

Open Source Business Intelligence for Cloud

An Open Source Cloud Initiative Domain (OSCi Domain3)

January 24, 2011

Parham Parvizi (pparvizi@talend.com)

<i>Revision</i>	<i>Date</i>	<i>Author</i>	<i>Description</i>
A	2011-01-24	Parham Parvizi (pparvizi@talend.com)	Document creation.

Introduction

Open Source Business Intelligence for Cloud project is initiated by the OW2 consortium as part of the Open Source Cloud Initiative (OSCi) project. This project is also known as Domain 3 or BI4Cloud project within OSCi. This project is a collaboration effort between open source leaders in Business Intelligence to provide a single and consolidated BI platform within the Cloud infrastructure. The intuitive would make it simple for users around the world to have quick and easy access to all the vendors' software and be able to deploy their own customized BI platform within minutes. By means of collaboration between major open source players in the market, the project would also bring forth a set of new and innovative offerings to the users within the BI and Cloud technologies.

Members & History

In mid-August of 2010, the main members of the BI4Cloud project started meetings to initiate their collaboration and set the goals of this project. At this point, the members started looking and comparing a series of Cloud providers and appliance building vendors to create a set of initial requirements and a media for the delivery of the project. As the result, the vendors choose UShareSoft, automated software appliance and deployment tool, to work closely with the other vendors to build

Cloud ready appliances of their software and provide deployable packages. The team initially chooses VMWare and Amazon EC2 to be the first media formats to deliver the appliance packages.

At the time of this paper, the current active members of the project are:

- Ingres: the leading Business Open Source Database
- Talend: Leading Open Source ETL and Data Integration
- JasperSoft: World's most widely known Open Source Business Intelligence tool
- Engineering featuring SpagoBI: World Wide Open Source Business Intelligence and Dashboarding Suite.
- BonitaSoft: Open Source Business Process Management (BPM) and Workflow management
- UShareSoft: Automated Software Appliance Creation and Deployment tool

Vision

In a typical Business intelligence project, users spend a substantial amount of time initially to setup and configure the software components required for their project. A Business Intelligence or data warehousing project has many different parts which at least

includes one or more of the following software components: Database, ETL, Data Quality, Data management console, Process management, Workflows, Reporting and Dashboard console. There are many choices of vendors for each of these components; and each vendor comes with a set of its own pre-requisites and requirements. Therefore a substantial amount of time is always spent during the setup phase of any Business Intelligence project.

In recent years, the emergence of Cloud and virtualization technologies has substantially cut down the cost and time of building the core infrastructural components needed to develop any project. Cloud technologies enable users to quickly setup servers pre-configured and ready to use for internal team members. Cloud has also cut down administration cost of such systems to almost nothing.

BI4Cloud domain of OSCi project aims to ease the setup phase of any Business Intelligence project by bringing together a set of appliances packages pre-configured and ready to use with leading open source BI software. BI4cloud members have joined forces to each provide their corresponding software component to build a full BI suite. BI4Cloud group aims to enable users to create a fully functional and customized BI Suite in minutes. With the help of the Cloud and appliances packages, each member of the BI4Cloud group would create an appliance containing their software and all of its pre-requisites and make it available publically. Subsequently the users are able to download and mix the components provided by each user to produce their own customized BI stack.

In September of 2010, the members of BI4Cloud project conducted a series of meetings and research to set an initial set of goals for this domain. These goals include:

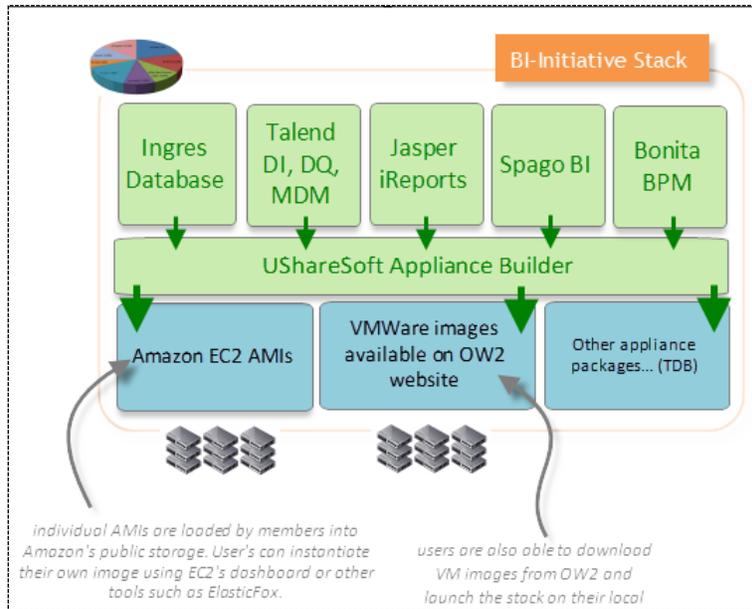
- Full collaboration and integration between all vendors' Open Source Software
- Providing a Complete BI Stack containing a:
 - Database layer
 - ETL and Data Quality layer
 - MDM or Master Data Management layer
 - Business Process Management (BPM) and Workflow Management layer
 - As well as Business analysis and Reporting layers
- Maximizing user experience and minimizing setup time by providing fully configured appliances
- Providing easy to deploy and generic appliance formats such as:
 - VMWare
 - Amazon EC2 AMI
 - Other common formats (TDB)
- Providing a landing page to direct users to manuals, tutorials and training resources, and sample projects.
- Providing easy path to upgrade each components with the latest version from the vendors.

- *What do we mean by an appliance?* In our words, an appliance is a machine containing an OS and a set of pre-configured and ready to use software packages and all of their pre-requisite components. For example a database appliance would include the main database component and all of its pre-requisites such as Java; and would already include all the security and configuration setup. Appliances are often packaged in a single deployable format such as VM (virtual machine) or Cloud instance. An example of a Cloud instance is Amazon EC2 machine image or AMI. An AMI is a complete virtual machine containing OS and other software packages that can be

replicated, instantiated, and ran on Amazon EC2 Cloud by users in few minutes.

Architecture

The main architecture of the BI4Cloud project is represented in the diagram below.



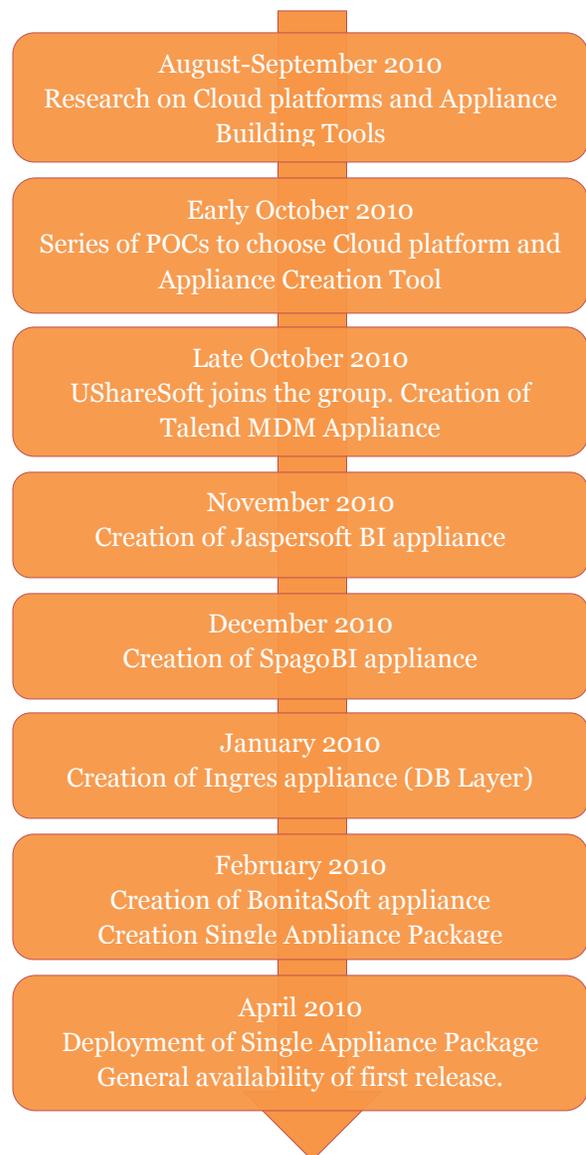
Through using UShareSoft, each vendor is able to package their software and all its prerequisites as an appliance. UShareSoft enables many different output formats for this appliance. The main output formats of this projects is currently set for Amazon AMI images that can be launched on EC2 Cloud; and VMWare images that can be downloaded and ran within the user's private Cloud or network.

The user first downloads and instantiates an appliance on either their Amazon account or through using VMWare. Then they are able to begin configuring the appliance remotely via using web browser. UShareSoft pre-installs a web wizard on each appliance. This wizard is used to configure and initialize the software with the appliance. Through the use of this wizard the user is also able to configure a

multi-node (multi-appliance) configuration. This is the case in which the user is able to instantiate two or more appliances, each containing a separate piece of the stack (ie: talend, ingres, and jaspersoft) and configure all the appliances together.

Timeline

As of the date of this paper, the following diagram presents the purposed timeline of the BI4Cloud project.



Currently, as of mid-January 2011, a set of

appliance are available. These appliances are Talend MDM appliance, JasperSoft BI & iReports appliance, and SpagoBI appliance. Currently the team is working on creating the Ingres appliance which will serve as the database layer of the BI stack.