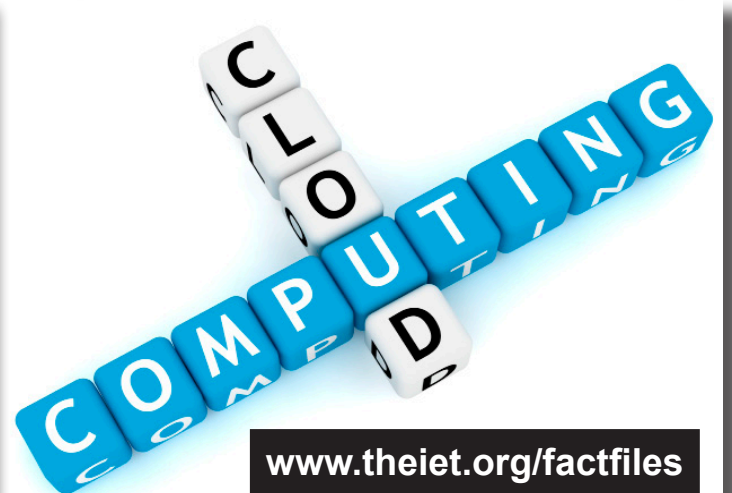


Cloud computing

Introducing Cloud computing and examining key factors in its design, implementation and adoption

A Factfile provided by the Institution of **Engineering and Technology**



About This Factfile

The Institution of Engineering and Technology acts as a voice for the engineering and technology professions by providing independent, reliable and factual information to the public and policy makers. This Factfile aims to provide an accessible guide to current technologies and scientific facts of interest to the public.

For more Position Statements and Factfiles on engineering and technology topics please visit <http://www.theiet.org/factfiles>.

The Institution of Engineering and Technology

The Institution of Engineering and Technology (IET) is a global organisation, with over 150,000 members representing a vast range of engineering and technology fields. Our primary aims are to provide a global knowledge network promoting the exchange of ideas and enhance the positive role of science, engineering and technology between business, academia, governments and professional bodies; and to address challenges that face society in the future.

As engineering and technology become increasingly interdisciplinary, global and inclusive, the Institution of Engineering and Technology reflects that progression and welcomes involvement from, and communication between, all sectors of science, engineering and technology.

The Institution of Engineering and Technology is a not for profit organisation, registered as a charity in the UK.

For more information please visit <http://www.theiet.org>

© The Institution of Engineering and Technology 2011

The Institution of Engineering and Technology is registered as a Charity in England & Wales (no 211014) and Scotland (no SC038698).

Enquiries

policy@theiet.org

Contents

Introduction	3
What is Cloud Computing?	3
Cloud Service Patterns	3
Cloud Service Provision	4
The Business and Commercial Case for Cloud Computing	4
The Private Cloud.....	4
The Public Cloud	5
The Commercial Case	5
The Issues with Cloud Computing	5
Performance	6
Companion factfiles for future publication	6

Introduction

This paper introduces Cloud Computing and examines key factors in the design, implementation and adoption of Cloud computing. It forms the core of a series of IET Factfiles on Cloud Computing and introduces the basic concepts and a taxonomy of topics which will be explored in more detail in subsequent papers.

What is Cloud Computing?



The term “Cloud Computing” stems from technical developments which remove the physical ties between the hardware and software components of computing systems.

Historically, systems were constructed around static designs where each application was mapped to a physical server and other related hardware components like disk storage, printers and network controllers. This system model was relatively easy to understand and build but suffered from inflexibility with respect to performance and capacity. As techniques evolved multi-programming operating systems, coupled with virtual memory, allowed more than one application to run on one physical server. System tasks and business applications could run alongside one another in dedicated partitions with the operating system allocating processor, memory and I/O resources as required. Practical implementations of multi-programming assumed that not all applications would require full use of system resources at the same time and so the systems manager gained improvements in capacity flexibility. Advanced designs of large multi-programming systems could push utilisation of the most expensive components like the main CPU and its main memory to 90% or more by judicious scheduling of workload and applications throughout the working day.

Even so, more flexibility was required, particularly in development environments, where new projects would need new customised hardware and software. The activities associated with procuring and commissioning these systems were expensive and time consuming. To address this challenge, virtual machines were introduced as early as the early 1970s. Virtual machines allowed a single mainframe computer to run many complete operating system instances, with associated I/O devices, at the same time. Developers and testers alike could be served by many tens or even hundreds of virtual machines within a single mainframe. Each machine would run independently of the others under the control of the virtual machine hypervisor.

This virtual machine concept has evolved through to the present day and forms the backbone of what we now call Cloud computing. Cloud computing also relies on a more recent innovation, that of automated provisioning and dispatching (or orchestration) of applications across a number of physical machines all running many virtual machine instances. Two elements of automated dispatching are

important. Firstly, the dispatcher (or orchestrator) maintains a real time awareness of system utilisation and can start new applications dynamically in systems that are lightly loaded.

Secondly, if an application requires a new virtual machine, that VM can be loaded and the application(s) started dynamically.

As workloads vary over time, the orchestrator can start and stop virtual machines as required and move applications between machines. Hence we arrive at the term “Cloud”, as the user

of computing resources has no detailed knowledge of where his applications are running or where his data are located. More recent developments in high speed peer-to-peer networking allow the pool of computers controlled by the orchestrator to be spread over many different physical locations, even in different parts of the world. A later fact file will be devoted to Cloud architectures and we shall see how these mechanisms of virtualisation and orchestration work in practice.



Cloud Service Patterns

Cloud services can be offered in a number of ways. The common patterns are:

1. A virtualised infrastructure, known as Infrastructure as a Service (IaaS), where basic virtual machines are configured together with “virtualised” data storage and network components and made available to users in that “raw” form for them to deploy and manage their own operating systems, middleware and applications. The Cloud provider has responsibility for managing the infrastructure services that he is contracted to provide.
2. A computing platform, known as Platform as a Service (PaaS), where the infrastructure is enhanced by the addition of preconfigured operating system and middleware. Users in this case will deploy and manage their own applications on the preconfigured software platform. The Cloud provider has responsibility for managing the platform software that he is contracted to provide.
3. An application service, known as Software as a Service (SaaS) where whole applications are provided on top of the managed software platform. Users in this case will be provided with an application, or applications, ready to run. The Cloud provider has responsibility for managing application service delivery to a defined user base. Those users will typically access the applications using browsers or client side web applications connected into an application web server operated by the SaaS provider. A special case of SaaS is the provision of typical desktop (PC) applications from within a Cloud infrastructure. Here the user will perform conventional office tasks (word processing, spreadsheets etc) via a browser connection into the Cloud.
4. A full business service, known as Business Process as a Service (BPaaS) where a suite of applications is provided within the Cloud infrastructure to provide the

end-to-end elements of a complete business process. Business processes are automated within the computing domain by linking different application components (application services) together to provide an end-to-end business function. This Cloud computing pattern overlaps substantially, in a commercial sense, with business process outsourcing offerings. Traditional business process outsourcing has been, and is, provided by specialist 3rd party companies who bring people, premises, technology and process management into the contract. BpaaS brings technology and process management into the agreement while the contracting company provides their own people operating out of existing premises.

Cloud Service Provision

Cloud services, defined in principle by the twin ideas of virtualisation and automated orchestration can be provided in two main ways, Public and Private. Public Clouds are provided by third parties while Private Clouds are enabled within an existing corporate IT infrastructure. Both types are enabled in the same way and contain the same types of components

in terms of infrastructure. All four Cloud service patterns can be instantiated into either public or private Clouds but commercial SaaS and BPaas

offerings are more aligned with public Cloud offerings as they offer significant business and commercial advantage to companies needing to adopt new business aligned services rapidly and easily. Private Cloud implementations are likely to be built around IaaS and PaaS to provide the IT department with greater flexibility in terms of scalability of the IT infrastructure and associated cost savings and deployment agility. Public and private Clouds each have their own advantages and disadvantages in terms of information security and privacy, performance, functionality and risk. As a result, it is typical that businesses will adopt a hybrid public/private implementation. We shall examine the business and commercial aspects of public and private Clouds together with an analysis of the advantages and disadvantages of each in summary in the rest of this article and in more detail in companion fact files.

The Business and Commercial Case for Cloud Computing

The business opportunities for Cloud computing is impressive. For many years now, IT managers have found it difficult to keep pace with the many and varied demands of their business colleagues to develop new applications and systems. The chief difficulty is that established IT systems consume an enormous amount of effort and cost just to maintain them; it is often reported that up to 70% of IT staff are allocated to maintenance and upgrade activities just to keep the existing

business systems running. IT managers are encouraged to build a closer alignment with the needs of the business that they support but just don't have the resources and funding to do that effectively. Hardware and software components have become very complex and difficult to re-engineer; new application projects need new development environments which can be costly and time consuming to acquire and set-up; new development methods need new skills and techniques.

The opportunity to procure prebuilt business systems and computing platforms from a third party Cloud provider is very compelling. The opportunity to build a more flexible and versatile in house computing environment through adoption of a private Cloud provides a new way for IT departments to be more responsive to the needs of the business. Some examples will help to illustrate this.



The Private Cloud

1. **Application Development.** Private Cloud implementation allows the IT department to provide rapid deployment of multiple development environments as and when they are required. New system images for development and testing can be set up on virtual machine partitions almost instantly without the need to procure and configure new hardware. As each new project completes, resources can be released ready for the next.
2. **Production Systems.** Production systems can be complex and inflexible in terms of server capacity and utilisation. Workloads tend to vary widely during the day and many servers can remain lightly utilised except for short peaks of demand. Virtualisation, coupled with automated dispatching and workload balancing, can reduce the number of physical servers required while at the same time increasing the average utilisation within the server pool leading to significant cost reduction while at the same time proving new ways to tackle "green" initiatives.
3. **Virtual Desktop.** Personal computers in the workplace, laptops, desktops and other mobile devices, have become expensive to own and operate but are an essential tool for a productive workforce. The initial cost of acquisition is high but the cost of managing these devices to ensure software integrity, information security and to provide an effective support service can be much higher. Often the annual cost of service and support can be up to 5 times the cost of the initial procurement. But in many cases, the range of applications is limited to internet, email and standard office functions. Virtualisation of the desktop, where network connectivity permits, can allow these applications to be server based and normally accessed via a web browser. The client devices can be of lower specification, the applications themselves can be more easily standardised across the enterprise, maintenance fixes and patches for the applications do not have to

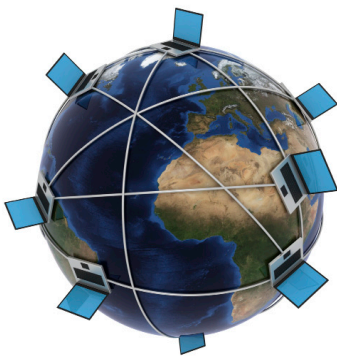
be distributed over the network and the server based application deployment aids interoperability.

The Public Cloud

In principle, the examples above, applied to the private Cloud can also be delivered by a public Cloud provider. Those examples above covered the patterns of IaaS, PaaS and SaaS. Additional services that are typical in public Clouds are, for example:

1. Applications. Many public Cloud providers offer applications like workforce automation, CRM, e-procurement and e-commerce. The functionality being offered will have been tried and tested by the provider and by existing customers. For a company wanting a rapid start to a new business initiative these solutions provide a significant advantage. Inevitably, there will be work involved in aligning data models and business processes and in practice the service provider design will need to be adopted with little modification. However this effort will be a fraction of that required to build an application from scratch or even purchase ready built software to run in house. Compromise over implementation detail will be inevitable but is eventually seen as a practical necessity.
2. Business Processes. For a new startup business, with no existing IT estate, BPaaS may provide a game changing approach. Provision of standard business processes like CRM, ERP, e-commerce, e-procurement on with rapid start-up can be very attractive. Inevitably, customers will have to accept the Cloud provider's process implementation with little flexibility for customisation but for a new business this issue may well be insignificant.

The Commercial Case



Cloud computing brings with it the ability to consume and pay for computing resources on demand. Both public and private Clouds bring this significant commercial opportunity to manage IT costs. Traditional IT deployment requires significant upfront capital expenditure based on a best estimate of projected usage.

Traditional system designs are normally sized to provide a specified performance level during the peak half hour within a given time period (day, week etc.). This can lead to significant over-investment and long periods where resources lie idle. Historically, multiprogramming techniques would address some of these issues but often, systems with critical performance requirements would require dedicated hardware. The Cloud computing business model is oriented to contracted service level agreements and usage based pricing. Capital investment can be minimised and, with resources shared across a large user base, performance service levels can be more easily managed and achieved. Clearly an internal private Cloud project will need initial investment to get started but

this can often be minimised by reusing existing equipment as part of a consolidation and rationalisation programme. Private Clouds bring the opportunity for significantly improved IT asset management; both software and hardware. A later fact file in this series will explore the business and commercial case for Cloud computing in more detail.

The case for Cloud computing seems extraordinarily attractive but we need to turn our attention to some potential issues which have inhibited the take-up of Cloud computing in some organisations.

The Issues with Cloud Computing

The delivery of IT services from within a shared environment inevitably gives rise to concerns about security, privacy and performance. In Cloud systems there are three main areas where the sharing of resources could potentially create problems. We shall examine briefly the kinds of security and privacy issues that might arise:

1. The Network. Users will typically connect to their applications over IP networks. Various techniques may be employed to isolate and protect user network traffic including Virtual Private Networks (VPNs) to isolate and protect data "on the wire", firewalls to restrict user-application connections and close disallowed TCP ports, proxy servers to mask real IP addresses and web servers to control authentication of users and their connection to authorised applications. In principle, well designed and tested systems can provide complete user isolation at the network level however many potential users of Cloud services will naturally be cautious about the prospect of malicious intrusion particularly in view of ongoing prominent publicity of successful hacking attempts. Network security within a private Cloud remains within the domain of conventional enterprise security policies and so inspires confidence. Business use of public Cloud based applications needs to be governed by similar security policies together with contract based agreements with public Cloud providers.
2. The Servers. Concerns about cross memory data leakage may well be raised but virtual machine technology is already mature from this point of view and the chances of intrusion very slight. The use of automated, dynamic dispatching might also present a potential integrity issue (for example connecting users to incorrect or outdated application services). From a practical point of view, rigorous data table maintenance combined with robust account management and regular integrity testing should provide adequate assurance.
3. Shared Storage. In a typical Cloud environment, user data space will be assigned to one or both of shared network attached storage devices (NAS) and local server attached discs. As with the previous discussion on virtual servers, the local disc software and hardware technology is mature and potential data leakage problems are unlikely to arise. NAS devices can be connected over IP or fibre channel. As with any networked technologies, security techniques are complex but software solutions are mature (see point 1 above). The key to good security is, as always, robust

policies coupled with rigorous administration of users, access permissions and data space management.

Security and privacy will be discussed a greater length in a forthcoming companion fact file.

Performance

As with all technologies, performance is governed by available resources and system loading. Computing resources naturally have finite capacity and device characteristics govern the maximum transaction loads that can be handled. Cloud computing servers rely on the notion that, by spreading the available resources across many applications and groups of users, device loading can be relatively evenly distributed during the working day. Inevitably some components will reach saturation before others and, as loads increase, bottlenecks will appear. The key to a successful Cloud service which meets user performance expectations is one which has a well designed automated dispatcher at the application level which also maintains awareness of device utilisation and performance. Of course the total available system capacity has to be able to handle expected peak loads but it is generally true that the performance of large transactional systems is easier to predict with simple mathematical models (compared with smaller single application systems) and for those models to provide confidence in the multiuser Cloud environment that contracted service levels can be met. At the network level, user perceptions of system performance, response times and latency will be limited by connection speeds. There is not much that Cloud providers can do about that. However, steps can be taken to optimise the two way interactive traffic over the end user network with various compression, caching and WAN optimisation techniques being offered. Cloud based virtual desktop services provide perhaps the most stringent challenge to restricted network bandwidth and component latency.



Often, Cloud providers address this issue by offering only a small subset of office functionality, compared with typical PC based office applications. That can be frustrating for users and also introduce interoperability problems between different organisations. This particular issue is one that has not been satisfactorily solved up to now but indications are that it is only a matter of time before this use of Cloud computing can be adopted with confidence.

A companion fact file will discuss performance, usability and interoperability issues in more detail.

Companion factfiles for future publication

The question now is “What can you do with Cloud based solutions?”. From that one should ask “How do you go about building or procuring a Cloud infrastructure?” and “What are the pitfalls and issues that must be solved?” Companion fact files will be published over time to address these questions under the following broad headings

- The business and commercial cases for Cloud computing
- Cloud computing architectures
- Security and privacy in Cloud computing
- Performance and usability of Cloud computing
- Planning for Cloud computing - Public and Private Cloud projects



Factfiles

The Institution of
Engineering and Technology

The Institution of Engineering & Technology
Michael Faraday House
Six Hills Way
Stevenage
SG1 2AY

01438 765690 - Policy Department
email: policy@theiet.org
<http://www.theiet.org/policy>
<http://www.theiet.org/factfiles>

© The IET 2011

Issue 1.0 - September 2011

The Institution of Engineering and Technology is registered as a Charity in England & Wales (no 211014) and Scotland (no SC038698).