International Journal of Engineering & Technology, 7 (4) (2018) 4850-4855

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



A cross study on apache hadoop and yarn schedulers

N. Deshai[1]*, B.V.D.S.Sekhar[1], S.Venkataramana[1], V.V.S.S.S.Chakravarthy[2], P.S.R.Chowdary[3]

[1]Department of Information Technology, S.R.K.R Engineering College, Bhimavaram, AP, India
[2]Department of Electronics & Communication Engineering, Raghu Institute of Technology, Visakhapatnam, AP, India
**Corresponding author E-mail:desaij4@gmail.com*

Abstract

Todays, digital world facing more challenges from processing data with general technologies on different real-time oriented applications. The specific reason that a fast-growing scale on the size of the datasets during continuously generated from heterogeneous applications in various industries and fields, while such rapidly expanding data size to handling, processing, computing and store efficiently with Existing techniques are extremely critical and difficult. Nowadays, computer world focuses the innovative direction for massive information process and storage about digital world activities is called big data (BD) because at present digital transaction world datasets have double in every second. So many fields, industries and applications are turned to big data methods and platforms. Core Hadoop open source community has most famous and advanced technologies which Assist efficiently process, organize also store huge length of datasets through popular components are Hadoop Distributed File System which is quickly stored peta and zetta-byte information and efficiently processing that petabyte and zetta-byte information by Map-Reduce (MR), but working with that hadoop1 version some restrictions on resource allotment, scalability and support only few applications. Therefore, we describe an efficient comparison with new MR is yet another Resource Negotiator to avoid Hadoop v1 efficient resource allotments issues. Because advance resource allotments are leading function for efficiently, process the jobs. Also, study default schedulers with advanced schedulers in their issues on basic Hadoop v1 MR and YARN.YARN presents advanced schedulers like fair and capacity schedulers are leads high utilization onresources,excellent sharing and more scalability.

Keywords:*Big Data; Hadoop; HDFS; Map Reduce; Schedulers; Yarn.*

1. Introduction

BD means an excessive quantity of DW data generated at enormous speed in various structures (structure data (2D data), unstructured data (combination ofmultiple things) also semi format data (xml). Its efficiently support to store, process, manage, examination on excessive size of thedissimilar format of enormously large and huge volume, good velocity and the broadmiscellany of universal data, at ample speed, in the limited deadline to done analysis [3]. The DW was created the latest era peta-to-zetta byte-oriented data, generated, gathered quickly and double every second, however modern and advanced distributed oriented data batch and streamorientedprocessing tools are highly demanded to operate today's excessive DW databases. BD means the latest approach to process DW data activities and issues, these are unable to resolve smoothly by traditional db and tools, BD not like the gigantic size of information also completely digital planet datasets are seriously challenging to store, operate and compete in a smooth way. An apache team presents tools for DW to process today's excessive measure of DW data by hadoop freely access software model regarding solving, efficiently keeping and working activities on DW information. BD is symbolizing the latest era in the data analysis field and functioning on huge databases in DW. The remaining of this survey and comparison paper focus as, in part II focuses thedetailed features of Apache hadoop tools, and Section III presents working of YARN compare with MR,in IVth Sectiondistinguish deep Conclusion

2. Apache hadoop

2.1. Hadoop

Hadoop implemented by Doug cutting in 2005[1]. Hadoop really developed to perform distributing elements for nutch web searching engine work. Hadoop is free also open access source powerful software approach for efficiently stored and perform well processing on very complex scale data sets on commodity hardware devices in the cluster. It is new era powerful popular BD tool. BD would be dominating in coming application by the support of information-processing strategies in the digital planet [3]. Afree access Apache open resource hadoop isthe latesttool, also it designsMR framework for information processing in BD. butthe starting version of hadoop working on the centralized model, because JT handles RM totasks and jobs processing .This implementation is sufficient for implementing, but intrinsic limitations becomechallenge as hadoop services move en era on millions of jobs perform simultaneously ontasks andmassivenodes of clusters in every system. Moreover, the JT is a master and face single point failure, means little failure will slow the system performance in JT as exposed in Figure 1 Hadoop Framework.

Copyright © 2018 N. Deshai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

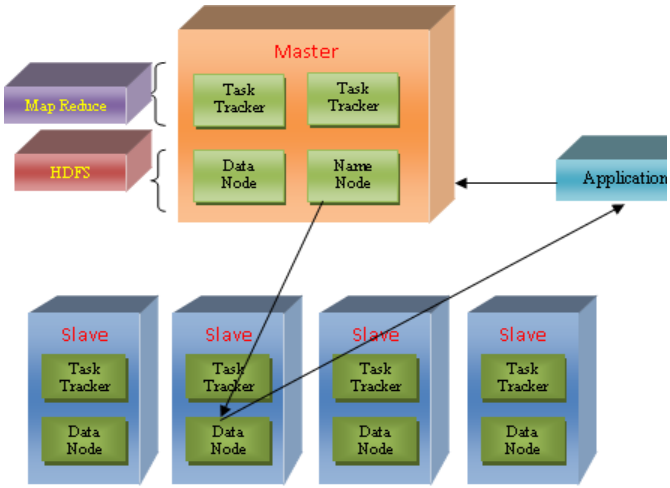


Fig. 1: Hadoop Framework.

Again it's problem like bottleneck about scalability because total tasks and jobs are controlled by the central point. Many traditional, general and common data integration techniques are joining with apache free access hadoop also BD helps to lead the further stage of the digital world. Table 1and Fig.2 explains features between Hadoop V1 and Hadoop V2 [15].

Features of hadoop: It has huge elasticity in processing various data syntaxes. It has vast scalability service. It has a high-level error tolerance. It has efficient information handing with much speed. Which are vigorous and fair effective Its latest open source, high reliability, replication, extremely available services, data locality, scalability, circulated method and simple to use.

2.2. HDFS

Main intention about HS component might flexibly store digital world data efficiently and reliably if the failure occurs on Name Node, Data Node and system separation [10]. HS is capable to powerfully store huge information storage device used by apache open resource tool hadoop computations. It is working based on Name node and Data Node to design a distributed File system. It offers and supports efficient performance, easily to use data on large clusters. HS can assist to control big data issues and support and helps efficiently transfer the information among computing nodes. It closely and tightly connects with MR for smoothly handle data processing strategies and computations. HS is mainly implementing for reach high fault tolerant. The working of HS architecture is explained in Fig.2.

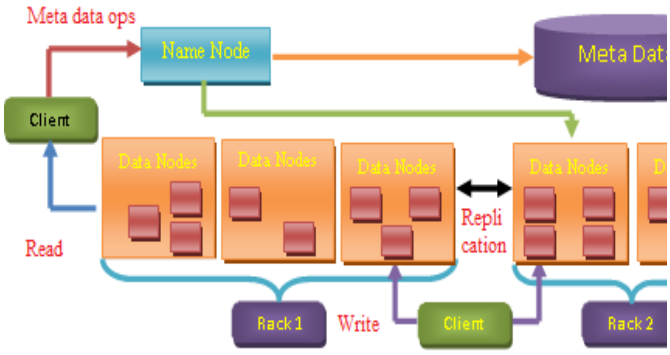


Fig. 2: HDFS Architecture.

2.3 Map reduce

In hadoop version 1 like MR a programming component also the most enabler for basic multiple BD platforms.MR like more famous function to handling BD aspects with parallel, distributed and high scalable strategies[7]. The Year of 2006 innovative, latest java programming framework called map reduce presented by hadoop version 1, but discomfort from serious limitation and restriction is accepted nodes maximum of four-thousand only. To determine concurrency depended on aspects on the system with working performance regarding dividing jobs into the number of tasks of equal size for this convenient hadoop MR component is used, while it supports to handle granularity with controlling total map tasks[11], [16].

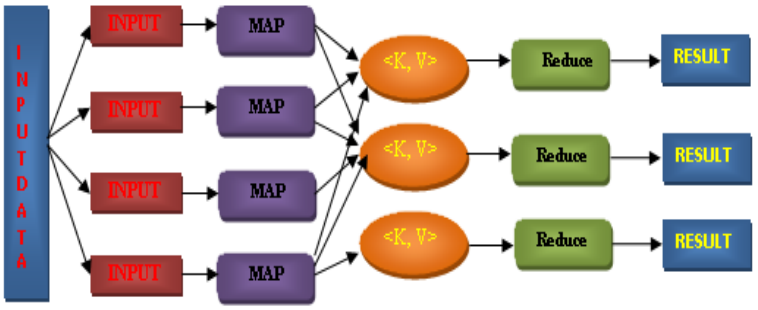


Fig. 3: Map-Reduce Framework.

MR has free access hadoop component invented by Google the year of 2004 to distributing and dealing huge information and files across a complex cluster in each system. To raise the performance of system regarding split the processing into the little part of dealing which must be executed in simultaneously across thousand and hundreds of nodes in each cluster.MR has programming popular framework follows dividing-and-conquering method for parallel working [10].

In MR compute set of operations for each job include various tasks working like automatically and parallel is hadoop MR. In many iteration all tasks which scheduled and spread on each container in the cluster. Hadoop partition input information into little chunks and all MR tasks must accountability to handle each chunk. The working of MR framework is explained in Fig.3. It is the latest and extremely developed an era of BD. BD has drive drastic changing happen in traditional data analysis. It can handle any sort of analysis techniques on large voluminous data but the imminent need is to increase system hardware facility and features. Selecting the suitable hardware and software techniques turn to a critical thing whenever users request must be fulfilled in deadline.BD has symbolized the latest era in data analysis and utilization.MR programming paradigm has most enabler for basic multiple BD platforms. MR is a more famous tool for handle BD in extra distributed and scalable. YARN introduces: In spite of an attractiveness of apache hadoop, its achievement could be restricted by some set of problems that are the failure on single a point, job allocation on centralized and support, accept only few programming frameworks compare to MR. apache hadoop version2 YARN has the latest generation of hadoop it presented to overcome issues of hadoop version v1 (MR) [11].

Table 1: Features of Hadoop V1 and Hadoop V2

Hadoop Version 1	Hadoop Version 2
It has MRV1 & HS	It has HDFS,MRV2 YARN
Single Map Reduce	Multiple models with YARN Map reduce iterative, streaming, graph, spark and storm.
MR done both processing and resource management	It done separately
One Name node to manage the whole namespace.	Multiple Name nodes, servers manage multiple namespace.
Low scalability	High Scalability
Doesn't support horizontal scalability	Support Horizontal scalability
Job Tracker handle resource allocation and task managing	AM handle scheduling and RM handle resources.
Bottleneck issue in Job Tracker	No bottleneck issue
No Multi-Tenancy	It has Multi-Tenancy
static size slots(slots are static)	Variable Size Containers(slots are dynamic)
Support up to 4000 nodes per cluster	More than 10,000 nodes per cluster
The restriction to treat as a ad-	The restriction to treat as a ad-

vanced policy especially to many event dynamic processing also advanced streaming and more real-time computing.	vanced policy especially to many event dynamic processing also advanced streaming and more real-time computing.
Name node Failure needs manual intervention to Recover.	Automatic Recover from Name node Failure
Does not support Microsoft Windows	Support Microsoft Windows

3. YARN architecture

For solving node restriction through well-organized more innovative resource distribution and huge scalable services are achieved by yarn, presented in hadoop version 2, also it offers exclusively assign a resource to MR jobs and efficiently increases range the dimensions of db, primarily focus performance growing[4]. Hadoop encompasses to accomplish an efficient file distributing and deal with massive information, storage on multiple domains and yarn for innovatively allocate and share required necessary resources to applications. Yarn authorize techniques as graph, interaction, stream and batch process to implement and store with HS [4] .Compare with MR, it makes possible purpose inbuilt processing system in yarn.

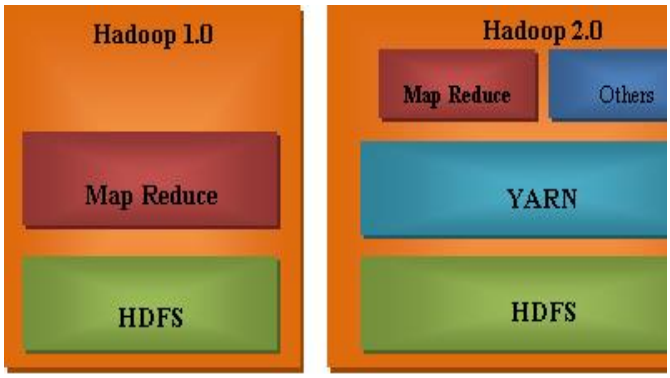


Fig. 4: Architectures of Hadoop Version 1 and 2.

Best services on yarn, it permits diversity of software frameworks to work single hardware when hadoop is implementing. Job tracker could be handle resource assigning and job scheduling aspects on hadoop v1. [2]Yarn divides the constantly running programs by job-tracker for digital planet resource-manager (RM) also Application Master (AM). In yarn environment framework functioning as a master slave where RM doing like a ruler and AM functioning like slaves. In Yarn, the RM handles and assigns resources to cluster application activities. Generally, Node Manager (NM) working like slave and single AM handle and assign digital planet cluster resources to each application; also it had more amounts of nodes and resources which are limited by RM. An AM can communicate with RM to accomplish resources and NM is accountable to manage client request per node also creates containers for every job. The Container is similar to map reduce, it has resource elements are a good memory, a central processing unit, hard disk and internet connection [17].

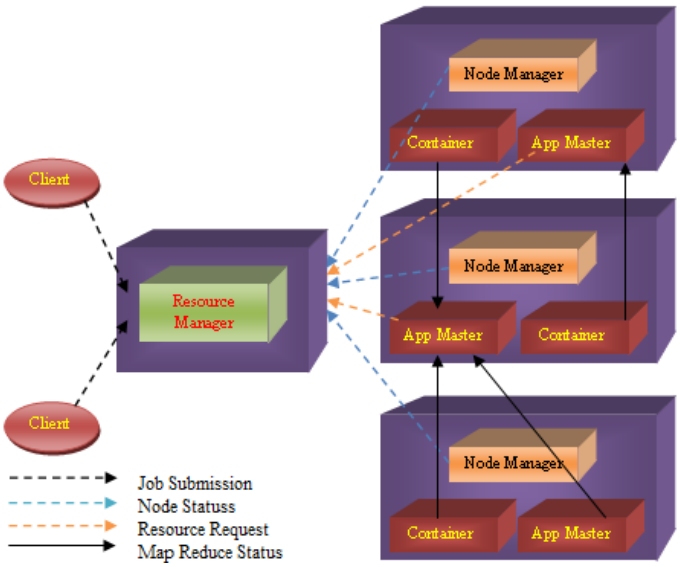


Fig. 5: Architectures of YARN.

[12], [15] In Yarn RM majorly had schedulers and AM are working regarding resource demands and job queues are recognize also schedulers are accountable for partitioning resourcesto the cluster. Yarn combined with FIFO, FAIR scheduler (FS), and Capacity scheduler (CS). In MR Version 2 is Yarn, FIFO does not follow optimally organize resources in the cluster, but in FS to make restrict the maximum executing applications per user and queue. In spite of attractiveness of apache free access hadoop, it's achievements are restricted by major problems in digital world datasets, that is failure on a single point, job allocation is done on centralized base and support and accepts few software frameworks than MR. Hadoop version 2 YARN means the latest technology of hadoop presented to overcome issues of hadoop version v1.

Client service: perform quickly responds to the user by Remote Procedure Call (RC) with job placing, Interrupting, Terminating and getting cluster data [2].

It had three essential components to construct the YARN are RM, NM and AM.

RM: Which is an advanced service for handle resources for every application in the cluster also control elements in yarn and doing like a master. It contains services like Client, admin, Resource Tracking, NM liveliness Monitor and Node list Manager Services.

Admin Service:to store working nodes list and Queue details, to separate the request between user and admin and assign the highest priority

Resource track service: it establishes a connection between RM and nodes respond RC from total nodes focus new nodes enrollments, reject the request on decommissioned nodes, getting and pass heartbeats to YARN schedulers from every node.

NM liveliness Monitor (NLM): it keeps live node status and marks dead nodes if a heartbeat is not moving in every interval then deleted from RM node list.

Table 2: Difference between YARN and Map Reduce

Features	YARN(MR v2)	Map Reduce
Responsibility	RM	RM & Data Processing
Execution model	More Generic	Less Generic
Application Execution	Well	By Own model
Architecture	Resource, Node and Application manager	Job and task tracker
Flexibility	More scalability	Less Scalability
Daemons	Name , Secondary Name & Data node, Resource & Node Manager	Name node, Data node, Secondary Name node, job tracker and task tracker
Limitation	No failure because it has multiple Masters if one got Failed another master will choose and resume the	Failure because poor resource usability Maximum of 4200 nodes in clusters

	execution.	
Size	128MB	64MB

Node List Manager: it stores collection of active and inactive nodes depend on NLM output. Table 2 and Fig.4 explain features between YARN and MR

Application Master Services (AMS): it can design latest AM, terminating if job completed, keep the status of required resource requests in AM also sends information to RM in yarn component.

AM liveliness Monitor (ALLM): it stores every AM also update present status regarding heartbeat iterations time. The RM removes AM container if it has no heartbeats and immediately create other container about AM tasks. Working process in YARN as Shown in Fig 5.

1) User place jobs.
2) RM allocates a specific container for run the AM.
3) AM communicating with the RM for required resource and Container services.
4) After finish efficient container allocation the AM connect with node manager to start the container.

RM manages required resources and assigns each task for efficiently running different applications in each cluster. In Yarn, the RM had a pluggable scheduler element which focuses efficiently resources allocation process regarding requirements. The RM only concentrates resource monitoring and utilization.

3.1. Schedulers in MR V2 YARN

The Feature of YARN schedulers explains details in Table 3.

Table 3: Explains the Features of YARN.

Scheduler	Preemption	Fairness	Homo & Heterogeneous	Assign jobs	Designed
FIFO	Not	Not	Homo	Static oriented	Designed
Fair	Yes	Yes	Homo	Static oriented	Designed
Capacity	Not	Yes	Homo	Static oriented	Designed
Label Based	Yes	Yes	Both	Both Static & Dynamic	Designed
HaSTE	Yes	Yes	Both	Both Static & Dynamic	Designed
Deadline Aware	Yes	Yes	Both	Only Dynamic	Not
Mercury	Yes	Yes	Both	Both Static & Dynamic	Designed
SLA	Yes	Yes	Both	Both Static & Dynamic	Designed
Delay Scheduler	Yes	Low	Homo	Only Static	Designed

3.1.1. Yarn default scheduler (DS)

HAdoop v1, v2 and Yarn packages have DS are FIFO, capacity and Fair schedulersFIFO:

Hadoop v1 had presents default scheduler is FIFO, It serves the jobs regarding job arriving order, which job arrive firstly execute firstly some limitations on FIFO. Efficiently handle little jobs. Simply and easy to implements each job in FIFO. It is critical to predicting processing time and utilization of resources and lack of performance in heterogeneous systems. It places the complex jobs firstly than little jobs in the queue the waiting time is more on small jobs it doesn't assign any priority level to the task, which not suitable bigjobs.

Fair scheduler (FS): Which is more famous scheduler in DS also it could handle both little and big clusters. A Major role is efficiently to share and assign resources to all application that means every user gets resources [3]. It follows queues to assign required resources between applications. Generally, all users use the same default queue, compare with fifo, the FS running applications like parallel on various queues. In FS it can request total cluster resources it needs for execution when a single application is executing on the cluster. When an individual application is running, if another application will join, the FS can assign the free resources to every application finally, all cluster applications get required resources regarding fairly distributing type. It provides multiple properties on queues like minimum, greatest distribute with weight. Basically cluster being at least five 5 number of queues and sixty quantity of containers, so each queue had 12 containers if queue weight is 2 then such queue had increase resource pool than other. It accepts a hierarchical queue, it brings nested queue services. In inside it divides the total resources between slave and master queues. It is well convenient to create queues for large clusters with good arranging structure and priorities on the requirements. It accepts different scalable memory into different container deeds. It accepts high weight queues for different applications if demand many resources.

Capacity Scheduler (CS):It is similar to DS setup in YARN [3], which is applicable for distributing computations on the big cluster in different organizations applications. CS allows various organization queues to allocate available resources in the cluster. Cs provides an efficient security by access control list is enable distributing cluster securely among clients and organizations. CS helps to memory exhaustive applications because huge demand also allocates containers to compute the resource needs and request.

3.2. Custom-built scheduler (CBS) in YARN

To growing performance for current world applications, we use an extended schedule (ES) against DS for particular applications. Because ES can help and support to raise the performance for coming large applications.

Label-based scheduler (lbs):LBS support an efficiently control jobs based on multi-occupant structure in apache hadoop. By using LBS we can manage jobs executions for different clients on node level .based on node and node groups the LBS support and precisely handle job executions.

HaSTE scheduler (HS):It helps to decrease the MR Job execution time, by modifying requests, strengths of resources and needs among tasks [2]. HS assigns working on initial task assignment (ITA) and real-time task assignment (RTTA) to influence the frequently use the resources in real world applications. RTTA allocating resources on heartbeat message from NM. it assigns necessary resources on needy and strength of the algorithm. This effective RTTA supports HS as a central part in the performance.

Mercury Scheduler (MS):It provides advanced resource managing framework that helps centralized distributing to schedules.

MS is the latest service to improve important variations of centralized distributing schedulers. [14]MS to handle the resource distributing efficiently by the centralized scheduler, multiple distributed scheduler and coordinatoralso designed these by MS .the centralized scheduler build in orthodox scheduler set more delay to every job with joining the housekeeping strategies, and be there for one scheduler. Distributed scheduler split the tasks by some scheduling strategies and support to decrease job-finishing time if complete the individual job.MS gets benefits from coordinator application running.

Deadline aware scheduler (DAS):DAS makes a deadline for each job execution regarding information processing and availability of resources. With the help of linear regression and machine learning to calculate the deadline.DAS is suitable to manage a heterogeneous environment.

Service aware constraint energy efficient scheduler (SACEC): Service level agreement (SLA) scheduler scheme allocate resources to MR jobs in YARN. In SLA constraint to avoid the MR network traffic by data locality.SLA follows a dynamic voltage and frequency scaling system to manage cpu clock frequency for coming tasks regarding real execution task completed time. SLA

adjusts frequency for decrease the read execution time and also raises the system energy.

Delay Scheduler:[3] Facebook designs it, the delay take place from the total quantity of time before you to obtain the information especially on the local nodes to pass over if not obtainable after that submit the jobs. The major intention is to beat the drawback of professional distribute locality. In typically your task might be scheduled and couldn't start duo to lack of many jobs, proper quantity of time and total price could be open in the place of its job, The main relational intent could be multiplex hafiz oriented clusters ragarding statistics and have minimal weight lying on equity very high data locality gain. The reallocation of resources is done by eliminating the remaining tasks of the previous work to organize a place for their work and waiting for a task to be completed in the assigned spaces for new work, but the disadvantage is the operational performance by eliminating the amount Tasks they do not arrive Useful because of waiting for jobs. From small jobs, many locality problems will occur because the file has low input data. So it contains a little number of blocks read, small drawback will happen whenever is the job follows ascending order list the head should be controlled one of the tasks is run on the remaining empty slot then whatever the node working is going on. The important problem with slot machines is that there is a possibility that a job is assigned to the same space that I will reuse. Finally, designers follow the schedule of delays because equity is free. Meanwhile, to increase the location of data in many jobs, they remain for the programming option with an efficient local data node. The Tosca assignment has a limitation on free job selection.

In HS the NN working is much similar to JT; HS handle total files also associated with blocks and locations. To effectively measure, those design issues, help, and support hadoop system implement and designing process. Different researchers frequently use simulator and emulators. Limited number of hadoop frameworks and simulation exist in every system. The simulation techniques used by MR simulator (MRS), SMR, PMR, MRSS, depend on discrete-sequence of event modeling and simulating. SF working an analytical modeling and simulating. In system, designing strategies on MR and SS are schedulers in Hadoop simulator for the handle on MR scheduler.Resource allocations, PMR, MRSS, MRSm concentrate on simulating the MR jobs execution tool. Simulations tools offer useful internal things on hadoop system working .mainly every simulator generates real-time and expensive design also valid incoming generation computations.

Table 4: Modeled System Process

Daemon Process	Abbreviations
Big Data	BD
Digital World	DW
Map Reduce	MR
HDFS	HS
Client	CLT
Resource Manager	RM
Application Manager	AM
Resource Scheduler	RS
Node Manager	NM
Map Task	MAP
Reduce Task	RED
Application Master	AMS
Resource Scheduler	RS
Node Manager	NM
Map Task	MAP
Reduce Task	RED
Application Master	AMS
Name Node	NN
Data Node	DN

4. Conclusion

We have studied and surveyed this entire paper on the major difference between classical MR and advanced new version YARN has done absolute research on Hadoop existing tools, which accomplish to digital planet different computation in Big Datasets. In addition, this complete research paper determines the issues, limitations, merits and difficulties about Hadoop components regarding different metrics like processing enormous data sets in digital world fault tolerance. Core project Hadoop free access apache turned to become more exclusively efficient technique v 2 is YARN. This survey paper centre of attention about the various scheduling, architectures and their facilities, services, improvements of YARN also examine YARN scheduler new benefits over hadoop1 version schedulers. Finally, this survey inspects the raising improvements on data facility is very close to computations by YARN compare with Hadoop default schedulers. computation genesis, basic Hadoop has turned to become, more advanced features asHadoop version 2 the map reduce 2 called YARN for increasing computational power. This paper discussed the MR v2 like YARN resource efficient allocation schedulers of Hadoop version1. Moreover, we analyze the many pluggable Hadoop excellent schedulers have presented on yarn with the benefits are more than the default oriented schedulers on Hadoop v 1.

References

[1] Jisha S Manjaly, T. Subbulakshmi "*A comparison Study and Performance evaluation of Schedulers in Hadoop YARN*", Proceedings of the 2nd International Conference on Communication and Electronics Systems (ICCES 2017), (2017), pp.78-83. http://dx.doi.Org/10.1109/CESYS.2017.8321202

[2] Yi Yao, Jiayin Wang, Bo Sheng, Jason Lin and NingfangMi, "Ha-*STE: Hadoop YARN Scheduling Based on Task-Dependency and Resource-Demand*," IEEE 7th International Conference on Cloud, (2014), pp. 184-191.https://doi.org/10.1109/CLOUD.2014.34.

[3] YehiaElshater, Patrick Martin and Dan Rope,"*A Study of Data Locality in YARN*", IEEE International Congress on Big Data, (2015) , pp.174-181.http://dx.doi.Org/10.1109/BigDataCongress. 2015.33

[4] XiaojunCai, Feng Li, Ping Li, Lei Ju and ZhipingJia, "*SLA-aware energy-efficient scheduling scheme for Hadoop YARN*", The Journal of Supercomputing,Vol,73, No.8,(2017), pp.623-628.https://doi.org/10.1007/s11227-016-1653-7.

[5] K. Kc, K. Anyanwu, "*Scheduling hadoop jobs to meet deadlines",* in Cloud Computing Technology and Science (CloudCom), IEEE Second International Conference on. IEEE, (2010), pp. 388-392https://doi.org/10.1109/CloudCom.2010.97.

[6] A. Verma, L. Cherkasova, and R. H. Campbell, "*ARIA: automatic resource inference and allocation for mapreduceenvironments*" in Proceedings of the 8th ACM international conference on Autonomic computing. ACM, (2011), pp. 235–244.https://doi.org/10.1145/1998582.1998637.

[7] J. Polo, D. Carrera, Y. Becerra, J. Torres, E. Ayguade, M. Steinder, I. Whalley, "*Performance-driven task co-scheduling for map reduce environments*," in Network Operations and Management Symposium (NOMS), IEEE, (2010), pp. 373–380.

[8] M. Zaharia, D. Borthakur, J. SenSarma, "*Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling*" in Proceedings of the 5th European conference on Computer systems.ACM, (2010), pp. 265-278. https://doi.org/10.1145/1755913.1755940.

[9] M. Isard, V. Prabhakaran, J. Currey, "*Quincy: fair scheduling for distributed computing clusters*" in *Proceedings of the ACM SIGOPS22nd symposium on Operating systems principles*. ACM, (2009), pp. 261-276.

[10] A. Verma, L. Cherkasova, R. H. Campbell, "*Two sides of a coin: Optimizing the schedule of mapreduce jobs to minimize their makespan and improve cluster performance*" in *Modeling, Analysis & Simulationof Computer and Telecommunication Systems (MASCO TS), 2012 IEEE20th International Symposium on* IEEE, (2012), pp.11-18.

[11] J. Dean and S. Ghemawat, "*Map Reduce: simplified data processing on large clusters*" Communications of the ACM, Vol. 51, (2008), pp.107-113.https://doi.org/10.1145/1327452.1327492.

[12] Vinod Kumar Vavilapalli, Arun C Murthy, Chris Douglas, SharadAgarwali,*"Apache Hadoop YARN: Yet Another Resource Negotiator"* in SoCC'13, Santa Clara, California, USA , (2013)

[13] B. Hindman, A. Konwinski, M. Zaharia, A. Ghodsi, A. D. Joseph, R. Katz, S. Shenker, I. Stoica,"*Mesos: a platform for fine-grained resource sharing in the data center*", In Proceedings of the 8th US ENIX conference on Networked systems design and implementation,NSDI'11, Berkeley, CA, USA, (2011), pp. 22-22.

[14] KonstantinosKaranasos, SriramRao, Carlo Curino, Chris Douglas, Kishore Chaliparambil, "Mercury: Hybrid Centralized and Distributed Scheduling in Large Shared Clusters", Proceedings of the 201 5 USENIX Annual Technical Conference, (2015), pp.485- 497.

[15] N.Deshai, G.P.S.Varma, S.V.Ramana, "A study on analytical framework to breakdown conditions among data quality measurements" in International Journal of Engineering & Technology, Vol 7(1.1), pp: 167-172, 2018.

[16] N.Deshai, S.Venkataramana, I.Hemalatha, G.P.S.Varma, "A Study on Big Data Hadoop Map Reduce Job Scheduling", International Journal of Engineering & Technology, Vol 7(3.31), pp: 59-65, 2017.

[17] N.Deshai, P. Swamy, G.P.S.Varma, "Big Data Challenges and Analytics Processing Over health Prescriptions", Jouonal of Advance Research in Dynamical & Control Systems, 15-Special Issue Vol 7(3.31), pp: 650-657, Oct'2017.