

Big Data Technologies - CSP 554
Assignment 10 - Apache Hbase
Prabhu Avula | A20522815
Illinois Institute of Technology, Chicago

I do not have output screenshots as I could not complete the hbase setup. I keep running into errors saying the HBase master node cannot be contacted, even though I made the right changes to the hbase-site.xml file and ensured all other instructions were followed correctly. For some odd reason, I think it is because hbase cannot connect to zookeeper (Though the cluster setup is as directed). I verified the related files to check that I made the necessary changes, so I am still determining what went wrong. I tried troubleshooting it multiple ways. I do know the code for each of the exercises, so I am going to type them here instead. I request that you understand and grade accordingly. Thanks.

Exercise 1:

```
create 'csp554Tbl', 'cf1', 'cf2'  
describe 'csp554Tbl'
```

Exercise 2:

```
put 'csp554Tbl', 'Row1', 'cf1:name', 'Sam'  
put 'csp554Tbl', 'Row2', 'cf1:name', 'Ahmed'  
put 'csp554Tbl', 'Row1', 'cf2:job', 'Pilot'  
put 'csp554Tbl', 'Row2', 'cf2:job', 'Doctor'  
put 'csp554Tbl', 'Row1', 'cf2:level', 'LZ3'  
put 'csp554Tbl', 'Row2', 'cf2:level', 'AR7'
```

Exercise 3:

```
get 'csp554Tbl', 'Row1', 'cf2:level'
```

Exercise 4:

```
get 'csp554Tbl', 'Row2', 'cf1:name'
```

Exercise 5:

```
scan 'csp554Tbl', {LIMIT => 2}
```

Exercise 6: Article Summary

In the paper "A Novel HBase Data Storage in Wireless Sensor Networks," the authors present a solution for efficient data storage in wireless sensor networks (WSNs) using HBase, a distributed, column-oriented database. The motivation behind this study arises from the growing need for real-time data analysis and storage in the Internet of Things (IoT) era, where WSNs generate massive amounts of data that traditional relational databases cannot efficiently handle.

The authors propose a real-time storage model that leverages HBase's capabilities to store and manage data from WSNs. This model includes a double-layer distributed storage architecture, which integrates heterogeneous datasets and optimizes data access and storage efficiency. The paper emphasizes the importance of real-time data storage and processing, as WSNs often require immediate data analysis and decision-making. The authors discuss various optimization strategies to achieve this, such as using Externics primitives to formalize data relationships and a consistent hashing algorithm for efficient data distribution.

The proposed system architecture includes several key components: a distributed caching system, a distributed file system based on Hadoop, and a multi-source data storage buffer. The distributed caching system temporarily stores query results to reduce database access time, while the distributed file system ensures high-throughput data access and fault tolerance. The multi-source data storage buffer processes incoming data before storing it in HBase, enhancing its efficiency.

Additionally, the paper discusses using filters and coprocessors in HBase to improve data retrieval efficiency. Filters allow for selective data access, while coprocessors distribute complex computation tasks across servers, reducing network transmission and client-side processing loads. The authors also explain HBase's physical and logical storage models, highlighting its suitability for unstructured data and real-time applications.

The authors propose a method for dynamically updating HBase cluster nodes to support dynamic data needs. This method efficiently scales the database as data volumes increase, ensuring the system can accommodate growing storage requirements without manual intervention.

The paper concludes with a performance analysis, demonstrating the proposed system's effectiveness in terms of write and read times compared to traditional systems. The results show significant data access and storage efficiency improvements, validating the proposed real-time storage model for WSNs.

Overall, the paper presents a comprehensive solution for addressing the challenges of real-time data storage in WSNs, leveraging HBase's capabilities to provide a scalable, efficient, and reliable system for managing large-scale sensor data.