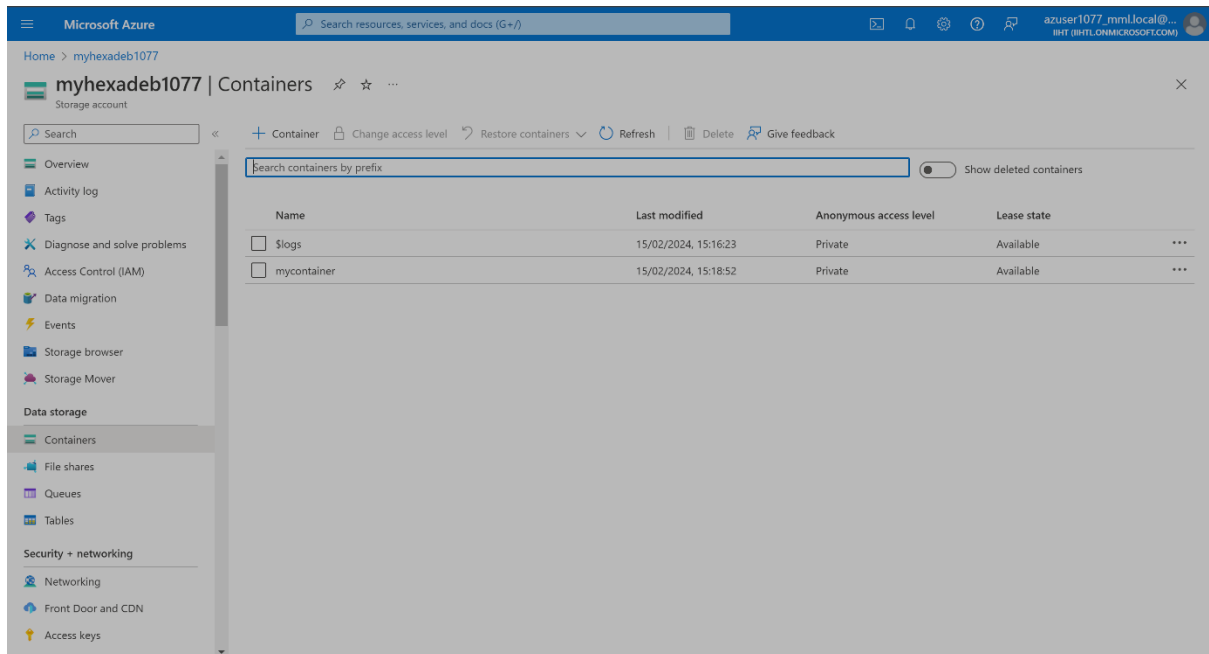
* Data Lake characteristics.

- store large amounts of structured, semi-structured and unstructured data.
- Data does not need to be transformed in order to be added to data lake which means data can be added (ingested) incredibly efficiently without upfront planning.
- Primary users of data lake can vary based on the structure of data.
- flexible nature of data lakes enables business ~~analysts~~ analysts and data scientists to look for unexpected patterns and insights.
- Data in data lake can be processed with a variety of OLAP systems and visualized with BI tools.

* Data Lake examples.

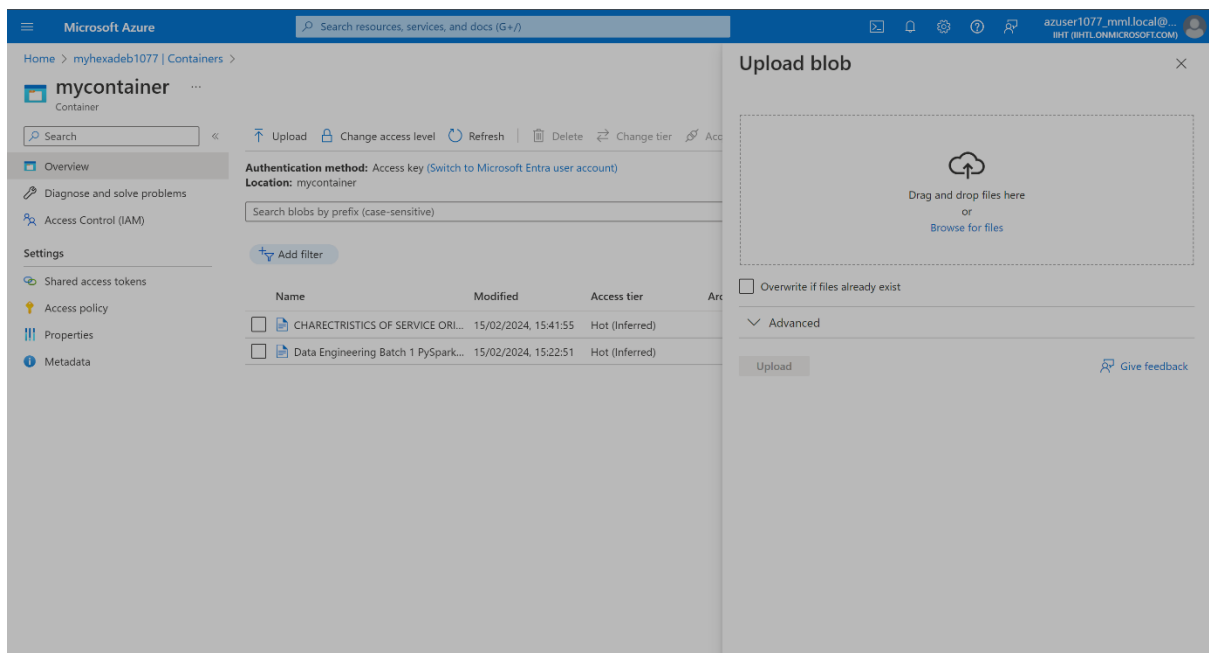
- AWS S3
- Azure Data Lake Storage Gen2
- Google cloud storage.

Azure Data Storage Creation



The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the Microsoft Azure logo, a search bar, and user information. The left sidebar contains a navigation menu with categories like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, Storage Mover, Data storage, Security + networking, and Networking. The main content area is titled 'myhexadeb1077 | Containers' and shows a list of containers. The 'Containers' section is active, displaying a table with columns: Name, Last modified, Anonymous access level, and Lease state. Two containers are listed: 'Slogs' and 'mycontainer'.

Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> Slogs	15/02/2024, 15:16:23	Private	Available
<input type="checkbox"/> mycontainer	15/02/2024, 15:18:52	Private	Available



The screenshot shows the Microsoft Azure portal interface with the 'mycontainer' container selected. The left sidebar shows the 'Containers' section. The main content area displays the 'mycontainer' details, including the authentication method (Access key) and location (mycontainer). Below this, there is a search bar for blobs and a table of blobs. The 'Upload blob' dialog is open on the right, showing a large dashed box for dragging and dropping files, a 'Browse for files' link, and an 'Upload' button.

Name	Modified	Access tier
<input type="checkbox"/> CHARECTISTICS OF SERVICE ORI...	15/02/2024, 15:41:55	Hot (Inferred)
<input type="checkbox"/> Data Engineering Batch 1 PySpark...	15/02/2024, 15:22:51	Hot (Inferred)

