Chapter 8
NoSQL Databases

Ganesh Chandra Deka
Government of India, India

ABSTRACT

NoSQL databases are designed to meet the huge data storage requirements of cloud computing and big data processing. NoSQL databases have lots of advanced features in addition to the conventional RDBMS features. Hence, the "NoSQL" databases are popularly known as "Not only SQL" databases. A variety of NoSQL databases having different features to deal with exponentially growing data-intensive applications are available with open source and proprietary option. This chapter discusses some of the popular NoSQL databases and their features on the light of CAP theorem.

INTRODUCTION

"NoSQL" is a breed of databases that are appearing in response to the limitations of existing relational databases (RDBMS). NoSQL databases are capable of handling large amounts of structured, unstructured, semi-structured and hybrid data with an amazing performance at reduced complexity and cost.

The foundation of NoSQL movement was laid by the following three major research papers:

1. Google Bigtable
2. Dynamo paper of Amazon (Gossip protocol, Distributed key-value data store and Eventual consistency)
3. CAP Theorem

Table 1 shows the chronology of the NoSQL movement (Noller, 2013), (Vasiliev, 2013).

DOI: 10.4018/978-1-4666-5864-6.ch008

Table 1. Chronology of Development of NoSQL

Year	Development
1998	**Carlo Strozzi** introduced the term NoSQL to name his lightweight, open-source relational database that does not render the standard SQL interface.
2000	Graph database Neo4j introduced
2004	Google Bigtable project started. The first paper published in 2006.
2005	CouchDB lunched
2007	Research paper on Amazon Dynamo released
2008	Facebook's open sources the Cassandra project started. Project Voldemort started
2009	The document database MongoDB started as a part of an open source cloud computing stack. The first standalone version released.
2009	The term NoSQL reintroduced in early 2009. Lots of commercial and open source NoSQL developed and floated in the market by various vendors and communities.

This chapter discusses about the NoSQL database features in general and features of mostly used 10 NoSQL in the light of CAP theorem (http://nosql-database.org/, 2011). Apart from these 10 NoSQL databases Microsoft Azure (SQL based) and IBM DB2 is also discussed with a focus on big data. Sufficient references are given for the benefit of readers. The important technical terms related to NoSQL are explained at the end of the chapter for ready reference.

NoSQL FEATURES

NoSQL databases provide:

- Scalability (can be scaled horizontally)
- High availability
- Optimized resource allocation and utilization
- Virtually unlimited data store capacity
- Multitenancy

Features of NoSQL are briefed below.

1. **High Scalability:** NoSQL does not support "Join" because joins makes databases unscalable. They are capable of handling large amounts of growing data. NoSQL databases use the concept of distributing database over multiple hosts for dealing with increasing load. Commodity hardware can be used cost effectively using NoSQL in the cloud or virtualized environment. NoSQL databases are designed to enlarge transparently for taking advantage of freshly added nodes using lower cost commodity hardware.
2. **Performance:** A growing number of people are joining the cloud for storing their data on different remote disks. More and more commodity servers were added by every passing day to enhance the performance of NoSQL for linearly balancing the loads at a minimal cost at the same time keeping performance higher to meet user's expectations. For enhanced performance NoSQL have:
   - No complex transaction support.
   - No constrains support.
3. **Availability:** NoSQL uses replication, i.e. storing multiple copies of data across the cluster and even at various data centers for ensuring high availability and disaster recovery. A properly managed NoSQL database is capable of continuous operations without ever going down. Another very important