SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NAAC – UGC with 'A' Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME : 19CS508 - BIG DATA ANALYTICS

III YEAR / V SEMESTER

Unit 5 - NOSQL DATA MANAGEMENT FOR BIG DATA AND VISUALIZATION

Topic 1 : NoSQL Databases: Schema-less Models : Increasing Flexibility for Data Manipulation

What is a NoSQL database?

- When people use the term “NoSQL database,” they typically use it to refer to any non-relational database.
- Some say the term “NoSQL” stands for “non-SQL” while others say it stands for “not only SQL.” Either way, most agree that NoSQL databases store data in a more natural and flexible way.
- NoSQL, as opposed to SQL, is a database management approach, whereas SQL is just a query language, similar to the query languages of NoSQL databases.

Cont..

Why use NoSQL?

- Customer experience has quickly become the most important competitive differentiator and ushered the business world into an era of monumental change. As part of this revolution, enterprises are interacting digitally – not only with their customers, but also with their employees, partners, vendors, and even their products – at an unprecedented scale.
- This interaction is powered by the internet and other 21st century technologies – and at the heart of the revolution of NoSQL are a company's big data, cloud, mobile, social media, and IoT applications.

1. these applications different from legacy enterprise applications like ERP, HR, and financial accounting etc.,

- Today's web, mobile, and IoT applications share one or more (if not all) of the following characteristics.
- They need to: Support large numbers of concurrent users (tens of thousands, perhaps millions)
- Deliver highly responsive experiences to a globally distributed base of users
- Be always available – no downtime
- Handle semi- and unstructured data
- Rapidly adapt to changing requirements with frequent updates and new features

Cont..

- Global 2000 enterprises are rapidly embracing NoSQL databases to power their mission-critical applications:
- **Tesco**, Europe's No. 1 retailer, deploys NoSQL for e-commerce, product catalog, and other applications
- **Marriott** deploys NoSQL for its reservation system that books $38 billion annually
- **Gannett**, the No. 1 U.S. newspaper publisher, uses NoSQL for its proprietary content management system, Presto
- **GE** deploys NoSQL for its Predix platform to help manage the Industrial Internet

Cont..

- **NoSQL databases are generally classified into four main categories:**

1.Document databases: These databases store data as semi-structured documents, such as JSON or XML, and can be queried using document-oriented query languages.

2.Key-value stores: These databases store data as key-value pairs, and are optimized for simple and fast read/write operations.

3.Column-family stores: These databases store data as column families, which are sets of columns that are treated as a single entity. They are optimized for fast and efficient querying of large amounts of data.

4.Graph databases: These databases store data as nodes and edges, and are designed to handle complex relationships between data.

Key-value database	Document database	Wide-column database	Graph database
Basic support for data visualization	Basic support for data visualization	Basic support for data visualization	Basic support for data visualization
Basic authentication, limited access control	Role-based access control, encryption at-rest, encryption in transit, and in use	Role-based access control (RBAC), encryption at rest	Role-based access control (RBAC), encryption at rest, in transit
Limited support for online data archival	Automatic online data archival	Limited support for online data archival	Limited support for online data archival
Basic key-based data retrieval	Full-text search, vector search	Limited search capabilities	Limited search capabilities
Basic CRUD operations, limited data manipulation capabilities	Advanced query and data manipulation capabilities	Data manipulation and CRUD capabilities	Supports advanced graph traversal and manipulation operations
Stores any type of data	Stores JSON/BSON data	Columns and rows matrix	Nodes, edges, and relationships
Limited support for complex analytics	Suited for time-series, IoT analytics, real-time analytics	Suited for time-series data, IoT analytics	Well-suited for graph analytics
Horizontal scalability and eventual consistency	Horizontal scalability and eventual consistency	Horizontal scalability and eventual consistency	Horizontal scalability and eventual consistency
Simple indexes	Indexes on fields	Secondary indexes	Indexes on nodes/edges
Limited by key	Rich querying capability	Limited by columns	Specialized graph query
Schemaless	Flexible schema	Flexible schema	Flexible schema
Simple key-value pairs	JSON/BSON documents	Column and row families	Nodes, edges, and relationships

Cont..

Key Features of NoSQL:

1.Dynamic schema: NoSQL databases do not have a fixed schema and can accommodate changing data structures without the need for migrations or schema alterations.

2.Horizontal scalability: NoSQL databases are designed to scale out by adding more nodes to a database cluster, making them well-suited for handling large amounts of data and high levels of traffic.

3.Document-based: Some NoSQL databases, such as MongoDB, use a document-based data model, where data is stored in a schema-less semi-structured format, such as JSON or BSON.

4.Key-value-based: Other NoSQL databases, such as Redis, use a key-value data model, where data is stored as a collection of key-value pairs.

5.Column-based: Some NoSQL databases, such as Cassandra, use a column-based data model, where data is organized into columns instead of rows.

Ex: Graph Databases: Examples – Amazon Neptune, Neo4j

Key value store: Examples – Memcached, Redis, Coherence

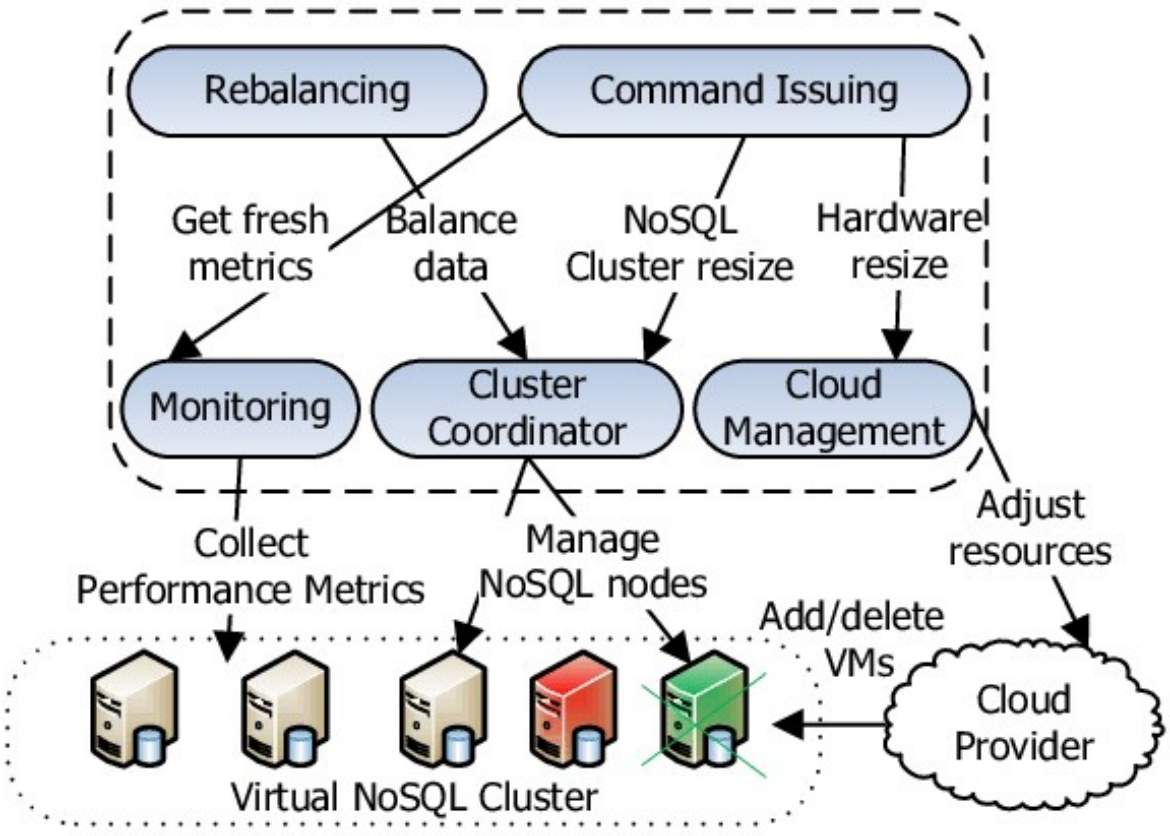
Column: Examples – Hbase, Big Table, Accumulo

Document-based: Examples – MongoDB, CouchDB, Cloudant

Cont..

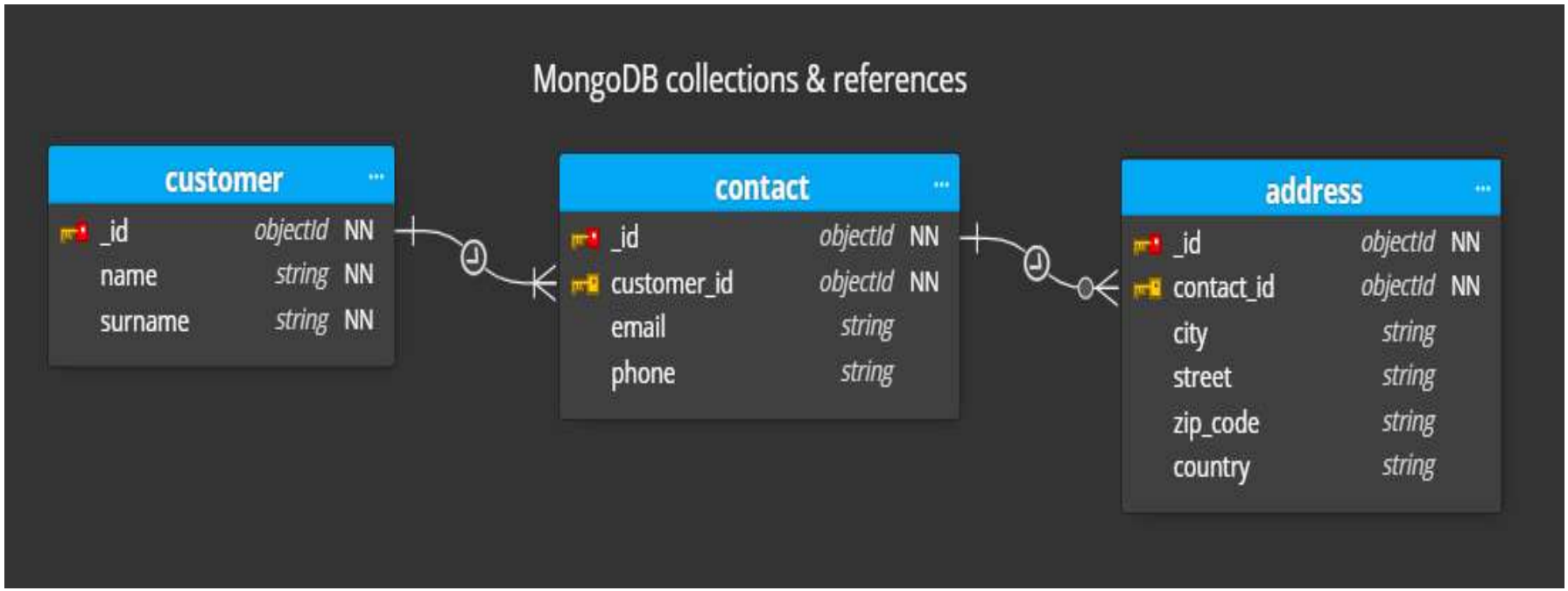
6. **Distributed and high availability:** NoSQL databases are often designed to be highly available and to automatically handle node failures and data replication across multiple nodes in a database cluster.
7. **Flexibility:** NoSQL databases allow developers to store and retrieve data in a flexible and dynamic manner, with support for multiple data types and changing data structures.
8. **Performance:** NoSQL databases are optimized for high performance and can handle a high volume of reads and writes, making them suitable for big data and real-time applications.

Cont..



Architecture of our Cloud-based NoSQL

Cont..


MongoDB collections & references
customer
_id objectId NN
name string NN
surname string NN
contact
_id objectId NN
customer_id objectId NN
email string
phone string
address
_id objectId NN
contact_id objectId NN
city string
street string
zip_code string
country string

Activity

Advantages

- High scalability
- Flexibility
- High availability
- Performance
- Cost-effectiveness
- **Disadvantages**
- Lack of standardization
- Lack of ACID compliance
- Open-source

Assessment 1

1. List out the advantages of No SQL
   a)_____
   b)_____
   c)_____
   d)_____

Identify the disadvantages of No SQL
   a)_____
   b)_____
   c)_____
   d)_____

REFERENCES

1. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012.
2. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", Morgan Kaufmann/Elsevier Publishers, 2013
3. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", Wiley publishers, 2015.

THANK YOU