NAME: BHOGADI NAGA ISWARYA LAKSHMI **SUBMISSION:**04-06-2022

BATCH: DXC-262-ANALYTICS-B12-AZURE **ASSESSMENT:**5

1. Explain various Difference between SQL & NoSQL DBs ?

SQL	NoSQL
Rational Database	Non- rational, Distributed Database
Vertically Scalable	Horizontally Scalable
Table Based Database	Document Based, Graph Based or Key- value Pair
Pre-Define Schema	Dynamic schema
Uses SQL	Uses UnQL (Unstructured Query Language)
Not Preferred for Large Datasets	Largely Preferred for large Datasets

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

Advantages of NoSQL:

- Cheap & easy to implement due to open source.
- Easy to distribute
- Scale to available memory
- Have individual query language rather than using a standard query language.
- Flexible data model.
- Developer friendly-More developer centric interface.

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

(empid:3,empname:'qa'}]

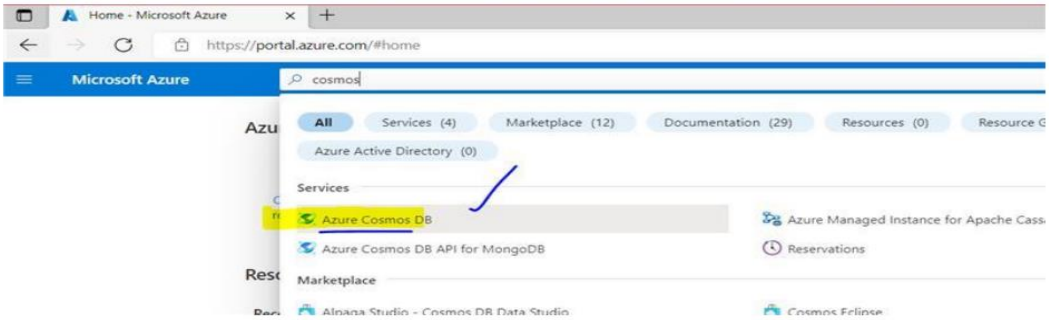
>db.newemp.insert(myemp);

>db.newemp.find()

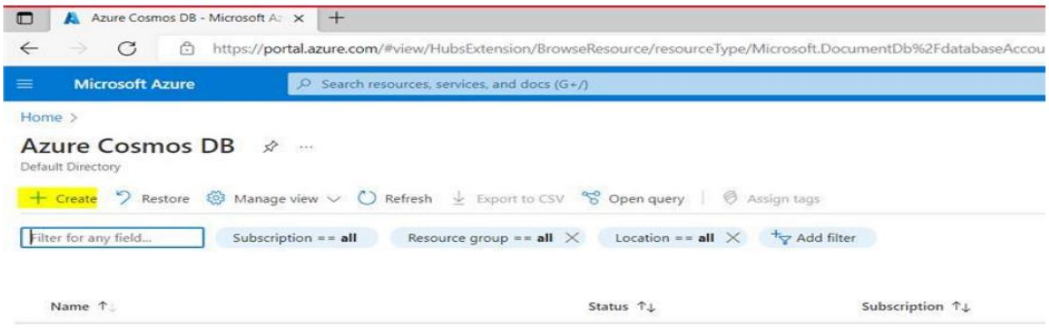
By using this commands we can insert data in MongoDB

3. Explain the steps - how COSMOS DB can be creatd with screens ?

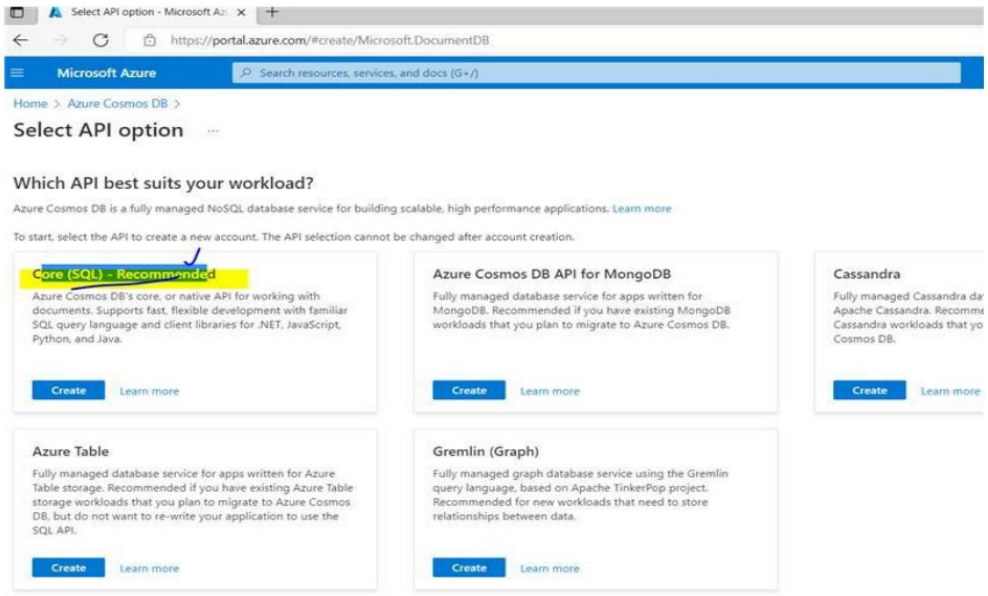
Go to portal.azure.com



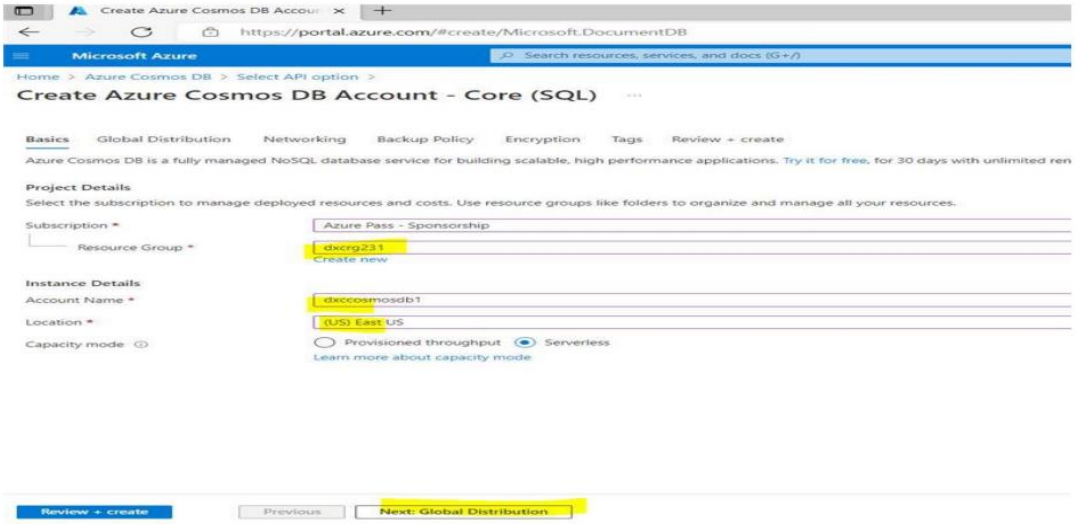
Click on create



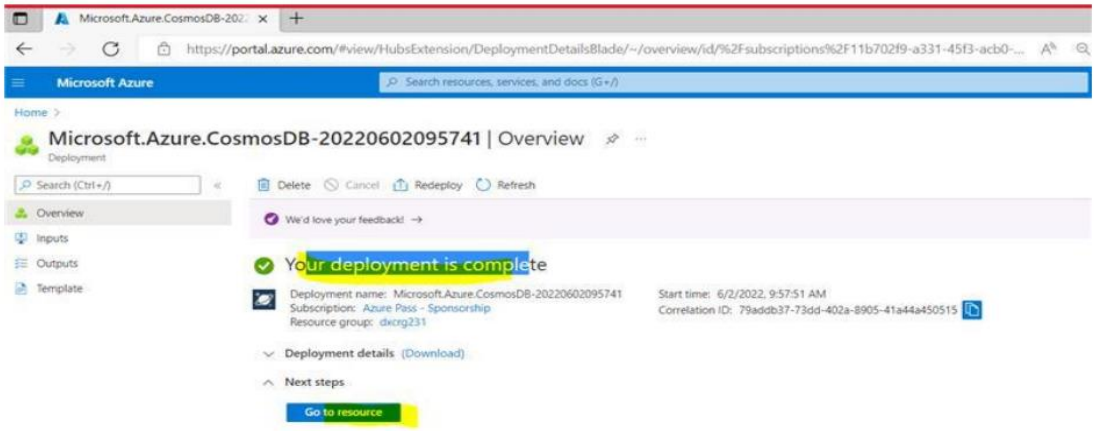
-it will show Core(SQL),Azure Cosmos DB API for MongoDB,Cassandra,Azure Table, Gremlin(Graph) options to create In that select Core(SQL) to create



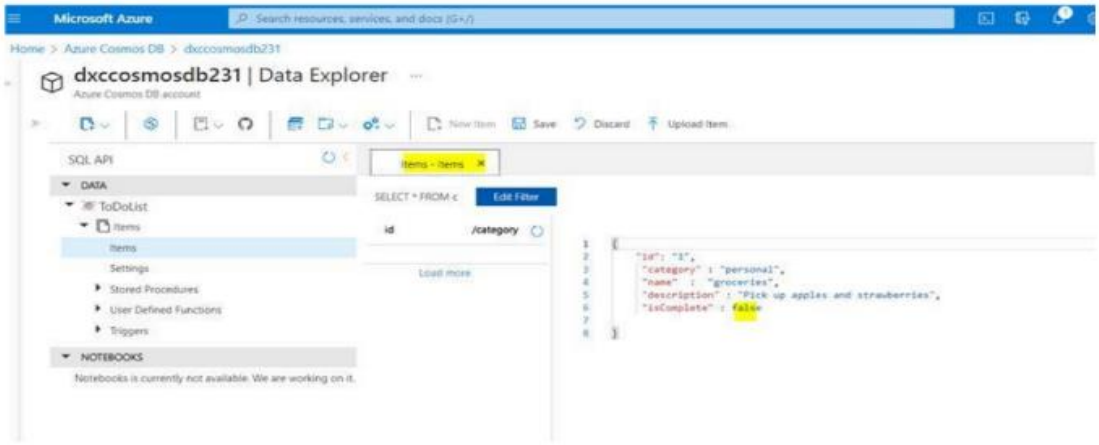
-give accountname dxc and create

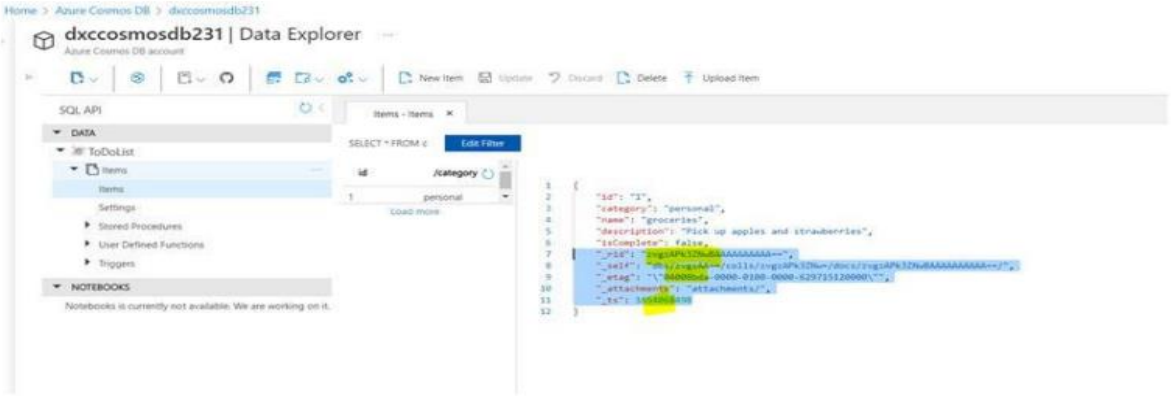


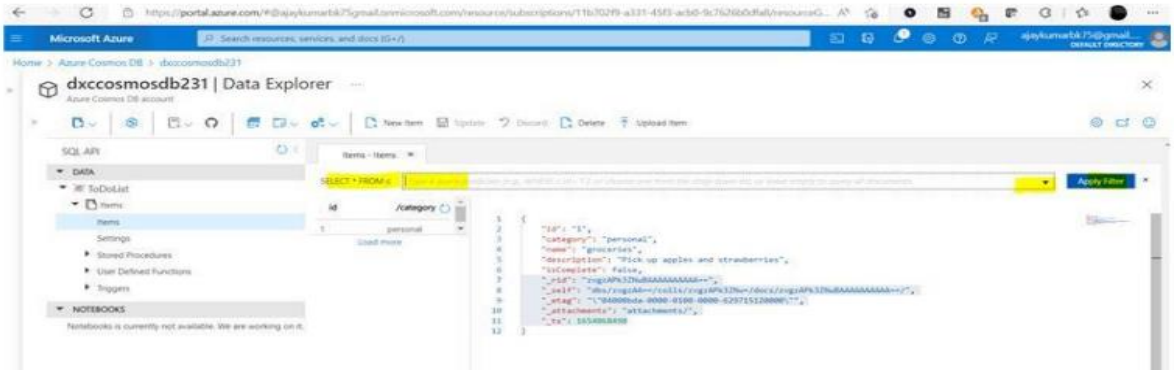
Click next till you get this page keep it all default and click next we will get Your deployment is complete



4. Explain how to write JSON query in COSMOS DB ?



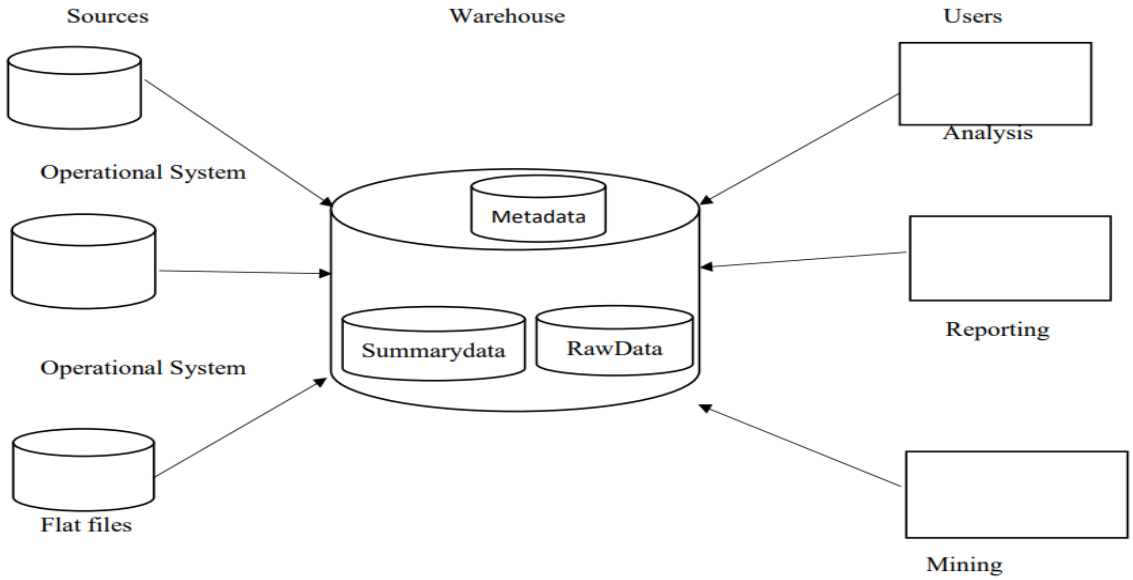




5. Explain major difference between databases & datawarehouses ?

DATABASES	DATAWAREHOUSES
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. Database
6.ER modeling techniques are used for designing	6.Data modeling techniques are used for designing Data Warehouse.

6. Explain the architecture of datawarehouses ?



A **data-warehouse** is a heterogeneous collection of different data sources organised under a unified schema. There are 2 approaches for constructing data-warehouse: Top-down approach and Bottom-up approach are explained as below.

1. Top-down approach:

The essential components are discussed below:

1. **ExternalSources** – External source is a source from where data is collected irrespective of the type of data. Data can be structured, semi structured and unstructured as well.

2. **StageArea** – Since the data, extracted from the external sources does not follow a particular format, so there is a need to validate this data to load into datawarehouse. For this purpose, it is recommended to use **ETL** tool.
    - **E(Extracted):** Data is extracted from External data source.
    - **T(Transform):** Data is transformed into the standard format.
    - **L(Load):** Data is loaded into datawarehouse after transforming it into the standard format.

3. **Data-warehouse** – After cleansing of data, it is stored in the datawarehouse as central repository. It actually stores the meta data and the actual data gets stored in the data marts.

4. **DataMarts** – Data mart is also a part of storage component. It stores the information of a particular function of an organisation which is handled by single authority. There can be as many number of data marts in an organisation depending upon the functions. We can also say that data mart contains subset of the data stored in datawarehouse.

5. **DataMining** – The practice of analysing the big data present in datawarehouse is data mining. It is used to find the hidden patterns that are present in the database or in datawarehouse with the help of algorithm of data mining.

2. Bottom-up approach:

1. First, the data is extracted from external sources (same as happens in top-down approach).

2. Then, the data go through the staging area (as explained above) and loaded into data marts instead of datawarehouse. The data marts are created first and provide reporting capability. It addresses a single business area.

3. These data marts are then integrated into datawarehouse.

6. Explain what are Datamarts & how different from DATABASES ? & 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. Metadata Summarydata RawData

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

7. 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

8. 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 Datawarehouses ?

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