

IT Operational Considerations for Cloud Computing

Donna Scott

Cloud computing market offerings increase the options available to source IT services. Critical services must have defined service-level agreements (SLAs), while less-critical services should set operational expectations on quality and support.

Key Findings

- Cloud computing offerings are increasing in number, bringing more choices for IT services.
- Most early cloud offerings will be for non-mission-critical IT services.
- Over time, more cloud offerings will become available for mission-critical IT services.
- If available, then penalties for not performing to SLAs will generally be weak or insufficient compared with the potential business loss (that is, penalties may offer a free day, week or month of use, but will not compensate for lost business or other costs).

Recommendations

- IT operations should assess not just the ability for the cloud service provider to meet business requirements, but also for operational management requirements, risks and process/tools integration points.
- Although most early cloud services will not be used for mission-critical work, expectations on quality and support should be defined in terms of SLAs.
- Cloud computing contracts should be no longer than annually to ensure that service levels and additional expectations can be built into the contract or re-evaluated as the value of the services change for customers.
- IT should be aware of the business use of cloud computing services, even for non-critical services. This will ensure the IT organization is meeting its customer requirements and not missing business requirements that result in a higher amount of "shadow IT" and increased overall business cost and risk.

ANALYSIS

The advent of cloud computing (see Note 1) may get enterprise IT excited that there are other or better ways to deliver IT services to customers. The good news is that there are an increasing number of options available — a service continuum between externally sourced and internally hosted. The bad news is that more analysis needs to be done to assess the myriad options out there, including cloud and software-as-a-service (SaaS) offerings, and the risks posed by new options that are just getting off the ground and have no business history or any ongoing commitments (for example, many Web hosting providers often gave 30 or fewer days notice when they were canceling a customer's contract due to bankruptcy).

You may find that your business customers are buying and using these services unbeknownst to the IT organization (for example, Facebook being used by the HR department) and expect the IT organization to provide service when incidents and outages occur (that is, there may be user assumptions that the internal service desk will handle support calls associated with these cloud services). As a result, now is the time to get ahead of cloud computing and set expectations with the customer, provider and IT organization, as well as manage any business risks that may be posed by it.

Although early cloud offerings will be offered for less-critical services, including consumer (vs. enterprise) services that have low expectations on service quality, if these early offerings are successful, then it will only be a matter of time before enterprises consider them for more-critical services as well. Moreover, you may find that cloud services are used for specific purposes in a portion of your user base (for example, to backup executive laptops).

From an IT service delivery perspective, for mission-critical services (or those that may become mission-critical in the next 12 months), it is vital that evaluations of cloud offerings take into consideration any SLAs, and if you can't get an offering at a level required for the business, then it should be eliminated from consideration. Even less-critical services should have defined expectations (or known lack thereof) of service quality and support. In addition, mission-critical service evaluations should include on-site reviews of the service provider's facilities and the IT service architecture (are they secure and protected against single points of failure?) and operational processes (such as change management and incident management) to instill confidence that SLAs are likely to be met. Customer references should be checked. The following should also be considered:

- **Planned downtime** — When will the service come down for maintenance? Does the customer have input into when to take planned downtime (for example, to avoid peak business times when downtime would be most inconvenient)?
- **Availability of end-to-end IT services** — The IT organization is responsible for the end-to-end experience received by its customers (for example, the end-to-end availability of an externally hosted application that may be integrated with an internally hosted business application — the combined grouping of which may represent the end-to-end IT service for the customer). Gartner defines best-in-class availability as 99.95% availability of the IT service or just five hours down per year. SLAs of the end-to-end IT service can only be as good as its piece parts. If the provider does not offer an SLA that meets your customer needs, then don't buy its services. For example, Facebook may have no cost for use, but it also has no SLAs. For SaaS vendors, the availability levels vary from no guarantees (Facebook), to negotiated guarantees (salesforce.com), to 99.5% availability guarantees during a quarter or month (for example, Amazon S3, Oracle/Siebel on Demand, RightNow Technologies and NetSuite; see "Look Beyond the Fees per User per Month When Comparing SaaS CRM vendors").

- **Capacity/demand** — If demand is greater than expected, then what are the SLAs to increase capacity (and associated cost of that increased capacity)? What expectations are there on timing for capacity increases? Although some services may dynamically expand (and charge more) for increasing business activity, others may be architected to ask for approval before adding additional capacity. From a longer-term perspective, what processes are put in place to assess demand to ensure the right level of capacity to meet business needs? There is a growing trend to over-purchase SaaS user fees just as IT organizations have historically over-provisioned infrastructure.
- **Service/support and mean time to repair** — Support expectations and SLAs should be established between the service provider and the customer. For example, in the event of an IT service outage, how does the customer/user contact the supplier's support organization? What are the proactive mechanisms that the supplier uses to communicate status to the customer/user? What are the expectations on support, including response time, and mean and maximum times for service repair? In addition, what are the expectations of the maximum number of service outages in a defined period (that is, monthly, quarterly and annually)?
- **Performance of the end-to-end IT services** — Access to the services over the Internet may provide variable performance (due to network latency, packet loss and Internet routing abnormalities) depending on the location of users to the hosting center and the architecture of the service. If you can't get a response time SLA that is acceptable to your business users, then don't buy the services.
- **Data recovery** — When data becomes corrupt or lost, what is the service level associated with getting the last-known good copy of data made available? How old will that last-known good copy be? In our analysis, we have yet to see recovery components in any cloud or SaaS SLA offerings. There are cases of backup being part of an SLA, availability being part of an SLA, and redundancy and geographical separation being part of an SLA, but we have seen few offerings (Oracle is one) with an SLA that guarantees recoverability or recovery-time objective (RTO) and recovery-point objective (RPO) metrics, and this is only for disasters, not for data corruption.
- **RTO and RPO** — In a defined disaster scenario (normally defined as multiple points of failure, natural disaster, fire, flood and so on) or logical/user/operator error recovery situation, what are the service levels associated with the recovery of data (RPO) and the time associated with the recovery of the service (RTO)? How is disaster defined?
- **Penalties, terms and conditions** — What are the penalties associated with all the SLAs listed above? The issue is that nearly all service providers lack meaningful penalties for missed SLAs. Enterprises don't want the penalty — they want the SLAs. The trick is to make the penalties large enough to get the service provider to do the right thing and meet the SLAs. Because the largest penalty tends to be "one free month of service for every missed SLA," there is little recourse available to the buyer of the services when the SLAs are missed. One recourse is to allow breaking the contract for missed SLAs; however, this means evaluating internal and external delivery options. Other terms and conditions should be considered related, for example, to the ownership of the data, accessibility and migration of the data, and service.

Where a combination of internal and external IT processes are necessary to deliver the end-to-end service, typically there are requirements for process integration and tool integration (that is, for incident management, change management and service-level management). Getting these defined and documented upfront in a contract will enable the IT organization to meet its requirements for service delivery. Without them, they will have to operate "in the dark" from the

perspective of outsourced service delivery, resulting in process silos, which prevents them from optimizing and enhancing end-to-end service delivery.

Cost Analysis

Another major aspect of cloud offerings are the costs: Know what you are buying, what service-level expectations are, and what the penalties and total costs are of buying and using the service over varying time periods. Ideally, when evaluating cloud services, you should benchmark multiple cloud providers, as well as the costs of delivering the service internally. This will help you determine internal vs. external delivery competitiveness, and should be used in sourcing decision analysis.

Because the cost and value of services change, you must reassess and recontract for service levels on an annual basis (that is, don't write multiyear contracts). During re-evaluation, the costs should be evaluated and benchmarked to determine whether the right level of value and quality are gained for the money being spent. If not, then reassessment (including changing service providers and internal delivery) should be done keeping in mind that there is a cost for service migration.

What You Need to Know

Cloud computing is an alternate form of IT service delivery. Just like other forms of delivery before it (including Web hosting, application service providers, management service providers and SaaS), IT operations must assess not only the ability for the service provider to meet business requirements, but also that the operational requirement SLAs, pricing, and terms and conditions be assessed for requirements and risks. The most important thing is for expectations to be set and known upfront by all parties. Contracts should be no longer than annually to ensure that service levels and additional expectations can be built into the contract or re-evaluated as the costs and value of the services change for customers.

Cloud services should not be bought by business users unbeknownst to the IT organization if the services are mission-critical or have the potential to become mission-critical in a 12-month time frame, because a lack of oversight and considerations of IT operational requirements could pose unacceptable risks and costs to the business. The IT organization should also be aware of the business use of cloud computing services, even for noncritical services. This will ensure the IT organization is meeting its customer requirements and not missing business requirements that result in a higher amount of shadow IT and increased overall business cost and risk.

Note 1

What Is Cloud Computing?

Gartner defines cloud computing as a style of computing where massively scalable, IT-enabled capabilities are delivered "as a service" to external customers using Internet technologies.

This research is part of a set of related research pieces. See "Cloud Computing Confusion Leads to Opportunity" for an overview.

REGIONAL HEADQUARTERS

Corporate Headquarters

56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

European Headquarters

Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters

Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters

Gartner Japan Ltd.
Aobadai Hills, 6F
7-7, Aobadai, 4-chome
Meguro-ku, Tokyo 153-0042
JAPAN
+81 3 3481 3670

Latin America Headquarters

Gartner do Brazil
Av. das Nações Unidas, 12551
9º andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509