e-Science Applications with Duckling Collaboration Library

Kevin DONG
CSTNET, Computer Network Information Center
Chinese Academy of Sciences

CANS 2013 @ Hangzhou
Outline

- Duckling Collaboration Environment
- Duckling Cloud Services
- e-Science Applications based on Duckling
  - Science Cloud Project @ CAS
  - Biomedical Research Cloud Service
  - An Experimental Data Pipeline for MTO LAB
  - .......
- Future
e-Science Collaboration Environment is a comprehensive and integrated collaboration platform for research groups.

Scientist Friendly – for daily use (easy-Science)
Extensible – exploring new way we do research
Cloud Computing

SaaS
Software as a Service

PaaS
Platform as a Service

IaaS
Infrastructure as a Service
We need cloud?

- Cloud
  - Cloud is more popular
  - Software as a Service (SaaS)
  - Cloud is making user easy-collaboration

- Cloud Applications
  - Software instantly creating! (Software as a Service)
  - Common Requirements

- Open Platform for Plug-ins/applications
  - Everyone is a developer!
  - To make their own application easy and scalable!
  - Statelessness and Scalability

- ✔ On-demand self-service
- ✔ Broad network access
- ✔ Resource pooling
- ✔ Rapid elasticity
- ✔ Measured service
Cloud Computing

SaaS
Software as a Service

PaaS
Platform as a Service

IaaS
Infrastructure as a Service
A SaaS software suite to build your collaboration environment with features of document collaboration, col-laboratory library, virtual organization management …

More than 230,000 researchers by August 2013

A cloud-enabling open platform to integrate resources and enrich it as a scalable web-based e-Science application as you want …

* An integrated SaaS research community around CAS
* An open platform for CAS applications as PaaS

http://www.escience.cn
Duckling Cloud Services
- Software as a Service (SaaS)

- Conference Service Platform (CSP)
  - A web-based platform to help organize conferences, meetings and workshops
- Duckling Document Library (DDL)
  - Help for team establishment and team collaboration and networks
- Duckling Homepage (dHome)
  - Self-service Homepage systems for researchers, connected with publications and social networks
- Duckling Chat Service (dChat)
  - Instant messaging for institutions, organizations and groups.
- CAS Mail Service/Video Conference
- ……

Research Online, http://www.escience.cn

CSTNET passport users have reached 230 thousands by August, 2013
Duckling Cloud Fundamental Services - Platform as a Service (PaaS)

- CSTNET Passport (UMT)
  - Unified User Identification

- Organization Management Tool (VMT)
  - Organization Map / Address book
  - Group and VO information

- Col-library Tool (CLB)
  - A high performance object data storage

- Cloud Operation Support System (COS)
  - Cloud registration/Audit/publish
Duckling Framework (FALCON)

Balancer / Scheduler (Nginx)

Web Container (Tomcat)

Services

CLB

UMT

DDAL

Cache

Session

Meta Data Service

Dynamic Data Access Layer

Open Platform for SaaS

Statelessness and Scalability

China Science & Technology Network
SaaS - Conference Service Platform

A Solution for Conference Management Informatization.

http://csp.escience.cn

By Sep. 6, 2013:
98 Institutions/Organizations
616 Conferences
35000+ Registrations

A collaboration platform for connecting organizers, attendees and decision maker.
SaaS - Document Library (DDL)

- A Wiki-based document collaboration environment for groups
- More than 2000 groups involved.

http://ddl.escience.cn

Project groups
Laboratory management
Students groups
A on-going project funded by CAS
An integrated environment of Network, Storage and HPC.
IaaS / PaaS /SaaS Services for CAS scientists.
A platform to extend more cloud applications for domains.
- Geography, High Energy Physics, Astronomy, Biology, ...

Portal of Science Cloud Service

Case I

* On Demand Self-Service
* Measured Service
* Service Market
Resource Pool
(Storage, Server, Blade, Super Computing, Data, Publication, Software)

Resource Management/Monitoring

Service Pool
(Cloud Data, Cloud Computing, Data Service, Collaboration Environment, other Application Services)

Cloud Management Tool/Self Service

Virtualization

Service-oriented

Portal
Application
Mobile App

Standard/Policy

Security

Infrastructures
Network
Storage
HPC

Science Cloud @
CAS

Publication, Module, Software
Biomedical Research Data Cloud Services with Duckling Collaboration Library

- Collaborated with NBCR, University California, San Diego
- Duckling Collaboration Library @ CNIC
  - As part of Duckling Collaboration Library, CLB is designed to manage and collate millions of data files by setting up a unified, robust, and scalable data repository, especially in support of experimental data collaboration and timeline-based data life cycle management.
  - CLB+ is implemented as CLB plugins that provide interfaces with biomedical research cloud services from a computer aided drug discovery (CADD) workflow for ensemble-based virtual screening.
- Opal Services @ NBCR
  - Opal is a toolkit which allows users to wrap scientific applications easily as web services without any modification to the scientific codes, by writing simple XML configuration files.
  - Selected application services are provided by NBCR.

Publication on 2013 IEEE e-science Conference, Beijing
* Kejun Dong, Ji Li, Kai Nan @ Computer Network Information Center, Chinese Academy of Sciences
* Wilfred W. Li @ San Diego Supercomputer Center, University of California, San Diego
Experimental Data Cloud Services

Architectures

- Distributed MongoDB GridFS, which breaks large files into manageable chunks.
- Composed of a web-based resource access method and WS-compatible RESTful data access interface.
- Compatible with VO system, and supports group-based authorization, as well as tag-based catalog, full-text search, and data versioning.

Data Timeline Map

- Improving the scalability of data synchronization
- Delivering a data snapshot mechanism, especially for workflow-based data-intensive applications.
Providing cloud data service for the CADD pipeline, which utilizes web-based Opal services to enable ensemble based virtual screening (EVS)
Biomedical Case Study

1. Dashboard jobs
   - Opal Service Portal
2. Submit jobs
   - NAMD Nodes
3. Data sync (W+R) (Real-time, explore-integrated)
   - CLB+ Data Repository
4. Data sync (R)
   - Visualization/Client
5. Get data
6. Manage/get data

... ... ...

... ... ...

... ... ...

... ... ...

AutoDock Nodes
An Experimental Data Pipeline for MTO LAB

- MTO LAB
  - Methanol to Olefins National Laboratory in DICP, CAS
- Based on Duckling Collaboration Library (DDL/CLB)
- Experimental Data Collaboration and Management
  - Automatic Data uploading
  - Unified Data Repository
  - Web-based User Interface and Workspace
  - Multiple Data Sharing and Collaboration
  - WebDAV-supported Data Synchronization Mechanism

- User Interface
  - PC Client
  - Web-based Interface
  - Mobile Client
Data Services for Haze Project

- Strategic pilot science and technology project funded by CAS
- Extension of “Atmosphere Data Monitoring and Processing”
- Forecasting and monitoring the City Haze
- Five institutes and UCAS are involved
- Collaborated with Universities
Beijing Synchrotron Radiation Facility (BSRF)

- Part of the Beijing Electron Positron Collider (BEPC) project
- Requirements by Institute of High Energy Physics (IHEP), CAS
- Features:
  - Resource Reservation and Application
  - Application Management and Approval
  - User Identification and Authentication
  - Collaboration Environment for Admin, Experts and End Users
  - Analysis and Audit System

Case V
Thanks!