CONCEPTUAL ARCHITECTURE OF HBASE REPORT

EECS 4314, York University

15th October 2018

Authors:

Akbar Khan (khanakb@my.yorku.ca)

Andrew Laniak-Fuoco (alaniakfuoco@gmail.com)

Anji Tong (……)

Asma Hassan (asmafrin12@gmail.com)

Gilbert Andze Tsoungui (gilberttsoungui@gmail.com)

Rabia Ejaz (rabia93@my.yorku.ca)

ABSTRACT:

The purpose of this report is to describe the conceptual architecture of HBase without discussing the details of its implementation. HBase is Open Source Column-oriented distributed database management system that runs on top of Hadoop Distributed File System (HDFS). The first part of report provides general introduction of HBase followed by the high level conceptual architecture and major components. The interaction between subsystems is also discussed in detail with appropriate figures. The second part of report provides the data flow from one component to another and concurrency present in HBase. Architectural styles and design patterns are also identified with the use of UML diagrams and use cases. The last part of report discusses the external interfaces of system, small number of essential use cases, and the noteworthy lessons that we learned while doing this report.

Table of Contents:

1. Purpose of Report:

This report was prepared for course Advance Software Engineering (EECS 4314), in fall 2018. The goal of this report is to describe the high level conceptual architecture of HBase, its major components and relationship of its subsystems using UML diagrams. Information derived for this report was acquired from various resources including HBase documentations, journal articles, academic books, and lecture notes for EECS 4314 provided by Dr. Zhen Ming (Jack) Jiang. Few documents contained conflicting information, but with detailed research, we gathered the information from most authentic resources.

2. Introduction:

HBase is one of the most popular non-rational (NoSQL) databases built on the top of Hadoop distributed file system (HDFS). It is open source distributed database modelled after Google's Bigtable. HBase is categorized as NoSQL database because it is schema-less, and it is good for both semi-structured and structed data. Since HBase is built on top of HDFS, it integrates very well with MapReduce framework.

2.1. Storage Mechanism in HBase

HBase is column-oriented database. The data is stored in tables that are sorted by rows. Each row is collection of column families and each column family can have any number of columns. These columns can be further divided into different versions. Each cell of the table ({row, column, version} tuple) has its own Meta data.

2.2. Features of HBase

It provides easy to use java API for clients.

It provides automatic failover support for region servers and HMaster.

It supports distributed storage and data replication across clusters.

It provides consistent read and write operations.

2.3. Applications of HBase

HBase is perfect for high-scale and real-time applications. It is used whenever there is a need to serve large amount of data and write heavy applications. It is also suitable when there is a need to provide random and real time access to available data, and a fault tolerant system is required.

3. Architecture of HBase

4. Data Flow

5. Concurrency

6. Architecture Style

7. Important Abstractions, Patterns, Classes, and Data structures

8. Diagrams

9. Dependency Diagram

10. External Interface

11. Use Cases

12. Data Dictionary

Term	Description
Open Source	
Distributed Database	

13. Naming Conventions

Abbreviations	Meaning
API	Application programming interface

14. Conclusion

15. Lesson Learned

16. References

1. Apache HBase Explained in 5 Minutes or Less. (2017, July 18). Retrieved October 09, 2018, from
   https://www.credera.com/blog/technology-insights/java/apache-hbase-explained-5-minutes-less/
2. T. HBase Overview. Retrieved October 09, 2018, from
   https://www.tutorialspoint.com/hbase/hbase_overview.htm

3. Pino, D. (2012, October 31). Introduction to HBase and NoSQL systems. Retrieved October 09, 2018, from https://blogs.igalia.com/dpino/2012/10/31/introduction-to-hbase-and-nosql-systems/

4. HBase Architecture, Data Flow, and Use cases. (n.d.). Retrieved from https://www.guru99.com/hbase-architecture-data-flow-usecases.html

Data Dictionary: Include a glossary that briefly defines all the key terms used in your architecture, giving when appropriate, the "type" of the item being explained. (Everyone will write their key terms here)

Open Source:

Distrubited Database:

Bigtable:

Bigtable is a distributed storage system that handles very large size of structed data.

Naming Conventions: List any naming conventions used in the described architecture. Explain any abbreviations that you use. (Everyone will write abbreviations here if they use any)

API

Lessons Learned: Document any noteworthy lessons: things you would do differently or things you wished you knew ahead of time. (Rabia)

Conceptual Architecture of HBase, Software-Group

References: List any documents that your reader may wish to or need to read in conjunction with your report. Since the report is to be web-readable, include links to references when appropriate. (Everyone, Please use IEEE format)