

# Introducing Data Management (a Cloud Service)

David Cohen, Architect and Senior Technologist  
Cloud Infrastructure Group  
EMC Corporation

# Data Storage as a Service (DaaS)

*(the delivery of virtualized storage on demand)*

Type		Presentation	Consumable
(1)	Object	REST (or SOAP) via HTTP	Blob
(2)	Distributed File System	CIFS, NFS, WebDAV, FTP, etc	File System (e.g. export, share, etc)
(3)	Block Device (used to provide local file system)	SCSI (or ATA) over IP	Volume suitable for use in Volume Manager
(4)	Tabular	REST via HTTP	Table
(5)	Relational	Product specific	SQL Database
(6)	Messaging	AMQP	Store-and-Forward Queue or Topic-based Publish/Subscriber

# Definitions

- 1. Resource** - Often referred to as a "server," this is a composition of I/O, RAM, and CPU.
- 2. Resource Pool** - A container of lots of servers.
- 3. Customer** – Procures resources for a specific use.
- 4. Service Provider** – The steward of the underlying resource pool; invoicing the customer on a subscription/usage basis.
- 5. Cloud Computing** – Resources are dynamically provisioned from a shared pool. This provisioning operation is initiated by the customer over a network via a self-service Portal.

# Service Models

## ***(1) Deployment Model***

(Mell et al, "The NIST Definition of Cloud Computing," 2009)

Type		Location	Relationship
(1)	Private	Customer Premise	Central IT
(2)	Public	Service Provider Premise	3 <sup>rd</sup> party
(3)	Community	Service Provider Premise	(n) customers/(1) 3 <sup>rd</sup> party; sharing
(4)	Hybrid	Both	Federation of Central IT and 3 <sup>rd</sup> Party

## ***(2) Platform Model***

(Mell et al, "The NIST Definition of Cloud Computing," 2009)

Model		Semantics	Community
(1)	Software as a Service (SaaS)	Defined by Application	Constrained to users of the application (end-user)
(2)	Platform as a Service (PaaS)	Defined by Programming model	Constrained to Application Developers (AppDev)
(3)	Infrastructure as a Service (IaaS)	Defined by "Resource" allocation/partitioning	Constrained to System Administrators (SA)

## Some Examples

Solution		Deployment Model	Service Model
(1)	Amazon Web Services (AWS)	Public and Hybrid	IaaS
(2)	EMC Cloud Infrastructure Services	Public	IaaS
(3)	Microsoft Azure	Public	PaaS
(4)	Google AppEngine	Public	PaaS
(5)	Facebook	Community	SaaS
(6)	Keas	Community	SaaS
(7)	SaleForce	Community	SaaS
(8)	Heroku	Private, Community	PaaS
(9)	3Tera	Private, Community	PaaS
(10)	Eucalyptus Systems	Private	IaaS

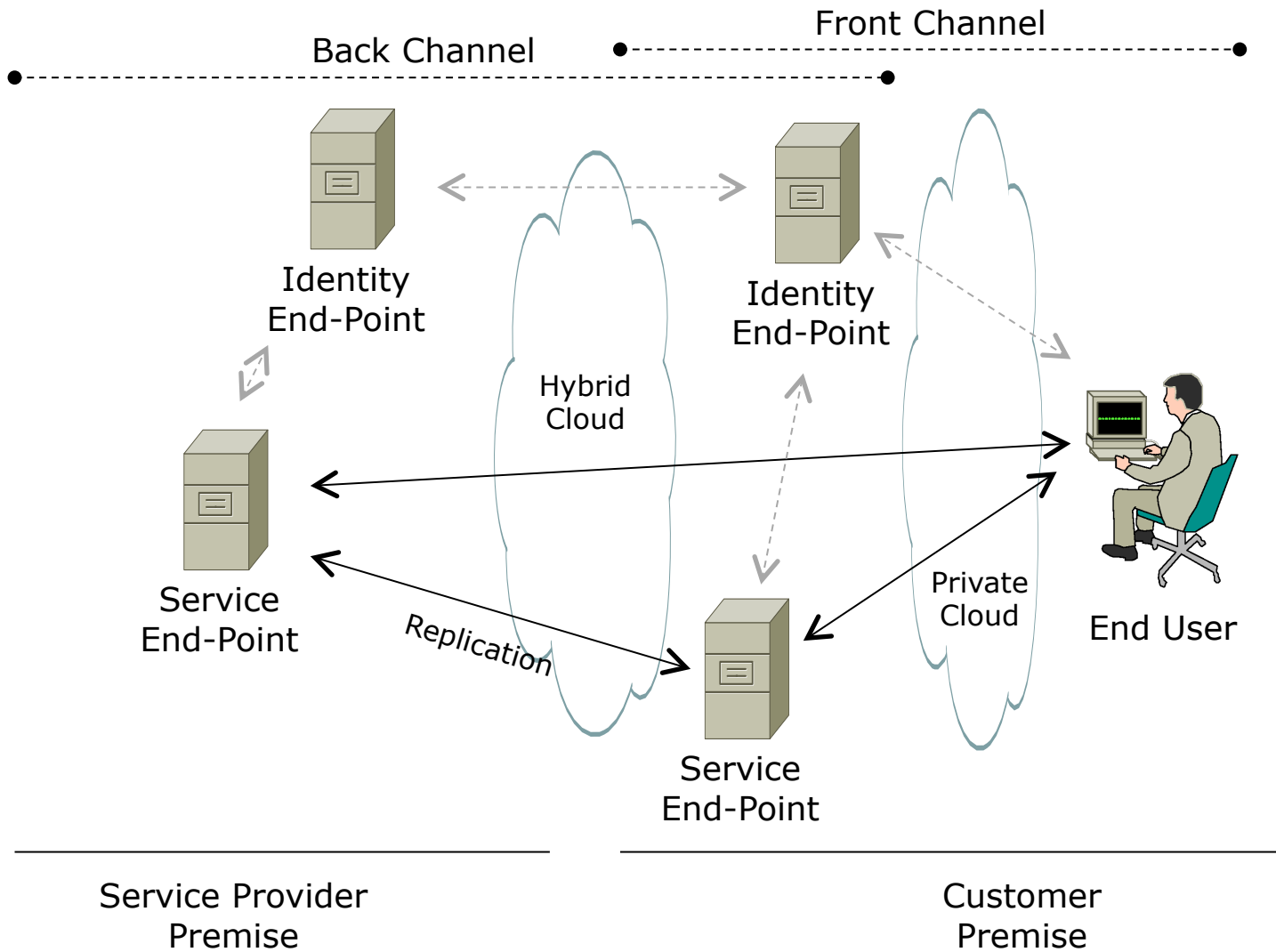
# Management Concerns/Considerations

Management Discipline		Description
(1)	<b>Configuration</b>	Build, Package, Deploy, Install, Configure, Verify
(2)	<b>Resource</b>	Logging/Monitoring, Event Capture, Propagation, Analysis/Reporting, Escalation
(3)	<b>Maintenance</b>	Scheduled downtime; Add new, remove/retire old
(4)	<b>Usage Reporting</b>	Management Dashboards, Billing/Chargebacks, long-term trending, etc
(5)	<b>Access</b>	Physical, KVM/Console, Crash Cart, etc
(6)	<b>Security</b>	Multi-Tenancy, Degrees of Isolation, Identity Management, Authentication/Authorization/Auditing
(7)	<b>Provisioning/ De-Provisioning</b>	Create new, flex/scale-out, flex/scale-in, locality-awareness of data
(8)	<b>Capacity</b>	Monitor disk space utilization, remove/replace breakage, forecast upcoming requirements
(9)	<b>Performance</b>	Throughput and Response-time
(10)	<b>Online Retention</b>	Data Migration, Tiering/ILM, CAR (Compliance/Archival/Retrieval)
(11)	<b>Backup/Recovery</b>	Replication, Snapshot, WAN Optimizations
(12)	<b>Integrity</b>	Tampering, Bit-Rot, etc

# Dive Deeper into Security Concerns for Cloud-based Data Management

Concern		Illustrative Example(s)	Method
(1)	Physical Access	Facilities, Cage, etc	Secured Access
(2)	Access via Network	KVM/Console, Service (IP/Port), etc.	VPN, PKI, SSH Keys,
(3)	Identity	End-User, Privileged Service Accounts	Central Management via LDAP/Kerberos, Other (e.g. OpenId)
(4)	Authorization	Privileged access to service capabilities	Group and/or Role based, OAuth, Friend-of-a-Friend (FoaF)
(5)	Logging	Capture Events	Distributed Data Capture, Central Aggregation, Analytics
(6)	Data Loss Prevention (DLP)	Identify sensitive data and control operations applied against it	Data Classification and Policy-based Enforcement

# Private and Hybrid Clouds





# Closing Observations

The undercurrent across all of these is "Internet Scale." This scale impacts both the platform and the platform's deployment model. As you can imagine, while we have some degree of consensus on deployment model (and topology) the platform has forked into tactical and strategic. The following six items are intended to highlight these dynamics:

Concern		Discussion
(1)	Capture financial benefit of OpEx	In the near term (6 - 36 mos) this is what is motivating customer demand for Cloud.
(2)	Private and Hybrid Cloud Deployment Models	Based on #1 (above) and for the foreseeable future (i.e. 6 to 36 mos), it is likely that the private and hybrid cloud models (as opposed to a "public" one) will be the dominate choices for small, medium, and large enterprise adopters.
(3)	Hosting Legacy Applications	Cloud Management platforms such as Amazon Web Services (AWS), Eucalyptus, OpenNebula, Vmware's vCloud, etc focus on enabling an organization to transform their IT operations into Infrastructure-as-a-Service (IaaS).
(4)	The economics of scale	Looking forward, the "scale" we are talking about is "multiple *5s* of data centers," "multiple *10s of thousands* of hosts," "a couple *100 thousand guests*," and "multiple *10s of petabytes* of storage."
(5)	Change represents opportunity	Conventional software design does not accommodate additive/scale-out resource management semantic (implied by #4 above) very well and this has given rise to application development paradigms, each replaced with something new/better at an increasingly rapid rate.
(6)	New Application Development Platform	This is the Platform-as-a-Service (PaaS) offering illustrated by Google's AppEngine and Microsoft's Azure.

# Questions

# References

1. Morton et al, "The Economics of Cloud Computing Analyzed," 2009  
<http://govcloud.ulitzer.com/node/1147473>
2. Golden, "Capex vs. Opex: Most People Miss the Point About Cloud Economics," 2009  
[http://www.cio.com/article/484429/  
Capex\\_vs.\\_Opex\\_Most\\_People\\_Miss\\_the\\_Point\\_About\\_Cloud\\_Economics](http://www.cio.com/article/484429/Capex_vs._Opex_Most_People_Miss_the_Point_About_Cloud_Economics)
3. Vaquero et al, "A Break in the Clouds: Towards a Cloud Definition," 2009  
<http://ccr.sigcomm.org/online/files/p50-v39n1l-vaqueroA.pdf>
4. Mell et al, "The NIST Definition of Cloud Computing," 2009  
<http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>
5. Pallis et al, "Cloud Computing," 2009  
<http://www.cs.ucy.ac.cy/~gpallis/publications/journals/editorial.pdf>
6. Gray, "Distributed Computing Economics," 2008  
<http://queue.acm.org/detail.cfm?id=1394131>
7. Greenberg et al, "The Cost of a Cloud: Research Problems in Data Center Networks," 2009  
<http://research.microsoft.com/pubs/79893/p68-v39n1o-greenberg.pdf>
8. Chappell, "A Short Introduction to Cloud Platforms," 2008  
<http://www.davidchappell.com/CloudPlatforms--Chappell.pdf>