**Troubleshooting**

1. Develop a patient history.
2. Develop a list of potential diagnoses.
3. Sort the list by likelihood.
4. Test!
5. Diagnosis.

The complete battery usually includes OS utilities such as top, vmstat, sar, iostat, and netstat, but also Hadoop-specific tools such as Hadoop dfsadmin, and hadoop fsck. Log files are how machines communicate with us, so make sure you’re always listening to what they have to say.

* **Humans (You)**
* **Misconfiguration Hadoop level**

1. Develop a solid understanding of the precedence of configuration value overrides.
2. Start with the basic parameters required for operation, such as storage locations, hostnames, and resource controls. Make sure that the cluster works with the minimal necessary set before graduating to the performance- and security-related parameters.
3. Before setting anything, make sure you have a good idea about what it does.
4. Double-check the unit of each parameter you set. Many parameters are expressed in bytes; a few are in megabytes.
5. With each release, look for any parameters that changed in meaning or scope. Audit your configuration files frequently.
6. Use a configuration management and deployment system to ensure that all files are up-to-date on all hosts and the daemons are properly restarted to affect changes.
7. Before and after making configuration changes, run a subset of MapReduce jobs to evaluate the performance and resource consumption impact of the changes.

* **Misconfiguration OS level:**

1. Incorrect permissions on log and data directories and resource limits such as the maximum allowed number of simultaneously opened files being set to their defaults.
2. This problem tends to happen when the initial setup and configuration of machines is not automated and the cluster is expanded.
3. Clusters configured by hand run a significant risk of a single configuration step being accidentally missed.
4. Always have it on your list of potential culprits when generating your list of potential conditions.

* **Hardware Failure**
* **Resource Exhaustion**

1. CPU cycles, memory, disk space and IO, and network bandwidth are all finite resources for which various processes contend in a cluster.
2. Resource exhaustion can be seen as a specialized subclass of misconfiguration.
3. Resource allocation can be seen as a hierarchy. A cluster contains many hosts, which contain various resources that are divided amongst any tasks that need to run.
4. You should measure and track task failures to help users identify and correct misbehaving processes. Repetitive task failures occupy task slots and take resources away from other jobs and should be seen as a drain on overall capacity.
5. Conversely, starving the Hadoop daemons for resources is detrimental to all users and can negatively affect throughput and SLAs.
6. Proper allocation of system resources to the framework and your users is just as critical in Hadoop as it is any other service in the data center.

* **Host Identification and Naming**

1. Like resource exhaustion, host identification and naming must be explicitly called out as a special kind of misconfiguration.
2. The way that a worker host identifies itself to the namenode and jobtracker is the same way clients will attempt to contact it, which leads to interesting types of failures where a datanode, because of a misconfigured /etc/hosts entry, reports its loopback device IP address to namenode and successfully heartbeats, only to create a situation in which clients can never communicate with it.
3. These types of problems commonly occur at initial cluster setup time.

* **Network Partitions**
* **E-SPORE**

1. Environment
2. Stack
3. Patterns
4. Output
5. Resources
6. Event correlation

* **Treatment and care**

Often there are multiple options for treatment, each with associated risk and side effects. Clearly the best scenario is one in which a problem is detected, diagnosed, and “cured” permanently. That, of course, is not always possible

1. Permanent eradication
2. Mitigation by configuration
3. Mitigation by architecture
4. Mitigation by process

* **On “Reboot It” Syndrome**
* **War stories**