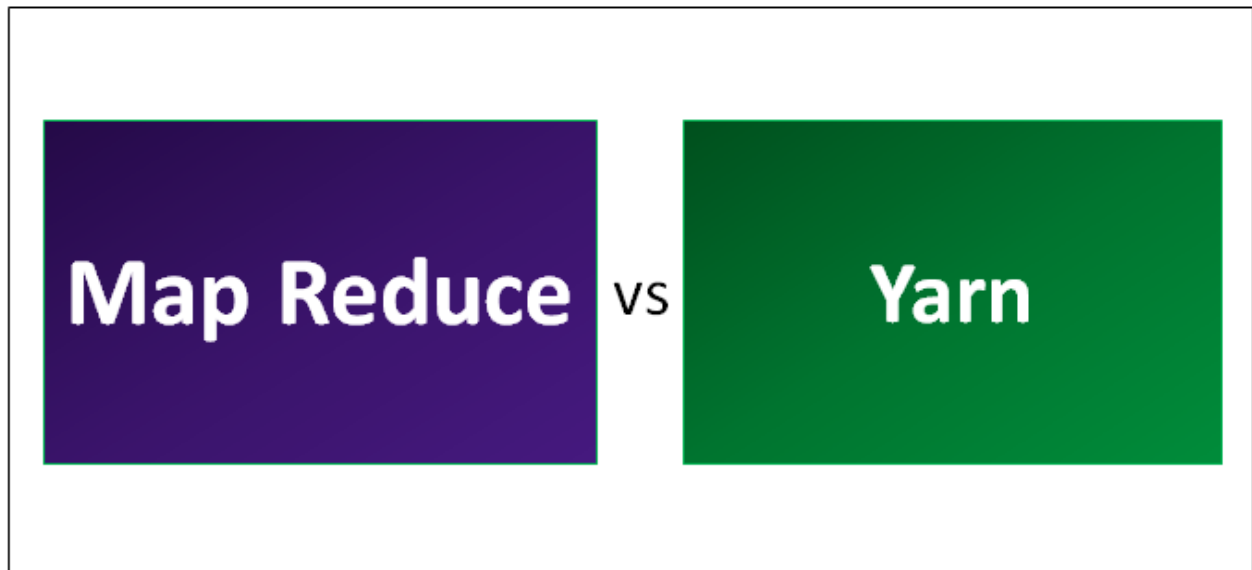


# MapReduce vs Yarn

 [educba.com/mapreduce-vs-yarn](https://educba.com/mapreduce-vs-yarn)

April 12, 2018



## Difference Between Map Reduce And Yarn

Yarn stands for Yet Another Resource Negotiator, it is the new framework to manage resources (Memory and CPU). It helps us in developing the distributed application of any kind, it provides us necessary daemons and APIs. Another important feature of YARN is, it handles and schedules resource request from the application and help the process to execute the request. YARN is a generic platform to run any distributed application, Map Reduce version 2 is the distributed application which runs on top of YARN, Whereas map reduce is processing unit of Hadoop component, it process data in parallel in the distributed environment. So basically map-reduce work on huge data component, it processes the data and store in HDFS in such a way that retrieval is easier than traditional storage.



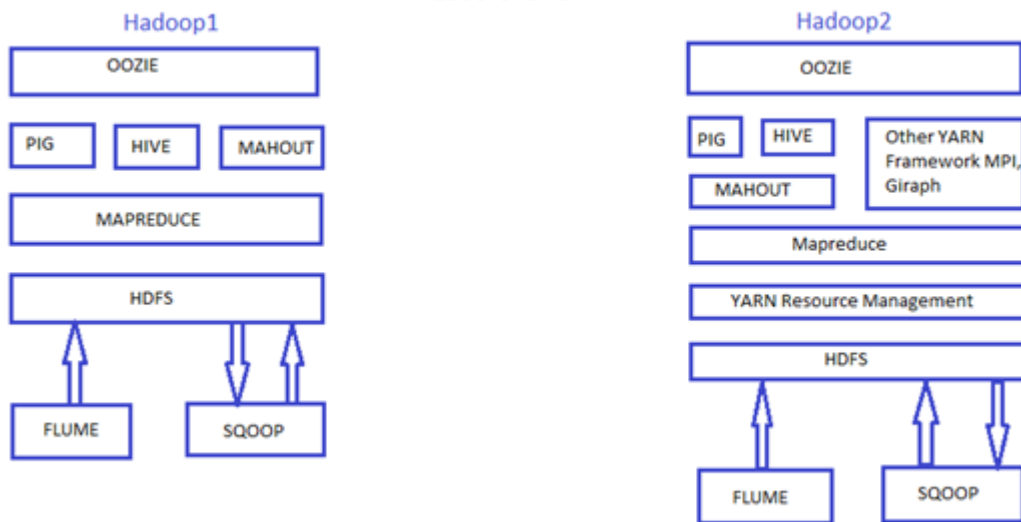


## Key Difference Between MapReduce and Yarn

---

1. In Hadoop 1 it has two components first one is HDFS (Hadoop Distributed File System) and second is Map Reduce. Whereas in Hadoop 2 it has also two component HDFS and YARN/MRv2 (we usually called YARN as Map reduce version 2).
2. In Map Reduce, when Map-reduce stops working then automatically all his slave node will stop working this is the one scenario where job execution can interrupt and it is called a single point of failure. YARN overcomes this issue because of its architecture, YARN has the concept of Active name node as well as standby name node. When active node stop working for some time passive node starts working as active node and continue the execution.
3. Map reduce has single master and multiple slave architecture, If master-slave goes down then entire slave will stop working this is the single point of failure in HADOOP1, whereas HADOOP2 which is based on YARN architecture it has the concept of multiple master and slave, if one master goes down then another master will resume its process and continue the execution.
4. As we can see in below diagram, the difference in both Ecosystems HADOOP1 and HADOOP2. Component wise YARN Resource Management interacts with Map-reduce and HDFS.

## ECOSYSTEMS



So basically YARN is responsible for resource management means which job will be executed by which system get decide by YARN, whereas map reduce is programming framework which is responsible for how to execute a particular job, so basically map-reduce has two component mapper and reducer for execution of a program.

5. In Map reduce each data node run individually whereas in Yarn each data node runs by a node manager.
6. Map reduce uses Job tracker to create and assign a task to task tracker due to data the management of the resource is not impressive resulting as some of the data nodes will keep idle and is of no use, whereas in YARN has a Resource Manager for each cluster, and each data node runs a Node Manager. For each job, one slave node will act as the Application Master, monitoring resources/tasks.

## MapReduce vs Yarn Comparison Table

Below are the comparison between MapReduce vs Yarn

Basis for comparison	YARN	Map Reduce
Meaning	YARN Stands for Yet Another Resource Negotiator.	Map Reduce is self-defined.
Version	Introduce in Hadoop 2.0	Introduce in Hadoop 1.0
Responsibility	Now YARN is responsible for Resource management part.	Earlier Map reduce was responsible for Resource Management as well as data processing
Execution model	Yarn execution model is more generic as compare to Map reduce	Less Generic as compare to YARN.

Application execution	YARN can execute those applications as well which don't follow Map Reduce model	Map Reduce can execute their own model based application.
Architecture	YARN is introduced in MR2 on top of job tracker and task tracker. In the place of job tracker and task tracker Application, the master comes into the picture.	In the earlier version of MR1, YARN is not there In the place of YARN job tracker and task tracker was present which help in the execution of application or jobs
Flexibility	YARN is more isolated and scalable	Less scalable as compare to YARN.
Daemons	YARN has Name Node, Data node, secondary Name node, Resource Manager and Node Manager.	Map Reduce has Name node, Data node, Secondary Name node, job tracker and task tracker.
Limitation	There is no concept of single point of failure in YARN because it has multiple Masters so if one got failed another master will pick it up and resume the execution.	Single point of failure, low resource utilization(Max of 4200 clusters by YAHOO) and less scalability when compare to YARN
Size	By default the size of a data node in YARN is 128MB	By default the size of a data node in Map reduce is 64MB.

## Conclusion

In Hadoop 1 which is based on Map Reduce have several issues which overcome in Hadoop 2 with Yarn. Like in Hadoop 1 job tracker is responsible for resource management but YARN has the concept of resource manager as well as node manager which will take of resource management. Map reduce has a single point of failure i.e. Job tracker, if job tracker stop working then we have to restart our entire cluster and executes our job again from Initial. In a real scenario, none of the organization don't want to take this kind of risk, especially in a bank defense sector. Such organization which works on streamline data will not ready to take this kind of risk. For the sake of few minutes, they are going to lose their data and may have some critical business impact. So YARN has a better result over Map-reduce.