ASSIGNMENT NO. 05

Name – Mohana Likhitha   **Roll Number - DXC262AB12019**
Batch – Dxc-262-Analytics-B12-Azure   **Company – DXC Technology**
Employee Domain –Azure Analytics   **Training Under – Manipal Pro Learn**
Trainer Name – Mr. Ajay Kumar

Q1- Explain various Difference between SQL & NoSQL DBs ?

SQL	NoSQL
1. RELATIONAL DATABASE MANAGEMENT SYSTEM (RDBMS)	1. Non-relational or distributed database system.
2. predefined schema	2. Dynamic Schema
3. not suited for hierarchical data storage.	3. These databases are best suited for hierarchical data storage.
4. best suited for complex queries	4. not so good for complex queries
5. Tables with fixed rows and columns	5. Document: JSON documents, Key-value: key-value pairs,
6. Vertical scale	6. Horizontal scale
7. Example—Mysql	7. Example—firebase

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

Flexible – Stores any type of data, both structured and un-structured. Unlike other DBs which save only structured data.
High Performance— NoSQL databases are built for great performance
Handle large amount of data.
Enable easy updates to schemas and fields.
Low Downtime due to cloud

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

Step1—
Go to microsoft Azure Cosmos DB official website -- https://azure.microsoft.com/en-us/services/cosmos-db/
Step 2—Create microsoft Azure Cosmos DB account/Sign In, select microsoft Azure Cosmos DB
Step 3—Enter project details, click on review and create resource

Step 4—Go to resource to go to the Azure Cosmos DB account page. Select Data Explorer from the left navigation on your Azure Cosmos DB account page, and then select New Container.

Step 5 – click on Add Container window.

Step 6— In Data Explorer, expand the ToDoList database, and expand the Items container. Next, select Items, and then select New Item.

Step7—Add Json code in document , select save

notebookaccount | Data Explorer ···
Azure Cosmos DB account

Search (Ctrl+/)

« | New Container ⌄ |

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Quick start
- Notifications
- **Data Explorer**

Settings

- Features
- Replicate data globally
- Default consistency
- Backup & Restore
- Firewall and virtual networks
- Private Endpoint Connections
- CORS
- Keys
- Add Azure Cognitive Search
- Add Azure Function
- Advanced security (preview)
- Locks

SQL API

▼ DATA
- ▶ DB1
- ▶ SharedDB
- ▶ SharedDB1

▼ NOTEBOOKS
- Gallery
- ▼ My Notebooks
  - How to use autoscale.ipynb
  - Untitled.ipynb

New Container

* **Database id** ⓘ
- ⦿ Create new   ◯ Use existing

| ToDoList |

☑ Share throughput across containers ⓘ

* **Database throughput (400 - unlimited RU/s)** ⓘ
- ◯ Autoscale   ⦿ Manual

Estimate your required RU/s with capacity calculator.

| 400 | *

Estimated cost (USD) ⓘ : **<price>**

* **Container id** ⓘ

| Items |

* **Partition key** ⓘ
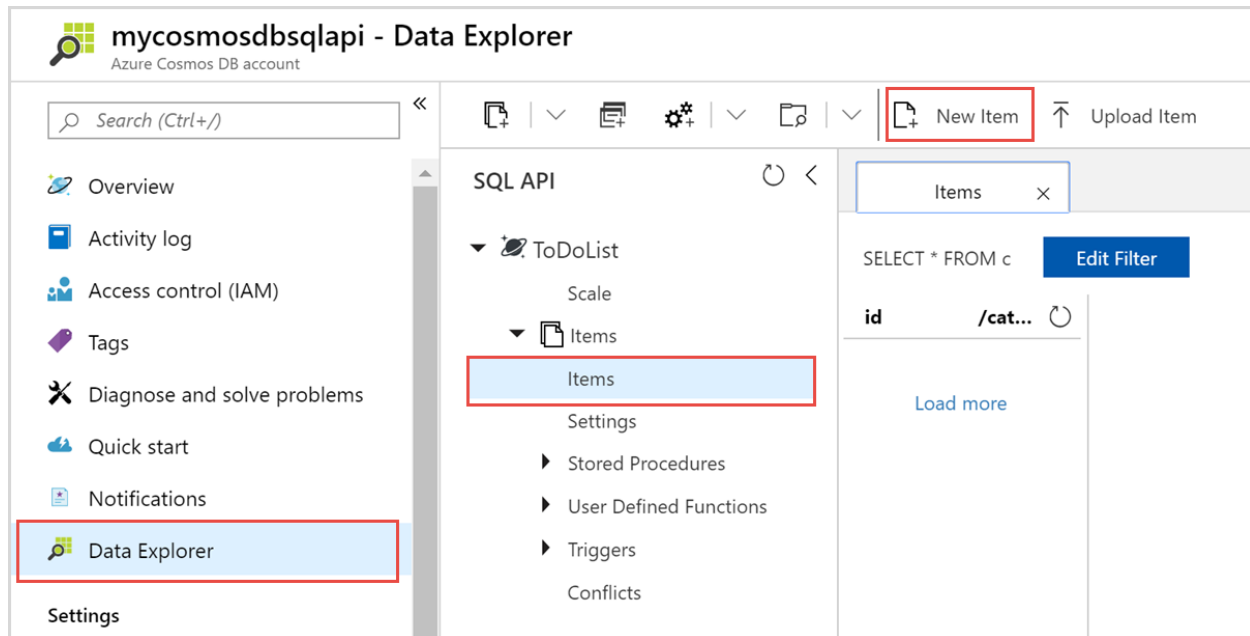
| /category |

Unique keys ⓘ

+ Add unique key

∨ **Advanced**
☐ My partition key is larger than 100 bytes

Analytical store ⓘ

| OK |

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

Step 1—Expand data under data explorer tab, click on data- items- new item
Step 2—start writing json code in document

5. Explain major difference between databases & datawarehouses ?

Database	DATA warehouse
1. Contains Current data	1. Contains Historical Data
2. Useful in running business	2. Useful in analyzing business
3. Based on entity relationship model	3. Based on star snowflake
4. Primitive and highly detailed data	4. Summarized and consolidated Data
5. 100MB-1GB	5. 100GB-1TB
6. High Performance	6. Flexible but not fast
7. All Bank transactions made by customer	7. Bank transactions made by user at particular time

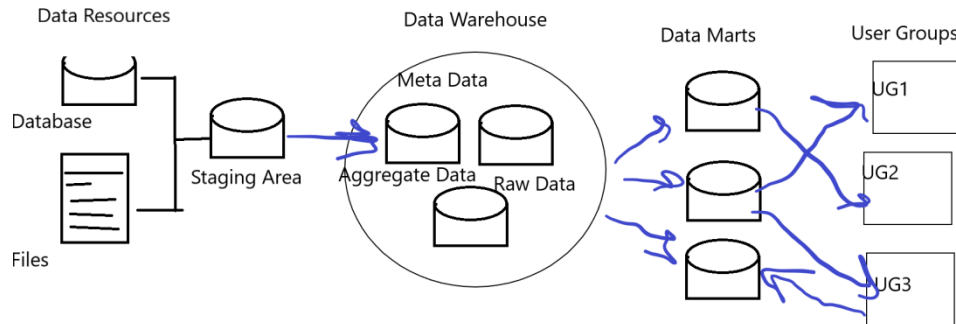**6. Explain the architecture of datawarehouses ?**

Fig 1: Data Warehouse Architecture

7. Explain what are Datamarts & how different from DATABASES ? And mention the types of Datamarts too

Database is Structured collection of data. It can be anything from list of names in a text file, to a relational database. It is commonly confused with the database management system .

Data mart = subset of the data warehouse structured to allow easy user access. Can be a subset of the accessible data, oraganised by department/domain such as sales relevant data or logistics data. Often contains [partial] aggregations.

Types Of data mart—

1 dependant data mart- the data is first extracted from OLTP systems then populated to central data ware house
From DWH, data travels to data mart.

2. independant data mart- the data is directly received from source
Suitable for small organizations or groups

3. hybrid data mart- the data is fed from both **OLTP** and **Data warehouse.**

8. Explain OLAP & OLTP with examples ?

OLAP	OLTP
Consists of historical data from various Databases.	Consists only operational current data.
It is subject oriented. Used for Data Mining, Analytics, Decision making,etc.	It is application oriented. Used for business tasks.
The data is used in planning, problem solving and decision making.	The data is used to perform day to day fundamental operations.
Large amount of data is stored typically in TB, PB	TB, PB The size of the data is relatively small as the historical data is archived. For ex MB, GB
100MB-1GB	100GB-1TB
Only read and rarely write operation.	Only read and rarely write operation.
Transaction details, movie ticket app, Bus booking app	Data analysis, financial analysis, budgeting, and forecast planning

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

BI is a process of transforming raw/operational data into useful information for business analysis
BI based on data warehouse technology extracts information from company's operational systems
The data is cleaned and loaded into data warehouse
Since data is credible it can be used for business insights.
Analyzing this data and using it for changing or setting new strategies, business gain, marketing and keeping up with the trends, using it to train machine learning models and using for predictions helps business to grow. This is how BI helps business to take intelligent decisions based on data.

10.Explain how ETL works with Datawarehouses ?

ETL – Extract, Transform, Load
ETL is process of extracting data from various resources, transforming this data to meet your requirement and then loading it into target data warehouse.
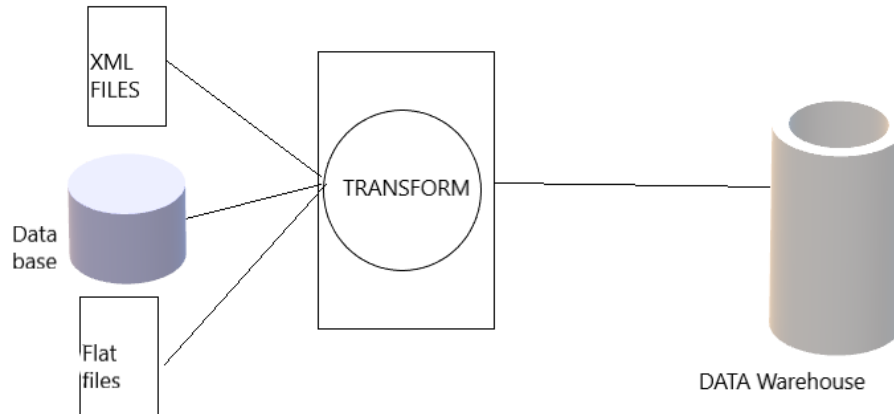
Extraction:
The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML, and flat files into the staging area.

Transformation:
 Filtering, Cleaning, Joining, Splitting, Sorting

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.