

BDA Tools

1. HDFS Commands

These commands interact with the Hadoop Distributed File System (HDFS).

File Management

- **Check HDFS status:**

```
hdfs dfsadmin -report
```

- **List directory contents:**

```
hdfs dfs -ls /path
```

- **Create a directory:**

```
hdfs dfs -mkdir /path
```

- **Upload a file to HDFS:**

```
hdfs dfs -put /local/path /hdfs/path
```

- **Download a file from HDFS:**

```
hdfs dfs -get /hdfs/path /local/path
```

- **Remove a file or directory:**

```
hdfs dfs -rm /hdfs/file  
hdfs dfs -rm -r /hdfs/directory
```

- **Copy files within HDFS:**

```
hdfs dfs -cp /src/path /dest/path
```

- **Move files within HDFS:**

```
hdfs dfs -mv /src/path /dest/path
```

File Information

- **Display file contents:**

```
hdfs dfs -cat /hdfs/file
```

- **Check file size:**

```
hdfs dfs -du -s /hdfs/path
```

- **Check file permissions:**

```
hdfs dfs -ls /hdfs/path
```

2. MongoDB Commands

MongoDB commands allow interaction with its NoSQL database.

Database and Collection Management

- **Show all databases:**

```
show dbs
```



- **Switch to a database:**

```
use dbName
```



- **Show all collections:**

```
show collections
```



- **Create a collection:**

```
db.createCollection("collectionName")
```



CRUD Operations

- **Insert a document:**

```
db.collectionName.insertOne({ key: value })  
db.collectionName.insertMany([{ key1: value1 }, { key2: value2 }])
```



- **Find documents:**

```
db.collectionName.find()  
db.collectionName.find({ key: value })
```



- **Update a document:**

```
db.collectionName.updateOne({ key: value }, { $set: { key: newValue } })
```



- **Delete a document:**

```
db.collectionName.deleteOne({ key: value })
```



Administration

- **Check database stats:**

```
db.stats()
```



- **Check collection stats:**

```
db.collectionName.stats()
```



3. Cassandra Commands

Cassandra's `cqlsh` provides commands for its NoSQL database.

Keyspace Management

- **List all keyspaces:**

```
DESCRIBE KEYSPACES;
```

- **Create a keyspace:**

```
CREATE KEYSPACE keyspaceName  
WITH replication = {'class': 'SimpleStrategy',  
  'replication_factor': 1};
```

- **Use a keyspace:**

```
USE keyspaceName;
```

Table Management

- **List all tables:**

```
DESCRIBE TABLES;
```

- **Create a table:**

```
CREATE TABLE tableName (  
  id UUID PRIMARY KEY,  
  name text,  
  age int  
);
```

- **Drop a table:**

```
DROP TABLE tableName;
```

CRUD Operations

- **Insert data:**

```
INSERT INTO tableName (id, name, age) VALUES (uuid(), 'John', 30);
```

- **Retrieve data:**

```
SELECT * FROM tableName;  
SELECT name FROM tableName WHERE id = <id>;
```

- **Update data:**

```
UPDATE tableName SET age = 31 WHERE id = <id>;
```

- **Delete data:**

```
DELETE FROM tableName WHERE id = <id>;
```

Administration

- **Check cluster status:**

```
nodetool status
```

- **Describe the schema:**

```
DESCRIBE SCHEMA;
```
