1. **OneDev** Hadoop Infrastructure Setup and Deployment Testing

To deploy a highly available, scalable, and secure Hadoop infrastructure to process and manage large-scale data efficiently, ensuring all the functionalities are working on the cluster.

### **Infrastructure Details**

| **Component** | **Details** |
| --- | --- |
| Hadoop Version |  |
| Cluster Size |  |
| High Availability | Enabled for NameNode and RM |
| Edge Node Setup | Configured using Airflow server |
| Operating System | [OS Version, e.g., RHEL 8 |
| Security | [OS Version, e.g., RHEL 8] |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

### **3. Test Cases and Results**

#### **3.1 Functional Test Cases**

| **Test Case ID** | **Description** | **Expected Outcome** | **Status** |
| --- | --- | --- | --- |
| TC-01 | Verify Hadoop cluster startup | All services (HDFS, YARN) start successfully | Passed |
| TC-02 | Test HDFS file creation and replication | Files are replicated per the replication factor | Passed |
| TC-03 | Verify YARN ResourceManager functionality | Jobs are scheduled and executed properly | Partially |
| TC-04 | Test high availability of NameNode | Automatic failover works during NN failure | Passed |
| TC-05 | Test high availability of ResourceManager | Automatic failover works during RM failure | Passed |
| TC-06 | Verify edge node access via Airflow server | Edge node allows secure client access | Passed |

#### **3.2 Performance Test Cases**

| **Test Case ID** | **Description** | **Expected Outcome** | **Status** |
| --- | --- | --- | --- |
| TC-08 | Test cluster performance with large file upload | Data upload speed matches expected throughput | Passed |
| TC-10 | Test cluster under high concurrent workloads | Cluster handles [x] concurrent jobs without failure | Passed |
| TC-11 | Test Spark sample job execution | Spark job executes successfully with correct output | Passed |

#### **3.3 Security Test Cases**

| **Test Case ID** | **Description** | **Expected Outcome** | **Status** |
| --- | --- | --- | --- |
| TC-12 | Test Kerberos authentication | Only authenticated users can access cluster | Passed |
| TC-13 | Verify HDFS encryption | Data at rest is encrypted | Passed |
| TC-14 | Test edge node firewall rules | Unauthorized access is blocked | Passed |

#### **3.4 Airflow Test Cases**

| **Test Case ID** | **Description** | **Expected Outcome** | **Status** |
| --- | --- | --- | --- |
| TC-15 | Test Airflow server startup | Airflow webserver and scheduler start successfully | Passed |
| TC-16 | Test DAG execution | Airflow DAGs execute successfully and produce correct output | Passed |
|  |  |  |  |
| TC-17 | Test Airflow integration with Hadoop | Airflow can submit jobs to Hadoop and retrieve results | Pending |
| TC-18 | Test Spark-submit operator in Airflow DAG | Airflow DAG runs Spark-submit operator and executes Spark job successfully | Partially |
| TC-19 | Test Spark-submit error handling in Airflow DAG | Airflow detects and logs errors from Spark-submit failures accurately | Passed |
| TC-20 | Validate full end-to-end Spark job workflow | Airflow submits Spark job, monitors execution, and verifies results end-to-end | In- Progress |

### **4. Deployment Steps**

1. **Hadoop Cluster Setup:**
   * Installed Hadoop on all nodes.
   * Configured HDFS and YARN.
2. **High Availability Configuration:**
   * Configured NameNode HA.
   * Configured ResourceManager HA.
3. **Edge Node Setup:**
   * Configured Airflow server as the edge node.
   * Need to restrict direct access to cluster nodes.
4. **Security Configuration:**
   * Enabled Kerberos for authentication.
5. **Monitoring and Management:**
   * Not Configured Yet.

### **5. Final Test Results**