

Command Line Web Portal for multiple physical and virtual clusters on Scientific computing Platform

Tuesday, 25 August 2020 14:40 (20 minutes)

In recent years, many supercomputers and clusters have been deployed and provide massive High Performance Computing (HPC) resources for users. As a virtual cluster can provide similar efficiency of physical computing resource, many users choose virtual clusters to execute HPC jobs for satisfying different kinds of requirements. Due to the advantages of low resource occupation, fast execution speed and general-purpose, Command Line (CL) is still an important interactive interface for accessing physical or virtual resources in scientific computing.

Different from desktop and Web graphical interfaces, it is difficult and annoying for users to learn, remember and use lot of commands properly, especially in a scientific computing platform composed of multiple management domains that have multiple supercomputers and clusters. There are 3 important and annoying issues with using CL in scientific computing. In network security, most HPC resources must be accessed through dedicated VPN networks, or that a SSH client must be running on the specified server whose IP is permissioned by all network firewalls that locate on the route path between the client and the targeted login server. In credential security, a user must install a client in different operating systems and platforms, configure a certificate in the target client, and input complex password information and the other authentication information from a smart card or other hardware devices. In configuration management, a user needs to repeat complicated and cumbersome works as done in network and credential security when the user wants to access a new cluster by CL. The frequent creation and destruction of virtual clusters makes it even more difficult.

Focusing on solving above issues, Command Line WEB Portal (CLWP) was proposed in this paper which provides easy-to-use access to multiple physical and virtual computing resources. The entire process of command execution consists of 3 interconnected message loops, a Web Socket (WS) circle, an Event Bus (EB) circle and a Secure Shell (SSH) circle. The WS-circle consists of xterm.js located on the browser, and a pair of WS client and server which are located separately on the front-end browser and on the back-end server. The WS-circle receives a command each time from a user and displays output of each command. The SSH-circle consists of a pair of SSH client and server which are located separately on the back-end server and a login server of physical or virtual clusters. The SSH-circle connects the specified login server by VPN network or satisfying certain firewall rules, then receives a command from the EB-circle and returns output of the command line by line. The EB-circle consists of an input chain and an output chain, which are core components in the entire process of command execution. The input chain is responsible for receiving commands letter by letter, modifying and checking commands on syntax, format and security attributes by different filters in the chain. The output chain is responsible for help users understanding output of commands from job life-cycle, cluster status and other aspects. Each command inputted from browser will go through the entire process consisted of 3 circles described above to provide easy-to-use command line service for users.

In addition, CLWP provides account registration service for users to add information about new HPC configuration and new account credential into persistence storage in an intuitive way. CLWP can help users authenticate into a login server by account/password and account/certificate credentials. CLWP also provides SSH login service for virtual clusters resided in HPC resources which are prohibited to access internet for security. In login process, SSH-cycle firstly through the fixed port logins to any login server on front of HPC computing resources that container virtual clusters, and then logins to the target virtual cluster from any specified port.

The anonymous account is the unique identify of each user in CLWP, which records different group of information, such as user, organization, research projects, and computing requirements. The anonymous account is mapped up to the 3rd Identify Provider (IdP), such as China Science and Technology network passport (CST-passport), wechat and similar public IdPs. The anonymous account is also mapped down to one or more accounts belong to a user from different HPC resources or virtual clusters. A user logins to CLWP by an up mapping account such as CST-passport but not an anonymous account. Then the user selects any physical or virtual HPC resources and a local account from the down mapping accounts on the target HPC resource, and accesses the target HPC resource based on web command line service provided by CLWP.

Summary

Referring to easy-to-use, security and configuration issues caused by command line in scientific computing platform, CLWP was proposed in this paper and its prototype was implemented and deployed based on Eclipse Vert.x and xterm.js open source frameworks. The prototype shows that CSWP is a simple and flexible web gateway that provides anywhere and anytime capability for users to use a shell in browser accessing multiple physical and virtual computing resources. In future, we will continue to extend the prototype CLWP especially filters in the input and output chains of EB-cycle to enhance availability and reliability further.

Primary author: Dr CAO, Rongqiang (Computer Network Information Center, Chinese Academy of Sciences)

Co-authors: Mr WANG, Xiaoguang (Computer Network Information Center, Chinese Academy of Sciences); Prof. CHI, Xuebin (Computer Network Information Center, Chinese Academy of Sciences); Dr WANG, jue (Computer Network Information Center, Chinese Academy of Sciences); Prof. WANG, yangang (Computer Network Information Center, Chinese Academy of Sciences)

Presenter: Dr CAO, Rongqiang (Computer Network Information Center, Chinese Academy of Sciences)

Session Classification: VRE

Track Classification: Virtual Reserach Environment (including Middleware, tools, services, workflow, ... etc.)