ISSN: 1813-4890

Research on Data Storage and Processing Methods of Cloud Computing Environment

Huiting Sun, Yanbin Long, Yuantao Li, Haoran Sun

University of Science and Technology Liaoning, China

Abstract

This design discloses the research of data storage and processing methods in cloud computing environment. The platform includes: a cluster computing framework for processing operations on jobs submitted by users according to different business processing requirements; Cluster resource manager, which is responsible for managing and scheduling cluster resources and jobs; Storage system, which is responsible for data storage management; Elastic cluster management center,Responsible for receiving user requests, managing user submitted workflows, monitoring the execution status of clusters and jobs, managing cluster configuration information, guiding cluster and application framework, and ensuring the reliability of the whole cluster; Application framework,It is used to build data warehouse, provide higher-level abstract processing and quasi-real-time processing for large data sets; This design can provide solutions for various data processing business scenarios. Compared with traditional data storage and processing methods, it is more efficient, more reliable, more flexible, simpler, safer and more economical.

Keywords

Cloud Computing; Big Data; Storage; Processing Methods.

1. Current Background Technology

The rapid development of cloud computing and big data has hatched a number of cloud service providers. According to the types of services they provide, cloud services can be divided into three levels: IaaS(Infrastructure as a Service), PaaS(Platform as aService) and SaaS(Software as a Service). For PaaS (Platform as a Service), it includes offline data processing, massive data online service and other application scenarios.It allows users to flexibly create cloud clusters and use data storage and processing services quickly and easily according to their needs, mainly including custom cluster software and hardware infrastructure, rich cluster job types, and custom cluster job execution strategies. Hosting the business cluster to the cloud can simplify the deployment, management and monitoring of the cluster.It can also improve the scalability, reliability and security of the cluster and reduce the cost of the enterprise. According to the flexible supply of resources and different types of data storage and processing requirements in the cloud computing environment,Building a big data platform based on cloud computing and providing various data storage services and processing conversion analysis services is the content that cloud service providers need to explore and study urgently.

2. System Design System

The main purpose of this design is to overcome the shortcomings of the prior art, provide research on data storage and processing methods in cloud computing environment, and realize unified storage and processing for offline data processing, real-time data processing and interactive query.

In order to achieve the above purpose, this design adopts the following technical scheme:

ISSN: 1813-4890

This design provides a platform for storing and processing big data in cloud computing environment, including cluster computing framework, cluster resource manager, storage system, flexible cluster management center and application program framework.

The cluster computing framework is used for batch processing, streaming processing and interactive processing operations on jobs submitted by users according to different business processing requirements;

The cluster resource manager is used for managing and scheduling cluster resources and jobs;

The storage system is used for storing and managing input data sets, processing results and log information;

The flexible cluster management center is responsible for receiving the user's request, managing the workflow submitted by the user, monitoring the execution status of the cluster and jobs, managing the configuration information of the cluster, guiding the cluster and application framework, and ensuring the reliability of the whole cluster;

The application framework is used to build a data warehouse, provide higher-level abstract processing and quasi-real-time processing for large data sets;

The flexible cluster management center includes six modules: user request processor, cluster configuration manager, cluster guidance manager, job flow manager, cluster status monitor and cluster management controller.

The user request processor is used for receiving requests from management clients, management interfaces and management command line users and forwarding them to corresponding components;

The cluster configuration manager is used for managing the configuration information of clusters, including cluster virtual machine configuration information, job configuration information, cluster operation strategy and job execution strategy;

The cluster boot manager is used for starting and closing the cluster virtual machine and the computing framework, as well as the boot operation of other application programs; in addition, the cluster boot manager is also used for calling the interface of cluster resource redistribution;

The job flow manager is used for managing and monitoring the whole job flow according to the job-related configuration submitted by the user to the cluster configuration manager;

The cluster status monitor is used for collecting the overall operation status of the cluster and performing simple performance analysis, evaluation and display;

The cluster management controller is responsible for monitoring the operation of five modules, namely, the user request processor, the cluster configuration manager, the cluster guidance manager, the job flow manager and the cluster status monitor, so as to ensure the normal work of the whole flexible cluster management center.

3. Specific Implementation Steps

This design also provides a processing method of big data storage and processing platform in cloud computing environment, which includes five flow steps: cluster configuration initialization and cluster creation, job flow submission to the cluster, job flow scheduling and execution, cluster resource redistribution, cluster termination and processing result acquisition. The specific implementation steps are as follows:

S1, initializing cluster configuration and creating a cluster: before logging in to the cluster management console, users need to create storage directories of input data sets and script files on the best external storage; Then, the user initializes the cluster configuration and sets the operation strategy of the cluster. The cluster configuration includes selecting different types of virtual machine nodes with different specifications and selecting the master node; At this point,If the user chooses to start the virtual machine cluster immediately, the cluster boot

ISSN: 1813-4890

manager of the flexible cluster management center will load the configuration information related to the cluster boot operation in the cluster configuration manager and boot the virtual machine cluster;

S2, submitting job flow to the cluster: after the virtual machine cluster is started, the user submits a series of job steps to the cluster, and when submitting job steps, it is necessary to specify the input data set, script file, the location of the cluster log and the running parameter information of the script file; At the same time, users can set the scheduling strategy of job steps,These job configuration information will be submitted to the user request processor of the flexible cluster management center, and finally will be handed over to the cluster configuration manager for storage management. After the job steps are configured, you can choose to schedule and execute immediately, then the cluster boot manager will start the cluster computing framework and optional applications.And notify the main node of the cluster to copy the input data set and script data stored in the external storage through the private network according to the cluster configuration information, and then, the cluster resource manager will allocate resources to each node and start scheduling and executing the current job steps;

S3. Job flow scheduling and execution: during the operation of the job, the cluster resource manager will feed back the resource load and job execution status of each node to the cluster master node, and this information will be finally reported to the job flow manager and cluster status monitor of the flexible cluster management center through the master node.The job flow manager monitors and manages the job flow submitted by users according to the job-related configuration information in the cluster configuration manager, which mainly focuses on the management and monitoring of the job flow and the macro scheduling of job steps, while the cluster control manager will ensure the normal operation of each module in the cluster, and restore and restart the failed module;

S4. Redistribution of cluster resources: users redistribute cluster resources in real time during job execution; The user submits the modified cluster configuration information through the console, and the modified cluster configuration information will be submitted to the cluster configuration manager through the user request processor, so as to redistribute resources to the cluster, and automatically adjust the resource distribution for the cluster.The cluster configuration manager, the cluster management controller, the cluster status monitor and the cluster guidance manager work together to achieve the goal of automatic redistribution of cluster resources.

S5, terminating the cluster and obtaining the processing result: after the scheduling of the whole job flow submitted by the user is completed, the cluster master node will delete the cache of the cluster virtual machine node, and copy the processing result of the job and the cluster log to the corresponding external storage set by the user.At this time, the job flow manager will load the configuration information of the cluster configuration manager to call the cluster boot manager to shut down the cluster or keep the cluster running continuously.

In step S1, the best external storage practice takes security, efficiency and scalability as the selection basis of cluster data sources, and selectively creates a cluster operation log storage directory.

In step S4, the redistribution of cluster resources is carried out automatically and manually, and each distribution mode includes coarse-grained telescopic cluster nodes and fine-grained adjustment of the number of processor cores, processing processes and corresponding threads of instance nodes.

By communicating with the cluster master node, the cluster guidance manager of the elastic cluster management center triggers the interface of adding or removing cluster nodes and adjusts the interface of cluster node resources, dynamically adds or removes virtual machines

ISSN: 1813-4890

in coarse granularity to the cluster resources through the debugging interface, or resets the memory capacity of cluster instance nodes and the number of processor cores.Even for some jobs, it can be refined to the number of worker processes and the number of threads contained in each worker process.

The cluster status monitor of the flexible cluster management center will monitor the health status of the cluster in real time, and perform simple analysis, evaluation and visual operation. Cluster nodes can be automatically expanded and contracted through the coordination work of four modules: cluster configuration manager, cluster status monitor, cluster guidance manager and cluster management controller.And dynamically change the memory capacity of cluster instance nodes and the number of processor cores.

The dynamic scaling operation of cluster nodes will take into account the consistency and balance of cluster node data. When there are a large number of idle nodes in the cluster, the data stored in the nodes will be transferred to storage before calling the interface for dynamically removing instance nodes from the cluster. When a new virtual machine node joins the cluster, by default,The cluster will also balance the cluster data.

In step S3, the job flow manager of the flexible cluster management center will schedule and execute the job steps submitted by users in turn, and will reschedule the failed job steps. For the job steps whose execution progress is stagnant, it will additionally schedule and execute the duplicate job steps, and for the jobs that repeatedly fail to execute, it will dump the intermediate execution results.And then mark it as a failure.

The user can make the cluster run continuously by setting the cluster running strategy, or automatically terminate the execution of the cluster after scheduling the whole job flow by setting the job running strategy, and the priority of the job execution strategy is higher than the scheduling strategy of the cluster, and the user can terminate the cluster at any time, provided that the execution of each job step must be cancelled in turn;All automatic or manual operations about jobs will be logged, and administrators can monitor the information through real-time logs to debug the cluster in a targeted manner.

If any of the six modules of the flexible cluster management center fails, the cluster management controller will recover and restart the failed module to ensure the normal operation of the whole cluster.

4. Advantages and Beneficial Effects

1). The platform is composed of the following parts: cluster computing framework, including but not limited to Hadoop map reduce and Apache spark. It is responsible for batch processing, streaming processing and interactive processing of data submitted by users according to different business processing requirements, and ensures that it is compatible with external storage systems (such as Amazon S3, aliyun OSS, cumulus and database, etc.)Efficient coordination of the work;The cluster resource manager can use the open source Apache yarn to manage cluster resources and schedule jobs. It will schedule the resources of each task in a coarse-grained way;The storage system can contain a variety of different file systems. In addition to HDFS, it can also include Amazon S3, aliyun OSS, cumulus, etc., as well as relational or non relational databases, such as HBase and mysql,It also includes the local file system inherent in the node; Flexible cluster management center: it is a combination of six management modules: user request processor, which is responsible for receiving requests from users (management client, management interface and management command line) and forwarding them to corresponding modules; Cluster configuration manager, which is responsible for managing the configuration information of the cluster,This includes cluster virtual machine configuration information, job configuration information, cluster operation strategy, job execution strategy, etc. Cluster boot manager, which is responsible for the startup

ISSN: 1813-4890

and shutdown of cluster virtual machines and computing framework, and the boot operation of other applications, in addition, it is also responsible for calling the interface of cluster resource redistribution; Job flow manager,It is mainly responsible for managing and monitoring the whole job flow according to the job-related configuration submitted by users to the cluster configuration manager. Cluster status monitor, which is responsible for collecting the running status of the whole cluster and performing simple performance analysis, evaluation and display; Cluster management controller, in order to ensure the reliability of the whole flexible cluster management center,It is responsible for monitoring the operation of the above five modules to ensure their normal operation. In addition, it will read the cluster status information of the cluster status monitor, and trigger the resource adjustment interface of the cluster boot management controller by modifying the configuration information of the cluster configuration manager to automatically expand or contract the cluster or dynamically adjust the node resources. Application framework,It includes, but is not limited to, Apache Hive, ApachePig, Spark Streaming library, etc., which are used to build data warehouse, provide higher-level abstract processing and quasi-real-time processing for large data sets, etc.

2). In the method of this design, the communication between the components in the cluster is asynchronous; In addition, it should be pointed out that the reason why external storage is used to store logs and analysis processing results (which does not seem to conform to the idea of data localization), such as Amazon S3, Aliyun OSS, Cumulus, etc., is that the storage cost of these external storage itself is lower than that of instance storage.Moreover, it has the characteristics of high reliability and flexible expansion, which can make users pay more attention to the computing itself, which is also in line with the idea of separation of storage and computing. Finally, the proprietary network of the cluster can accelerate the data transmission efficiency and be safer. This big data processing platform provides various types of data processing and analysis services,And users are allowed to manually scale the cluster and adjust the resource allocation of each node of the cluster according to business requirements, so that the actual data processing process is more efficient, more reliable, more flexible, simpler, safer and more economical.

5. Detailed Description of the Preferred Embodiments

As shown in Figure 1, the main components of this method are described. This big data processing platform consists of the following parts: the cluster computing framework, including but not limited to Hadoop MapReduce and Apache Spark, is responsible for batch processing, streaming processing and interactive processing of data submitted by users according to different business processing requirements.And ensure efficient coordination with external storage systems (such as Amazon S3, Aliyun OSS, Cumulus and database, etc.); Cluster resource manager, which can use open source Apache Yarn to manage cluster resources and schedule jobs, will schedule coarse-grained resources for each task of jobs; Storage system, which can contain many different file systems, besides HDFS,It can also include Amazon S3, Aliyun OSS, Cumulus, relational or non-relational databases, such as HBase, MySQL, etc., and the local file system inherent in the node. Elastic cluster management center: it is a combination of six management modules: user request processor, which is responsible for receiving requests from users (management client, management interface and management command line).And forward it to that corresponding module; Cluster configuration manager, which is responsible for managing cluster configuration information, including cluster virtual machine configuration information, job configuration information, cluster operation strategy, job execution strategy, etc. Cluster boot manager, which is responsible for startup and shutdown of cluster virtual machines and computing framework, and boot operation of other applications. In addition,It is also responsible for calling the interface of cluster resource redistribution; Job

ISSN: 1813-4890

flow manager, which is mainly responsible for managing and monitoring the whole job flow according to the job-related configuration submitted by users to the cluster configuration manager; Cluster status monitor, which is responsible for collecting the running status of the whole cluster and performing simple performance analysis, evaluation and display;Cluster management controller, in order to ensure the reliability of the whole flexible cluster management center, is responsible for monitoring the operation of the above five modules to ensure their normal operation. In addition, it will read the cluster status information of the cluster status monitor,By modifying the configuration information of the cluster configuration manager, the resource adjustment interface of the cluster guidance management controller is triggered to automatically expand or contract the cluster or dynamically adjust the node resources. Application framework, which includes Apache Hive, Apache Pig and Spark Streaming library, etc., is used to build data warehouse, provide higher-level abstract processing and quasi-real-time processing for large data sets, etc.

This method mainly describes the process steps of cluster configuration initialization and cluster creation, job flow submission to cluster, job flow scheduling and execution, cluster resource redistribution, cluster termination and processing result acquisition.Fig. 1 describes the position of each component in a cluster and the cluster configuration consisting of their mutual communication, such as initializing and creating a cluster, submitting a job flow to the cluster, scheduling and executing the job flow, redistributing cluster resources, terminating the cluster and obtaining the processing results.

The purpose of this design is to build a large number platform that provides unified storage, processing and conversion analysis services for the flexible supply of resources in the cloud computing environment and the demand of data storage and processing with different types and functions.

After the user logs in to the cluster management console (interface call, management interface), the flexible cluster with master-slave architecture is created. At the same time, the user needs to provide the location of the input data set and create an external storage directory for storing processing results and log files. In addition, the user needs to initialize the node configuration of the virtual machine cluster,And set the running strategy of the cluster. After that, users can start the cluster, submit a series of job steps (of various types) to the cluster, and set the operation parameters and scheduling strategies of jobs. After that, the flexible cluster management center will be the core to perform a series of data storage and processing, conversion and analysis operations.This request will be submitted to the user request processor module of the flexible cluster management center, and finally the cluster configuration manager will manage the storage. After that, the cluster boot manager will start the cluster computing framework and notify the cluster master node to remotely copy the relevant data files. Before job step scheduling, the cluster master node will design resources to the cluster resource manager,In the process of job step execution, the cluster resource manager will periodically feed back the load information and task information of its node to the cluster master node, and the master node will further send these information to the cluster status monitor and the job flow manager respectively, and the cluster status monitor will simply analyze and quantify the collected information.The job flow manager focuses on the management and scheduling of jobs. For example, once the current job step is finished, it will load the configuration information related to the job step in the cluster management configurator module to decide the scheduling of subsequent job steps, or instruct the cluster boot manager to shut down the cluster.The job manager will repeatedly try to schedule those failed jobs, and provide the functions of competitive jobs and job data consistency. According to the specific situation of cluster job execution,The cooperation of cluster management controller, cluster status monitor, cluster configuration manager and cluster guidance management controller can realize automatic expansion and contraction of cluster nodes and

ISSN: 1813-4890

dynamic adjustment of resources. Finally, when all job steps are scheduled and executed, the whole flexible cluster will automatically shut down the cluster after dumping the output results and log information.

6. Research on Data Storage and Processing Methods in Cloud Computing Environment

Comprises the following steps:

1) Cluster configuration initialization and cluster creation:

Before creating a cluster, users need to create storage directories of input data sets, script files, etc. on external storage, and take security, efficiency and scalability as the best practices of external storage. The purpose of storing cluster input data to external storage is to separate storage from calculation.At the same time, external storage has higher reliability, security and scalability. You can also create a cluster log storage directory, which is optional. If you don't save the cluster logs to external storage, you can view the cluster logs in real time through WebUI during the cluster operation. They will be stored in the temporary directory of the main node of the cluster by default. When the cluster is terminated,Will be deleted. In addition, the database can also be used as the data source and output location of the cluster. For example, you can write MapReduce program to read the unstructured data in HBase, or export the structured data with small amount of data to MySQL.

Then, the user also needs to initialize the cluster configuration and set the running strategy of the cluster, which includes selecting different types of virtual machine nodes with different specifications and selecting the master node, and setting the running strategy of the cluster. Cluster nodes can be divided into three types. The master node is responsible for managing the metadata information of the cluster.In addition, it will save the execution status of tasks and the load information of clusters. The core node is responsible for data calculation and data storage of the cluster. Computing nodes are optional and are used to provide computing resources. When creating a cluster, you can initialize a certain number of core nodes and select the master nodes according to business requirements. In addition, in the specification selection of nodes,It needs to be determined according to the type of job, such as CPU-intensive, memory-intensive or ordinary. Regarding the operation strategy of the cluster, such as the behavior of the cluster after the job flow is scheduled, you can choose to shut down the cluster immediately or keep running continuously.

These initialization information will be submitted to the cluster configuration manager module for storage and management through the user request processor module of the flexible cluster management center.

If the user chooses to start the virtual machine cluster immediately at this time, the cluster boot manager will load the configuration information about the cluster boot operation in the cluster configuration manager at this time, so as to start the virtual machine cluster, allocate the resources of the instance node, and select the master node.

2) Submit job flow to the cluster:

After starting the virtual machine cluster, users can submit a series of job steps to the cluster. When submitting job steps, you need to specify the storage locations of input data sets, script files, cluster logs (optional), and the running parameters of script files. At the same time, users can also set the execution strategy of job steps.This information will be submitted to the user request processor module of the flexible cluster management center, and finally will be handed over to the cluster configuration manager for storage management. About the execution strategy of job step, such as the behavior of cluster when a job step of job flow fails to execute. At this time, the cluster will reschedule this job step by default.If it is the logic error or hardware

ISSN: 1813-4890

failure of the user program that causes the repeated scheduling and execution of the job step to fail, the execution status of this step will be marked as failed. In this case, the default execution strategy of subsequent steps is Cancel. Moreover, at this time, the cluster will dump the intermediate results of the job execution to the designated external storage and make a log record.So that the cluster can continue to schedule execution from this step next time. You can also choose to reset the default policy to cancel dumping the intermediate execution result of the job. In this case, the intermediate execution result of the job will be deleted after the cluster is terminated.

After the user submits the job steps to the cluster, on the one hand, the cluster boot manager module of the flexible cluster management center will transmit the data/script copy command to the cluster master node to copy the externally stored data sets, script files, etc. to all nodes of the cluster, and will also start various background processes of the cluster computing framework and other applications (ApacheHive, Apache pig, etc.).If specified by the user. On the other hand, the job flow manager module loads the job-related configuration information of the cluster configuration manager and communicates with the cluster master node to start scheduling the first job step. At this time, the cluster master node will design resources to the cluster resource manager, which will create an application proxy object for this job.The application proxy object created by will regularly collect the load status and task execution of each node and report it to the cluster master node, which will classify and store the received information in the temporary directory of this node.

3) Job flow scheduling and execution:

When we submit multiple job steps at the same time, they will be scheduled by the job flow manager module of the flexible cluster management center in turn. Generally speaking, the output of the previous operation step will be used as the input of the next step. In the process of job execution, the master node will actively report the job execution information to the job flow manager module.And report the load information of the cluster to the cluster status monitor, which will perform simple analysis, evaluation and visual operation for the administrator to debug the abnormal situation of the cluster, and also serve as the basis for automatic expansion and contraction of the cluster and dynamic adjustment of node resources.

The job manager module will manage and monitor the job flow submitted by users according to the job configuration information in the cluster configuration manager. For example, the intermediate result of a job execution will be directly stored in the file system or memory of this node. Once that schedule of the current job step is successfully executed,The cluster boot manager module is notified to boot the next job step. The cluster boot manager will communicate with the cluster master node to guide it to perform corresponding operations, such as copying the relevant input data of this job step, and then design resources to the cluster resource manager. On the contrary, if the job fails to execute, it will reschedule the job steps by default.After repeated attempts, if the execution still fails, the operation step will be marked as failed, and the subsequent operation steps will be marked as cancelled, and the intermediate results of operation step execution will be dumped. In addition, there may be situations where the execution of job steps stagnates, that is, the job execution progress obtained by the master node from the resource manager has not been refreshed for a long time.At this time, by default, the master node will copy the schedule of this step, that is, redesign resources to the resource manager to reschedule this step, and as long as any one of these steps is completed, the cluster will explicitly terminate the execution of another step. Finally, if all the job steps are scheduled,The job manager will inform the cluster boot manager to terminate the cluster, otherwise the whole cluster will be in a waiting state if the user allows it.

When any sub-module of the flexible cluster management center fails, the cluster control manager will restart the failed module and restore the normal operation of each processing module.

ISSN: 1813-4890

4) Redistribution of cluster resources:

The running nodes in the cluster will periodically report the resource consumption and load of this node to the master node through the resource manager. When the cluster is running, the log information will be stored in the temporary directory of the master node, and will eventually be analyzed, evaluated and displayed by the cluster status monitor of the elastic cluster manager.We can view the load of the cluster in real time through the cluster management and monitoring interface, such as using ApacheGanglia to generate log reports and visualize the performance index data of the whole cluster or a single node instance.

If the workload of jobs submitted by users increases significantly and the load of nodes in the cluster reaches the threshold, the cluster master node will catch that the load of the cluster is abnormal at this time. At this time, the cluster provides an automatic or manual way to reallocate the resources of the cluster. On the one hand, if cluster resources are redistributed manually,The user can request the processor to resubmit the configuration information of cluster resources to cover the relevant configuration information of the cluster configuration manager, and trigger the cluster boot manager to redistribute the interface of cluster resources. If it is the operation of adding nodes, the cluster boot manager will call the interface for dynamically adding virtual machine nodes to the cluster (such as Amazon EC2, OpenStack, Nimbus, etc.).The redistribution of cluster node resources can be regulated from two levels. At the fine-grained level, the memory capacity of nodes and the number of processor cores can be changed. For some jobs, it can even be refined to the number of job processes in nodes and the number of threads contained in each job process. When the node's resource redistribution is completed,The related configuration information of the configuration manager changes, which will trigger the interface of the cluster boot manager to reallocate resources, and it will inform each node to make corresponding resource adjustment. In case of insufficient memory, temporary data exchange will be considered to the file system of this node. At the coarse-grained level,You can manually add or remove cluster nodes, which usually needs to consider the type of job, so as to expand and contract cluster nodes, including core nodes and computing nodes. Newly added nodes need to be registered in the master node.And the application agent object in the resource manager should be used to periodically report its job execution and node load to the master node. On the contrary, the removed node needs to delete its registration information in the master node, and dump its node data, including memory and disk data, before removing in the cluster.This premise is that there are reserved memory or disk resources in the whole cluster. On the other hand, users can also set the cluster to automatic scaling adjustment. The specific process of automatic scaling adjustment is that the cluster control manager will regularly read the execution status of cluster jobs and the load status of the cluster collected by the cluster status monitor.If a certain performance index of the cluster exceeds the threshold set by the user, the cluster control manager will trigger the call of the related interface of the cluster guide manager by modifying the cluster-related configuration information in the cluster configuration manager, and finally achieve the purpose of scaling the cluster or adjusting the cluster resource allocation.By automatically or manually scaling the cluster and adjusting the resource allocation of cluster nodes, the efficiency of job processing and the utilization of resources can be improved, and the scaling operation of the cluster or the adjustment of node resources will be recorded as a log.

5) Terminate the cluster and obtain the processing result:

After all the job steps submitted by users are scheduled, the job flow manager module of the flexible cluster management center will read the job execution strategy and cluster operation strategy of the cluster configuration manager module, and the job execution strategy takes precedence over the cluster operation strategy. Therefore, in general, in order to avoid wasting cluster resources,It can be set to terminate the cluster operation after all job steps are executed.

ISSN: 1813-4890

7. Conclusion

The boot manager will inform the master node of the cluster to release the resources of the cluster instance, and notify the cluster resource manager to stop working, and each node instance will delete the meta information of its own node from the master node. Once the cluster is terminated, the information of all modules in the flexible cluster management center will be deleted. In addition, if the cluster is scheduling jobs,Or there is a job waiting to be scheduled. At this time, the cluster cannot be terminated directly. You must first terminate the execution of the job or cancel the submission of the job steps.

We can obtain the analysis results and log information of job flow processing from external storage, or hand it over to front-end components for visual processing.

References

[1] Research on application strategy of cloud computing in computer data storage technology [J]. Zhou Feng. China New Communication. 2020(05).

[2] Research on data storage technology based on computer cloud computing [J]. Li Gen. Fujian Tea. 2020(01).

[3] Research on data storage technology based on cloud computing [J]. Shang Hua. Information and Computer (Theoretical Edition). 2020(15).

[4] Analysis of data storage technology in computer cloud computing [J]. Chang Chunyan. Electronic Production .2020 (22).

[5] Research on data storage technology of twin-screw ship positioning [J]. Bai Yanqing, Lu Xiaofang. Ship Science and Technology .2020 (04).

[6] Development of data storage technology in the era of big data [J]. Bai Hongjun, Xia Jian, Lynne. Electronic Technology and Software Engineering. 2017(04).

[7] Research on data storage technology based on cloud computing [J]. Liu Yuchang, Xie Lei. Digital technology and application. 2017(06).

[8] Analysis and Application Research of Network Data Storage Technology [J]. Sun Wei. China New Communication. 2014(06).

[9] Talking about network data storage technology and its application [J]. Kong Qingyu, Yu Qiang, Yan Binbin. Computer CD software and application. 2014(12).

[10] Analysis of data storage technology in cloud computing environment [J]. Chen Fengping. Silicon Valley. 2013(17).