NAME: SHAIK ABDUL KHADAR JILANI   ROLLNO: DXC262AB12038

BATCH: DXC-262-Analytics-B12-Azure   SUBMISSION: 3-6-2022

COMPANY: DXC TECHNOLOGY

DAY-5

1.Explain various Difference between SQL & NoSQL DBs ?

SQL DB	NoSQL DB
1.Database are categorized as relational Database Management System (RDBMS)	1.NoSQL databases are categorized as non-relational or distributed database system
2. SQL database have fixed or static or predefined schema.	2.NoSQL databases have dynamic schema
3.SQL database display data in form of tables so it is known as table-based database	3.NoSQL databases display data as collection of key-value pair, documents, graph databases or wide-column stores
4.SQL databases are vertically scalable	4.NoSQL databases are horizontally scalable.
5.SQL databases are best suited for complex queries	5.NoSQL databases are not so good for complex queries because these are not as powerful as SQL queries.
6.Eg: MySQL, Oracle, SQLite, PostgreSQL and MS-SQL are examples of SQL data base	6. Eg: MongoDB, BigTable, RavenDB, Cassandra, Hbase, Neo4j, CouchDB are the examples of NoSQL database

2.Explain advantages of NoSQL DBs? Explain how MongoDB data will be inserted?

Advantages of NoSQL DBs

1.It supports query language

2.It provides faster performance

3.It provides horizontal scalability

4.It support high scalable Data applications (Big Data & Realtime Data storage)

>var myemp = [{empid:1, empname:'admin'}, {empid:2, empname: 'manager'},
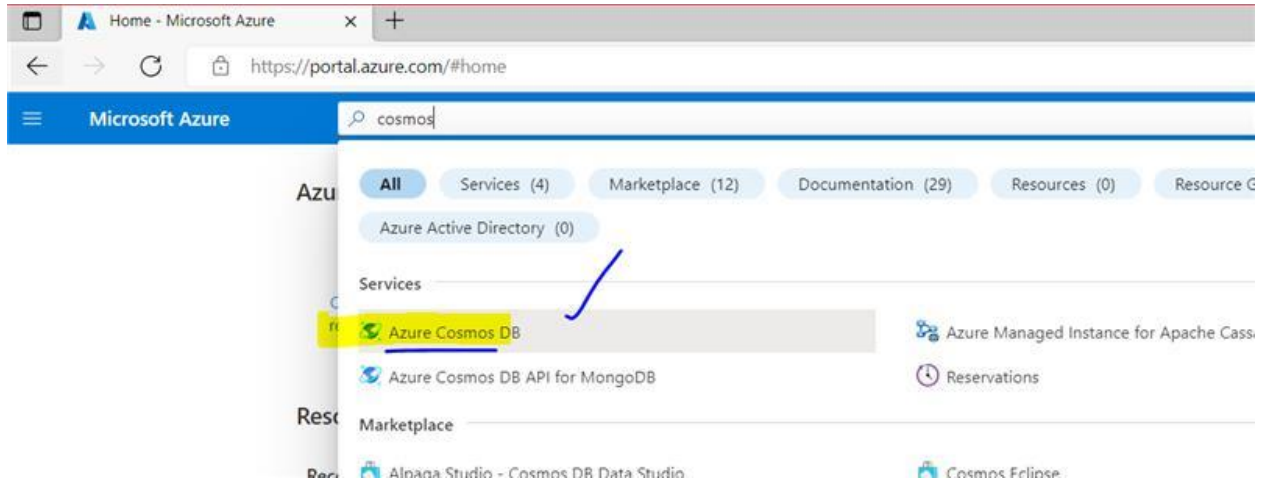
 {empid:3,empname:'qa'}]

>db.newemp.insert(myemp);

>db.newemp.find();

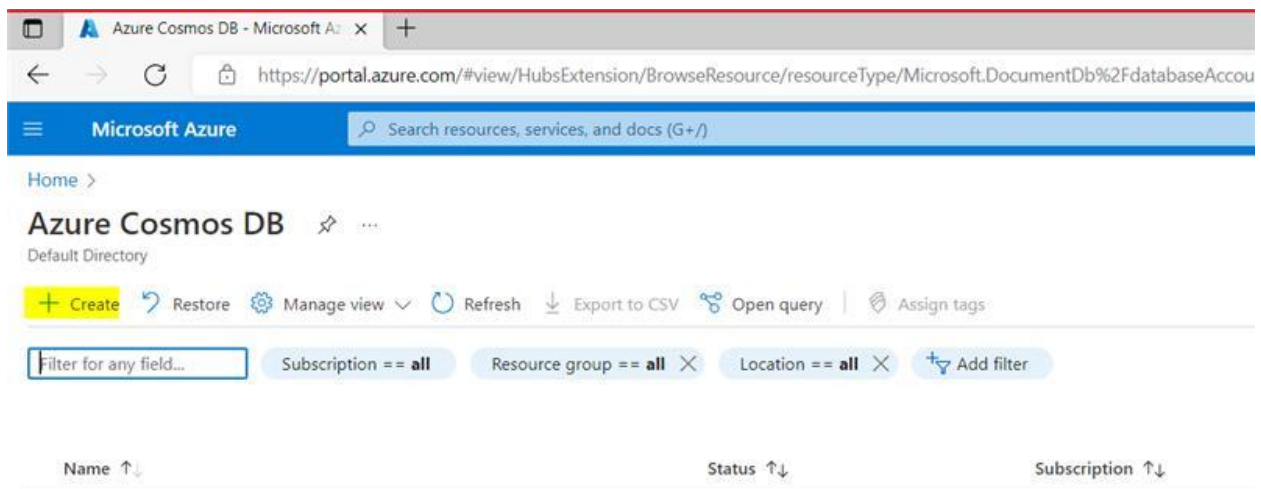By using this commands we can insert data in MongoDB

3. Explain the steps – how COSMOS DB can be created with screens ?

Go to portal.azure.com

Search for cosmosDB and select Azure Cosmos DB

Click on create

-it will show Core(SQL),Azure Cosmos DB API for MongoDB,Cassandra,Azure Table,
Gremlin(Graph) options to create

In that select Core(SQL) to create

-give accountname dxc and create

Click next till you get this page keep it all default and click next we will get Your deployment is complete
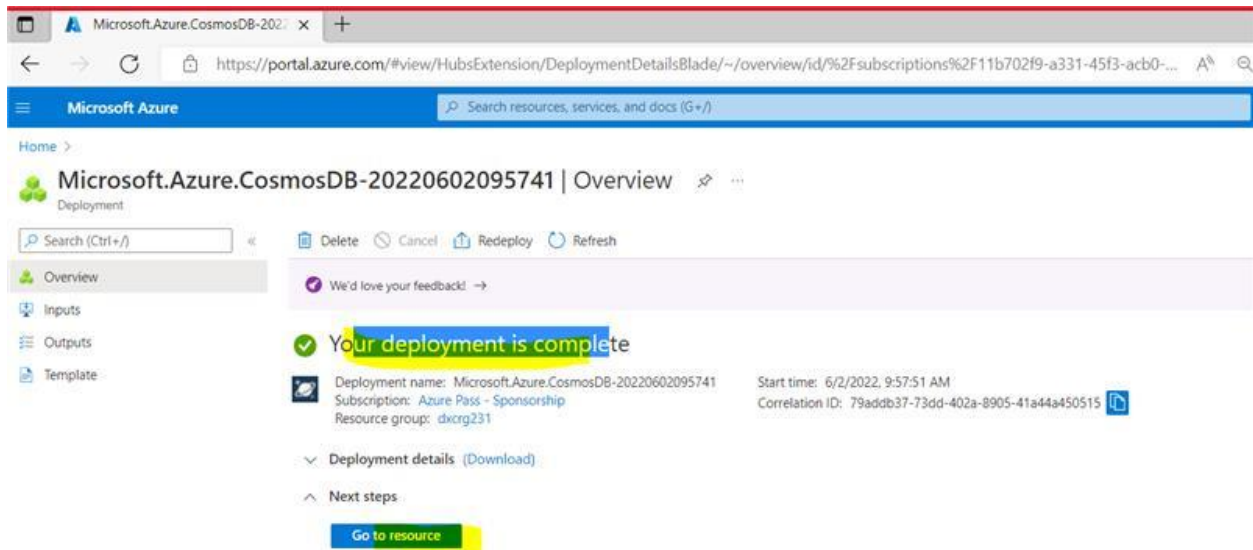
4.Explain how to write JSON query in COSMOS DB?
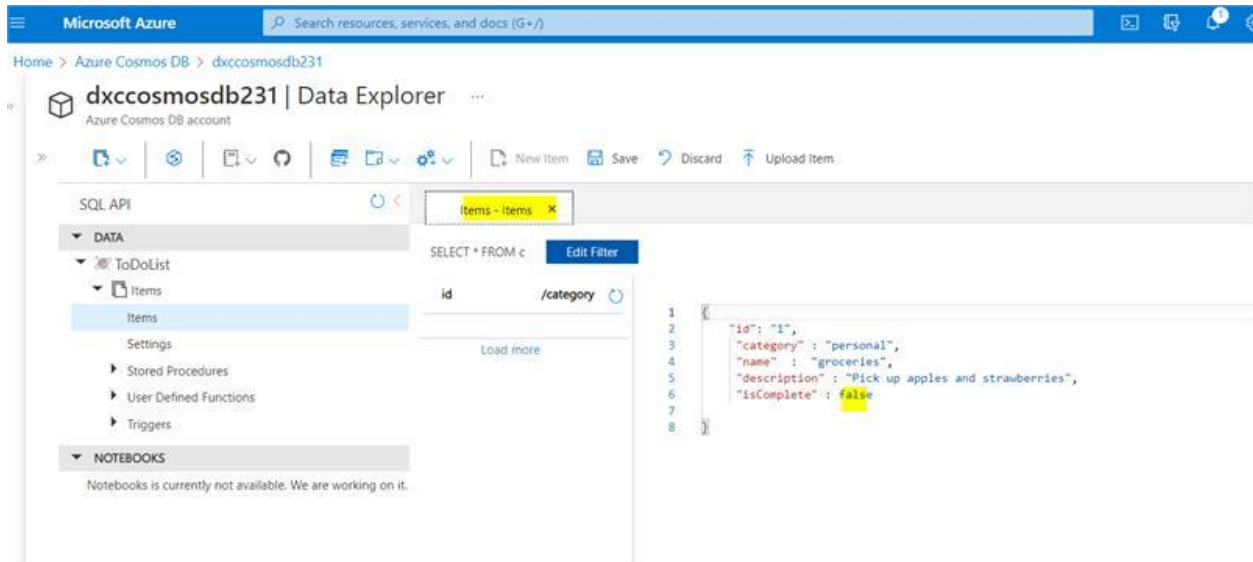
dxccosmosdb231 | Data Explorer ···
Azure Cosmos DB account

New Item | Update | Discard | Delete | Upload Item

SQL API

DATA
- ToDoList
  - Items
    - Items
    - Settings
  - Stored Procedures
  - User Defined Functions
  - Triggers

NOTEBOOKS

Notebooks is currently not available. We are working on it.

Items - Items ✕

SELECT * FROM c   Edit Filter

id	/category
1	personal

Load more

```
1  {
2    "id": "1",
3    "category": "personal",
4    "name": "groceries",
5    "description": "Pick up apples and strawberries",
6    "isComplete": false,
7    "_rid": "zvgzAPk3ZNwBAAAAAAAAAA==",
8    "_self": "dbs/zvgzAA==/colls/zvgzAPk3ZNw=/docs/zvgzAPk3ZNwBAAAAAAAAAA==/",
9    "_etag": "\"04008bda-0000-0100-0000-629715120000\"",
10   "_attachments": "attachments/",
11   "_ts": 1654068498
12 }
```

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

ajaykumarbk75@gmail...
DEFAULT DIRECTORY

dxccosmosdb231 | Data Explorer ···
Azure Cosmos DB account

New Item | Update | Discard | Delete | Upload Item

SQL API

DATA
- ToDoList
  - Items
    - Items
    - Settings
  - Stored Procedures
  - User Defined Functions
  - Triggers

NOTEBOOKS

Notebooks is currently not available. We are working on it.

Items - Items ✕

SELECT * FROM c   [Type a query predicate (e.g., WHERE c.id = '1'), or choose one from the drop down list, or leave empty to query all documents.]   Apply Filter ✕

id	/category
1	personal

Load more

```
1  {
2    "id": "1",
3    "category": "personal",
4    "name": "groceries",
5    "description": "Pick up apples and strawberries",
6    "isComplete": false,
7    "_rid": "zvgzAPk3ZNwBAAAAAAAAAA==",
8    "_self": "dbs/zvgzAA==/colls/zvgzAPk3ZNw=/docs/zvgzAPk3ZNwBAAAAAAAAAA==/",
9    "_etag": "\"04008bda-0000-0100-0000-629715120000\"",
10   "_attachments": "attachments/",
11   "_ts": 1654068498
12 }
```

5. Explain major difference between database & Datawarehouse?

Database	Datawarehouse
1.Database is a collection of related data that represents some elements of the real world	1.Datawarehouse is an information system that stores historical and commutative data from single or multiple sources
2.Database is designed to record data	2.Data warehouse is designed to analyze data.
3.It is application-oriented collection of data	3.It is the subject-oriented collection of data
4.Database uses OLTP	4.Data warehouse uses OLAP
5.Database tables and joins are complicated because they are normalized	5.Data warehouse tables and joins are easy because they are denormalized.
6.ER modeling techniques are used for designing Database	6.Data modeling techniques are used for designing Data Warehouse.

6.Explain the architecture of Datawarehouse?
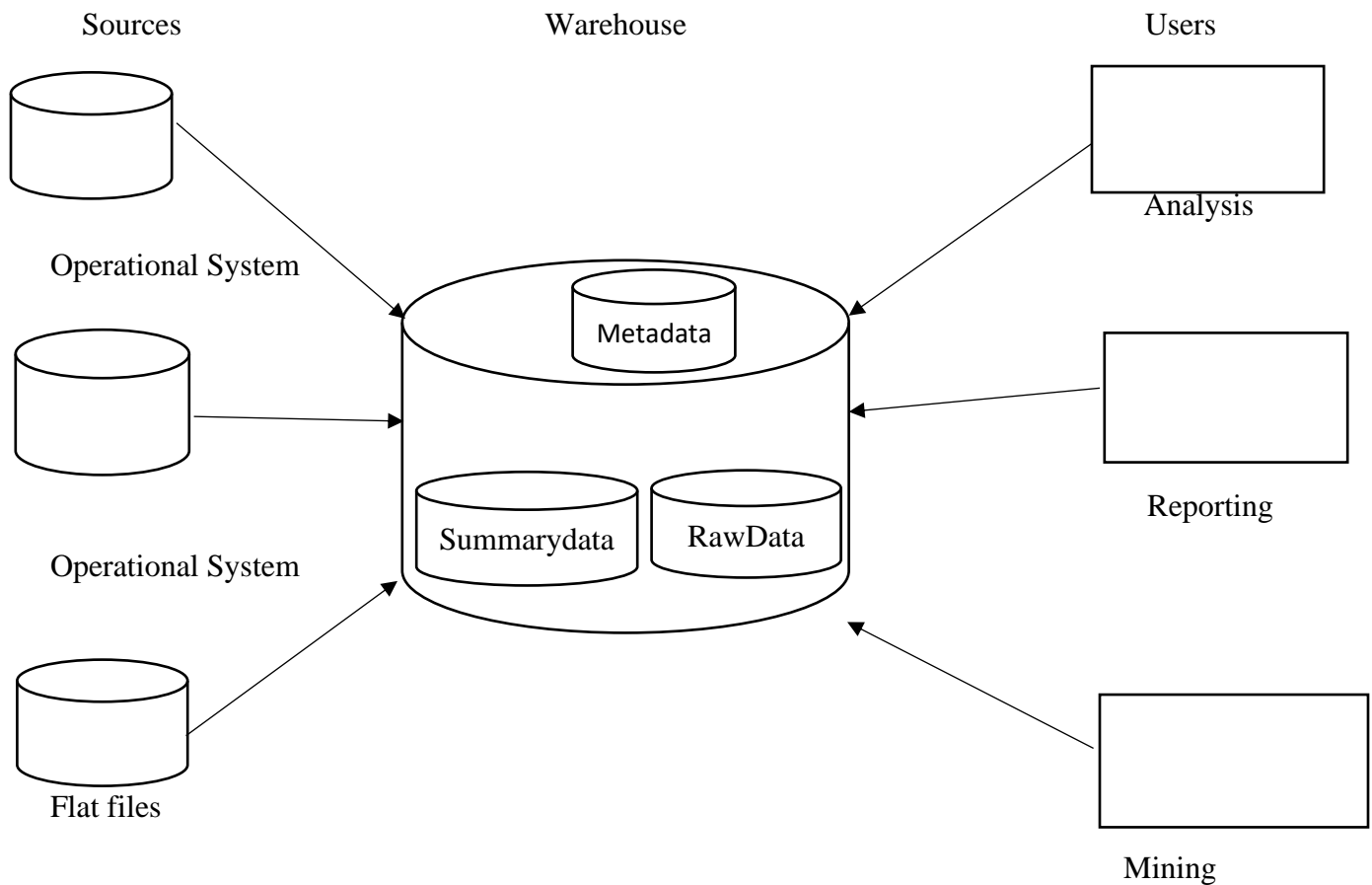
7.Explain what are Datamarts & how different from Databases and mention the types of Datamarts too?

DATAMARTS

- DM is a smaller version of the data Warehouse which deals with a single subject.
- DM are focused on one area. Hence, they draw from a limited number of sources.
- Time taken to build Data Marts is very less compared to time taken to build a Datawarehouse.

A database is a transactional data repository (OLTP) where a data mart is an analytical data repository(OLAP)

A database captures all the aspects and activities of one subject in particular. A data mart will house data from multiple subjects.

Types of Data Mart

1.Dependent Data Mart

- The data is first extracted from the OLTP systems and then populated in the central Datawarehouse
- From the Datawarehouse, the data travels to the Data mart

2.Independent DataMart

- The data is directly received from the source system
- This is suitable for small organizations or smaller groups within an organization

3.Hybrid Data Mart

- The data is fed both from OLTP systems as well as the Data Warehouse

8.Explain OLAP & OLTP with examples?

OLAP

OLAP stands for On-Line Analytical Processing. It is used for analysis of database information from multiple database systems at one time such as sales analysis and forecasting, market research, budgeting. Data Warehouse is the example of OLAP system.

- It is used for data analysis
- It uses data warehouse
- It manages all insert, update and delete transaction
- Processing is little slow
- Tables in OLAP database are not normalized.

Examples

1.Bank Manager wants to know how many customers are utilizing the ATM of his branch. Based on this he may take a call whether to continue with the ATM or relocate it

2.An insurance company wants to know the number of policies each agent has sold. This will help in better performance management of agents.

OLTP

OLTP stands for On-Line Transactional processing. It is used for maintaining the online transaction and record integrity in multiple access environments. OLTP is a system that manages very large number of short online transactions for example, ATM.

- It is used to manage very large number of online short transactions
- It uses traditional DBMS
- It is mainly used for data reading
- Responses In Milliseconds
- Tables in OLTP database are normalized.

Examples

1.A supermarket server which records every single product purchased at that market

2.A bank server which records every time a transaction is made for a particular account.

3.A railway reservation server which records the transactions of a passenger.

9. Explain what is BI & how BI helps business to take intelligent decisions ?

Business Intelligence is the activity which contributes to the growth of any company

Planning->Data Gathering->Data Analysis->Business Action->>Business Growth

BI is the act of transforming raw/operational data into useful information for

business analysis

1.BI based on Datawarehouse technology extracts information from a company's os.

2.The data is transformed (cleaned and integrated) and loaded into Data Warehouses.

3.Since this data is credible it is used for business insights.

10.Explain how ETL works with Datawarehouse ?

ETL is a process in Data Warehousing and it stands for Extract, Transform and Load.

It is a process in which an ETL tool extracts the data from various data source

systems, transforms it in the staging area, and then finally, loads it into the

Data Warehouse system.

Extraction:

The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML, and flat files into the staging area.

It is important to extract the data from various source systems and store it into the staging area first and not directly into the data warehouse because the extracted data is in various formats and can be corrupted also. Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult.

Therefore, this is one of the most important steps of ETL process.

Transformation:

The second step of the ETL process is transformation. In this step, a set of rules
 or functions are applied on the extracted data to convert it into a single standard
format. It may involve following processes/tasks:

Filtering – loading only certain attributes into the data warehouse.

Cleaning – filling up the NULL values with some default values, mapping U.S.A, United States, and America into USA, etc.

Joining – joining multiple attributes into one.

Splitting – splitting a single attribute into multiple attributes.

Sorting – sorting tuples on the basis of some attribute (generally key-attribute).

Loading:

The third and final step of the ETL process is loading. In this step, the transformed
 data is finally loaded into the data warehouse. Sometimes the data is updated by
loading into the data warehouse very frequently and sometimes it is done after
longer but regular intervals. The rate and period of loading solely depends on the
 requirements and varies from system to system.

ETL process can also use the pipelining concept i.e. as soon as some data
is extracted, it can transformed and during that period some new data can be
extracted. And while the transformed data is being loaded into the data warehouse,
 the already extracted data can be transformed