ResearchOnline: a Universal Collaborative Service for Distributed Large Scale Scientific Data based on Cloud Computing

Jianjun Xie, Jianjun Yu and Fan Yang
Computer Network Information Center, CAS

Integrating heterogeneous scientific dataset and providing universal collaborative services is an important goal for e-Science informatization project of Chinese Academy of Sciences (CAS). In this paper, we first introduce the concept of collaborative environment for scientific data, analyze the reason why collaborative environment adapts cloud computing as its service model. We develop a collaborative environment named ResearchOnline (http://www.escience.cn) based on cloud computing. ResearchOnline includes two core components: the open cloud infrastructure and the collaborative services. The open cloud infrastructure is design with the idea of virtualization technology, pararrl computing and distributed storage based on the Hadoop cluster. We provide a resources pool container to support distributed heterogeneous scientific resources integration and continuous expansion with this infrastructure. The collaborative services are built on the self-developed collaborative suite-Duckling, and provide open collaborative functions, like instant message, collaborative editing, distributed storage and so on. Also ResearchOnline supports third-party to develop customized collaborative applications base on the open cloud interface with scientific data and computing resources, which aims to support cloud service invocation applying back-end transparent resources pool. Finally we present typical applications of collaborative environment created by ResearchOnline.

Keywords: Scientific Data Processing, e-Science, Cloud Computing, Data Integration