**Report**

**Criteria to select hadoop distribution**

1. **Performance:**

Many people use Hadoop as their data platform because of its excellent performance, especially when MapReduce is swapped out for Apache Spark's data processing engine in the Hadoop stack.

**Rating:** 5

1. **Dependability:**

It's unfortunate that reliability hasn't been a higher priority for data management companies given that data has increasingly been the foundation of many businesses today. Dependability is a key distinction when choosing a distribution.

**Rating:** 4

1. **Availability:**

Hadoop distributions offer high availability. A system availability of 99% can only be promised by a select few implementations. Verify that your provider is providing you with actual HA functionalities. Consider following these characteristics

* No Downtime Upon Failure
* Tolerate Multiple Failures
* Easy to use
* Data Protection
* Self-Healing

**Rating:** 5

1. **Manageability:**

A user-friendly data management solution has emerged with Hadoop. Through various administrative tools, many implementations have contributed to improving Hadoop's manageability. To administer, troubleshoot, place employees, and monitor your system, choose a distribution that offers user-friendly administrative tools.

**Rating:** 4

1. **Data Access:**

Data collection and storage are only the start of what Hadoop is all about. It's crucial that your data be both conveniently and securely available if you want to take use of all of its insightful capabilities.

* Access to file system API
* Read, write and update access to files
* Developer control over resources

**Rating:** 4.5

**Decision:**

MapR M-series