JISC

Technical Review of Using Cloud for Research

Guidance Notes to Researchers

June, 2010

Introduction

There is a rapidly increasing range of research activities which involve the outsourcing of computing and storage resources to public Cloud Service Providers (CSPs), who provide managed and scalable resources virtualised as a single service. For example Amazon Elastic Computing Cloud (EC2) and Simple Storage Service (S3) are two widely adopted open cloud solutions, which aim at providing pooled computing and storage services and charge users according to their weighted resource usage. Other examples include employment of Google Application Engine and Microsoft Azure as development platforms for research applications.

Despite a lot of activity and publication on cloud computing, the term itself and the technologies that underpin it are still confusing to many. This note, as one of deliverables of the TeciRes project[1], provides guidance to researchers who are potential end users of public CSPs for research activities. The note contains information to researchers on:

- The difference between and relation to current research computing models
- The considerations that have to be taken into account before moving to cloud-aided research
- The issues associated with cloud computing for research that are currently being investigated
- Tips and tricks when using cloud computing

Readers who are interested in provisioning cloud capabilities for research should also refer to our guidance notes to cloud infrastructure service providers. This guidance notes focuses on technical aspects only. Readers who are interested in non-technical guidance should refer to the briefing paper produced by the "using cloud computing for research" project.

Relations to pre-existing research computing models

So, what is cloud computing? How is it different from previous computing models?

In order to better understand the essence of cloud computing, we use an analogy to Web 2.0. Web 2.0 caused many arguments when it was firstly proposed, and the cloud is in the same situation. First of all, cloud or cloud computing is a business term, as is Web 2.0, and it refers to the way we deliver and use computing services rather than to a new technology. Both Web 2.0 and cloud are underpinned by existing computing technologies and communication protocols. The emergence of Web 2.0 brings us a new RESTful design pattern. Similarly, we need to rethink how we build cloud applications in order to get the full benefits of cloud computing.

In principle, cloud computing can deliver existing types of computing services, including grid computing services, high-performance computing services, etc. This does not mean that cloud computing is a replacement for grid computing, high-performance computing or other existing computing models. Instead, cloud computing in essence a new business term that aims at highly scalable and highly flexible service delivery. The difference between cloud computing and other existing computing models can be better demonstrated by reference to the TeciRes "towered architecture" and its key enabling technology, virtualisation. Although

[1] http://tecires.ecs.soton.ac.uk

other key enabling technologies may also employed by current computing models, they are not fundamental technologies which differentiate cloud from other computing models.

What are the current issues with Cloud Computing that I should be aware of?

According to the example cases classified in the TeciRes project, two scenarios have been identified where researchers use public cloud services, being cloud outsourcing and cloud bursting. More details of the example cases can be found in the TeciRes final report. The following provides an outline of the technical issues of which a researcher needs to be aware.

Capacity planning

Capacity planning and resource allocation are traditional issues for parallel and distributed computing. Traditionally, researchers had to consider and define capacity planning and resource allocation strategies, depending on their use cases and the local deployment environment. When outsourcing to third-party cloud computing facilities, the burden of capacity planning is, to some extent, shifted to CSPs. However, there may be limited functionality offered by public CSPs. Researchers as end users still need to consider how much computing and storage capabilities to reserve to ensure reasonable system performance. Some public CSPs provide auto-scaling facilities either through dedicated service APIs or programming APIs. A cloud computing application still needs to define custom auto-scaling policies to respond to application-specific fluctuating workloads. The TeciRes final report gives more details on auto-scaling features of individual public CSPs.

Security and data privacy

Data privacy is considered by many researchers to be a principal requirement. Concerns come mainly from two considerations. First of all, data stored in a Cloud is potentially distributed to a worldwide infrastructure, and the location of the data may be unknown to end users. This relates to existing concerns over UK academic data being used and stored outside of the UK. Second, the underlying virtualisation technologies of the Cloud may exhibit potential security holes for malicious users to compromise data privacy from the internal Cloud environment. Although data privacy can be defined as a quality of service (QoS) metric in the Service Level Agreement (SLA), many researchers argue against its usefulness since the outcome of security fraud is often unquantifiable.

Virtualisation issues

Virtualisation is a key enabling technology for cloud computing and scalability. Virtualisation can take place at both the platform and application level, making it easier for researchers to develop and use new applications while hiding the complexity of the low-level infrastructure and reducing manageability overheads.

Researchers debate the performance of Virtual Machines (VM), which are perhaps not suitable for high performance applications, especially for those applications involving closely coupled communications.

Monitoring and logging issues

Current offerings of public CSPs provide limited functionality for monitoring and logging virtual resources. These facilities can be employed by a cloud application to monitor resource utilisation, disk usage, and network traffic. However, these limited monitoring functions focus on single virtual machine instances at the infrastructure level, making it hard for them to be used for complex and dynamic infrastructures involving multiple virtual resources, as in a workflow for example.

In order to enable monitoring and management of cloud resources like other physical resources deployed in a local environment, there are an increasing number of third-party cloud monitoring tools emerging which leverage existing monitoring functionality provided by public CSPs, while providing advanced features for monitoring and logging cloud services at different levels. More details can be found in the TeciRes final report.

Although there are also many scientific computing monitoring tools available, their effectiveness and applicability in cloud computing environment needs to be demonstrated.

Programmability, debugging, and profiling

At present, cloud applications can be deployed to a cloud environment in two ways. For cloud applications targeted as Infrastructure-as-a-Service (IaaS) CSPs, applications along with runtime dependencies are packaged as a Virtual Machine Image (VMI), which is attached and deployable to a virtual machine instance. Alternatively, most Platform-as-a-Service (PaaS) CSPs provide a development framework (i.e. a Software Development Kit) allowing end users to develop, debug, and deploy cloud applications by exploiting dedicated programming APIs.

These two deployment models exhibit certain advantages and limitations. For the VMI deployment model, a researcher can employ various programming abstractions such as Messaging Passing Interfaces (MPI) and OpenMP, and can debug applications in the local environment before deploying into the cloud. However, this model requires additional tasks for researchers. In addition, the local debugging process makes it difficult to guarantee the same effectiveness in the target cloud environment. Compared to the VMI deployment model, a PaaS CSP provides cloud-friendly programming and debugging facilities, but with limited programming abstractions (e.g. simple C/S and multi-threading) making it difficult to develop complex scientific applications.

Federation and interoperability

Most Cloud application developers consider it best that applications be hosted by a Cloud provider that is interoperable with other Cloud providers. The consensus is that a standards-based API for manageability is desirable to save re-engineering efforts when moving an application from one Cloud infrastructure to another. Standardisation is also important for the federation of infrastructure resources from multiple Cloud providers. Interoperability is more important for open-source Cloud providers than for commercial Cloud providers at present.

Tips and tricks

The following provides some tips and tricks derived from two case scenarios of cloud outsourcing and cloud bursting which are of most relevance to researchers. More details of these two case scenarios and example cases can be found in the TeciRes final report.

How to select a public CSP?

Consider sourcing to a public CSP that is focused on the level of the stack you are interested in. For example:

If you are looking for infrastructure as a service (IaaS), cloud providers that focus on this level of the stack may be worth evaluating. For example, the various Amazon services (EC2, S3 etc) are currently focused on providing IaaS.

If you are looking for platform as a service (PaaS), cloud providers that focus on this level of the stack may be worth evaluating. For example, Microsoft Azure provides a suite of PaaS functionality. Google provides various PaaS offerings e.g. Google App Engine and Google Docs to share any file (i.e. a cloud storage-type offering).

If you are looking for Software as a Service (SaaS), cloud providers that focus on this level of the stack may be worth evaluating. Further, at this level of the stack it may be particularly useful to choose a research application-focused offering. For example, e-Science Central (www.esciencecentral.co.uk) is one research-specific offering.

Things to be aware of when bursting to public CSPs

For HE sectors that have local research computing facilities, but would like to burst to public CSPs to accommodate fluctuating workloads, it is important to be aware of the fact that it is difficult to run high performance closely coupled distributed applications on many cloud providers. Many research applications are structured as a high performance closely coupled distributed application. That is: the application is actually a number of jobs spread across two or more distributed nodes; and the jobs require a high bandwidth and/or low latency connection between the nodes so the jobs can rapidly exchange data, for example via MPI.

If your application matches this description, be aware that many public cloud providers do not currently focus on making this class of application high performance. For example, they may not guarantee predictable high bandwidth and/or low latency connections between the nodes. That is not to say this will not be provided, but it is may not be guaranteed to the same extent that would likely be the case in a modern HPC resource in a University. Indeed, if you have a closely coupled application that you currently run on a HPC resource you may be better off without using a cloud.

Conclusion

This guidance note provides information related to cloud computing including its relations to existing research computing models; considerations and decisions that have to be made when using cloud computing for research; technical issues associated with cloud computing for research from users' perspectives; and tips and tricks for the realisation of case scenarios of using cloud for research as observed during the course of the TeciRes project.

For more Information

- **TeciRes Project**
  http://tecires.ecs.soton.ac.uk
- **TeciRes Report**
  http://tecires.ecs.soton.ac.uk/documents
- **Review of Using Cloud Computing for Research**
  http://www.jisc.ac.uk/fundingopportunities/funding_calls/2009/09/cloudcomputing.aspx
- **Review of Environmental and Organisational Implications of Cloud Computing for HE and Further Education**
  http://www.jisc.ac.uk/whatwedo/programmes/greeningict/environmentalreviewcloudcomp.aspx

This document was produced by David Bacigalupo, Xiaoyu Chen, Gary B. Wills, and Lester H. Gilbert, from the Learning Societies Lab of the University of Southampton during the course of the TeciRes project funded by JISC.