

THE OPEN GROUP SERIES

# Cloud Computing for Business

The Open Group Guide



Chris Harding and members of  
The Open Group Cloud Work Group



THE *Open* GROUP

## **4 Buying Cloud Services**

***The following excerpt is from Chapter Four – Buying Cloud Services***

### **4.1 Determining Fit**

In establishing your cloud vision, you have achieved an understanding of the business context, and made an assessment of cloud suitability. At this point you should have a good idea of whether cloud services will be part of your solution and what kind of cloud services to consider. The next stage is to see whether those services fit your functional requirements and solution models.

For IaaS or PaaS, the functional requirement is simply for infrastructure or platform; unless you need special features – in which case you should probably not be considering a cloud solution – functional fit is not a problem. For SaaS, you must establish that there are cloud services that meet your needs for application processing. For all kinds of cloud service, you then assess fit with the solution models.

Developing the models for your solution will take time and work. Do not expect them to appear magically as a result of your cloud architecture vision. You should develop the models in collaboration with your enterprise architect and the IT team. The enterprise architect will probably take the lead in this. Expect several iterations, with different versions of the models, before you are satisfied with them.

The important models for choosing and monitoring cloud services are the financial, workload, and risk models. The workload model and the financial cost model are intimately involved in the buying lifecycle. The other financial models, and the risk model, are less directly related.

This section describes the workload and cost models, and how they and the other models are used to determine fit. The descriptions are illustrated by the Consort-Prinz and ViWi examples. (Sam Pan Engineering is developing a cloud platform rather than buying cloud services, so is generally not relevant here.)

#### **Workload and Cost Models**

Your workload model is crucial. It enables you to estimate your cloud service costs. It also enables you to determine your throughput and configuration-speed requirements.

The model describes, in business terms, the amount of processing that the service must handle, and how this varies with the number of users and over time. It also describes, in technical terms, the resources that are used to perform this processing, and it shows how the resources used relate to the processing load.

There is no standard workload model. It is different for every enterprise, and for each service that the enterprise uses.

You can estimate the service costs from the resource usage in the workload model. The calculations will unfortunately be different for each supplier, because suppliers package and

## **Cloud Computing for Business**

describe their resources in different ways. This is particularly so for processing resource; the processing unit offered (and charged for) by each supplier may be different, and it can be hard to relate these units to each other. You have to do this, though, if you want to compare the different offerings. This means that you need a cost model for each supplier.

It is a good idea to note how sensitive the models are to changes in different factors. For example, storage costs may be easier to predict than IO costs. This information will help you to assess risk, as described in the next Chapter (Chapter 5).

### **Workload Factors**

When modeling workload, you should consider variability and predictability, as well as the average or “steady state” value.

Average levels might for example be 20 transactions per second and 500,000 transactions per day, with variability between zero and 20 transactions per second, depending on the time of day.

Predictability can be:

- Planned and scheduled – for example, financial management mini-peaks at end of quarter and major peak at end of fiscal year;
- Planned and unscheduled – for example, pharmaceutical certification on average five times a year, Florida hurricane response; or
- Unplanned and unscheduled – for example, air traffic control response to ash cloud.

Where predictability is planned and scheduled, you should note its periodicity. This is the cycle of requirements, including average, peak, and off-hours. Cycles may be on a calendar basis (including hourly, daily, weekly, monthly, quarterly, and annually), or on an event basis. Examples are: monthly periodicity, with peaks at the same time each month, and annual periodicity, with seasonal peaks.

You should also consider events with predictable capacity, which don't occur at the same time each month or year. Examples are weather patterns such as blizzards, unpredictable fads or events, occasions such as weddings, and disasters such as fires.

Think about the growth or shrinkage of your typical steady state usage, including such things as:

- Increase or decrease in capacity at the beginning of each steady state contract year;
- One-time changes; and
- Acquisitions, divestitures, regulatory changes, and new markets.

Consider the average and peak number of users by type. Types of user include business users, “power” users, and process administrators.

### **Workload Allocations**

## Cloud Computing for Business

Be aware that there are different types of workload that scale in different ways. This must be taken into account when modeling your workload.

Applications with a large amount of shared memory, many interdependent threads, and tightly-coupled interconnections cannot easily be partitioned between processing units, and only scale vertically. Applications with independent threads that do not share memory and have loosely-coupled interconnections can be partitioned, and scale horizontally. Other applications scale to some extent in both directions.

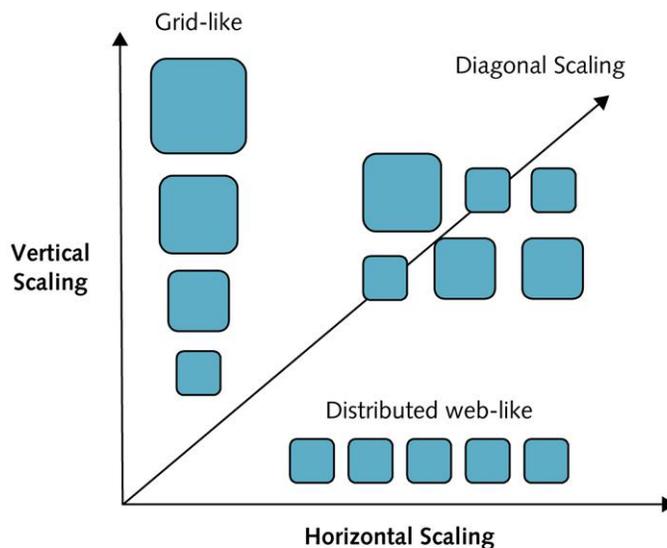


Figure 4.2 Types of workload allocation

The figure below shows some examples of the different types.

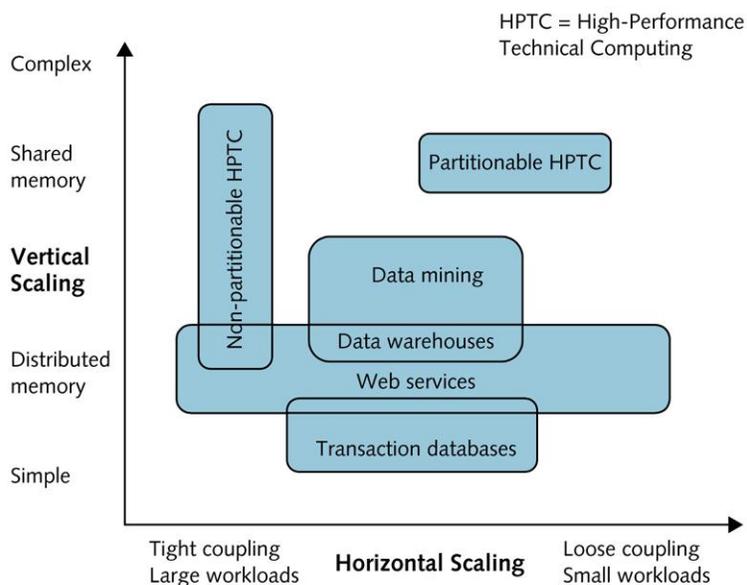


Figure 4.3 Examples of workload allocation types

## **Cloud Computing for Business**

Principles to follow in allocating workload include:

- Centralize key services for maximum pooling of total capacity available;
- Deploy solutions based on standard commoditized modular building blocks for maximum flexibility and re-use potential;
- Automate provisioning of standard images for maximum agility;
- Abstract the workload, where appropriate, from infrastructure using virtualization; and
- Identify workloads that can benefit from moving short-term burst loads to a cloud service.

### **Pay-As-You-Go Versus Ownership**

Cloud costs are typically lower with short term burst use but, for a steady or moderately-variable load, buying an on-premise system will probably be cheaper than using cloud services in the long term.

There are pricing options that can make cloud attractive in some long-term scenarios. Reserve cloud instances – cloud resources that are contracted for on a longer-term basis and typically for a lower price than those provided on-demand – lower the cost of long-term use. Burst services enable peak operational demand to be diverted to Cloud capacity.

Companies should assess their demand usage models in terms of actual workload profiles versus capacity usage. Many operational workloads are not 24x7x52 and can be accommodated by a combination of reserve instances and spot instances or “cloudburst” facilities. The figures below illustrate this. (Note that 30% OPEX is assumed for hybrid cloud and 25% for on-premise. In a pure on-premise solution OPEX is likely to account for 25% of the annual cost with 75% being interest and depreciation – the CAPEX element. The hybrid cloud solution has a lower proportion of CAPEX and a higher proportion of OPEX because of its public cloud element. These amounts are illustrative of what might be encountered.)

## Cloud Computing for Business

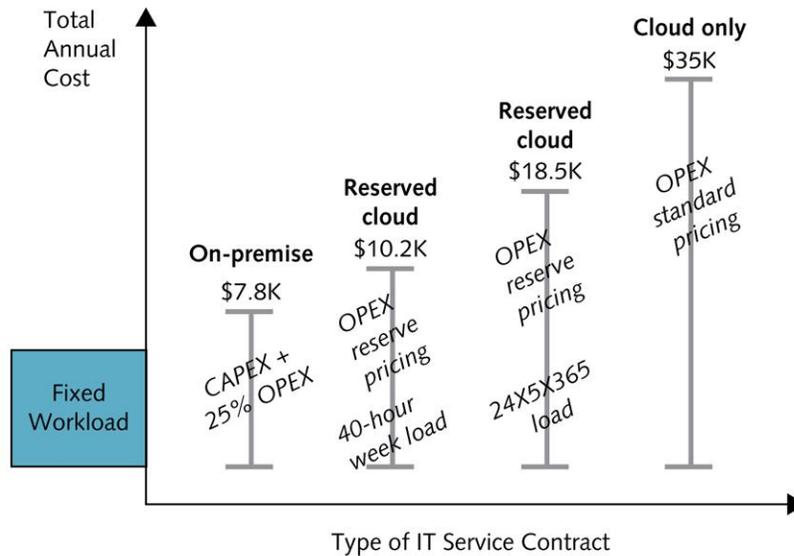


Figure 4.4 Example fixed workload costs

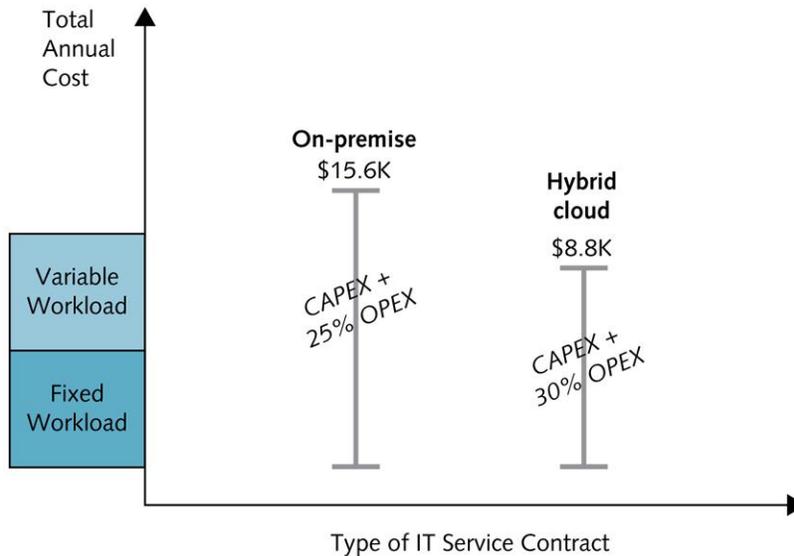


Figure 4.5 Example variable workload costs

With IaaS and PaaS, you can have a hybrid system, with some processing performed on-premise and some on the Cloud. (This is not such a practical possibility with SaaS.) The challenge is to define variable workloads and to be able to turn capacity on and off and switch between off- and on-premise services. There are costs associated with this that need to be assessed to “plug into the Cloud”.

### Modeling Resources and Costs

If you are using SaaS, and the charges relate to business units (such as number of users) rather than computing resources (such as Megabytes of memory), there is no need to build the resources into your model. You build a cost model for each supplier by mapping your business units to the supplier’s charging units. This model enables you to determine your probable costs for a given load.

## **Cloud Computing for Business**

For PaaS and IaaS, or for SaaS if the supplier charges by resource rather than by business unit, you should model use of processing power, memory, storage, and IO. These are what most suppliers charge for.

Consider the:

- Peak resource level – the capacity you require the supplier to be able to provide;
- Usage per time period – which is what you will be billed for; and
- Rate of resource change – which determines your requirement for speed of provisioning and de-provisioning.

Try to express these in standard terms that you can map to the units used by the suppliers you are considering. There may be benchmark figures available that can help you. This is a difficult area, which requires specialist technical expertise. If you do not have this in your IT team, consider employing an external consultant.

Mapping your units to the units used by the suppliers will give you the cost models that you need.

It is sensible to validate the models before relying on them for big decisions. You might develop your own benchmark program and run it with test data using each of the services that you are considering. One of the advantages of cloud computing is that on-demand self service makes it easy to do this.

### **Example Workload and Cost Models – Konsort-Prinz**

Konsort-Prinz develops a simple workload model. The load is measured by the number of products, the number of customers, and the number of orders. The processing power, memory, storage and IO required are calculated from these factors by simple algebraic formulas.

The company has estimates, based on experience, for typical, maximum, and minimum loads for each month of the year. Feeding these into the model gives estimates for the computing resources required, including the maximum at any time and the expected total for a year.

There are large variations in load, but the levels can be predicted with reasonable accuracy, at least a day in advance. The architects plan to provision capacity twice per day, and set a requirement to be able to provision a new resource within one hour. They are confident that this will meet the needs, and do not see a need to model rate of change.

The modelers express processing resource in “computer resource” units based on the existing in-house system. They develop a simple benchmark program based on their model, and calibrate it on the in-house system. They then run it on the IaaS services that they are considering, to validate the model and determine how the services’ units relate to theirs. This gives them a validated cost model for each IaaS supplier.

**Cloud Computing for Business**

The end result is that they can calculate the expected annual resource cost for each supplier, giving a vital part of the information the company needs to make the supplier choice.

**Example Workload and Cost Models – ViWi**

ViWi’s founder draws up a different model. He measures the load in terms of the number of virtual widgets in existence and the proportion of them that are active. Again, the processing power, memory, storage and input/output (IO) required are calculated simply from these factors. Each widget is estimated to require 10 Megabytes of disc storage all the time, and to require 0.1 processing units, 100 Megabytes of memory, and an average of 25 2-kilo-byte data transfers per second when it is active.

He has no real idea of how many virtual widgets the company will sell, or what proportion of the time they will be active. For supplier selection purposes, he decides to assume 100,000 widgets active 5% of the time on average and 20% at peak, giving a maximum of 2,000 and average of 500 processing units, with corresponding figures for memory, storage and IO.

The predictability is planned and unscheduled. He anticipates that the workload will change dynamically during the day, with sudden, unexpected peaks. The system will provision resources automatically in response to actual and anticipated demand. He sets a requirement to be able to provision or de-provision a new compute unit plus memory within 2 minutes, and the system will provision processors and memory at 10% above required capacity to allow for unexpected surges.

He captures the model in a simple spreadsheet.

	A	B	C
1			
2	Storage Megabytes per Widget		10
3	Processing units per active Widget		0.1
4	Memory Megabytes per active Widget		100
5	IO Transfers per second per active Widget		25
6	Average IO transfer size (Kilobytes)		2
7			
8	Average proportion of active Widgets		5%
9	Peak proportion of active Widgets		20%
10	Spare widget provisioning		10%
11			
12	Number of Widgets		100,000
13			
14	Peak processing power	$B12 * B9 * B3$	
15	Peak memory usage (Gigabytes)	$B12 * B9 * B4 / 1000$	
16	Peak Storage usage (Gigabytes)	$B12 * B9 * B2 / 1000$	
17	Peak IO transfers per second	$B12 * B9 * B5$	
18			
19	Average processing power	$B12 * B8 * B3 * (1 + B10)$	
20	Average memory usage (Gigabytes)	$B12 * B8 * B4 * (1 + B10) / 1000$	
21	Average storage usage (Gigabytes)	$B12 * B2 / 1000$	
22	Average IO transfers per month (Gigabytes)	$B12 * B8 * B5 * B6 * 3600 * 24 * 30 / 10^6$	
23			
24			

Figure 4.6 ViWi workload model

## ***Cloud Computing for Business***

The model expresses processing resource in the units used by one particular cloud provider. ViWi's founder (and, at that point, the sole developer) deploys an initial proof-of-concept version of the software on that provider's resources to validate the model, and on other providers' resources for comparison.

The measure derived from the model that he will use to compare suppliers is cost per widget per month.

### **Using the Models**

Having created a cost model for each service under consideration, you determine the expected cost of using it in the solution, and feed this cost into the overall financial model. You should also use the risk model to assess the risk profile of the solution, as described under Understanding Cloud Risk (Chapter 6).

These financial and risk assessments may result in you rejecting some cloud services. Indeed, you may reject all the candidate services, and conclude that cloud computing is not the answer after all.

But if you have made a thorough assessment of cloud suitability it is likely that you will have some candidate services at this stage. You have determined fit, and proceed to establish your requirements.