Ethics Pledge

Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature: Haodong Zhao Date: Mar 11th. 2019

Please note that assignments in this class may be submitted to

www.turnitin.com, a web-based anti-plagiarism system, for an evaluation of their originality.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Reading review**

**A Brief Introduction to Apache Cassandra**

This article provides an overview of non-relational databases and Apache Cassandra and describes many of the top Cassandra concepts and how it operates and manages data.

Many applications now have high database requirements that go beyond the capabilities of traditional relational databases. This resulted in a non-relational database category. In order to meet the requirements of modern applications, the new databases (NoSQL) needs to be weighted according to the CAP theorem.

Apache Cassandra is a large-scale open source non-relational database developed by Facebook. It has following key features and benefits:

1. Large scale scalable architecture
2. Ubiquitous design
3. Linear expansion performance
4. Continuous availability
5. Transparent fault detection and recovery
6. Flexible dynamic data model
7. Powerful data protection
8. Adjustable data consistency
9. Multiple data center replication
10. Data compression
11. CQL (Cassandra Query Language)

The top use cases of Apache Cassandra:

1. IoT applications
2. Product catalogs and retail applications
3. User activity and integrated monitoring
4. Messaging
5. Social media analytics and recommendation engine
6. Other time-series based applications

The article introduced the architecture of Apache Cassandra. Instead of using traditional master-slave equipment or a manual and difficult-to-maintain shard design, Cassandra has an elegant, easy-to-set and easy-to-maintain ‘no-ring’ architecture (all nodes play the same role, without the concept of a master node, all nodes communicate with each other through a distributed and extensible protocol called ‘gossip’).

The article also describes the following operations of Cassandra:

1. Writing and reading data
2. Data distribution and replication
3. Multi-Data center and cloud support

Cassandra is a wide-row storage database that uses highly denormalized models designed to capture and query data in an efficient manner. And it provides a familiar security paradigm for anyone from a relational database. And Cassandra offers many backup options to ensure data protection and recovery in the event of data loss to ensure data consistency.