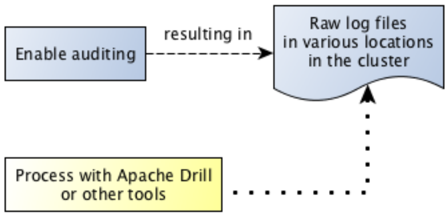
Audit Architecture: Operations

This section describes auditing architecture in MapR.

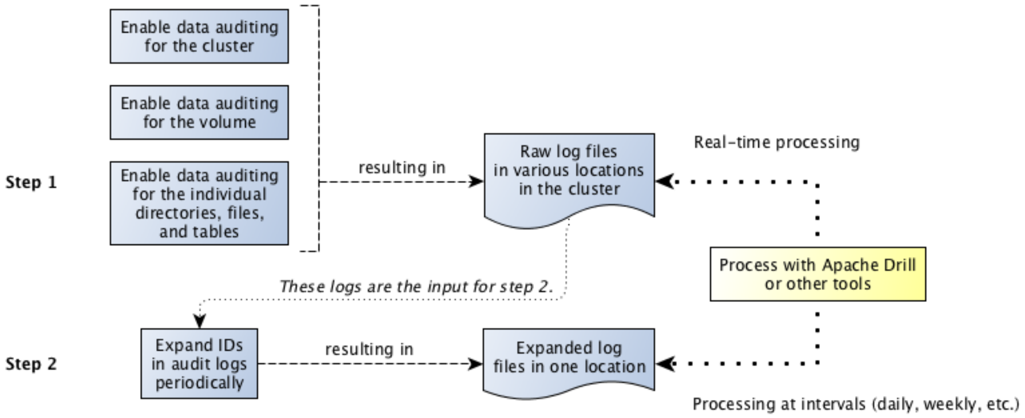
The auditing features in MapR let you log audit records of cluster-administration operations and operations on directories, files, tables, and streams.

Auditing of cluster-administration operations creates audit records of maprcli commands being executed, REST API calls, and actions performed on a cluster by means of the MapR Control Service (MCS). Auditing of directories, files, tables, and streams creates audit records of sixteen types of operations on directories and files, twenty-three types of operations on MapR-DB tables, and ten types of operations on MapR streams.

After you enable auditing, audit records immediately start to be recorded in audit logs. You can use Apache Drill or other tools to process these logs. The following diagram shows the workflow for processing audit logs of cluster-administration operations:


Enable auditing
resulting in
Raw log files in various locations in the cluster
Process with Apache Drill or other tools

The next diagram shows the workflow for processing audit logs of filesystem and table operations. The step "Expand IDs in log files periodically" refers to the use of the expandaudit utility. Raw audit logs contain file identifiers, volume identifiers, and user identifiers. The expandaudit utility looks up the names that are associated with those identifiers and puts them in new copies of the audit logs.


Step 1
Enable data auditing for the cluster
Enable data auditing for the volume
Enable data auditing for the individual directories, files, and tables
resulting in
Raw log files in various locations in the cluster
Real-time processing
Process with Apache Drill or other tools
These logs are the input for step 2.
Step 2
Expand IDs in audit logs periodically
resulting in
Expanded log files in one location
Processing at intervals (daily, weekly, etc.)

By analyzing audit records, security analysts can answer questions such as these:

- Who touched customer records outside of business hours?
- What actions did users take in the days before leaving the company?
- What operations were performed without following change control?
- Are users accessing sensitive files from protected or secured IP addresses?
- Why do my reports look different, despite sourcing from the same underlying data?

Data scientists can analyze audit records to find out the answers to questions such as these:

- Which data is used most frequently, is therefore of high value, and should be shared more broadly?
- Which data is least commonly used, is therefore of low value, and could be purged?
- Which data should be used more, is therefore underused, and needs better advertising?
- Which administrative actions are most commonly performed and are therefore candidates for automation?