Enabling Computational Geoscience Software Requirements Specification

Version 1.0

Group Name -Dhara
Group Number - 21

Group Members

A.I. Ranatunga - 090423F
M.H. Kumara - 090269L
L.N.P.T Perera - 090377P
M.S.H Jayaratna - 090205N

Revision History

Date	Version	Description	Author
31/05/2013	1.0	Initial requirement specification	Dhara

Table of Contents

Software Requirements Specification

1. Introduction

The aim of this document is to gather and analyze system related knowledge and give an in-depth insight of project Dhara, by defining the problem statement in detail. Nevertheless, it also concentrates on the capabilities required by stakeholders and their needs while defining high-level product features. The detailed requirements of project Dhara are provided in this document.

1.1 Purpose

The purpose of this document is to present a detailed description of project Dhara, which focuses on enabling computational geoscience using Apache Airavata. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react. This document is intended for both the stakeholders and the developers of the system.

Further it focuses to collect and analyze all assorted ideas that have come up to define the system, its requirements with respect to consumers. Also, we shall predict and sort out how we hope this product will be used in order to gain a better understanding of the project, outline concepts that may be developed later, and document ideas that are being considered.

In short, the purpose of this SRS document is to provide a detailed overview of our software product, its parameters and goals. This document describes the project's target audience and its user interface, hardware and software requirements. It defines the functionalities and how end users experience the product.

1.2 Scope

Primarily, the scope pertains to building geoscience gateway using Apache Airavata to enable the construction of geoscience workflows that leverage standardized, publicly available geoscience services and geo data sources. The system will be designed to maximize the productivity of geoscience experiments by providing tools to assist in automating the workflow exposing process, which would otherwise have to be performed manually.

This SRS is also aimed at specifying requirements of the software to be developed. More specifically the system will focus on geoscience standards and their engagement with the web services.

1.3 Abbreviations

OGC – Open Geospatial Consortium
WPS – Web Processing Service
WSDL – Web Service Description Language
GUI – Graphical User Interface

1.4 References

Dhara Literature Review – Enabling Computational Geoscience

1.5 Overview

The rest of this document is structured into 3 main sections. Section 2 includes project perspective, basic functionality of the project, characteristics of the target users, assumptions and dependencies regarding the project implementation. . Section 3 gives specific requirements which the software is expected to deliver. It includes functional requirements given by various use cases and non-functional requirements. Furthermore it includes system requirements for proper functionality of the implementation as well as design and implementation constraints we have identified. Finally, the Section 4 gives information about any other document which includes details about our project including information about system documentation.

2. Overall Description

2.1 Project Perspective

Project Dhara is an ultimate effort for enabling computational geoscience leveraging existing tools and technologies available in the open source software industry. Dhara is a geoscience gateway which enables geoscientists to perform their experiments without understanding underlying complex high performance computational resources. Apache Airavata which is a scientific workflow management system and OGC's geoscience standards are extensively used in the Dhara project. OGC's Web Processing Service (WPS) is a widely adopted geo-processing standard in the geoscience world. Deploying complex geoscience algorithms in supercomputers and wrapping them as a web service to create workflows are major objectives of a science gateway. Dhara provides the capability of creating workflows using available WPS services. Geoscientists create workflows as part of their experiments and Dhara is capable of providing support to deploy developed geoscience workflow as a WPS services in WPS implementations. Deploying workflows created in Apache Airavata as a WPS services enable high degree of reusability of geoscience workflows which can be used by other geoscientists in their experiments.

Visualization aspects and monitoring of executing workflows are major functions in a geoscience gateway. Project Dhara provides a portal which has a service layer to publish geospatial data on the portal for visualization aspects.

2.2 Product functions

Dhara is a geoscience gateway which wraps the complex computational resources and provides a convenient interface for gateway users. Key features are listed below.

- Creating workflows using available WPS services
- Publishing created workflow as a WSP service
- Visualizing output of WPS services
- Monitoring execution of workflows
- Publishing services to the gateway for use in workflow creation

2.3 User Characteristics

Dhara focus on providing a user friendly platform for geoscientists to create workflows and expose them as WPS services. Geoscientists are the mainly focused users of the system. Students who are willing to explore geoscience computations also can be identified as potential users. Ultimately gateway developers, WPS or other service providers are also users of the system. Users of the system need to have an understanding about the fundamentals of the web services and workflow creation using Apache Airavata. Knowledge on OGC's WPS standard and its operations are also required. Deployment of scientific applications in supercomputers is also beneficial for users to get the maximum advantage of the system.

2.4 Assumptions and Dependencies

Nature of standards is known to be changing with the time. Since system going to use OGC's WPS, it assumes that standard won't change significantly in near future. Since Apache Airavata is built using Java, it's required to have machines with Java Runtime Environment (JRE)

- It is assumed that system will be installed with Java runtime environment and Java development kit (1.6 or above but depends on Apache Airavata version used).
- Tomcat version 6 or above is installed.
- Apache Airavata is successfully installed and geoscientist is fully capable of dealing with Apache Airavata.
- OGC standards will be consistent throughout and any alterations will not affect the current standards but introduction of new standards will occur.

3. Specific requirements

3.1 Functional requirements

This section identifies the stakeholders of the geoscience gateway and respective use cases in detail.

3.1.1 The Use-Case-Model Hierarchy

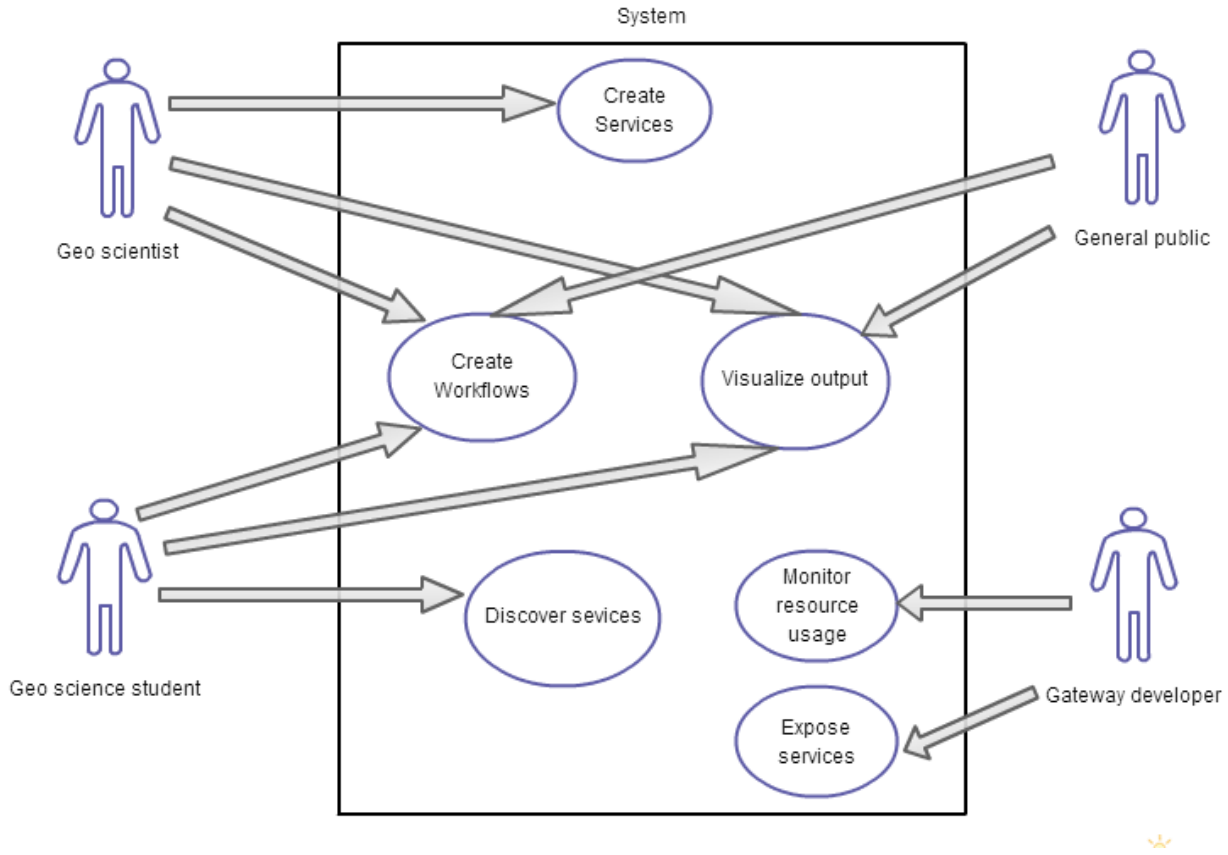
Actors:

- Geoscientist: This actor conducts researches regarding geoscience field. Mostly this actor is attached with academic institutes like universities or other research institutes. Thus they always do researches collaboratively forming various groups.
- Geoscience student: Unlike geoscientists, students have much simple needs. And they always look at the geoscience at a perspective of learning new things and would like to gain knowledge.
- Gateway developer: Gateway developers are mainly engaged with developing front end and backend components of a science gateway. Their main objective is to customize the components to align with the objectives of a science community. Mainly they can be categorized into two: instance developers and framework developers. Framework

developers are involved with developing back end for instance developers to customize the gateways as they wish.

Ordinary user: They are interested in prediction results like weather forecasting and disaster prediction. They are usually interested in the final outcome of a research in an understandable manner.

3.1.2 Diagrams of the Use-Case Model



3.1.2.1 Primary use cases

- Create workflows

  Precondition: All components are loaded (using WSDL)
  Post condition: Ready to be hosted in WPS server

  Main stakeholders associated with this project are geoscientists. They intend to create workflows using Xbaya GUI using its capability of simple drag and drop. Anyway this project should go beyond these phenomena and allow geoscience students and general public to create workflows. Components are listed at Apache Airavata after loading the WSDL. After dragging and dropping components they will try to link these components.

Each of these components describes a web service. And also there are input and output components which are inherently defined by Airavata which will be used to link with WPS services.

- Create services

  Precondition: Workflow execution completed.
  Post condition: Ready to provide WPS services locally

  After analyzing the result (basically after the experiment) using workflows geoscientist needs to give the output as another WPS. So that should be able to reuse by other scientists. One who exposes the services might have to deal with WPS servers. Basically here the service might be created locally while that has to be hosted somewhere accessible which will be discussed in exposing the service.

- Visualize output

  Precondition: Data is available in desired file format (i.e KML)
  Post condition: Ready to reuse same set of data in another workflow

  Any way workflow output is shown as table structure in Airavata for any scientific experiment. But from the perspective of geoscientists and geoscience students the requirement becomes different. Many geoscience experiment data output is related to a map. So they wish to push the workflow output data to a map and analyze them. Output does not mean the output of the entire experiment. Even the workflow steps in between the experiment are desired to be monitored via a map. Thus they will dynamically decide whether to go ahead on the current workflow.

- Discover services

  Precondition: Apache Airavata ready to accept services.
  Post condition: Load components via WSDL

  Geoscience students learning perspective emerges need of finding wider scope of WPS services to be used in their academic activities. But some of the students might be interested to go beyond consuming services to create own experiments where it would be a point where their role in this scenario will be changed to a scientist. Discovering services can be done via the portal where the other scientists register their services. Or they would search for any other WPS services which are not registered in portal.

- Expose services

  Precondition: Experiment execution is over
  Post condition: WPS services are web accessible

  Services created by geo scientists should be hosted in computational grid. Domain knowledge regarding the computational grids and cloud computing resources are very less in geoscientists. Gateway developers host these services in a manner that it can be

consumed by another scientist. Services are discovered via a wsdl and the scientist who loads previously hosted services will obtain them by just dragging and dropping the service component in x baya GUI without using any underlying knowledge about where the computation takes place.

- Monitor resource usage

  Precondition: Services are exposed
  Post condition: Ready for the necessary services to be altered

  Since gateway developers are mainly involved with developing backend structure for instance developers to customize science gateways they often need to look at resources usage of the hosted WPS services. And then make accurate decisions about the management of grid services.

3.2 Non-functional requirements

3.2.1 Ease of use

Nowadays sophisticated high level tools to deal with complex computations and experiments are available for scientists. However complexity and difficulty in adapting to these technologies are areas that need to be addressed. Generally the resource sharing and scalable facilities, accessing Grid infrastructure require significant learning. Thus while expanding the facilities and capabilities of the applications, they should provide ease of use through convenience and intuitiveness especially regarding the programming models and user interfaces.

In an ideal gateway researchers should be able to focus on their research experiments with little concern on the underlying technology. To achieve this status the application need to hide the complexity and expose user to desired set of features in an accessible manner. In other words, scientist should be able to create, manage and execute a process for an experiment in a convenient manner.

3.2.2 Scalability

Scalability is a fundamental requirement for large-scale scientific experiments and has a great impact on the application performance and completion. The recent progress in modeling and visualization technologies have significantly focused on utilizing resource pools for on demand and scalable scientific computing. Utilized resource quantities and characteristics of geoscience experiments vary at runtime. Thus cloud computing enabled workflow engines would be a desirable feature for applications to facilitate dynamic scaling up and down. In addition, frameworks excessively need to access a range of computational resources in order to manage distributed geoscience applications.

3.3 Performance requirements

Hardware Requirements

- Intel Pentium Dual Core processor or above
- System memory of 1GB or above.

- A hard disk space of 20GB or above.

Software Requirements

- Java Runtime Environment 1.6 or above

3.4 Design and implementation constraints

Apache Airavata selected as the workflow management system from the available systems. Each workflow system has its own strengths and weaknesses. It's required to use available WPS implementation software in the industry for provide WPS services. Each WPS standard implementation software has not fully supported the all the operations mention in the WPS specification document.

4. Supporting Information

- Project proposal
- Literature review
- Research proposal
- Project website (http://projectdhara.net78.net/)

4.1 System Documentation

Since the system design is at the research level, there are an interesting discussions going on Apache Airavata mailing list and other mailing lists such as WPS implementation software like PyWPS and 52 North WPS which are publicly visible for anyone who interesting in developing geoscience gateways. System use free and open source software. So every discussion going on the mailing lists are publicly visible for anyone.

Several research papers will coming out with the project. Also there will be documentations for system setting up and the usage. Project website is developing with the project and it contains a full set of documents and resources use in the project.