Construction and Key Technology of Coal Geology Cloud Based on Openstack

Lintian Miao\textsuperscript{1,2,*}, Zhonghui Duan\textsuperscript{1}, Tingting Lv\textsuperscript{1}, Jianjun Zhang\textsuperscript{1} and Yang Yang\textsuperscript{3}

\textsuperscript{1}Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi’an 710021, China
\textsuperscript{2}College of Geology and Environment, Xi’an University of Science and Technology, Xi’an 710054, China
\textsuperscript{3}Beijing Longruan Technology Co., Ltd., Beijing 100190, China

*Corresponding author email: mlt@mtdz.com

Abstract. In order to promote the information and digital construction of coal geological work, promote the rapid development of green coal mines and smart coal mines. In this study, we developed the Coal Geology Cloud (CGC) using key technologies such as elastic computing, mass storage of map data, and cloud message queues. The research and development process is based on the secondary development of Openstack and mature cloud computing platform technology. The researchers adopted the four-layer technical architecture of IaaS, PaaS, SaaS and Client. At the same time, it combines coal mine GIS series software and cloud computing IaaS platform to form an object-oriented Coal Geology Cloud GIS solution. CGC realizes the collection, aggregation, storage, analysis, processing and display of coal geology and related information, and provides convenient online cloud services for the exploration, development, research and management of coal resources.

Keywords: Coal geology cloud; Internet + coal geology; Openstack; Cloud computing.

1. Introduction
Coal geology provides resource guarantee and geological guarantee for the development of coal industry. With the rapid development of Internet, cloud computing and big data, it is an inevitable trend for coal geology to enter the new era of "Internet +". More and more scholars pay attention to the importance of the combination of information technology and geological industry. China Geological Survey established Geocloud \cite{1,2}, comprehensively studied the characteristics of geological data and key technologies for storage management \cite{3,4}, deeply explored the construction of geological cloud framework supported by big data, as well as the application of Internet of things and cloud computing in China's geological survey work, and put forward a tentative idea and development vision of data collection, geological big data analysis and geological cloud computing service in future geological survey work \cite{5,6}. With regard to the application of high and new technologies such as the Internet of things, big data, cloud computing and artificial intelligence in coal mine work, Lu Xinmin \cite{7} studied the construction architecture, key technologies, service modes and status functions of cloud computing platforms in the Internet of things in mines. Wuqiang \cite{8} academician team adopted hybrid cloud architecture to build a mine flood intelligent emergency rescue service system. However, a complete cloud computing platform specifically for coal geological work has not been reported. Based on this, the key Laboratory of Coal Resources Exploration and Comprehensive Utilization of the Ministry of
Natural Resources of China has built a coal geological cloud. Based on unified coal geological data warehouse and cloud computing technology, coal geological cloud provides a variety of cloud services for coal geological work and realizes information, networking and intelligent services. It mainly serves the coal industry management departments, geological survey units, mining companies and scientific research institutes. CGC adopts OpenStack cloud computing framework and OGC standard, follows the technical system of cloud platform, adopts the design idea of virtualization, flexibility and service, and designs according to the "four-tier" cloud platform architecture based on the data model of spatio-temporal information database. Based on mature cloud computing platform technologies such as Fusion Compute, vSphere, OpenStack, Docker, together with coal mine GIS application server software, coal mine GIS distribution server software, coal mine GIS mapping software, coal mine GIS end software, cloud computing GIS platform, etc., constitute an enterprise-oriented mine cloud GIS solution. China's first "Internet + Coal Geology" cloud service platform has been completed, which is of great practical significance to improve the modern management level of coal mine safety and production technology, reduce the occurrence of production safety accidents, and accurately manage coal resources.

2. System Architecture and Functional Modules

2.1. Overall Design Requirements
Based on the reality of coal geological work, as far as possible to meet the information needs of coal geological work, and fully considering the needs of future development, a unified planning of the system has been made, the specific requirements are as follows:

2.1.1. Progressiveness. Choose mainstream GIS technology, computer network technology, development technology and tools, and database storage technology.

2.1.2. Reliability and Stability. The design takes into account the fault-tolerant design and error handling mechanism, so that the data can be obtained stably and the information can be expressed reliably.

2.1.3. Easy to Operate and Maintain. The system provides a good interactive, easy to use visual user interface.

2.1.4. Extensibility. Considering the development trend and needs in the future, the system construction can meet the needs of function expansion and version upgrade in the future.

2.1.5. Openness and sharing. CGC is an open, shared system. Meet different application needs. CGC uses internationally recognized standardized technologies.

2.1.6. Security. The system is planned and deployed in strict accordance with the provisions and requirements of the state on information security. Use advanced anti-virus technology and products to monitor the activities of computer network viruses and prevent the destruction of viruses. All network security software and hardware must be certified by the national security agency to ensure the security of the system and data.

2.2. System Logic Architecture
The logical architecture of the system is divided into four layers from the bottom up: resource layer, virtualization layer, management layer and service layer[9]. This is shown in figure 1.
The resource layer is the material basis for the construction and operation of the software platform, and provides the physical infrastructure for the operation of various platforms and service components, including servers, networks, storage and other hardware devices. The virtualization layer adopts KVM, Xen, LXC, QEMU, vSphere, hyper-v, Docker and other technologies to provide virtual cloud server, which can provide rich instance specifications (CPU, memory) and bandwidth, storage disk selection, and upgrade bandwidth, CPU and memory at any time, so as to support the sustainable development of business. For high concurrent read-write and high performance requirements, I/O optimization using SSD cloud disk technology can also be selected. Management layer is to achieve the physical resource and the bearing of the unity of the virtual resources, efficient and intelligent management, the administrator can in addition to efficient management on a unified interface to physical equipment resources, also can each mainstream virtualization platform of virtual resource management, at the same time can cloud infrastructure services in resource scheduling, planning, alarm, since the restoration of provide Suggestions based on the strategy of a certain. Service layer to provide users with all kinds of business online customization platform application system development, regardless of business users during the deployment or run business applications, all don't need to care about the underlying server, operating system, the GIS platform of the construction of the infrastructure and operational, also need not care about data, analyze the function of creating and updating resources, can put more energy into the upper business application innovation.

2.3. System Technical Architecture

The technical architecture follows the technical system of cloud Platform and adopts the design ideas of virtualization, flexibility and servitization. Based on the data model of spatio-temporal information database, the design is carried out in accordance with the "four-layer" cloud Platform technical architecture system[10]. From bottom to top, the technology architecture is Infrastructure as a service(Iaas), Platform as a service(Paas), Software as a service(Saas) and Client. And through the management and scheduling system through multiple layers to achieve the collaborative work between the layers. The application integration of the platform is realized to provide data resource services and functional services for various business applications. This is shown in figure 2.
2.4. Deployment Architecture
CGC is developed based on Openstack secondary. The basic cloud has 6 nodes, including 1 controller node, 1 network node, 3 compute nodes and 1 cinder node, and the capacity can be dynamically expanded at any time as needed. The deployment architecture is shown in figure 3.

Control node (Controller): this node runs identity authentication, image management, compute node control, network broker, and Dashboard related services. To support these services, this node installs services such as SQL, MQ, and NTP.
Network node: this node mainly runs Network agents and manages the services of the host Network.
Compute nodes: compute nodes run cloud host instances, using the KVM virtualization engine by default, and Compute nodes have network proxy services installed to connect instances to the virtual network through network proxies.
Block storage node (Cinder): this node provides block storage or file sharing services for each instance, such as the ability to add hard disk space to the host, which is partitioned from this node.

2.5. Function Module
Coal geological cloud (CGC) service platform is divided into two parts: cloud management platform and cloud service platform. The cloud management platform includes three parts: project, administrator and identity management. The cloud service platform includes ten parts, such as CGC navigation, coal management, geological service, resource sharing, software service, mobile APP, CGC cloud disk, equipment and instrument, information release and cloud community. The functional modules of each part are shown in figure 4.
3. The Construction of Coal Geological Cloud Service Platform
The coal geological cloud service platform refers to the technical platform that supports the operation of the coal mine GIS kernel with cloud service and big data computing architecture and the serialized coal mine geological guarantee application system.

The business application product model based on the geological cloud service platform is shown in figure 5.

**Figure 4.** Coal geology cloud function module.
CGC can simplify the business environment of GIS and the establishment of the load balancer, the operational process, reduce the GIS professionals deployment, and management of GIS business environment and the difficulty of the load balancer, can real-time monitoring the GIS business environment of CPU, memory, and network I/O usage, through the topology and monitoring, alarm mechanism, can real-time monitoring the running state of GIS environment. The platform also provides function of load balancing to improve the performance of the common platform. In addition, the platform can integrate with customers' existing information platform solutions to help users quickly build their own GIS private cloud system.

The hardware requirements required to deploy a basic OpenStack platform experimental environment are shown in figure 6.

**Figure 5.** Cloud service platform system service pattern.

**Figure 6.** Hardware requirements of OpenStack.
To deploy a runnable OpenStack experimental environment, one control node and one computing node should be configured at least. The control node should be configured at least with 1-2 cpus, 8GB of memory, 100GB of disk storage space, and 2 network cards. The computing node should be configured with 2-4 cpus, more than 8GB of memory, more than 100GB of storage space, and 2 network CARDS. The remaining block storage nodes and object storage nodes are optional. CGC has deployed 1 block storage node and 2 object storage nodes, corresponding hardware configuration is 1-2 cpus, 8GB of memory, more than 500GB of storage space, and each node has a network card.

Controller: the control node runs the identity authentication service, the mirror service, the management part of the computing service, the management part of the network service, various network agents and the dashboard. Support services such as SQL databases, message queues, and NTP (Network Time Protocol) also need to be included.

Calculation: runs the hypervisor portion of the management instance in the calculation service on the calculation node. By default, computing services use KVM. More than one compute node can be deployed, requiring at least two network cards per compute node.

Block storage: block storage nodes contain disks, block storage services, and Shared file systems. Block storage contains a separate storage network to enhance network performance and security.

Object storage: the optional object storage node contains the disk. The object storage service uses these disks to store accounts, containers, and objects. A separate storage network is deployed by CGC to enhance performance and security. This service requires two nodes, Each node requires at least one network card. More than two object storage nodes can be deployed.

Network: the public network USES the simplest possible way to deploy OpenStack network services, primarily through "bridge/switch" services and the partitioning of VLAN networks. Essentially, it is a bridge that builds virtual networks to physical networks and relies on the physical network infrastructure to provide routing services.

OpenStack provides a web-based self-service portal. After the installation and configuration of all the components are completed, all the services can be started and the OpenStack platform information can be viewed on the Web interface.

4. Key Technologies

Elastic computing is a simple and efficient computing service with flexible and scalable processing capacity, which can quickly build more stable and secure applications to improve operation and maintenance efficiency, reduce IT costs, and enable users to focus more on core business innovation [11].

4.1. The Technology of Elastic Computing

Elastic computing is provided to users in the form of cloud hosts (cloud servers). Cloud host is an important part of computing in infrastructure application, located at the bottom of the cloud computing industry chain pyramid, and its products are derived from the cloud computing platform. IT integrates computing, storage and network resources with IT infrastructure capacity rental services, providing on-demand use and pay-as-you-need server rental services based on the cloud computing model. Customers can deploy the required server environment through a self-service platform with a Web interface.

Cloud host is realized through virtualization technology. Currently, virtualization technologies mainly include: KVM, Xen, LXC, QEMU, vSphere, hyper-v, and Docker. Cloud server can provide rich instance specification (CPU, memory) and bandwidth, storage disk choice, also can upgrade bandwidth, upgrade CPU and memory at any time, to support the continuous development of business. For high concurrent read-write and high performance requirements, I/O optimization using SSD cloud disk technology can also be selected. The user has complete control over the cloud server, and can connect the terminal remotely to solve the system problems and carry out various operations just like the traditional server, also can restart the server, reset the login password, network isolation and other operations through the administrative console, API, command line and similar tools.
4.2. The Technology of Mass Storage of Map Data
Openstack block storage and SQL Server database distributed storage are adopted in the construction of coal geological cloud platform, and interface access is provided through LRSDE.

4.2.1. Cinder Block Storage. Cinder block storage is the basis for storing virtual machine image files and data used by virtual machines. The Cinder interface allows you to create and attach block devices to virtual machines, such as "create volumes", "delete volumes", and "attach volumes". And supports the ability to expand capacity, snapshot and create virtual machine image clones. CGC map files are stored by using OpenStack Cinder component blocks to achieve elastic storage. When the current resource is insufficient, the LRSDE low-level call creates Cinder dynamically and links to the resource free cloud host [12].

4.2.2. Master Data. Compared with operation type data, master data is relatively stable data, which is used in GIS to define geometric objects, layer information, attribute information, etc. Map data indexes are stored in master data, such as geometric entity ID index, attribute ID index, map index, topological relationship, etc.

4.2.3. Metadata. Metadata is mainly used to store the data with a large increase in the map, such as geometric data, change data, attribute data. When resources are tight, LRSDE calls to create a cloud host interface, and creates a cloud host with SQLServer to store the data in the future.

4.3. The Technology of Cloud Message Queuing
Message queuing (MQ) is an application-to-application communication method. Messaging refers to the communication between programs by sending data in messages, while queuing refers to the communication between applications through queues [13]. The cloud platform product develops a set of cloud service framework based on the excellent open source component RabbitMQ, which can quickly realize cloud computing for synchronous requests and greatly save server response time for asynchronous processing, thus improving the throughput of the system.

The RabbitMQ simple architecture is shown in figure 7. The Client and the Server do not communicate directly, but send messages through the Client to the switch Exchange in the RabbitMQ Server. After the Exchange receives the messages, it forwards them to the message Queue via the corresponding Routing rules, and the Queue forwards the received messages to the Server that subscribes to the Queue.

![RabbitMQ architecture diagram](image)

RabbitMQ Server: also known as broker Server, it is a transport service. His role is to maintain a route from Producer to Consumer, ensuring that data can be transferred in a specified way.

Client: also called Producer, the sender of data. A Message has two parts: payload and label. Payload is the data that's being transmitted. Label is the name of exchange or a tag that describes the payload, and RabbitMQ also USES this label to decide which Consumer to send the Message to. The AMQP only describes the label, while RabbitMQ determines how to use the rules of this label.

Figure 7. RabbitMQ architecture diagram.
Server: also called Consumer, the receiver of data. Think of a queue as a mailbox with a name. When a Message arrives in a mailbox, RabbitMQ sends it to one of its subscribers, the Consumer. Of course, you might send the same Message to many consumers. In this Message, there's only payload, the label has been deleted. For Consumer, it doesn't know who sent the message, and the protocol itself doesn't support it. But if the payload sent by a Producer contains information from the Producer, that's another story.

Queue: is the internal object of RabbitMQ, which is used to store messages, Messages in RabbitMQ can only be stored in Queue. The producer produces the message and sends it to the Queue, from which the consumer can get the message and consume it.

Exchange: a producer sends messages to an Exchange, which routes the messages to one or more queues (or drops them). When sending a message to Exchange, the producer specifies a routing key to specify the routing rules for the message. RabbitMQ limits the length of the message to 255 bytes. If the Exchange Type is fixed with a binding key, the producer can specify a routing key when sending a message to Exchange to determine where the message goes.

Through RabbitMQ, the system provides two task distribution mechanisms, one is a simple task polling distribution and the other is a fair distribution based on the working status of a single service. Through the task distribution mechanism, the system can achieve good resource elasticity expansion. Amount when the task is too much, the existing model service load increase, the system can allocate corresponding resources according to the requirements to run a new service model. The new model service only needs to bind the message queue corresponding to the model to obtain the model calculation request, so as to reduce the load of other model services and speed up the task processing [14].

5. Conclusion
The "Internet + coal geology" professional cloud service platform (CGC) was established based on the secondary development of Openstack. It realizes the collection, aggregation, storage, analysis, processing and display of coal geology and related information, and provides convenient online cloud services for the exploration, development, research and management of coal resources. The coal geological cloud builds a data model based on spatio-temporal information database, and applies some key technologies such as elastic computing, massive storage of map data and cloud message queue technology according to the "four-tier" cloud platform architecture system. The cloud service platform provide a full range of online applications for coal enterprises, coal geological exploration units and scientific research institutes through CGC navigation, coal management, geological services, resource sharing, software services, mobile APP, CGC cloud disk, equipment and instruments, information release and other functional modules. With the continuous maturity and development of openstack and big data analysis technology, CGC is also constantly upgrading. The dynamic updating of coal geological data driven by energy demand will become the driving force for the sustainable development of CGC. The sustainable development of openstack provides a technical guarantee for the sustainable development of CGC.

Acknowledgments
This research was supported by Innovation capability support Plan Project of Shaanxi Provincial, program No.2018PT-23, and major science and technology project from the Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, China, program No.ZD2018-2, and Special Foundation of technology innovation guidance of shaanxi province, program No.2019CGHJ-20. We thank the academic editors and anonymous reviewers for their kind suggestions and valuable comments.

References
[1] Wang Yongzhi, Bao Xiaodong, Miao Jinli, et al. Construction and application of GeoCloud monitoring platform based on big data[J]. Progress in Geophysics. 2018. 33(2): 0850–0859.
[2] He Wenna, Wang Yongzhi. Prototype system of geological cloud computing[J]. Progress in Geophysics, 2014, 29(6): 2886–2896.
[3] Chen Jianping, Li Jing, Cui Ning, et al. The construction and application of geological cloud under the big data background [J]. Geological Bulletin of China, 2015, 34(7):1260–1265.

[4] Li Jing, Chen Jianping, Wang Xiang. A study of the storage technology of geological big data [J]. Geological Bulletin of China, 2015, 34(8):1589–1594.

[5] Tan Yongjie, Qu Honggang, Wen Min. On Big Data of Geological Survey [J]. Geomatics World, 2018, 25(2):7-11.

[6] Tan Yongjie. Architecture investigation of the construction of geological big data system [J]. Geological Survey Of China. 2016, 3(3):1-6.

[7] Lu Ximing, Kan Shuting, Zhang Xingli. Cloud computation and platform technology for mine Internet of things [J]. Industry and Mine Automation, 2018, 44(2):1-5.

[8] Wu Qiang, Xu Hua, Zhao Yingwang, et al. Cloud-based smart emergency rescue system and its application in mine water disaster [J]. Journal of China Coal Society, 2018, 43(10):2661-2667.

[9] YU Nai-Bo. Investigation of IaaS Mode [J]. Telecommunications Science, 2011, 27(S1):39-43.

[10] Yang Guang. Design and build an IaaS platform basing on the Cloud computing [D]. Chengdu University of Technology, 2012.

[11] Li YingZhuang, Jiang Nan, Li XianYi. Design and Implementation of Developed Cloud Platform Based on OpenStack [J]. Journal of Wuhan University (Natural Science Edition), 2012, 58(S1):58-60.

[12] Xiong ZhenHua. The Research of Cloud Storage Technology Based on OpenStack [D]. Jilin University, 2014.

[13] Cheng DaYu, Chen Jun, Han Gang, Wang FaLiang. Study on interoperability in synchronous cooperative GIS based on message queue [J]. Journal of China University of Mining & Technology, 2012, 41(01):127-132.

[14] Xu Zhen, Jiao WenBin. Optimization of Tiny Message Acknowledgement in RabbitMQ [J]. Computer Systems & Applications, 2018, 27(03):252-257.