Real-Time Search in Clouds

Misbah Uddin, Amy Skinner, Rolf Stadler
ACCESS Linnaeus Center
KTH Royal Institute of Technology
Email: {ahmmud,skinner,stadler}@kth.se

Alexander Clemm
Cisco Systems
San Jose, California, USA
Email: alex@cisco.com

We developed a novel approach for management of networks/networked systems based on *network search* [4]. Network search provides a simple, uniform interface, through which human administrators and management applications can obtain network information, configuration or operational, without knowing its schema and location. We believe that the capability of network search will spur the development of new tools for human administrators and enable the rapid development of new classes of network control functions and applications.

In [5] we present an object model, a query language and distributed query processing functions for network search. We model network objects as bags of attribute-value pairs and identify them by unique names. The model allows search queries with minimal knowledge about information structure. The query language is composed of a few-easy-to understand rules for search queries, from simple free text queries to complex ones that include operators for projection and aggregation. In addition, it provides operators to find related objects by discovering their link relations. Apart from projection and aggregation queries, all search queries return both exactly and approximately matched objects, which are ranked by a scheme that is based on boolean information retrieval [2]. The queries are matched and ranked in a distributed fashion, using an echo protocol [3], which allows a network search system to scale to beyond 100,000 nodes.

In this demonstration, we present CloudSearch, a search system for an IaaS cloud. The system implements the models and functions mentioned above and is based on the architecture in Figure 1. The search system is realized as a network of search nodes with identical capabilities, whereby the nodes run as daemons in the servers that provide cloud services. The nodes form a search plane, which realizes in-network distributed search using an echo protocol. An administrator can access configuration/operational data in the cloud using a web-based search interface in a management station.

CloudSearch is realized on a federated cloud that provides IaaS (see Figure 2) using OpenStack [1]. The cloud spans to three remote sites in Stockholm (two in different campuses of The Royal Institute of Technology and one at the Swedish Institute of Computer Science). The cloud infrastructure is used for cloud-based research projects. Collectively, the sites include 24 high-end servers (each with 24 cores and 40-64 GB memory). Our search system contains 24 search nodes, one per server. Each node stores configuration and operational information about cloud components, which are retrieved by invoking search queries to a web-based search interface (see Figure 3), which runs on a management station that has access to the servers.

The purpose of this demonstration is to illustrate that compared to existing technology, our system enables easier access and more flexible discovery of information, which enables a better understanding of the operational status of the cloud infrastructure and thus helps the development of novel management functions for clouds. We will demonstrate this by the following use cases.

- A researcher is running an elastic key-value storage in the cloud. She inquires about the number of virtual machines of the application to estimate the resource demand.
- A programmer is running a client-server application in the cloud. He inquires about the packet exchanges between the client and the server to study the communication load of the application.
- A customer complained about degrading performances. To find the root cause of the problem, a system administrator inquires about servers that run virtual machines of the customer.

REFERENCES

[1] openstack. "http://www.openstack.org", January 2013.

[2] Joon Ho Lee. Properties of extended boolean models in information retrieval. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '94, pages 182–190, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[3] Rolf Stadler. Protocols for distributed management. Technical Report 2012:028, KTH, Communication Networks, 2012. QC 20120604.

[4] Misbah Uddin, Rolf Stadler, and Alexander Clemm. Management by network search. In *Proceedings of the 2012 IEEE Network Operations and Management Symposium (NOMS 2012), Maui, Hawaii, April 16 - 20*, 2012.

978-3-901882-50-0

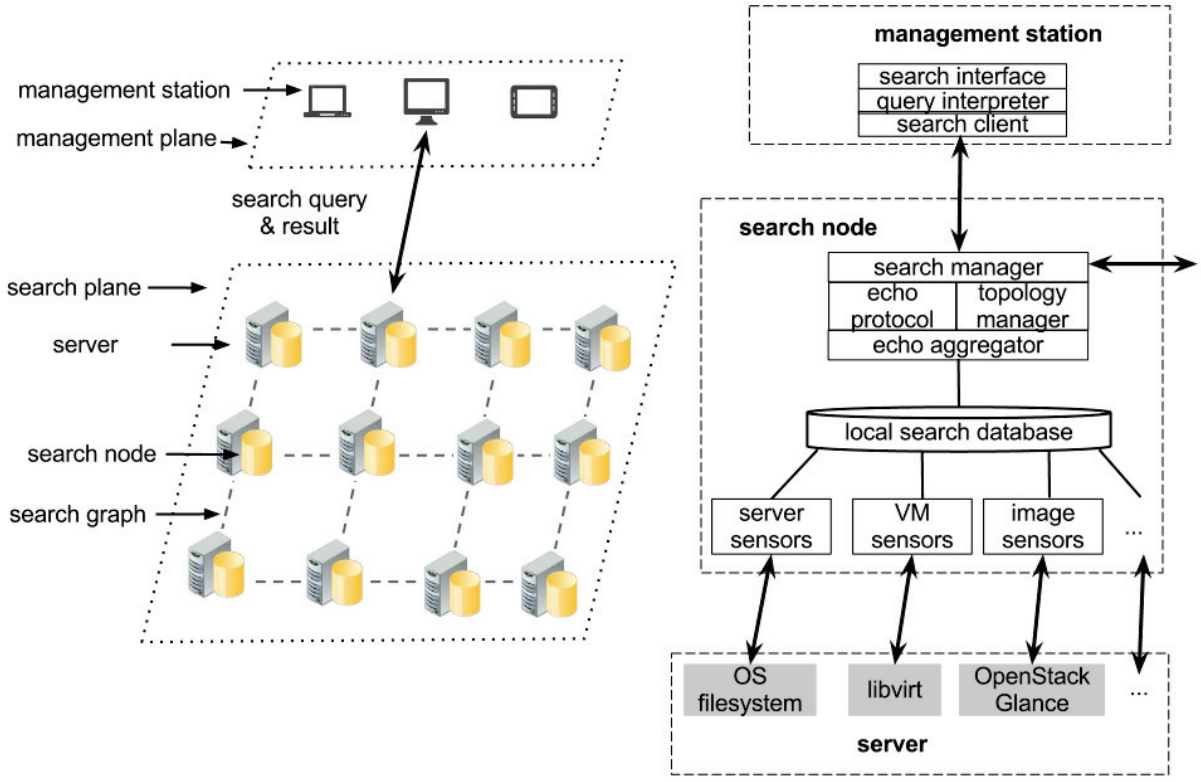

Fig. 1. Architecture of CloudSearch

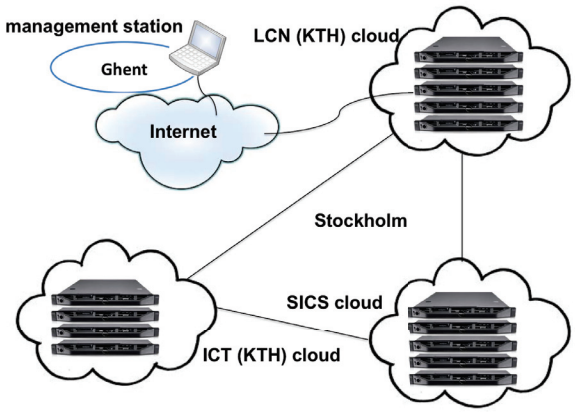

Fig. 2. The testbed for CloudSearch and demo setup

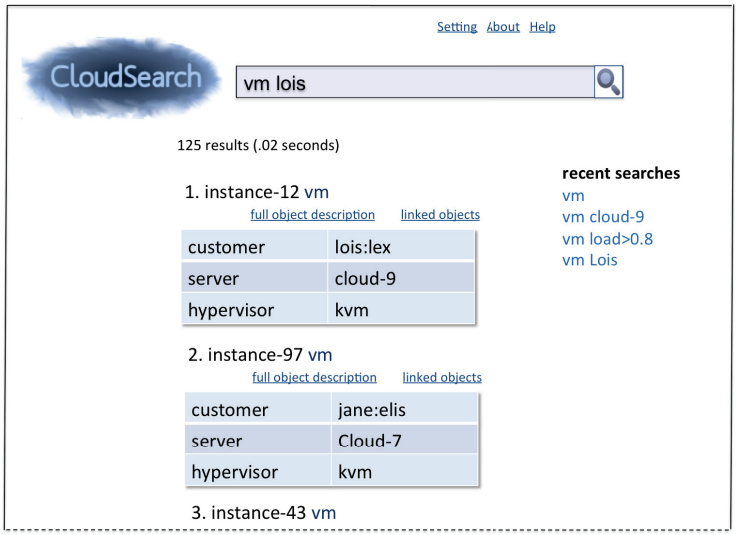

Fig. 3. CloudSearch interface

[5] Misbah Uddin, Rolf Stadler, and Alexander Clemm. A query language for network search. In *Proceedings of the 2013 IFIP/IEEE Symposium on Integrated Network Management (IM 2013), Ghen, Belgium, May 27-31*, 2013.