International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014)

Parameter Sweeping Programming Model in Aneka on Data Mining Applications

P. Jhansi Rani[1], G. K. Srikanth[2], Puppala Priyanka[3]
[1,3]*Department of CSE, AVN Inst. of Engg. & Tech., Hyderabad.*
[2]*Associate Professor, Department of CSE, AVN Inst. Of Engg. & Tech., Hyderabad.*

***Abstract*— Data mining applications and techniques are used in many areas as a required knowledge discovery from large data sets. Cloud computing is one of the prevailing models based on IP architecture. Cloud computing is nothing but the delivery of the computing services over the internet to improve the business of many organizations. Cloud systems which can be effectively handle parallel mining since they provide scalable storage and processing services, software platforms for developing and running data mining applications. In this paper, we present a data mining application in .NET frame work that supports the execution of parameter sweeping programming model on cloud. Parameter sweeping is an important task in the domains of the system modeling and optimization. The frame work has been implemented using Aneka platform. Parameter sweeping applications can be highly computing demanding, since the number of single tasks to be executed increases with the number of swept parameters and the range of their values. In this paper the parameter sweeping model is implemented on a data set by using the design explorer user interface.**

***Keywords*- Data mining, Cloud computing, Parameter sweeping; Aneka**

I. INTRODUCTION

Cloud computing refers to the delivery of computing services allow business to use software and hardware that are managed by the third parties at remote locations on pay-as-you-go basis. Many companies are delivering the cloud services, for example Google, Microsoft, Salesforce.com. Clouds can be classified as public, private, hybrid depending on the model of deployment. A public cloud is a cloud available on pay-as-you-go manner to the general public. A private cloud is a data center for a particular organization. The hybrid cloud is a seamless use of public cloud along with private cloud. The computation resources in a cloud are serviced to the end users in different ways as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). SaaS refers to applications delivered as cloud services and end users can use the applications that are accessible at any time and from anywhere. PaaS refers to environment for the application development through users can create their own applications that will run on clouds.

IaaS refers on demanding computing capacity from a service provider which contains the virtualized hardware and storage. There are several solutions available in PaaS, few are Google App Engine, Microsoft Windows Azure, Force.com and Manjrasoft Aneka. In this paper we are considering the PaaS as Aneka.

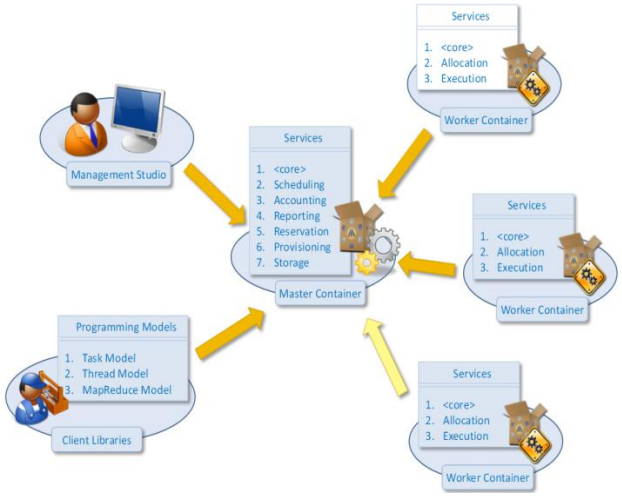

Fig-1 Basic Architecture of Aneka

The above figure-1 shows that the basic architecture of Aneka. The system includes 4 key components including Aneka master, Aneka worker, Aneka management console and Aneka client libraries. The master runs the scheduling, accounting, reporting, reservation, and provisioning and storage services. While the workers run execution services.

Different kinds of cloud computing programming models are available for the Aneka PaaS, those are Task Programming, Thread Programming, MapRecduce Programming and Parameter Sweeping Programming models. In this paper mainly we are concentrating on parameter sweeping programming model on Aneka PaaS.

Advantages of Aneka PaaS are Aneka provides a more flexible model for developing distributed applications and provides integration with external clouds such as Amazon EC2 and Gogrid.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014)

Design and implementation in the Aneka (PaaS) is the integration of the Aneka PaaS and Windows Azure platforms. Aneka is built on a solid .NET service oriented architecture allowing seamless integration between public clouds and main stream applications. It harnesses the computing resources of a heterogeneous network of desktop pc's and servers on demand. Aneka provides rich set of APIs for transparently exploiting such resources and for executing different applications by using a variety of models and abstractions. Aneka acts as a middle man integrating the access to the public clouds from user applications.

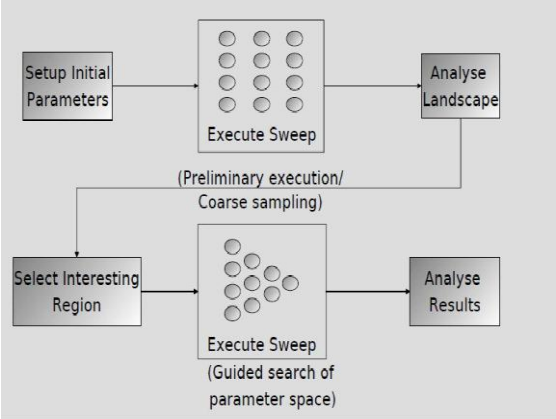

Fig-2 Basic Parameter sweeping System

Parameter sweeping is an important task in the domains of system modeling and optimization. Parameter sweep experiments are performed as subsequent batch job submissions as shown in fig-2. Scientists wait until the entire experiment finished before analyzing and doing further refinements/changes on the experiments settings. In Aneka many programming models has been designed to be extensible and these classes can be used as a starting point to implement a new programming model. This can be done by using the base classes to define new models and abstractions. The parameter sweep model is a specialization of the task model and it has been implemented in the content of management of applications on Aneka. It is achieved by providing a different interface to end users who just need to define a template task and the parameter that customize it. In the below section-II we are explaining the user interface by taking one data set and in the section-III we are mentioning the application scenario.In the section-IV we are concluding the paper.

II. User Interface

The design explorer is a visual environment that helps user to quickly create parameter sweeping applications and run it in few steps. More precisely those are

- Identify the executable required to run the application.
- Define the parameters that control application executable and their domains
- Provide the required input files for running the application.
- Define all the output files that will be produced by the application and made available to the user.
- Define the sequence of commands that compose the task template that will be run remotely.

Once the template is completed the design explorer allows the user to run directly it on Aneka's cloud by using parameter sweeping API's.

III. Application Scenario

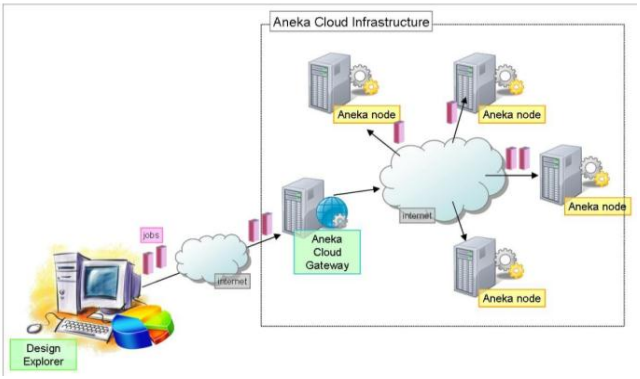

Figure 3. Application Scenario

The Design Explorer is integrated environment for quickly prototyping Parameter Sweeping applications, controlling and monitoring their execution on Aneka Clouds.

The Aneka Design Explorer is locate in the bin directory of the Aneka installation ([Programs Folder]\Manjrasoft\[Aneka Version]\bin) and it is accessible from the Start→All Programs Manjrasoft →[Aneka Version] →Design Explorer menu item.

The Aneka PSM APIs provide the logic for creating the sequence of task instances (jobs). They automatically submit these tasks to the Aneka Cloud and collect back their results that are then presented to the user through the Design Explorer from a template task given the parameters domains.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014)

In order to create a new Parameter Sweeping application it is necessary to create a new project. This can be done by clicking the leftmost icon in the toolbar representing a blank sheet or selecting the File →New...menu item.

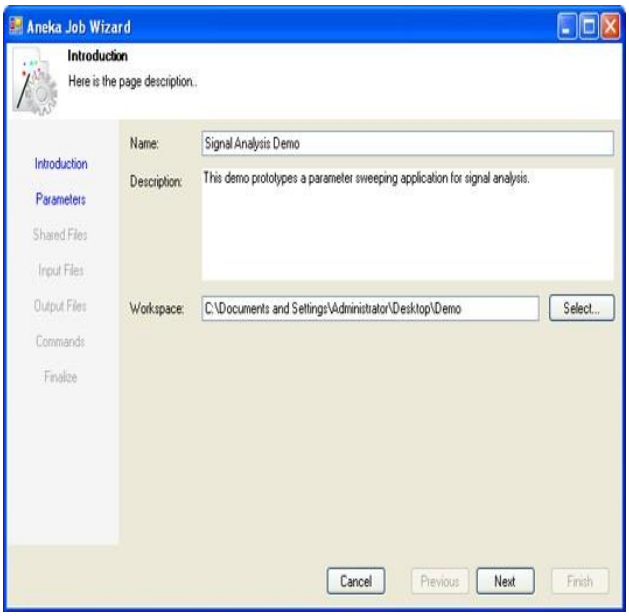

Figure 4. Aneka Job Wizard: Application Details.

Figure 4 shows the first page of the Aneka Job Wizard that is activated by the previous operation. In this page the user is requested to enter some general details of the application being created such as a name, a description, and the workspace directory. On the left side of the wizard it is possible to see all the steps that will be covered in order to define the task template of the Parameter sweeping application.

Once the user has successfully entered the detail of the application can press the Next button and move to the Parameters page where he/she can define all the parameters that control the application.

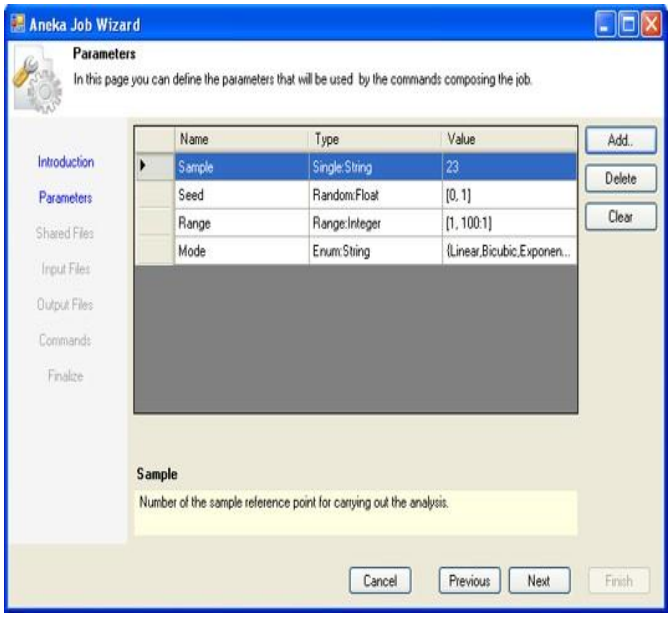

Figure 5. Aneka Job Wizard: Parameter Definition.

Figure 5 shows the Parameters page. It shows the list of parameters currently defined for the application. A parameter is defined by three elements:

The Design Explorer allows defining four different types of parameters:

a. *Single:* represents a parameter that can assume one single value. The underlying type of the parameter is string.
b. *Random:* represents a parameter that can assume a random value within a range limited by a lower and an upper bound. The parameter is a real number.
c. *Range:* represents a parameter that can assume a discrete set of values within a limited range and that are generated by starting from the lower bound and adding a step. The parameter is an integer number..
d. *Enum:* represents a parameter that can assume a discrete set of values that are defined by the user.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014)

The underlying type of the parameter is string.

For all the parameters described above a name is mandatory while the user can enter an additional comment that specifies the role of the parameter.

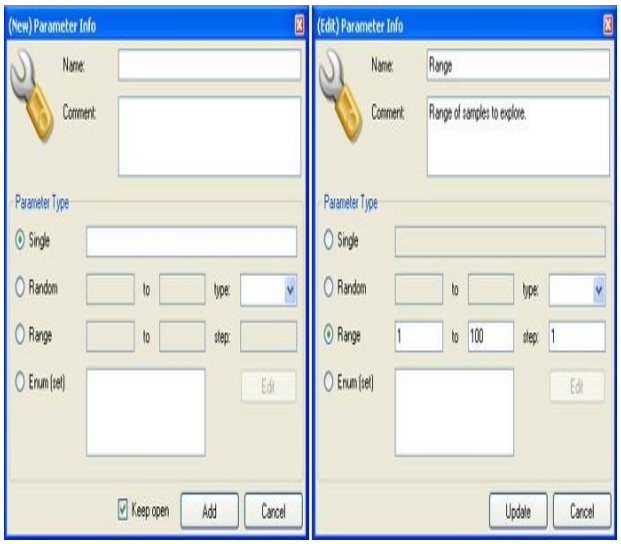

Figure 6. Aneka Job Wizard: New and Edit Parameter Modes.

The next two steps allow users to specify input and output files for each of the job instances. Differently from the shared files, input and output files can be specialized with parameters. This means that the real name of the file is generated and checked at runtime by the PSM engine. It is possible to have three different views for the parameters:

- *All parameters:* shows all the available parameters.
- *User parameters:* shows only the parameters defined by the user in the task template.
- *Special parameters:* shows only the system parameters that are available by default for each job instance.
- At the moment only the Task Id (Job identifier) parameter is available in this list. Special parameters are characterized by a leading $ in the parameter name that makes them reserved words.

The first option shows both users and special parameters. Once the user has entered all the files, by pressing the Next all the files are checked and an error message box is displayed for those that are not valid. Figure 8 shows the Input and Output files pages.

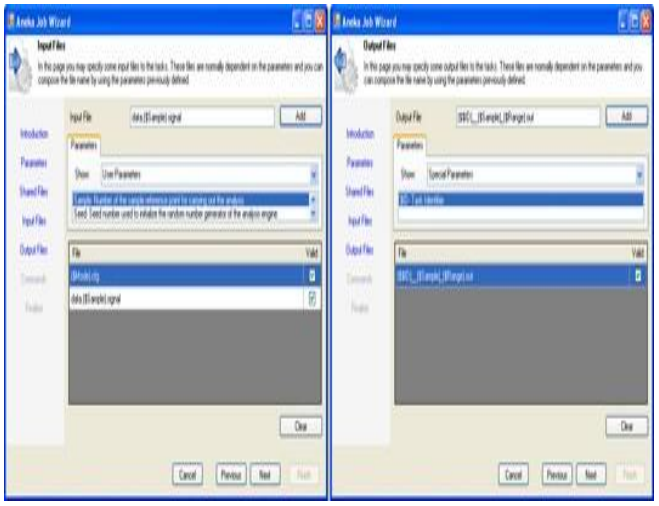

Figure 8. Aneka Job Wizard: Input and Output Files pages.

The final step for defining a task template is specifying the sequence of operations that characterize will be executed on the remote node for each of the job instances. This is the last step because the sequence of commands can make use of all the previous elements: parameters, shared, input, and output files.

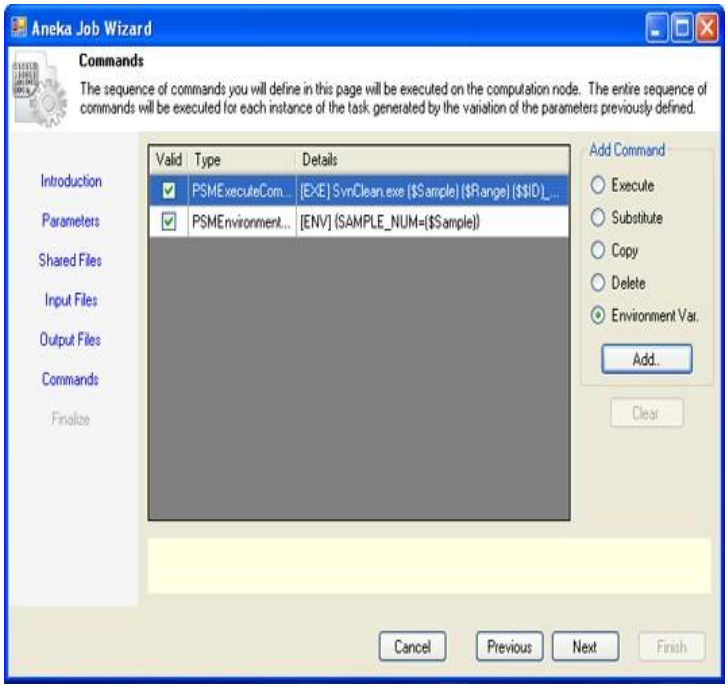

Figure 9. Aneka Job Wizard: Commands Page.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014)

Figure 9 shows Commands page. The users can select among five different ready to use commands:

- Copy command (CPY): this command completely executes on the remote node and copies a file to another file under a different name but always on the same node. Other implementations of the parameter sweeping model use the copy command to move files from the local client machine to the remote node. With Aneka this task is transparently done and there is no need to do that explicitly in the task template.
- Delete command (DEL): deletes a file on the remote node.
- Execute command (EXE): executes a shell command or console application on the remote node.
- Substitute command (SUB): substitutes the occurrences of the parameters with their run time values into a file.
- Environment command (ENV): sets a collection of environment variables in the shell used to execute the template task on the remote node.

The creation of commands is the last step for creating the template.

Figure 10. Aneka Job Wizard: Job Completion Page.

Figure 10 shows the Job Completion page. The user is presented with different options:

- It is possible to save the task template into an XML file. This is accomplished by providing a name into the Save path text box or by pressing the Browse button to look for an existing file. Once the name is set, it is possible to press the Save button.
- It is possible to directly edit the XML source file of the task template. This is accomplished by clicking the Edit button. This feature is only available on the .NET/Windows version; when the code is compiled for the Mono environment an informative message is displayed in place of the XML editor that allows modifying the source of the template.
- It is possible (default action) to open a project and run the parameter sweeping application into the Design Explorer. This option is checked by default and opens up a Project Window through which the users can monitor and execute and modify the template.

In order to execute a project it is necessary to authenticate against the Aneka Cloud. The user has to provide the access point to the cloud and valid user credentials. Once the user has opened or created a new project he or she can run It by clicking the play icon in the project window toolbar. As long as the project is running the run icon shows the stop symbol and by clicking on it is possible to terminate the execution of the project. Once the user run the project and there are no problems in connecting with the Aneka Cloud the Parameter Sweeping application starts and the two tabs on the right pane are filled with information about the running application.

IV. Conclusion

Cloud computing infrastructures can be effectively used to run data intensive applications. This paper presents the efficient execution of parameter sweeping data mining applications in Aneka PaaS framework. The advantage of this frame work is integration of more two platforms and it supports many programming models. This integration of more than one platforms would give numerous benefits to not only the users of Aneka but also the windows Azure platform. The user interface is very simple and hides the complexity of the Cloud infrastructure used to run applications.

The experimental results discussed in the paper demonstrates the effectiveness of the proposed framework, as well as the scalability that can be execute the parameter sweeping applications on a pool of virtual servers. Other than supporting users in designing and running parameter sweeping data mining applications we intend to exploit Cloud computing platforms for running service oriented knowledge discovery processes designed as a combination of several data analysis steps to be run in parallel on Cloud computing elements. By using the console tool we can analyze all the tasks. Particularly in this paper we are not explaining how to compose by using the API a task template and generate and run tasks from it.

Acknowledgement

We would like to thank everyone who has motivated and supported us for preparing this manuscript.

REFERENCES

[1] Fabrizio Marozzo, Domenico Talia , Paolo Trunfio, DEIS, University of Calabria, Rende (CS), Italy -A Cloud Framework for Parameter Sweeping Data Mining Applications,

[2] A.Wibisono,D.Vasyunin,V.Korkhov,F.Terpstra(Virtual laboratory for e-science)-Towards a system for interactive Parameter Sweep Applications on grid.

[3] Dr.Rajkumar Buyya,Karthik Sukumar- Platforms for building and deploying applications for cloud computing.

[4] Ruxandra-Ştefania PETRE Bucharest Academy of Economic Studies-Data mining and cloud computing

[5] B.Suresh Kumar, GirishPaliwal ,Mr.Manish Raghav,Mr. Sudeep Nair –Aneka(PaaS)

[6] Yi Wei,Karthik Sukumar, Christian Vecchila,Dr.R.K.Buyya-Aneka cloud Application platform and it's integration with windows Azure.

[7] Liu, Yuan, Zhang, Chen, Yang-Cloud work flow system Architeture.

[8] Dr.Raj kumar Buyya- Cloud Application Programming and the Aneka platform.

[9] Dr.Raj kumar Buyya for Aneka APIs-Manjrasoft Aneka.

[10] Christian Vecchiola, Suraj Pandey, and Rajkumar Buyya (Cloud computing and Distributed Systems (CLOUDS) Laboratory -High performance cloud computing

BIOGRAPHY

P. Jhansi Rani[1] is pursuing Post Graduate in Master of Technology with specialization of Computer Science & Engg. at AVN Inst. of Engg.& Tech, Hyderabad, AP, India. I am interested research area is Data Mining and Cloud Computing.

G.K.Srikanth[2] is working as Associate Professor in the department of Computer Science & Engg. at AVN Inst. of Engg.& Tech, Hyderabad, AP, India. He is interested research area is Data Mining, Network Security and Networking Technology.

Puppala Priyanka[3] is pursuing her Post Graduate in Master of Technology with specialization of Computer Science & Engg. at AVN Inst. of Engg.& Tech, Hyderabad, AP, India. Her interested research area is Data warehousing & Data Mining, Network Security and Data Structures.