1. Explain various Difference between SQL & NoSQL DBs ?

| SQL | NoSQL |
| --- | --- |
| RELATIONAL DATABASE MANAGEMENT SYSTEM (RDBMS) | Non-relational or distributed database system. |
| These databases have fixed or static or predefined schema. | They have dynamic schema. |
| These databases are not suited for hierarchical data storage. | These databases are best suited for hierarchical data storage. |
| These databases are best suited for complex queries | These databases are not so good for complex queries |
| SQL databases are vertically scalable. | NoSQL databases are horizontally scalable. |
| SQL databases display data in form of tables so it is known as table-based database. | NoSQL databases display data as collection of key-value pair, documents, graph databases or wide-column stores. |

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

(i) Flexible Data Model:

NoSQL databases are highly flexible as they can store and combine any type of data, both structured and unstructured, unlike relational databases that can store data in a structured way only.

(ii) Evolving Data Model :

NoSQL databases allow you to dynamically update the schema to evolve with changing requirements while ensuring that it would cause no interruption or downtime to your application.

(iii) Elastic Scalability:

NoSQL databases can scale to accommodate any type of data growth while maintaining low cost.

(iv) High Performance:

NoSQL databases are built for great performance, measured in terms of both throughput (it is a measure of overall performance) and latency (it is the delay between request and actual response).

(v) Open-source:

NoSQL databases don’t require expensive licensing fees and can run on inexpensive hardware, rendering their deployment cost-effective.

Inserting Data into MongoDB

> db.COLLECTION\_NAME.insert(document)

Example

> db.users.insert({

... \_id : ObjectId("507f191e810c19729de860ea"),

... title: "MongoDB Overview",

... description: "MongoDB is no sql database",

... by: "xyz",

... url: "http://www.xyz.com",

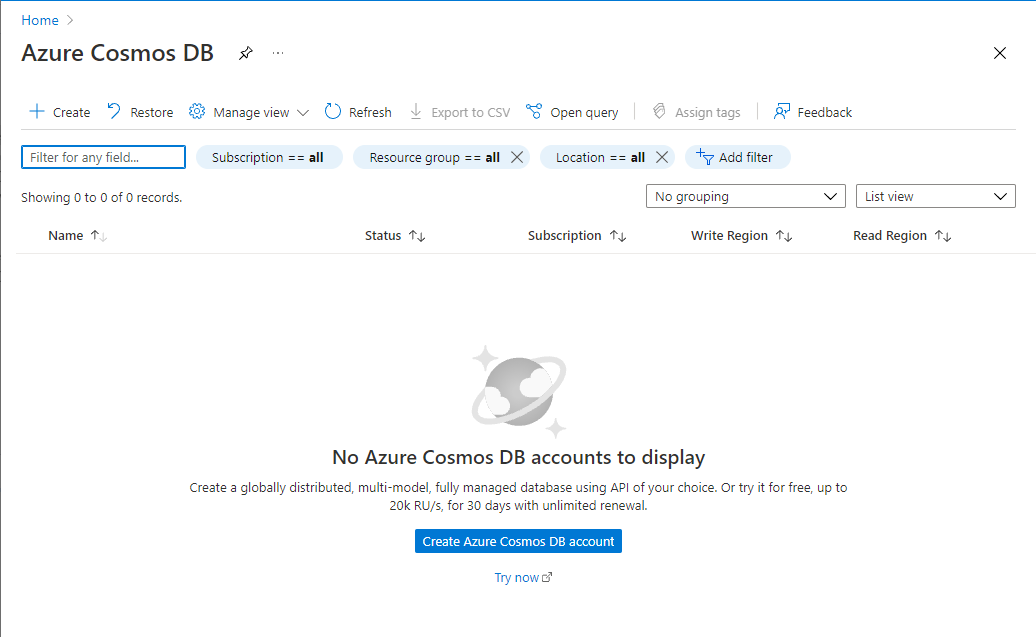
... tags: ['mongodb', 'database', 'NoSQL'],

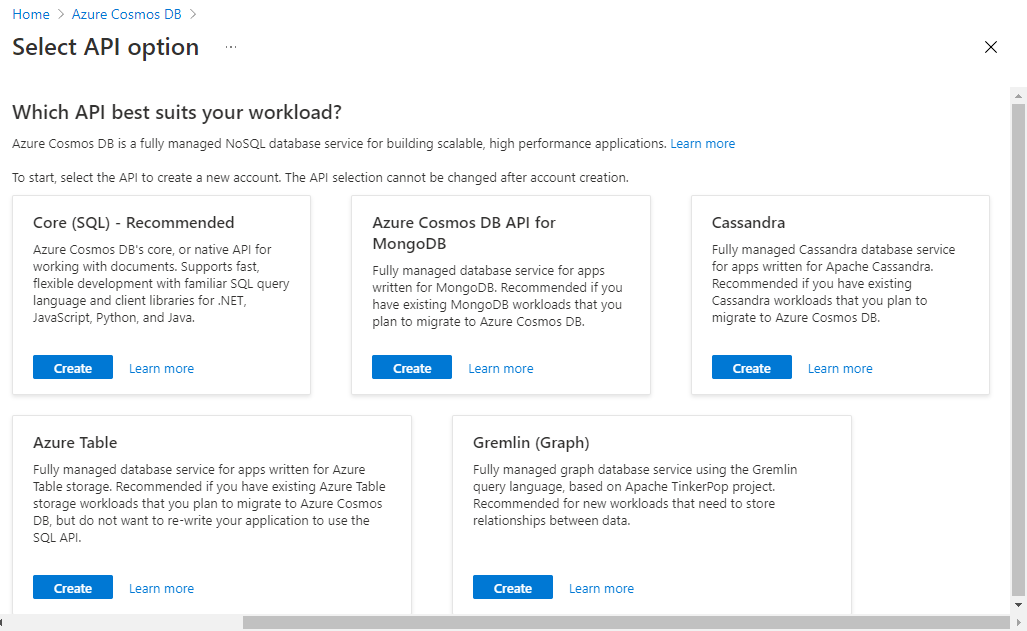
... likes: 100

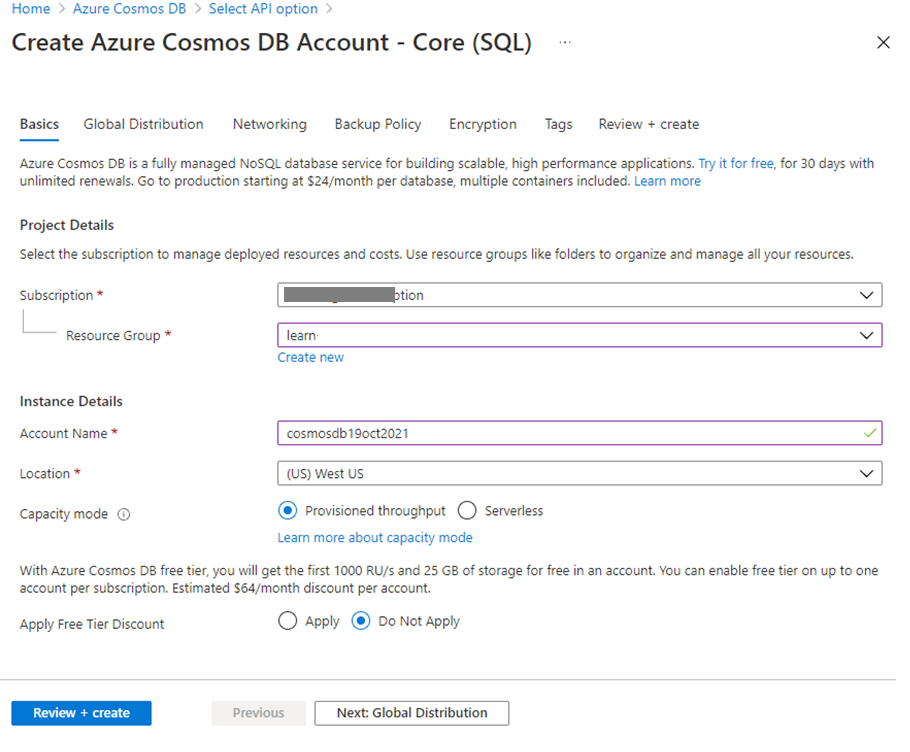
... })

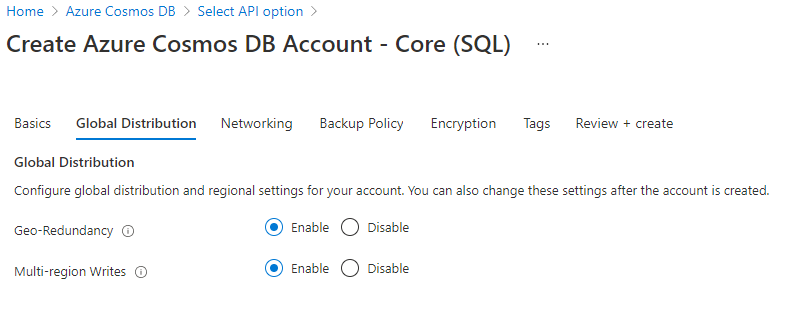
WriteResult({ "nInserted" : 1 })

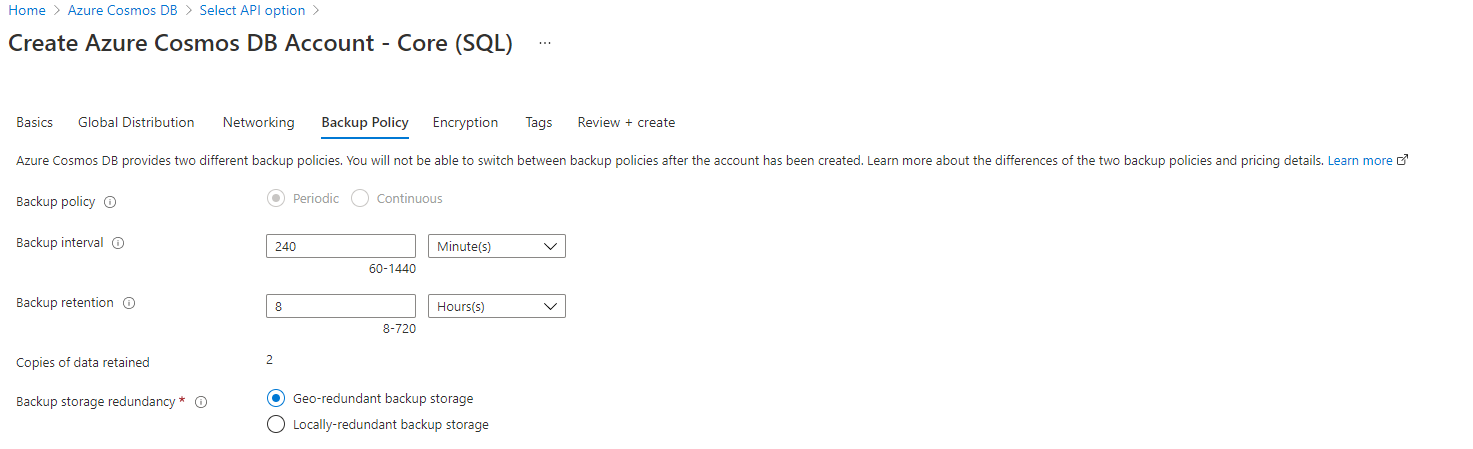
1. Explain the steps - how COSMOS DB can be creatd with screens ?

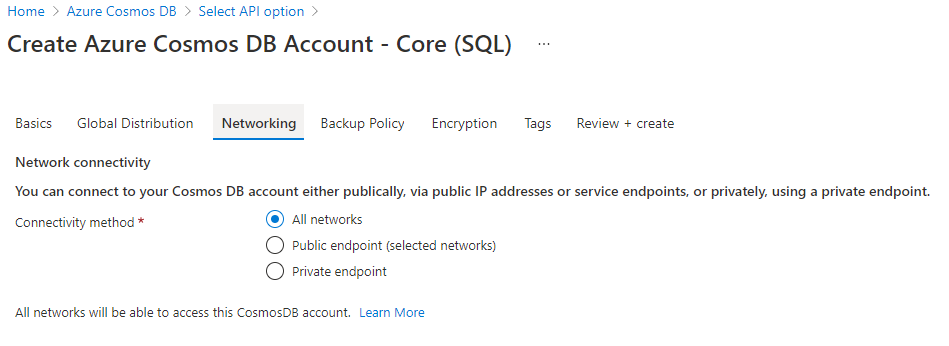


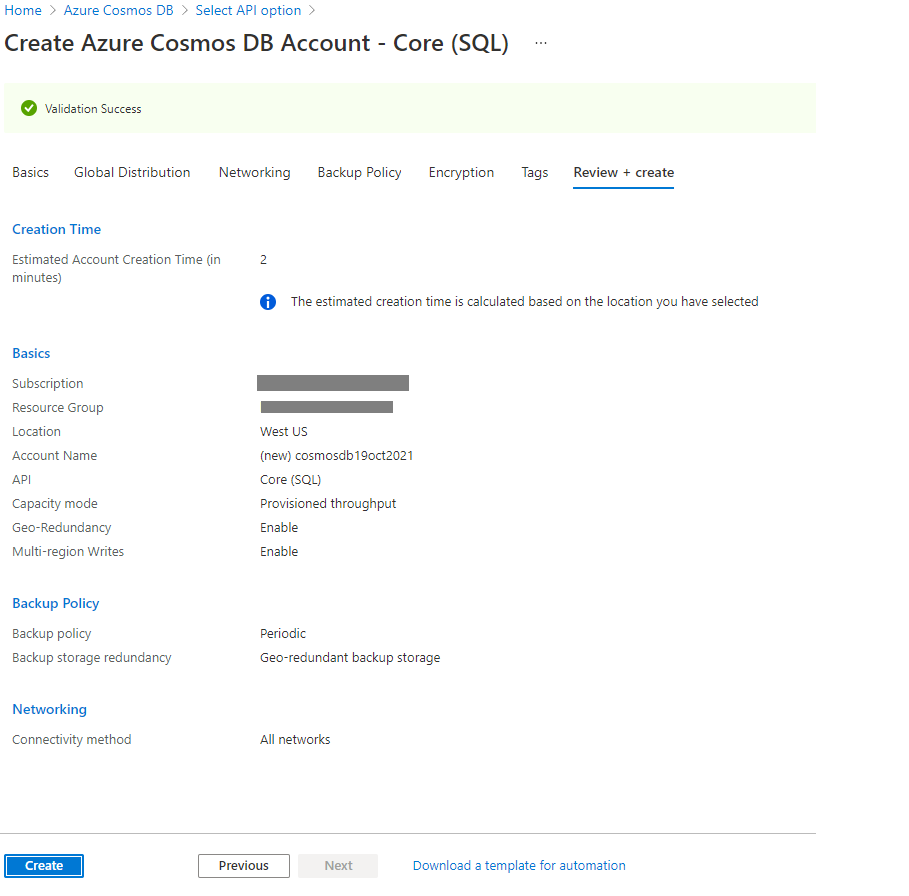


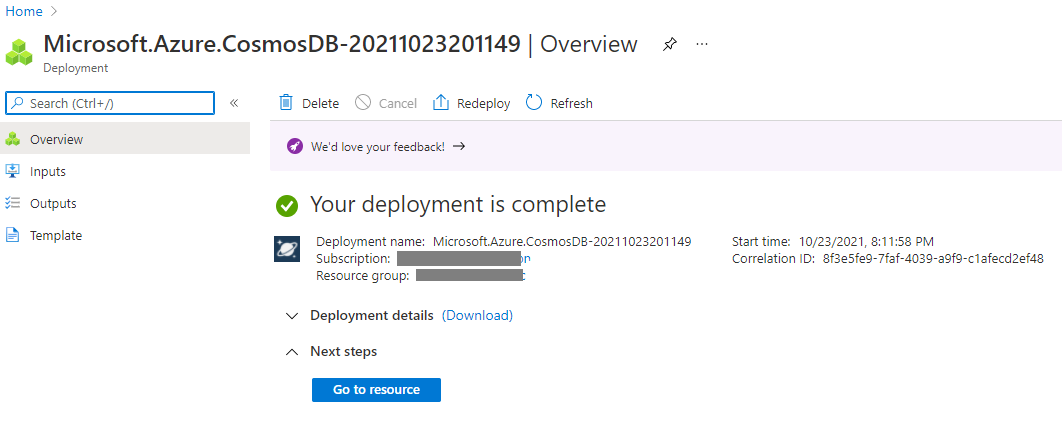












1. Explain how to write JSON query in COSMOS DB ?

JSON

{

"id": "AndersenFamily",

"lastName": "Andersen",

"address": {

"state": "WA",

"county": "King",

"city": "Seattle"

},

"creationDate": 1431620472,

"isRegistered": true

}

SQL

SELECT f.address.state, f.address.city

FROM Families f

WHERE f.id = "AndersenFamily"

JSON

[{

"state": "WA",

"city": "Seattle"

}]

5.Explain major difference between databases & datawarehouses ?

|  |  |
| --- | --- |
| Database  It supports operational processes. | Data Warehouse  It supports analysis and performance reporting. |
| Capture and maintain the data. | Explore the data. |
| Current data. | Multiple years of history. |
| Data is balanced within the scope of this one system. | Data must be integrated and balanced from multiple system. |
| Data is updated when transaction occurs. | Data is updated on scheduled processes. |
| Data verification occurs when entry is done. | Data verification occurs after the fact. |
| 100 MB to GB. | 100 GB to TB. |
| ER based. | Star/Snowflake. |
| Application oriented. | Subject oriented. |
| Primitive and highly detailed. | Summarized and consolidated. |
| Flat relational. | Multidimensional. |

6. Explain the architecture of datawarehouses ?

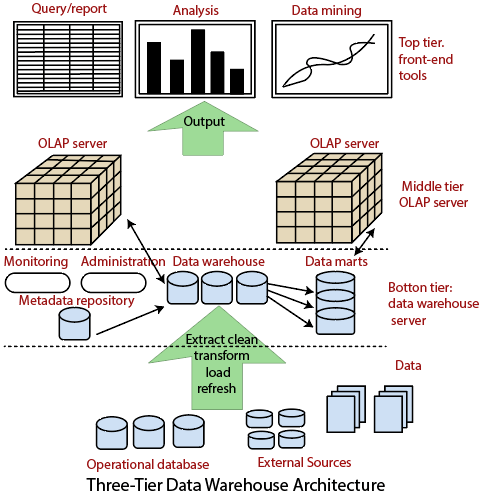
**Bottom Tier** − The bottom tier of the architecture is the data warehouse database server. It is the relational database system. We use the back end tools and utilities to feed data into the bottom tier. These back end tools and utilities perform the Extract, Clean, Load, and refresh functions.

**Middle Tier** − In the middle tier, we have the OLAP Server that can be implemented in either of the following ways.

By Relational OLAP (ROLAP), which is an extended relational database management system. The ROLAP maps the operations on multidimensional data to standard relational operations.

By Multidimensional OLAP (MOLAP) model, which directly implements the multidimensional data and operations.

**Top-Tier** − This tier is the front-end client layer. This layer holds the query tools and reporting tools, analysis tools and data mining tools.



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

A data mart is a subset of a data warehouse oriented to a specific business line. Data marts contain repositories of summarized data collected for analysis on a specific section or unit within an organization, for example, the sales department.

Difference

Data warehouses are built to serve as the central store of data for the entire business, whereas a data mart fulfills the request of a specific division or business function. Because a data warehouse contains data for the entire company, it is best practice to have strictly control who can access it. Additionally, querying the data you need in a data warehouse is an incredibly difficult task for the business. Thus, the primary purpose of a data mart is to isolate or partition a smaller set of data from a whole to provide easier data access for the end consumers.

Types of Data marts

**1. Dependent Data Marts**

A dependent data mart is created from an existing enterprise data warehouse. It is the top-down approach that begins with storing all business data in one central location, then extracts a clearly defined portion of the data when needed for analysis.

To form a data warehouse, a specific set of data is aggregated (formed into a cluster) from the warehouse, restructured, then loaded to the data mart where it can be queried. It can be a logical view or physical subset of the data warehouse:

* **Logical view** - A virtual table/view that is logically—but not physically—separated from the data warehouse
* **Physical subset** - Data extract that is a physically separate database from the data warehouse

Granular data—the lowest level of data in the target set—in the data warehouse serves as the single point of reference for all dependent data marts that are created.

**2. Independent Data Marts**

An independent data mart is a stand-alone system—created without the use of a data warehouse—that focuses on one subject area or business function. Data is extracted from internal or external data sources (or both), processed, then loaded to the data mart repository where it is stored until needed for business analytics.

Independent data marts are not difficult to design and develop. They are beneficial to achieve short-term goals but may become cumbersome to manage—each with its own as business needs expand and become more complex.

3. **Hybrid Data Marts**

A hybrid data mart combines data from an existing data warehouse and other operational source systems. It unites the speed and end-user focus of a top-down approach with the benefits of the enterprise-level integration of the bottom-up method.

8. Explain OLAP & OLTP with examples ?

OLAP

Online Analytical Processing, a category of software tools which provide analysis of data for business decisions. OLAP systems allow users to analyze database information from multiple database systems at one time.

OLTP

Online transaction processing shortly known as OLTP supports transaction-oriented applications in a 3-tier architecture. OLTP administers day to day transaction of an organization.

## **Example of OLAP**

Any Datawarehouse system is an OLAP system. Uses of OLAP are as follows

* A company might compare their mobile phone sales in September with sales in October, then compare those results with another location which may be stored in a sperate database.
* Amazon analyzes purchases by its customers to come up with a personalized homepage with products which likely interest to their customer.

Example of OLTP

An example of OLTP system is ATM center. Assume that a couple has a joint account with a bank. One day both simultaneously reach different ATM centers at precisely the same time and want to withdraw total amount present in their bank account.

However, the person that completes authentication process first will be able to get money. In this case, OLTP system makes sure that withdrawn amount will be never more than the amount present in the bank. The key to note here is that OLTP systems are optimized for****transactional superiority instead data analysis.****

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

The term Business Intelligence (BI) refers to the technologies, applications, strategies, and practices used to collect, analyze, integrate, and present pertinent business information. The entire purpose of Business Intelligence is to support and facilitate better business decisions. BI allows organizations access to information that is critical to the success of multiple areas including sales, finance, marketing, and a multitude of other areas and departments. Effectively leveraging BI will empower your business with increased actionable data, provide great insights into industry trends, and facilitate a more strategically geared decision-making model.

Business intelligence can help companies make better decisions by showing present and historical data within their business context. Analysts can leverage BI to provide performance and competitor benchmarks to make the organization run smoother and more efficiently. Analysts can also more easily spot market trends to increase sales or revenue. Used effectively, the right data can help with anything from compliance to hiring efforts. ****A few ways that business intelligence can help companies make smarter, data-driven decisions:****

* Identify ways to increase profit
* Analyze customer behavior
* Compare data with competitors
* Track performance
* Optimize operations
* Predict success
* Spot market trends
* Discover issues or problems

1. Explain how ETL works with Datawarehouses ?

**1.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.

**2.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).

**3.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.

