1. Cluster Setup

The initial step in configuring a Cloudera cluster involves defining the cluster's size and resources.

- **Nodes:** Decide on the number of nodes. This will depend on your data size and processing needs.
- **Memory and CPU Cores:** Allocate memory and CPU cores for each node. Cloudera Manager allows you to configure these settings under the "Hosts" tab, where you can view and edit each host's configuration.

  Example:

```
# Script to automate the setup of nodes
for host in host_list:
    ssh user@${host} "sudo yum install -y cloudera-manager-daemons cloudera-manager-age
```

2. HDFS Configuration

HDFS configuration is crucial for data storage and access.

- **Block Size:** The default block size in HDFS is 128 MB. You can adjust it in the `hdfs-site.xml` through Cloudera Manager.
- **Replication Factor:** Determines the number of copies of data. Adjust this setting based on data criticality and storage capacity.

  Example:

```
<!-- hdfs-site.xml snippet -->
<property>
  <name>dfs.blocksize</name>
  <value>134217728</value> <!-- 128 MB -->
</property>
<property>
  <name>dfs.replication</name>
  <value>3</value>
</property>
```

3. YARN Settings

YARN (Yet Another Resource Negotiator) manages and allocates cluster resources.

- **Resource Allocation:** Define CPU and memory for YARN containers.
- **Scheduling:** Choose a scheduler (like the Capacity Scheduler or Fair Scheduler) and configure it for workload management.

  Example:

```
<!-- yarn-site.xml snippet -->
<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
  <value>8192</value>
</property>
<property>
  <name>yarn.scheduler.maximum-allocation-mb</name>
  <value>4096</value>
</property>
```

4. Security Settings

Ensuring the security of your cluster is paramount.

- **Kerberos Authentication:** Configure Kerberos for secure access.
- **Encryption:** Enable HDFS encryption for sensitive data.

  Example:

  ```
  # Kerberos principal and keytab setup script
  kadmin.local -q "addprinc -randkey hdfs/hostname@YOUR.REALM"
  kadmin.local -q "xst -k hdfs.keytab hdfs/hostname@YOUR.REALM"
  ```

5. Monitoring and Logging

Monitoring the health and performance of your cluster is essential.

- **Alerts:** Set up alerts for critical events in Cloudera Manager.
- **Log Management:** Configure log directories and retention policies.

  Example:

  ```
  # Log rotation script example
  logrotate /etc/logrotate.conf
  ```

Integration of the Components

Integration of these configurations is typically done through Cloudera Manager's web interface, but automation scripts can be written to handle some of the configurations, especially when dealing with a large number of nodes.

Example Scenario

Consider a scenario where you are setting up a 10-node cluster with a focus on high availability and data processing efficiency.

1. **Cluster Setup:**
   - Define 10 nodes with specific CPU and memory allocations.
   - Automate the installation of Cloudera Manager agents on all nodes.

2. **HDFS Configuration:**
    - Set a custom block size of 256 MB for large files.
    - Set replication factor to 3 for fault tolerance.
3. **YARN Settings:**
    - Allocate 10 GB of memory per node for YARN.
    - Configure the Capacity Scheduler for better resource management.
4. **Security Settings:**
    - Set up Kerberos for all Hadoop services.
    - Enable encryption for sensitive data in HDFS.
5. **Monitoring and Logging:**
    - Set alerts for node failures, high memory usage, etc.
    - Configure log rotation to manage disk space.