19 Commercial Use of WS-PGRADE/gUSE

Tamás Kiss, Péter Kacsuk, Éva Takács, Áron Szabó, Péter Tihanyi, and Simon J. E. Taylor

Abstract. Although originally an academic and research product, the WS-PGRADE/gUSE framework is increasingly applied by commercial institutions too. Within the SCI-BUS project, several commercial gateways have been developed by various companies. WS-PGRADE/gUSE is also intensively used within another European research project, CloudSME (Cloud-based Simulation Platform for Manufacturing and Engineering). This chapter provides an overview and describes in detail some commercial WS-PGRADE/gUSE-based gateway implementations. Two representative case studies from the SCI-BUS project, the Build and Test portal and the eDOX Archiver Gateway, are introduced. An overview of WS-PGRADE/gUSE-based gateways for running simulation applications in the cloud within the CloudSME project is also provided.

19.1 Introduction

The core developer team of the WS-PGRADE/gUSE framework comes from an academic research institute, MTA SZTAKI from Hungary. Early collaborators and adaptors, for example, the University of Westminster or the Middle Eastern Technical University, are also academic and research institutions. Therefore, the first use-cases and targeted end-users of the WS-PGRADE/gUSE technology were researchers.

On the other hand, accessing distributed computing resources in a flexible way and from user-friendly interfaces is not a requirement that is restricted to academia. Many industry and business applications require access to DCIs, and many companies would like to offer services on top of these resources to their employees or customers as end-users.

Although the majority of SCI-BUS related use-cases described in this book are academic and research applications, there are a growing number of companies also interested in this technology. This chapter provides an introduction to some of the commercial use-cases implemented on top of WS-PGRADE/gUSE. All these use-cases target SMEs in various application domains and implement user-friendly access to mainly cloud computing resources. While the majority of the academic use-cases still utilize cluster or grid computing infrastructures, not surprisingly, the commercial implementations focus on easily accessible cloud resources that do not require large up-front costs from the company.

The three examples presented in this chapter come from diverse application domains. The Build and Test Portal is the product of a software company which provides a flexible build and test environment for software developers. The eDOX Archiver Gateway is a solution for electronic document management, replacing and processing current paper-based document stores. Finally, CloudSME [Taylor/2014], a European project, and some of its commercial use-cases are introduced that implement a cloud-based simulation platform and one-stop-shop solution for manufacturing and engineering SMEs.

19.2 The Build and Test Portal

The Build and Test Portal is an out-of-the box web application for
- building/testing software projects continuously, making it easier for developers to integrate changes to the project and for testers to run tests just after the build,
- monitoring jobs and detailed reporting of results,
- providing a repository for software packages and reports.

The Build and Test Portal, developed and operated by 4D Soft(http://www.4dsoft.hu/), a Hungarian SME (small and medium enterprise), existed before the SCI-BUS project as a GWT-based (Google Web Toolkit) web application submitting jobs to a local cluster served by Condor/NMI (NSF Middleware Initiative) [Thain/2005] job submission middleware. Within SCI-BUS, the portal has been integrated with the services provided by WS-PGRADE/gUSE to extend the existing job submission capabilities towards a more flexible, user-centric cloud infrastructure. The major motivation when moving to cloud is that the Gateway will be operated as a Software as a Service (SaaS) serving software developers. In the case of Build and Test Portal, the SaaS or "on-demand" software delivery means that the software and its associated data are hosted locally. At the same time, the build servers, on which the tests and builds run, are dynamically hosted and distributed in the cloud. The Build and Test Portal is accessed by users via a web browser. The targeted users of this service are small and medium-sized IT companies and individual, freelancer programmers who do not want to operate their own build and test infrastructure. For these professionals, the cloud approach is probably the most cost efficient method to use, maintain and upgrade.

Slightly similar continuous build automation tools are available both on the open-source and on the commercial markets. On the one hand, open source solutions are CruiseControl (http://cruisecontrol.sourceforge.net/) for Java and .NET, and Jenkins (http://jenkins-ci.org/), the extensible continuous integration engine. Regarding commercial products, Microsoft's Team Foundation Server (http://msdn.microsoft.com/en-us/vstudio/ff637362.aspx) and JetBrain's Team City (http://www.jetbrains.com/teamcity) are the most popular ones. However, the market potential of the Build and Test-portal is in its intuitive user interface and the dynamic utilization of cloud computing resources.

19.2.1 Role of WS-PGRADE/gUSE

gUSE is an open-source science gateway framework, empowered with a workflow toolset and giving users to convenient and easy access to various DCIs, including cloud infrastructures. In this respect, gUSE was a good candidate for the Build and Test Portal when implementing it as a service in the cloud.

The Build and Test Portal submits only jobs to the build servers As a consequence, the workflow capabilities of WS-PGRADE/gUSE are currently not utilised. Job submission is via the Remote API to cloud computing resources through the DCI Bridge and CloudBroker components. The Remote API is a convenient way to manage workflows (in this case, only one-job workflows) between the local environment and cloud infrastructures for various reasons. Among them, it allows running and managing workflows from custom user interfaces, in this case from the Build and Test Portal. The API exposes gUSE workflow (job) submission services through a simple web service interface. In that way it integrates well with the existing GWT-based portal implementation.

Please note that the terms workflow and job are sometimes used interchangeably. The reason is that the Build and Test portal operates with jobs. At the same time, the Remote API works with workflows. In order to facilitate the advantages of Remote API, jobs were transformed into one node workflows having as input the job identifier and as output the resulting software packages and test reports. This simple data transformation seamlessly resolved the data transfer between Build and Test Portal and the Remote API.
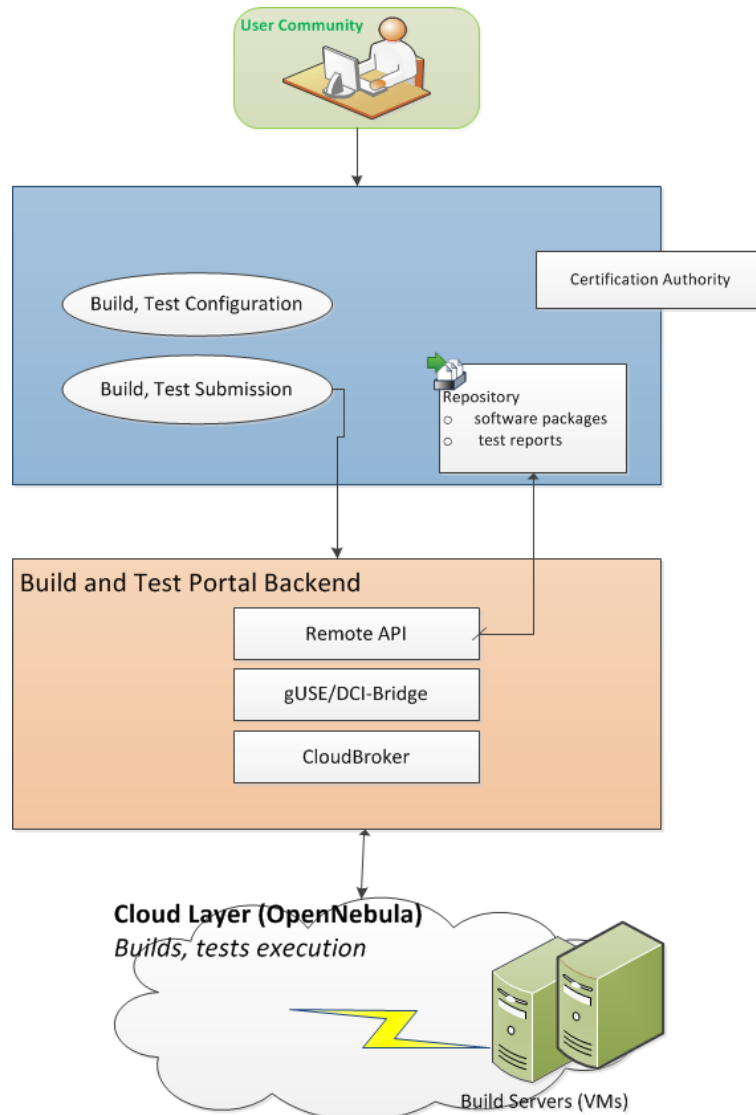
19.2.2 Architecture of the Build and Test Portal Extended with Cloud Resources

As stated before, the Build and Test Portal has a web (GWT) based front-end from where users can configure and submit their build and test jobs. As back-end service, a WS-PGRADE/gUSE portal and its Remote API are utilized. The WS-PGRADE portal is used for setting up the cloud environment for the Build and Test job submission, and defining the basic workflow that governs the build job submission.

The Build and Test jobs run as workflows in the build server (an Ubuntu-based server which resides in SZTAKI based OpenNebula cloud). This execution environment is managed by and accessed via the CloudBroker platform interface. The CloudBroker plugin of the DCI-Bridge is used in WS-PGRADE/gUSE in order to pass job parameters submitted via the Remote API. The results of the Build and Test jobs are the generated software packages and test reports. The Remote API is used to transfer the output back into the local environment. Finally, these outputs are stored in a local repository service.

Figure 19.1 illustrates the architecture of the system. Please note that relying on the CloudBroker platform in this architecture provides flexibility regarding the cloud resources utilised. While currently only resources of the SZTAKI

OpenNebula cloud are used, extending or replacing these resources with other clouds is fully supported by the CloudBroker platform.

Fig. 19.1 Architecture of the Build and Test Portal

A typical user scenario is illustrated in Figure 2. Software engineers can set up a build and test environment for their software stack using the front-end of the portal. These build and test jobs can then be executed either on remote virtual machines residing in the SZTAKI OpenNebula cloud, or in the local cluster governed

by Condor. After execution the results of build jobs, the created software packages and reports are sent back to the portal.
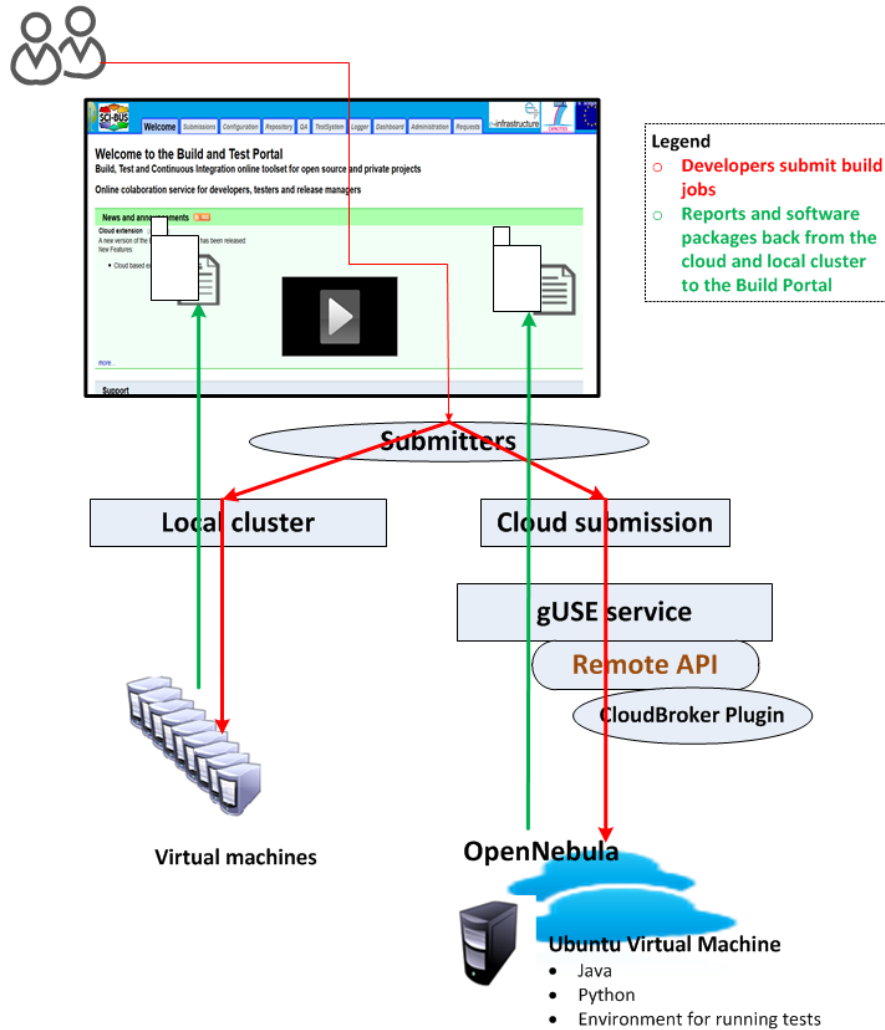

Fig. 19.2 Typical user scenario of the Build and Test Portal

19.2.3 Usage of the Build and Test Portal

The Build and Test portal is designed for software developers to perform their daily software build and test work seamlessly without the need of operating large software engineering systems. Cloud-based build servers enable software professionals to use the build infrastructure in an "on-demand" manner.

In the SCI-BUS project, the Build and Test Portal served as quality assurance portal. Several build jobs for gUSE were run, and automated web tests for the various science gateways were also performed.

Having a cloud extension to the Build and Test Portal enabled users to run multiple jobs at the same time. Also, the maintenance of the build servers was seamless due to the professional service provided by cloud operators.

As for user statistics, the Build and Test Portal has 25 registered users currently. Usually, a job instance average execution time is approximately two hours. Job execution time strictly depends on the types of jobs submitted. Test jobs run approximately one hour, build jobs approximately three hours. The total number of job instances is in the range of 10,000.

19.2.4 Further Development Plans

As for further development plans, the usage of cloud resources could be extended to other software stacks of the Build and Test Portal, not just for build servers. For example, the repository service (the service which is responsible for the storage of software packages and test reports) is a good candidate to be operated on cloud resources. As the number of users and the number of build jobs increase, there is a need for further storage capacity for this repository service. Finally, as the gUSE/DCI Bridge component is evolving, more and more cloud resource types and submission mechanisms are available for the extension of the Build and Test Portal.

19.3 Electronic Document Management – eDOX Archiver Gateway

Despite of the electronic nature our world, a huge amount of information is still to be found in paper based, so called unstructured documents. This type of information is not very easily searchable. Even after digitalization (scanning), some metadata still need to be attached either manually, or if a good optical character recognition (OCR) engine is available then full-text search can also be tried. The problem with manual processing is that it requires far too much human resources, and the full-text search in most cases does not provide typical characteristics. An automatic method finding the characteristic metadata of documents could make the unstructured information in paper-based documents very easily searchable.

Another problem with paper is their storage. Eliminating all paper based information could lead to huge savings. For example, in Hungary there is a legal background supporting this, Decree No. 13/2005. [X. 27.] of the Minister of Information and Communications, Decree No. 14/2013. [IV. 15.] of the Minister of National Development, and Decree No. 83/2012. (IV. 21.) of the Government. Similar legislation is becoming more and more common in other European countries too, for example, the e-government law in Germany [Bundesgesetzblatt/2013]. Enabling electronically signing of paper-based documents

after converting them into electronic versions could eliminate papers and renting storages completely.

Based on the above, the major aim of the eDOX Archiver Gateway is to automatically attach some characteristic metadata to the converted documents, after which these will be electronically signed. eDOX is document and records management system developed by E-Group, a Hungarian-based multinational SME, that provides all the necessary functionality of this application, except of categorization and authentication. The eDOX Archiver Gateway has a supplemental role in whole system.

Automatic categorization is not an easy task. It requires a good OCR engine, a module which finds the roots of the words (in the Hungarian language this is a complex function), and finally a categorization module, which has to find characteristic metadata of the specific documents. In the first release, the aim of the gateway was to collect a general set of metadata of all documents (document type, a characteristic date, company, and person). All of these modules are rather time- and calculation-consuming tasks, requiring DCI (cloud) resources as well as a general interface to the DCI systems. WS-PGRADE, the DCI-Bridge and gUSE have been proven as good candidates for this.

19.3.1 Role of WS-PGRADE/gUSE

WS-PGRADE/gUSE is a complex system for supporting workflows that use grid or cloud resources in the background. In the case of the eDOX Archiver Gateway such functionality is desired. The eDOX Gateway can access the gUSE system via the Remote API component (REST API), providing access to WS-PGRADE/gUSE workflows. However, the biggest advantage in this case is provided by the DCI Bridge component. The DCI-Bridge gives a common interface for several grid and cloud systems, making job submission to any of the supported DCIs straightforward.

The DCI Bridge, together with its CloudBroker plugin, also acts as a resource broker. The eDOX Archiver Gateway is a commercial product (and its developer is an SME), which means that, beyond ensuring the expected level of quality, cost-effectiveness is an important requirement. The management interface of the DCI Bridge enables setting configurations that support this objective.

19.3.2 Architecture of the eDOX Gateway

The eDOX Archiver Gateway supports the digitization and processing of paper-based documents. The architecture of the gateway is illustrated in Fig. 19.3.
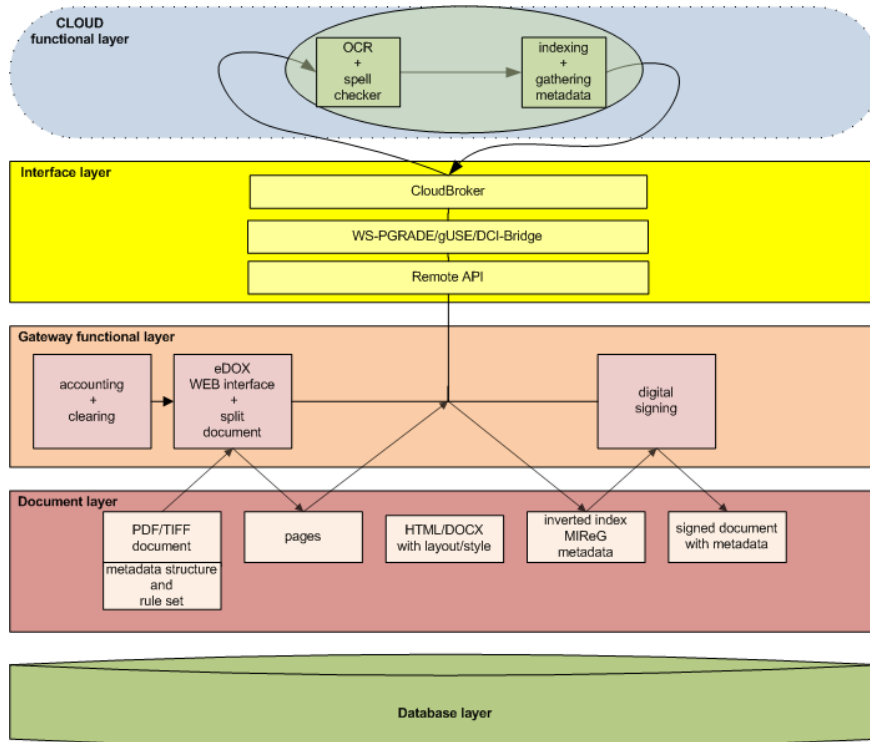
Fig. 19.3 Architecture of eDOX Archiver Gateway

The eDOX Archiver Gateway document management system (Gateway functional layer) uses a database (database layer) to store contents and metadata of documents and workflows (document layer). The cloud management system (interface layer) is provided by SZTAKI to access the chosen cloud service (cloud functional layer) via the CloudBroker plugin.

The input of the workflow is the scanned, digitized document (either in PDF or multi-paged TIFF format). This image content is uploaded (via web form, FTP/SCP etc.) to the portal, and the server side starts processing. The OCR module covers the following functionalities:

- OCRing: recognition of characters (e.g., Asian font sets), and keeping layout layer (e.g., creating XPath (XML Path Language) based rules for stylesheets,
- Spellchecking: post-processing of OCR-enabled content; approximately 200 languages are supported currently.

The goal of content processing is to support classifying, labeling, and metadata handling functions, such as:

- extracting: dictionary-based and regular expression-based (regexp) rules in order to retrieve predefined expressions, words, data structures;

- indexing: canonicalization of content by transforming expressions to the roots (available only for Hungarian at the moment);
- ranking: enumerating roots, assigning counters and other ranking values based on a ranking algorithm;
- classifying: categorization of document based on algorithms that analyze content values and content types (e.g., the document is probably an invoice if it contains the string "Invoice" as title, and contains date data types).

Based on the legal background and business needs, metadata and labels are automatically generated and assigned to the content:

- labeling: retrieving and assigning the most relevant expressions (results of ranking) as a "tag cloud" to the processed document,
- metadating: the construction of XML-based (Management Information Resources for eGovernment, MIReG) metadata structure based on retrieved metadata of processed document.

The cryptographic module is invoked to create an electronic signature on the whole data package. This functionality is required by the legislation in order these digitized documents are acceptable in a legal mean:

- digital signing: creation of XML-based (XML Advanced Electronic Signatures, XadES) electronic signature on a processed document and its related outputs by using a cryptographic private key (e.g. RSA), retrieving timestamp and revocation information (e.g. Certificate Revocation List, CRL), or Online Certificate Status Protocol, OCSP) response.

The eDOX Archiver Gateway as a commercial product manages user accounts and supports clearing functionality. The integration of a payment solution is not in the scope of the product. There are several methods for credit top-up of a user account (e.g., money transfer to a central bank account via netbanking solution), but these methods shall be supported by the environment:

- accounting: calculating accountable values based on the real usage of the service (e.g., processed page per job).

The output electronic data package is in a legal sense equivalent to the input paper-based document.

19.3.3 Usage of the eDOX Acrhiver Gateway

During workflow design, the execution statistics of each step were analyzed, as shown in Table 19.1. Based on these statistics, computation-intensive and parallelizable functions, such as recognizing characters (OCRing) and finding the roots of the words (indexing), were mapped into the cloud in order to reduce execution time of these jobs.

The eDOX Archiver Gateway is a commercial product, where users can set priorities and decide the level of parallelization based on the precalculated cost of allocating cloud resources and executing the jobs. In principal, the best throughput

time could be achieved by allocating a separate cloud resource to each document page. However, in practice this could be different due to the default boot and setup time (~10 minutes) of virtual machines. For example, in the case of a larger book that has 830 pages, it is possible to dedicate a virtual machine for each page for OCRing, finding the roots and indexing. In this case, processing time is measured at approximately 90 seconds, with an additional 10 minutes to set up all the virtual machines. This compares rather well to the almost 20-hours processing time on a single processor.

Users can access the portal via a web-based GUI. After successful login, users can upload the documents to be processed, and start processing them by accepting the calculated cost. The processed contents, including signatures, OCRized contents and metadata, can be downloaded from the portal.

Table 19.1 Execution statistics of eDOX workflow components

#	Step	Measuring unit	Amount	Note
1	Retrieve document (from database or file system)	sec/piece [document]	0.50 – 10.00	Depends on file size, (< 10 MB) and connection speeds
2	OCRing Spellchecking	sec/piece [document]	4300.00	In case of avg. 100 pages/document
		sec/piece [page]	43.00	Depends on file size and resolution (dpi)
		sec/piece [word]	0.22	In case of avg. 200 words/page
3	Extracting Indexing Ranking Classifying Labeling Metadating	sec/piece [document]	3000.00	In case of avg. 100 pages/document
		sec/piece [page]	30,00	In case of avg. 200 words/page
		sec/piece [word]	0.15	–
4	Digital signing	sec/piece [document]	2.50	Depends on file size (< 10 MB) and connection speed

19.3.4 Further Development Plans

The first release of the eDOX Document Archiver Gateway collects document type independent sets of the metadata, as was explained earlier. Future releases will consider more sophisticated approaches, such as form recognition, for example. This method can identify metadata that is dependent on various document types and will be more helpful searching information in documents, such as:

- Invoices [invoice no., issuer, company, price per unit, total price],
- Salary documents [name, birthday, year],
- Sickness/benefit documents [name, birthday, year],
- Employment contracts [name, year, job],
- Delivery notes [date, ID, order ID, company name],
- Ledger documents [year, period],
- Balance sheets [year, period],
- VAT documents [form ID, year, period],
- Environmental taxes [form ID, year, period],
- Local taxes [form ID, year, period].

Also, in the current release only one virtual image is utilized in the cloud environment. However, as the number of clients is expected to grow, the number of images will also be expanded in the future.

19.4 Gateways for Manufacturing and Engineering Simulations

Simulation is widely used in both academic research and industry practice to model phenomena that could be too expensive, complicated, or unfeasible to study otherwise. Simulation can be used to analyze a wide range of physical and chemical processes, manufacturing systems, logistics, transportation networks, and supply chains, for example. Larger companies widely utilize various simulation software products supporting their design or manufacturing operations. However, the take-up of simulation solutions at SMEs is still relatively low.

The main reason why SMEs do not engage in simulation is the high entry barrier to these solutions. Not only must the company purchase an expensive software license, but they must also have or pay for specific expertise in designing and implementing the simulation model applied. Large-scale simulation also requires robust computing resources to run parallelized or parameter sweep simulation scenarios. Purchasing the necessary licenses, expertise and equipment is way beyond the capabilities of an average SME.

Cloud-based Simulation Platform for Manufacturing and Engineering (CloudSME) is a European research and development project that started in July 2013 with the aim of building a cloud-based simulation platform and one-stop-shop solution for manufacturing and engineering SMEs. Cloud computing has the potential to provide the necessary resources to run simulation applications on a pay-as-you-go basis, significantly reducing entry barriers. Moreover, cloud computing also provides the necessary service provisioning models to build and consume simulation solutions in a feasible way for SMEs. End-users of the CloudSME Simulation Platform, manufacturing and engineering SMEs, can access the solutions as Software as a Service (SaaS). The SaaS solution represents a ready-to-use simulation product and model that fits to their requirements and that can be purchased from the cloud on demand. Simulation software provider com-

panies, typically also SMEs, can exploit the CloudSME Simulation Platform as Platform as a Service (Paas). These simulation providers can use a cloud-based toolkit to build new SaaS solutions for their end-users.

The CloudSME Simulation Platform is built on top of the WS-PGRADE/gUSE framework and the CloudBroker platform. In the first phase of the project, four different simulation software packages are ported to the cloud-based platform supporting various use-cases, including discrete event simulation for supply chain optimization, capacity management and maintenance planning, data mining for the optimization of aircraft maintenance, fluid dynamics simulation, and the simulation of tailored insoles for shoe manufacturing. An additional six to eight use-cases will be incorporated to the project during the summer of 2014 as a result of an open call. The rest of this section first explains the CloudSME Simulation Platform architecture and the role of WS-PGRADE/gUSE in this, and then the implementation of a related use-case based on WS-PGRADE/gUSE is introduced.

19.4.1 Role of WS-PGRADE/gUSE in the CloudSME Simulation Platform

The generic architecture of the CloudSME Simulation Platform is illustrated in Fig. 19.4. The platform is developed in a layered architecture.

The top layer, simulation and application layer, provides the SaaS services through custom user interfaces that enable the execution of predefined models and scenarios utilizing various simulation software products. CloudSME currently supports four simulation soft-ware solutions:

- Simul8 by Simul8 Corporation (a UK-based SME), a commercial off-the-self discrete event simulation package,
- TransAT by Ascomp GmbH (a Swiss SME), a multiphysics, finite-volume code for fluid dynamics simulations,
- 3D Scan Insole Designer by INGECON (a Spanish SME), a design and simulation platform for the design of tailored insoles for footwear used for sports are by people with foot problems,
- BFly by 2MoRO Solutions (a French SME) that offers simulation environment to structure heterogeneous data for optimizing aircraft maintenance planning and generating alerts.

The middle layer, Cloud Platform Layer, consists of WS-PGRADE/gUSE and the CloudBroker Platform. The CloudBroker Platform offers access to multiple heterogeneous cloud resources, assuring that the solutions will not be tightly coupled to a particular cloud. This flexibility is especially important for simulation software provider SMEs which do not want to be locked-in with a specific cloud infrastructure or technology. WS-PGRADE/gUSE, on top of the CloudBorker platform, offers workflow creation and execution capabilities that are specifically required for more complex scenarios, including multiple steps or parameter sweeps, for example.

The Cloud Platform Layer supports various scenarios for the implementation of the actual use-cases. Software vendors can connect existing user interfaces directly to the CloudBroker platform using its REST or Java API, or can utilise the Remote API of WS-PGRADE/gUSE to implement this connection. The ASM API of WS-PGRADE/gUSE can also be applied in developing custom interfaces for end-users. The project also develops several commercial components at this layer that will enable accounting and billing functionalities and their seamless integration into the SaaS solutions.

Finally, the Cloud Resource Layer is a test-bed of commercial and academic cloud resources that enables the execution of the simulation applications. The Cloud Resource Layer consists of heterogeneous cloud resources, including OpenStack and OpenNebula-based academic clouds, CloudSigma (an independent cloud resource provider from Switzerland), and Amazon commercial cloud resources.
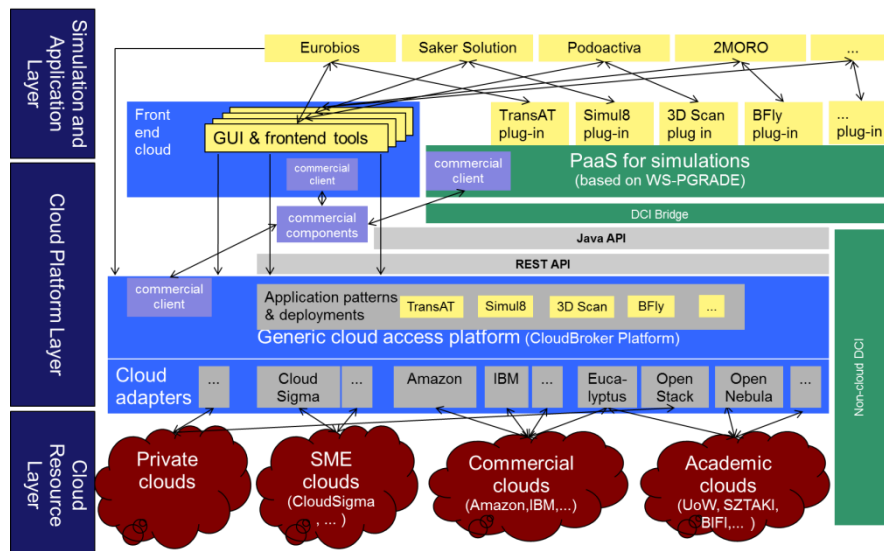

Fig. 19.4 Architecture of the CloudSME Simulation Platform

19.4.2 WS-PGRADE Gateways in CloudSME

As the CloudSME Simulation Platform provides flexibility regarding the actual architecture of the SaaS solutions, not all implementations and use cases will necessarily utilize WS-PGRADE/gUSE. At the time of writing of this book, three out of the four simulation software vendors are working on WS-PGRADE-based implementations.

Ascomp and Simul8 are using the Remote API to execute WS-PGRADE workflows from their existing user interfaces. On the other hand, for the INGECOM use-case a custom WS-PGRADE portlet has already been implemented using the ASM API. This first prototype implementation of this portlet is illustrated on Fig. 19.5.

The INGECON 3D Scan Insole Designer is based on the 3D Scan Sport Podoactiva process that is a unique method to obtain a 3D scanned image from the foot using a laser. The objective of the cloud-based solution is to make the validation of the scan available in specialized surgeries, with no geographical boundaries. The patients' feet are scanned, and the obtained images can be uploaded to the cloud for validation directly from the surgery. The validation finishes in a few minutes, after which the podiatrist can decide whether any rescan or modification is required. Currently, this validation is performed off-line, making the process time-consuming. A cloud based SaaS solution, (Fig. 19.5), can significantly speed-up the process and make it available from thousands of surgeries in Europe. The upper part of the window is the upload portlet, while the results are shown in the lower part.
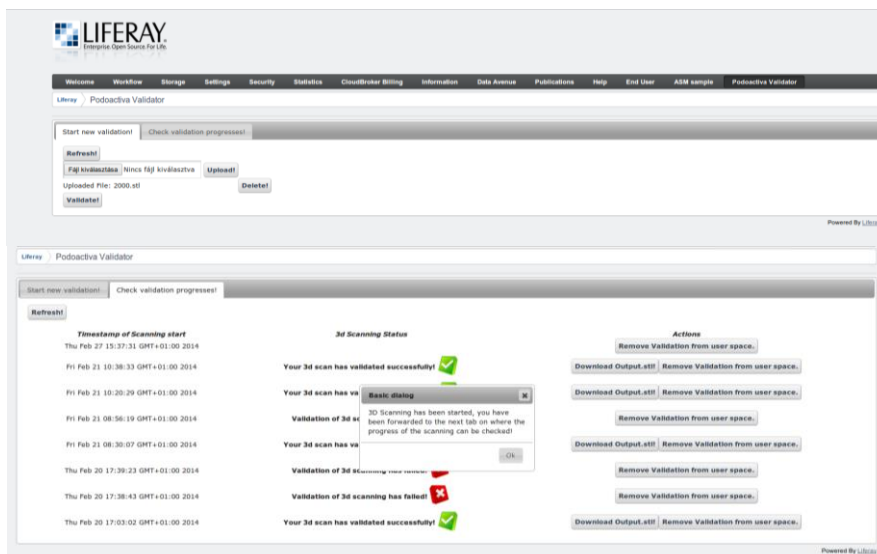

Fig. 19.5 3D Scan Insole Designer portlet in WS-PGRADE/gUSE

19.4.3 Further Development

The CloudSME project is only in its first year at the moment. The implementation of five core use-cases based on the four incorporated software solutions is a work in progress, with early prototypes, as it was illustrated in Sect. 19.4.2, al-

ready available. These prototype implementations are scheduled to be ready by autumn 2014 with fully operational solutions by the end of 2015.

Besides the core partners and application scenarios, the project also implements several other use-cases that not only demonstrate the wider applicability of the technology but that will also provide valuable feedback for the further development of the platform. These new use-cases were selected in an open call process during the summer of 2014.

19.5 Conclusions

For small and mid-sized companies, the advantages of using cloud infrastructures are evident. Cost effectiveness as well as low operational and administrative overhead are the most important factors. Leveraging the competence of cloud operators, backup and recovery of the system are straightforward. Additionally, these companies could have access to almost unlimited computing and storage capacity.

However, the access to cloud infrastructures should be facilitated via used-friendly interfaces that enable the utilization of cloud resources in a user-transparent way. In this chapter three examples for providing commercial solutions for SMEs via the WS-PGRADE/gUSE framework, and its capability to access heterogeneous cloud resources through the CloudBroker platform, were illustrated. Utilizing the Remote API or the ASM API of the framework enables companies to develop commercial solutions that run computation, or data-intensive applications as SaaS on various clouds. As these first examples illustrate, the WS-PGRADE/gUSE framework is not only an academic solution but can also be successfully exploited in commercial scenarios.