**HDFS Online Reconfiguration Error Detect Reading and ToDos**

**Task 1: Background**

* **Week 0:** Knowledge ofJava, Maven, Linux are required for this project.
* **Week 1:** Learn how to use Hadoop:
  + Follow the Official Guideline of Hadoop to set up a single-node Hadoop cluster in your local machine.: <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html>
* **Week 2:** Learn how to compile Hadoop from source:
  + Follow this instruction to build a Hadoop program from its source code: <https://pravinchavan.wordpress.com/2013/04/14/building-apache-hadoop-from-source/>
* **Week 3:** Learn how to use Cloudlab (a Cloud server platform)
  + You can first create a Cloudlab account (<https://cloudlab.us/signup.php>) and you should join an existing project. A further explanation should be shown in the meeting.
* **Week 4:** Learn how to use log4j:
  + Follow this instruction to know how to use log4j for logging data: https://www.tutorialspoint.com/log4j/index.htm

**Task 2: Modify the Hadoop code**

* Since the whole project is only related to HDFS, we only need to focus on the code under “hadoop-3.3.1-src/hadoop-hdfs-project/hadoop-hdfs” directory.
* **Week 3:** Learn how to Log data within the HDFS.
  + Modify the DataNode.java in “hadoop-3.3.1-src\hadoop-hdfs-project\hadoop-hdfs\src\main\java\org\apache\hadoop\hdfs\server\datanode”. Try to add a log code in the initialization of the datanode and check whether it works.
* **Week 4:** Learn the official online reconfiguration code of both datanode and namenode
  + **Focus on the usage of** hdfs dfsadmin [-reconfig <namenode|datanode> <host:ipc\_port> <start |status |properties>]: <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html>
  + **Read the source code of both datanode and namenode in this part:** Correlated variable
    - private static final List<String>
    - *RECONFIGURABLE\_PROPERTIES*
    - public String reconfigurePropertyImpl(String property, String newVal)
  + **Explanation:** By adding properties in RECONFIGURABLE\_PROPERTIES and adding a switch case for the properties it should work well for our customize reconfiguration. Besides, it should be notice that the parameters are stored in DFSConfigKeys.java
* **Week 5:** Learn the Dtest of the Hadoop
  + Learn how to use mvn test to test all the testcases for your new HDFS code. It would take up to 5 hours to run a complete test.
  + Focus on the MiniDFSCluster.java in “hadoop-3.3.1-src\hadoop-hdfs-project\hadoop-hdfs\src\test\java\org\apache\hadoop\hdfs.“ It is the simulate cluster that would be used in every testcase.
* **Week 6:** Understand the reconfigure model
  + **Idea:** We first start a socket program([BroadCastClient.java](https://github.com/sinscry/minicluster_hongzhen/blob/main/BroadCastClient.java)) that would hold the new parameter and its value. Then we add a new thread in MiniDFSCluster.java which would ask [BroadCastClient.java](https://github.com/sinscry/minicluster_hongzhen/blob/main/BroadCastClient.java) for a new parameter and value with a random delay time. And then modify the parameter either in datanode or namenode.
  + **Verify:** In order to exclude the result of the incorrect value, we also support initializing the new value before the MiniDFSCluster creates datanodes or namenode. We only need to modify the conf value in the initialization function of MiniDFSCluster.
  + **Github storage:** [**https://github.com/sinscry/minicluster\_hongzhen**](https://github.com/sinscry/minicluster_hongzhen)
* **Week 7:** Understand the result of the former experiment
  + **Hongzhen** Liang’s: Final Result.md in <https://github.com/sinscry/minicluster_hongzhen>
  + **Sixiang Ma’s:** Please check the test list for hdfs in the following link: <https://github.com/StarThinking/ZebraConf/tree/pytest/app_meta/final_hdfs/ultimate>
  + **What you should do:** Keep expanding the testing parameters list for the online reconfiguration, and categorize them all.
    - **There are five categories:**
      * **Error:**
        + **The new parameter you choose is not valid**
        + **The parameter would be used multiple time**
      * **Not Error:**
        + **The parameter would be used multiple times**
        + **The parameter would be used only one time**
        + **The parameter would not be used**