Certified MongoDB Professional Sample Material

V-Skills Certifications

A Government of India
&
Government of NCT Delhi Initiative

V-Skills

1. INTRODUCTION TO MONGODB

MongoDB is an open-source document database which is a NoSQL database developed in c++ programming language. MongoDB is a cross-platform, document oriented database that provides, high performance, high availability, and easy scalability. MongoDB works on concept of collection and document.

NoSQL database refers to non-relational, distributed, flexible, and scalable database paradigm which grew in prominence due to the need of storage and processing of unstructured data like e-mail, video and especially social media related information. It is widely used by social media websites.

1.1. MongoDB Evolution

Development of MongoDB began in 2007, when the 10gen, a company was building a PaaS (Platform-as-a-service) software similar to Windows Azure or Google App Engine. 10gen focused on the database that became MongoDB as users were more interested in it.

In 2009, MongoDB was open sourced as a stand-alone product with an AGPL license. 10gen continues to sponsor the database's development as an open source project and the source code is publicly available.

In March 2010, MongoDB version 1.4 was launched and was considered production ready. The latest stable version, 2.4.9, was released on January 10, 2014.

The project's roadmap continues to be determined by the needs of its user community

1.2. Need for Document-Oriented Databases

Relational databases are used to store and retrieve data. But, with the growth of internet, the number of reads and writes a typical application needed to perform grew rapidly needing scaling. Traditional RDBMSs were hard to scale as SQL operation or transaction spanning multiple nodes doesn't scale well. Thus, leading to many companies abandoning traditional RDBMSs for "noSQL" data stores which use low cost compute or storage boxes for scaling.

Document-Oriented Databases provide more scalability and functionality as, the system is aware of the stored object structure as, documents can contain varying named attributes of different data types (integers, strings), objects can grouped into collections and the system has a query mechanism to search collections for objects with particular attribute values. Document stores can also partition the data over many machines, replicate data for automatic recovery and persist the data. Examples of document stores are SimpleDB, CouchDB, MongoDB, and Dynamo.

The advantages of using document-oriented database are:

- ✓ Documents (i.e. objects) correspond to native data types in many programming languages.
- ✓ Embedded documents and arrays reduce need for expensive joins.
- ✓ Dynamic schema supports fluent polymorphism.

Document-oriented databases have special characteristics compared to RDBMS as

- ✓ Objects stored as documents - Object serialization is in-built hence, no serialization is to be applied.
- ✓ Complete object access - Entire objects can be read & written at once without group of inserts or stored procedures.
- ✓ Documents are atomic - Documents are independent of each other thus, improving performance and decreasing concurrency side effects.
- ✓ Standardized Data Layout - Documents are usually described by standardized JSON or XML which are self-describing.
- ✓ Schema Independence - Documents being schema free provides flexibility for evolving system without forcing the existing data to be restructured.

Key Features of MongoDB

High Performance - MongoDB provides high performance data persistence. In particular,

- ✓ Support for embedded data models reduces I/O activity on database system.
- ✓ Indexes support faster queries and can include keys from embedded documents and arrays.

High Availability - To provide high availability, MongoDB's replication facility, called replica sets, provide

- ✓ Automatic Failover.
- ✓ Data Redundancy.

A replica set is a group of MongoDB servers that maintain the same data set, providing redundancy and increasing data availability.

Automatic Scaling - MongoDB provides horizontal scalability as part of its core functionality.

- ✓ Automatic sharding distributes data across a cluster of machines.
- ✓ Replica sets can provide eventually-consistent reads for low-latency high throughput deployments.