Attempt of the implementation of spatial information extension for the data catalog server CKAN and release its service using cloud computing environment

Yoichi. Kayama[1],, Shingo. Otomo[1], Dai. Yagishita[1]

[1]Geospatial Information Laboratory, Aeroasahi Corporation, Kawagoe, Japan

Keywords: Spatial data infrastructure, Cloud computing, Meta data catalog

1. About CKAN

CKAN is an open source system for searching the metadata catalog. CKAN has been used as a portal site to publish the government open data in many countries, including the United Kingdom and the United States. Then CKAN has been used for the site such as data.go.jp and clearing house of GSI in Japan. For CSV and PDF format data in CKAN has a preview function. And there are also a lot of spatial data that can be previewed in CKAN. Spatial search can also be performed by creating a schema of spatial information in the search engine Solr used in CKAN. In this presentation in order to contribute to the further utilization of spatial information, we write about the spatial information extension of CKAN , made with reference to the space information expansion of the UK and the Geographical Survey Institute site. The extension has functions of spatial data processing and spatial data preview and WEB map cooperation. Most of the functions of CKAN can be used from the API. There is a plugin that calls CKAN API for QGIS and can search the catalog and use the CKAN spatial information. It is possible to construct a spatial information infrastructure by publishing catalogs of spatial data and endpoints for archiving in CKAN.

2. Construction of G-space information center

We have used these functions to construct the G-space information center which is the SDI of Japan.The G-space information center was developed as a project of the Japanese government. It is running in Amazon cloud environment. We were in charge of development and operation of this platform. In this case, catalog server using CKAN, data storage using PostGIS / PostgreSQL, and WEB map service using GeoServer are implemented. We also implement tile map delivehry service using Amazon S3 and CloudFront.

3. Use cloud computing environment

By using Amazon's cloud computing, the G-space information center has enjoyed various conveniences. The amount of computer resources (CPU, memory, disk) required by the G-space information center does not know how much things should be arranged at the start of service. It is difficult to estimate how many people will access per unit time or how much disks should be prepared for future data growth at the start of service. In the cloud computing environment, such computer resources are virtualized and it is possible to flexibly change after the service is started. Particularly, if you prepare map tiles data for distribution in a wide area with high resolution, you occupy a large amount of disk. Therefore, even if a large capacity disk is prepared, most of the disk usage fee is wasteful expense because the tiles are not maintained at the beginning so the amount of disks to occupy is small. Even if you prepare a large capacity disk for tile delivery, you will run out of space on the disk you prepared if the type of drawing you want to distribute tiles increases in the future. In the cloud computing environment, there are multiple disk usage methods As in the conventional method, when using the file system which prepares the disk space from the beginning, estimate the future necessary area as the file system and create the virtual disk Initially it occupies unnecessary areas and it also costs unnecessary costs. When using this method the capacity of the allocated disk is also finite, so if the file capacity increases the disc

[1] Corresponding Author:Yoichi Kayama
Email address: yoichi.kayama@gmail.com

may become full in the future. In the object storage provided in the cloud computing environment, the fee is charged according to the disk capacity used by the file, so there is no charge for the unused area. Also, there is no upper limit on the amount of available disk space in object storage, so we do not need to prepare disks assuming future disk usage when building the system. In the cloud computing environment, it is also easy to duplicate the created virtual machine, so it becomes easier to perform virtual machine multiplexing according to load increase. When software maintenance of the virtual machine is performed it is useful to perform maintenance on the duplicated virtual machine instance. By switching the connection destination of the load balancer on the front of the production environment instance from the production environment to the maintenance instance duplicated from the production environment when maintenance on the replication machine is completed, it is possible to make the maintenance virtual machine into the production environment. By using this method, you can perform virtual machine maintenance work without stopping the service. There are many advantages to using the cloud environment for publishing Internet services. In this case, using open source software to configure the server enhances the advantage of using the cloud computing environment. The virtual server image can be easily duplicated, but if it contains licensed software that is not open source, you may need to procure the license each time you copy the virtual server. If you are building a virtual server with open licensed software, software licensing problems are less likely to occur when creating a new replication instance.

The cost of constructing a computer system in recent years is steadily decreasing. Open source software and cloud computing are great tools as a tool to build sophisticated systems at low cost.

4. References

CKAN, the world's leading Open Source data portal platform. Retrieved from https://ckan.org/

G-space information center. https://www.geospatial.jp/gp_front/

Kayama, Y.Otomo,S. Yagishita,D.Takahashi,Y. Kitashima,S & Sudiu,T . (2016). Attempt of the implementation of spatial information extension for the data catalog server CKAN. *Academic conference proceedings of Geographical Information System Association Japan 2016*

5. Author/s Biography

Yoichi Kayama is a steering committee member of OSGeo.JP, a charter member of OSGeo.org. For more than 30 years I have been doing program development for GIS for local governments in Japan. Since joining Aeroasahi corporation in 2006, I was involved in the investigation of FOSS4G and promotion of use in Japan. We localize QGIS to Japanese and disseminate it in Japan. And I am working at the SIG of FOSS4G in the GIS Society Japan. Every year I am actively involved in domestic and world FOSS4G conferences. I was in charge of development and operation of G-space information center and learned system construction in the cloud computing environment. As a member for Code for Japan, we are promoting open data, open goverment, civic tech . As a member of ITDART, I am also involved in activities to support backwards using IT at the time of disaster. Facebook page https://www.facebook.com/yoichi.kayama