

The Evolution of Cloud Storage - From "Disk Drive in the Sky" to "Storage Array in the Sky"

Allen Samuels

Co-founder & Chief Architect



Cloud Storage Definition

- ❑ Not clustered NAS rebadged
- ❑ Not private cloud
- ❑ Utility public service – pay as you go
 - ❑ Amazon S3 is the prototypical example

- ❑ For our discussion:
 - ❑ Cloud storage is geographically remote from compute

Cloud Storage is Great!

- ❑ **Top names:** Amazon, AT&T, Iron Mountain, Google, Microsoft, Rackspace & more
- ❑ **Cheap:** ~\$0.30 / GB / month and declining
(In-house storage > \$1 / GB / month, all-in)
- ❑ **Off-site:** Guaranteed business continuity
- ❑ **Never lost:** Up to 11-nines durability guarantee
- ❑ **No CapEx:** Pay only for what you need



But Cloud Storage Is Also...

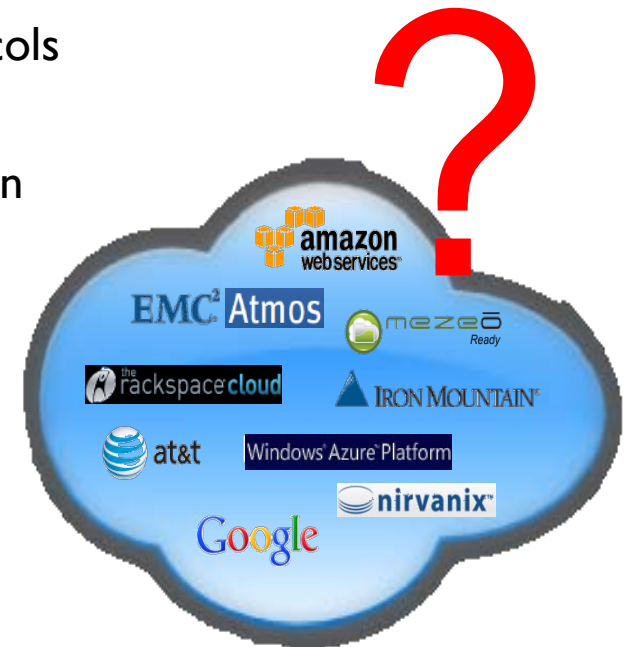
 **Incompatible:** Doesn't use enterprise storage protocols

 **Insecure:** Shared infrastructure, #1 enterprise concern

 **Slow:** Internet vs. high speed SAN

 **Incapable:** no thin provisioning, snapshots, etc.

 **Costly:** Interactive workloads drive transfer costs up



Result: Enterprises can't use cloud storage "as is"!

What does this mean?

- ❑ Cloud storage is a great “bit bucket” if:
 - ❑ You can custom write your app to it
 - ❑ You have low expectations
- ❑ Cloud storage is analogous to a disk drive in the sky
- ❑ Why does the enterprise buy storage arrays when they can buy cheap disk and RAID cards at Fry's?
- ❑ That same “value add” is what's missing from the cloud

Fearless Prediction

- ❑ The enterprise cloud storage market *will not expand rapidly* without catalyst technology.
- ❑ Must preserve the benefits of cloud storage
 - ❑ Cheap
 - ❑ Elastic
- ❑ Must plug the holes
 - ❑ Secure
 - ❑ High Performance
 - ❑ Fully compatible with existing apps
- ❑ Must appear transparent to end-users – is it cloud or is it local?

Getting from here to there

- ❑ Catalyst technology must do several things
 - ❑ Interact with applications using standard storage protocols
 - ❑ Must be multi-cloud lingual – preserve choice of provider
 - ❑ Deliver high performance
 - ❑ Not require anything to change at the cloud provider
 - ❑ Minimize cloud storage service fees
 - ❑ Overcome security concerns
 - ❑ Provide on-site array features
 - ❑ Provide familiar management & administration

Cloud Storage Tower of Babel

- ❑ Every provider speaks a different language
- ❑ Every provider has different rules
- ❑ Not all providers are created equal – a partial sampling
- ❑ What's an app developer to do?



	Eventual Consistency	Object replace operations	Object naming	Directory Levels Supported	API
Provider A	Immediate for objects and directories	Permitted	Permitted	3	Proprietary
Provider B	Immediate for objects; up to 30 minutes for directories	Not Permitted	Only in object metadata	0	Proprietary
Provider C	Immediate for objects and directories	Not Permitted	Not Permitted	Unlimited	Proprietary

Would data storage be a \$20B market?

- If you had to custom write each app to a specific array vendor?
- If you couldn't move data from one array to another?
- If there were no standards like SCSI and POSIX?
- If you had to intimately understand the nuances and characteristics of your storage system before you could do anything?

Catalyst technology Requirement #1

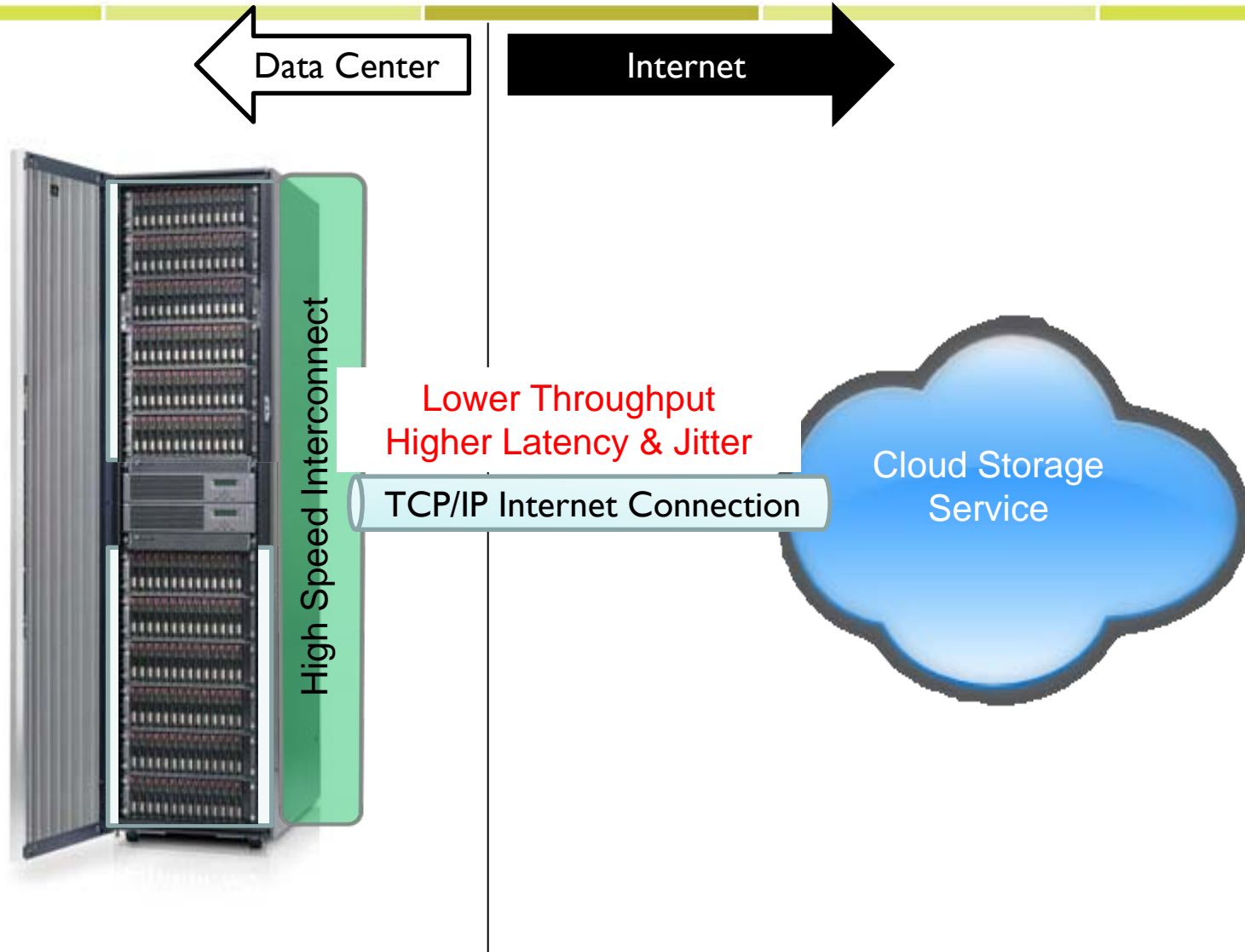
- ❑ Cloud API abstraction layer
- ❑ No knowledge of provider semantics required
- ❑ Maintain choice
- ❑ Mask complexity and limitations

- ❑ What about CDMI?
 - ❑ CDMI defines common API sets
 - ❑ Cloud object stores are different under the covers
 - ❑ Unique behavior manifests even with a common API

Getting from here to there

- ❑ Catalyst technology must do several things
 - ❑ Interact with applications using standard storage protocols
 - ❑ Must be multi-cloud lingual – preserve choice of provider
 - ❑ Deliver high performance
 - ❑ Not require anything to change at the cloud provider
 - ❑ Minimize cloud storage service fees
 - ❑ Overcome security concerns
 - ❑ Provide on-site array features
 - ❑ Provide familiar management & administration

The WAN is not a SAN!

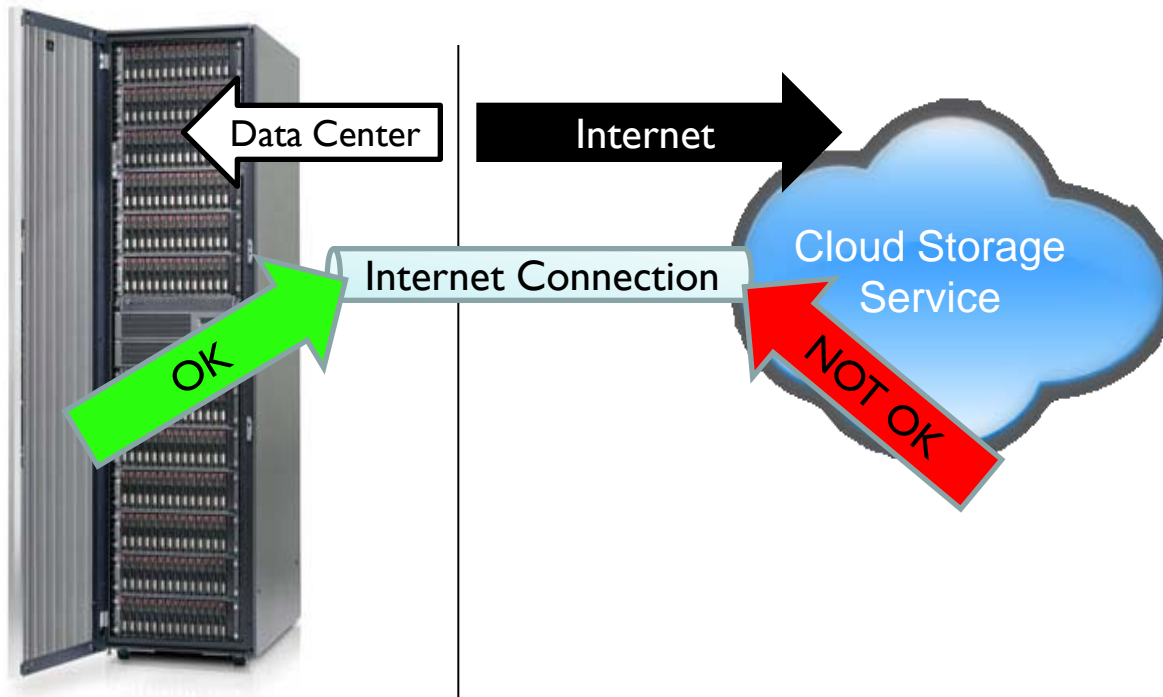


How do you overcome bandwidth bottlenecks?

- ❑ Assume you can't make the pipe bigger
- ❑ Assume you can't change the speed of light
- ❑ Many time-tested, proven techniques
 - ❑ Avoid transfers - Caching (e.g. L1/L2/L3 CPU caches)
 - ❑ Transfer less - Compression (e.g. MPEG4)
 - ❑ Transfer more efficiently - Protocol optimization (e.g. TCP acceleration)
- ❑ Companies have made a lot of money leveraging these technologies:
 - ❑ Riverbed....Cisco
- ❑ In short....WAN Optimization

But how does WANOp work?

- ❑ Puts a box on each end of the pipe



- ❑ What about a box in the cloud?
 - ❑ Not feasible most of the time
 - ❑ Leads to vendor lock-in.
 - ❑ Difficult when you don't control both ends
 - ❑ Expensive (even when you can do it)

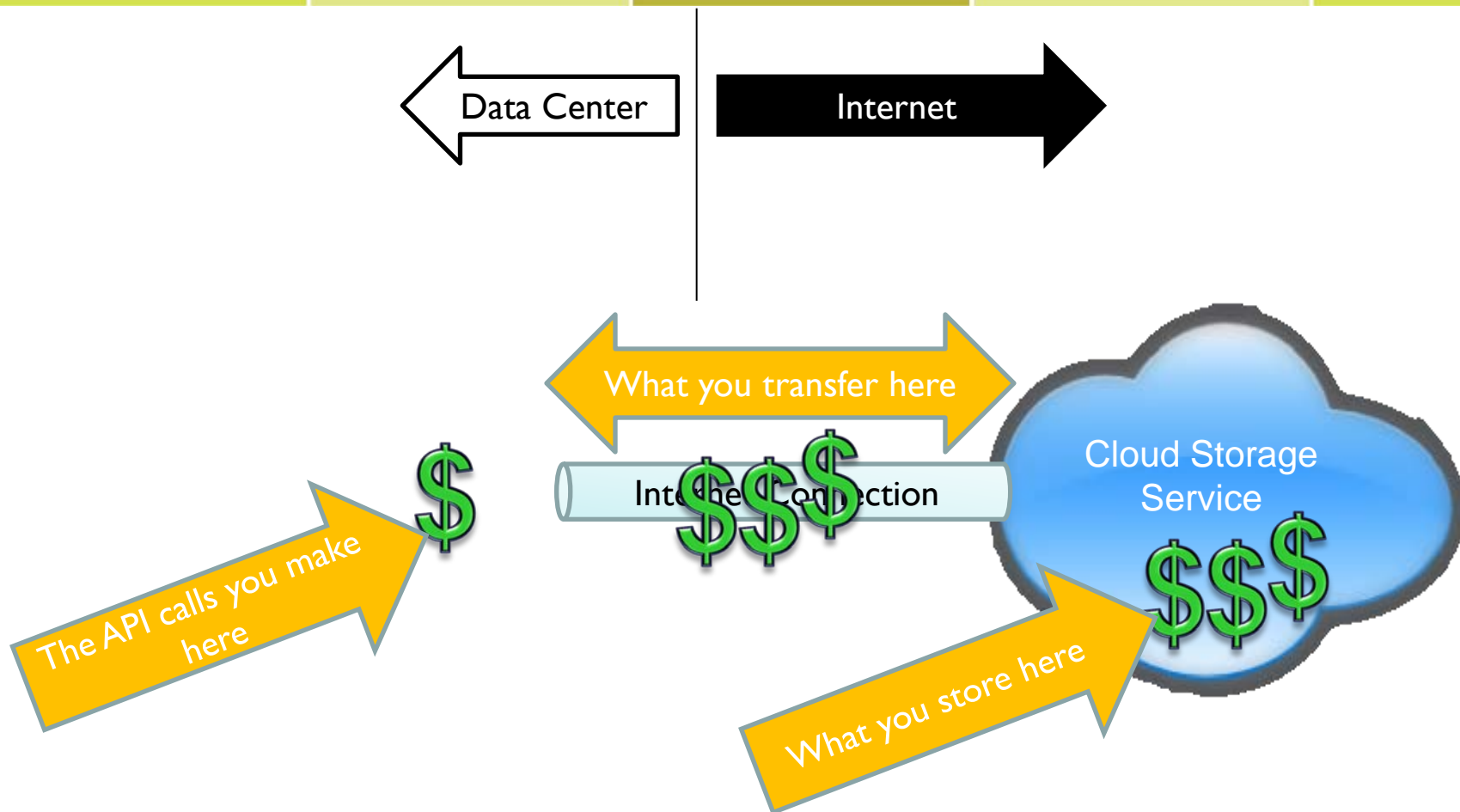
Catalyst technology Requirement #2

- ❑ Implement acceleration techniques that:
 - ❑ Improve WAN transfer rates
 - ❑ Mask/avoid WAN transfers
 - ❑ Optimize WAN transfers that must be made
- ❑ Require nothing in the cloud – physical or virtual
- ❑ New, asymmetric techniques must be developed

Getting from here to there

- ❑ Catalyst technology must do several things
 - ❑ Interact with applications using standard storage protocols
 - ❑ Must be multi-cloud lingual – preserve choice of provider
 - ❑ Deliver high performance
 - ❑ Not require anything to change at the cloud provider
 - ❑ Minimize cloud storage service fees
 - ❑ Overcome security concerns
 - ❑ Provide on-site array features
 - ❑ Provide familiar management & administration

What do cloud providers charge for?



Catalyst technology Requirement #3

- ❑ Store less data in the cloud
- ❑ Move less data into the cloud
- ❑ Move less data out of the cloud
- ❑ Minimize API calls

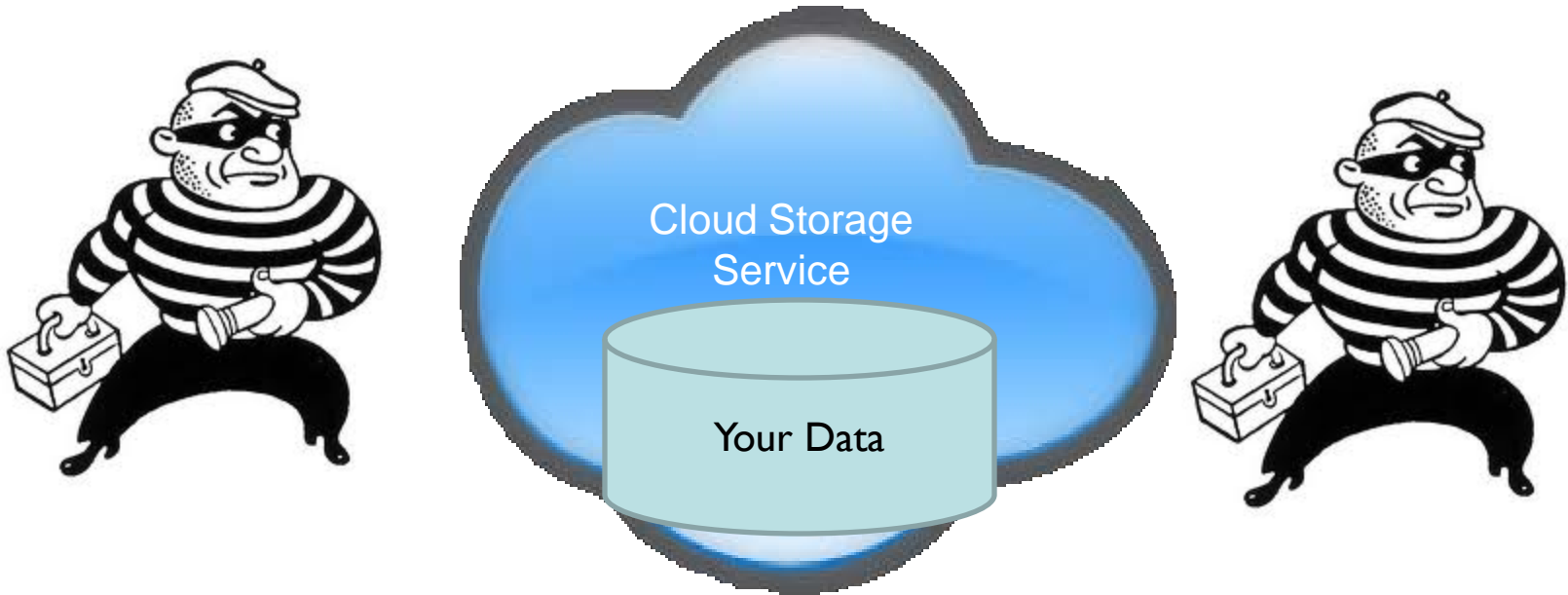
- ❑ How do you do this?
 - ❑ Luckily asymmetric WANOp techniques help here
 - ❑ Dedupe/compress data prior to transmission to the cloud
 - ❑ Leave data in deduped/compressed state in the cloud
 - ❑ Optimize transfer sizes to balance cost vs. access latency

Getting from here to there

- ❑ Catalyst technology must do several things
 - ❑ Interact with applications using standard storage protocols
 - ❑ Must be multi-cloud lingual – preserve choice of provider
 - ❑ Deliver high performance
 - ❑ Not require anything to change at the cloud provider
 - ❑ Minimize cloud storage service fees
 - ❑ Overcome security concerns
 - ❑ Provide on-site array features
 - ❑ Provide familiar management & administration

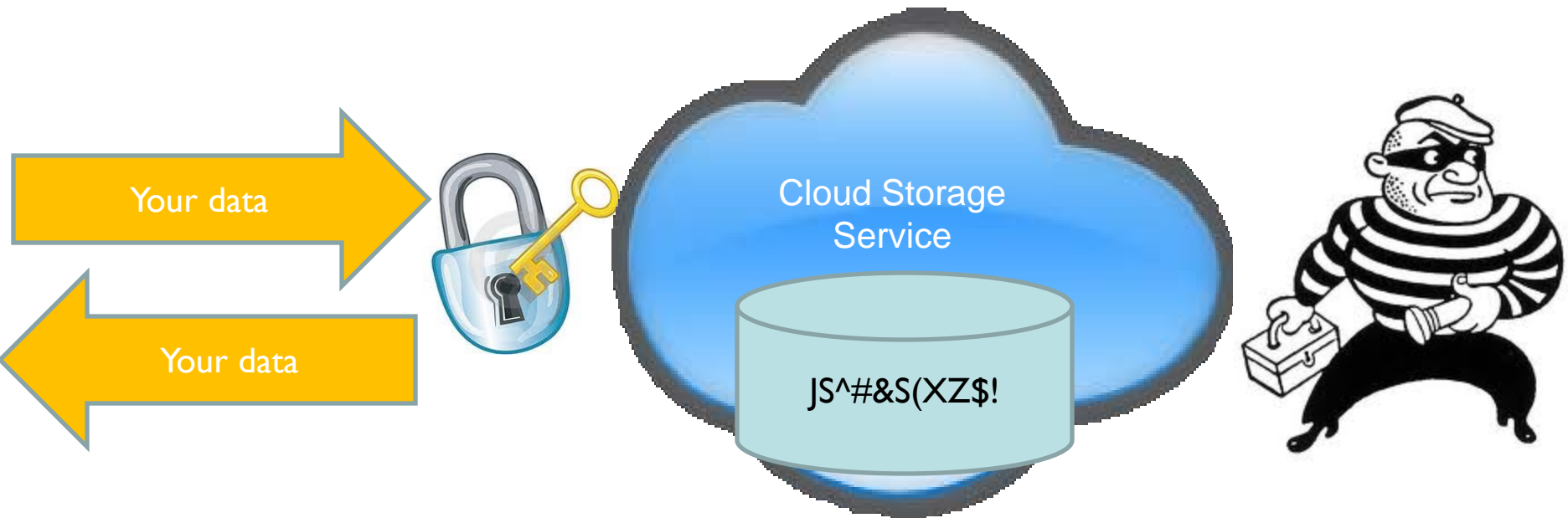
Is this secure?

- ❑ Clear text at the provider?



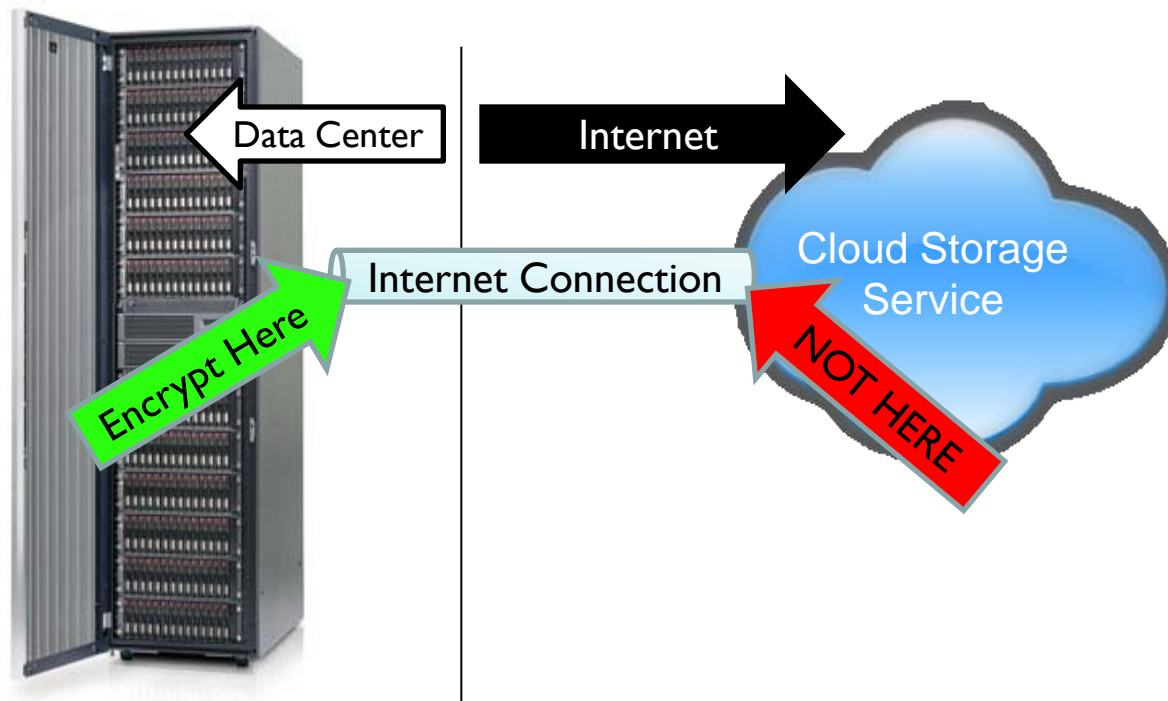
Is this secure?

- ❑ Provider encrypts your data?
- ❑ Provider has the keys!
- ❑ Only protects against one side of the vulnerability



Catalyst technology Requirement #4

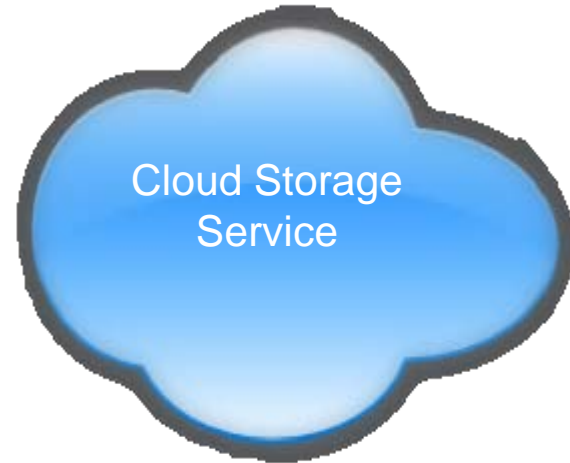
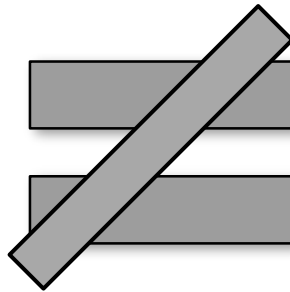
- ❑ Must encrypt data prior to transmission to the cloud
- ❑ Cloud provider must not have the keys
- ❑ Customer must control the keys



Getting from here to there

- ❑ Catalyst technology must do several things
 - ❑ Interact with applications using standard storage protocols
 - ❑ Must be multi-cloud lingual – preserve choice of provider
 - ❑ Deliver high performance
 - ❑ Not require anything to change at the cloud provider
 - ❑ Minimize cloud storage service fees
 - ❑ Overcome security concerns
 - ❑ Provide on-site array features
 - ❑ Provide familiar management & administration

The Cloud is not a Storage Array!



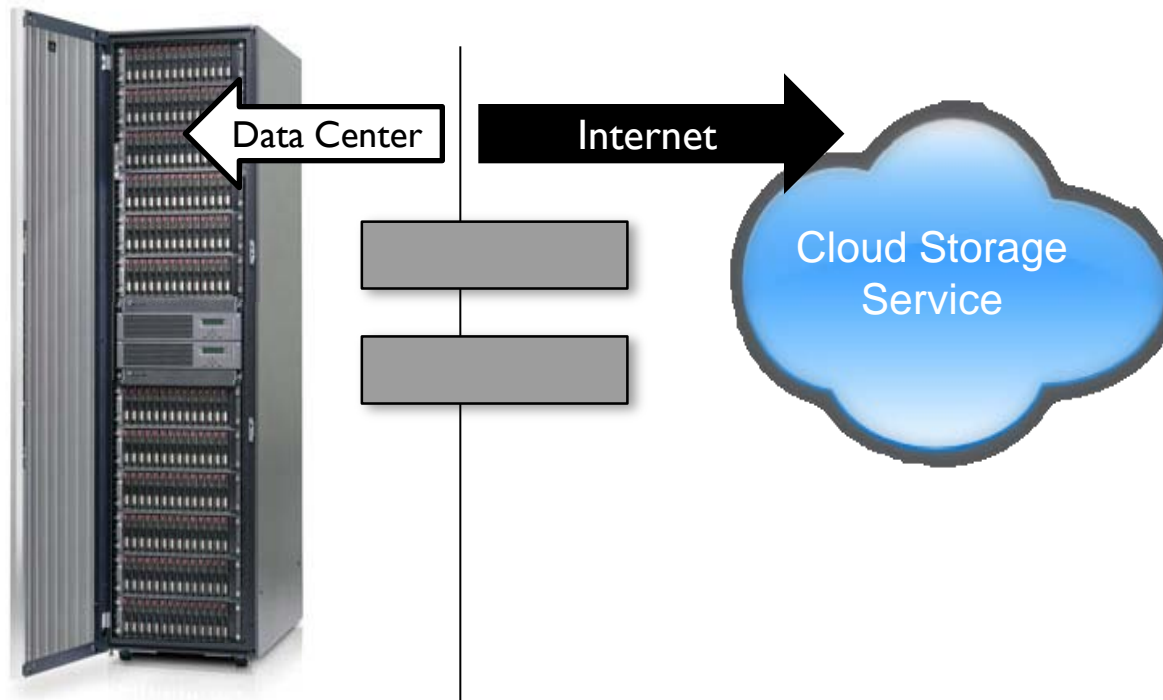
The Cloud is not a Storage Array!

Capability	Storage Array	Cloud
Physical Data Protection	Yes	Yes
iSCSI/FC/NFS/CIFS Access	Yes	No
Mountable Logical Volumes	Yes	No
Thin Provisioning	Yes	No
Snapshots	Yes	No
Automated Tiering	Yes	No
Remote Replication/Backup	Yes	No
Standard SCSI/POSIX semantics	Yes	No

- ❑ Without these capabilities, existing applications, processes, and management practices break down
- ❑ Relegates the cloud to custom application development, not storage array alternative

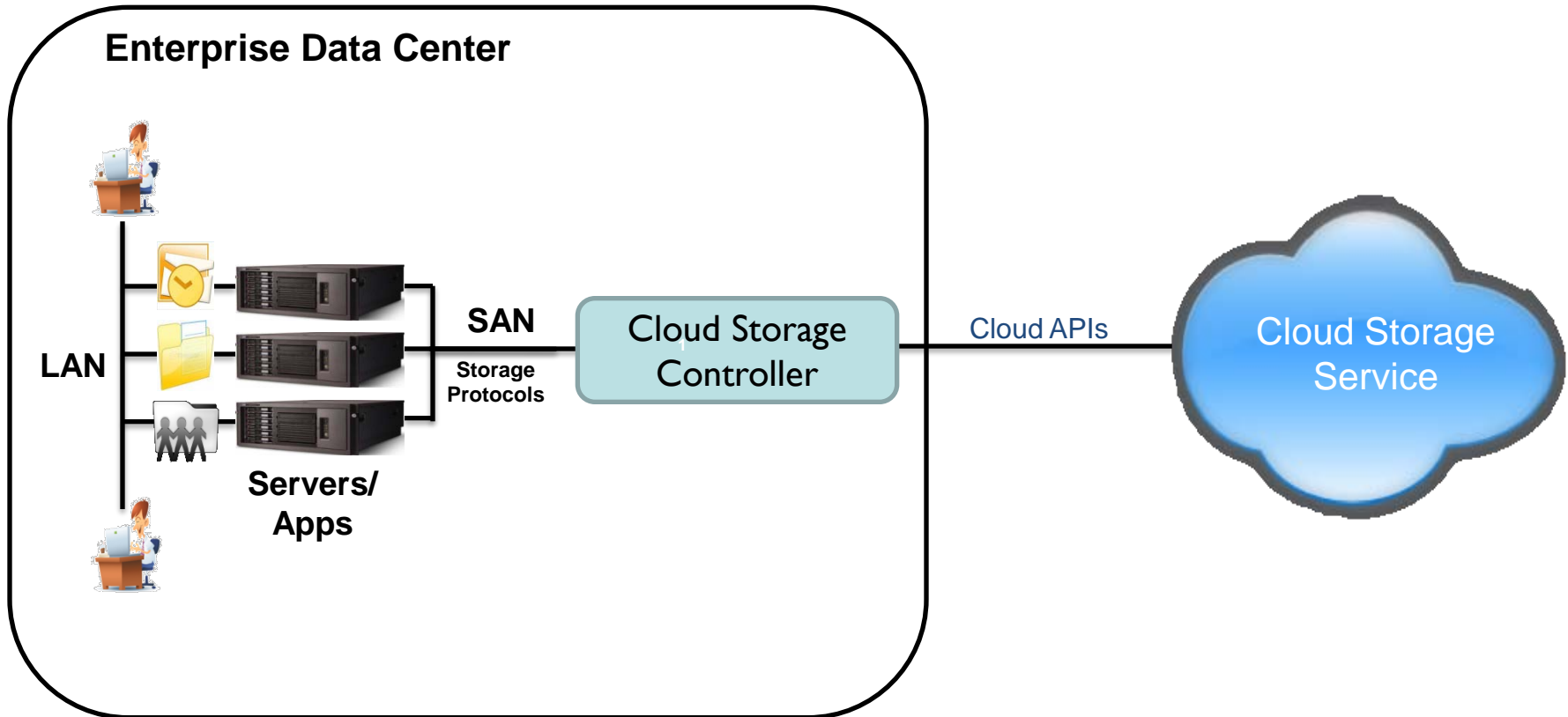
Catalyst technology Requirement #5




- ❑ Must overlay local storage features onto the cloud
- ❑ Transparency is key – the user shouldn't know...or care





- ❑ For cloud storage to evolve from “disk capacity” in the sky to “storage array” in the sky, we need:
 - ❑ Cloud provider independence
 - ❑ Standards-based access
 - ❑ Asymmetric WAN acceleration technology
 - ❑ Cloud provider cost optimization
 - ❑ Encryption prior to transmission to the cloud
 - ❑ Full compliment of storage array features

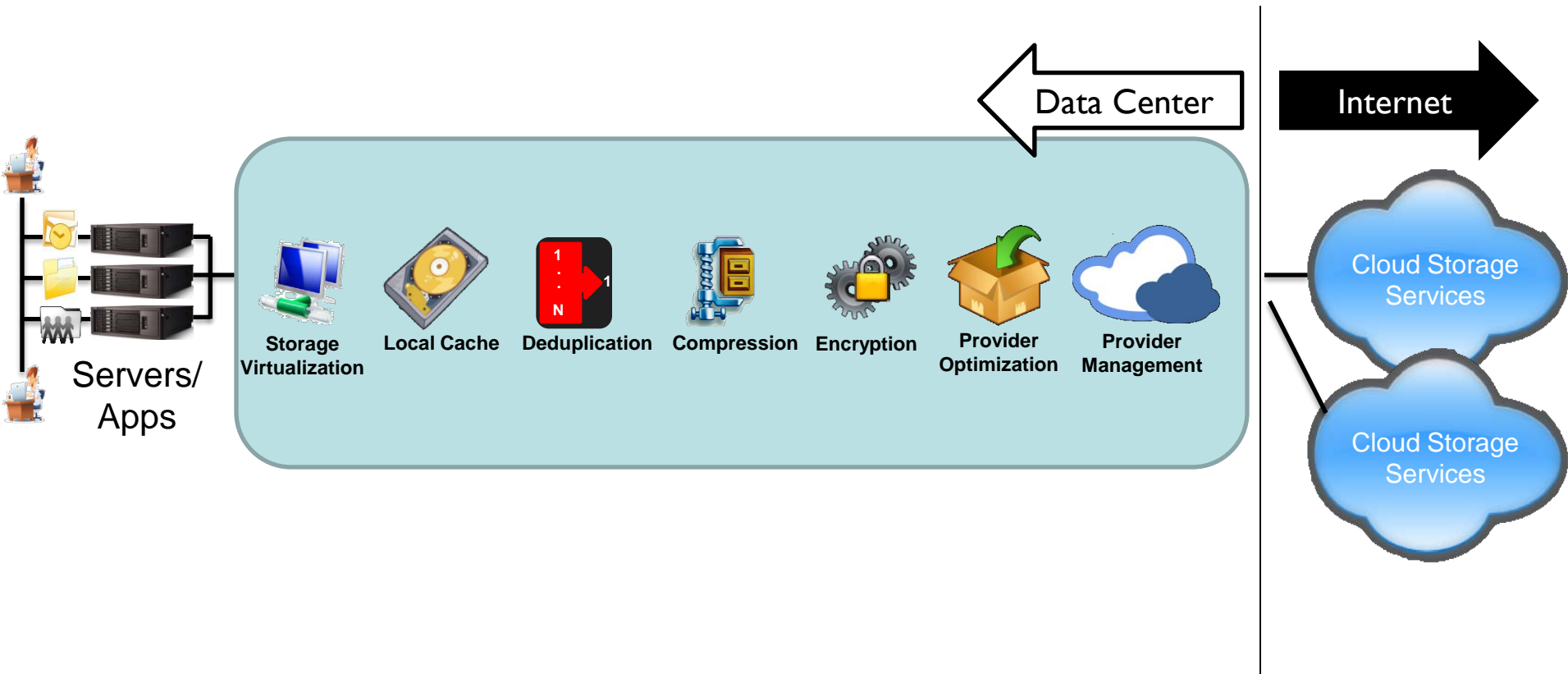
Cloud Storage Controller



-  Use cloud storage for any application
-  Local SAN performance
-  Fully secure

-  Enterprise storage management features
-  Reduced cloud storage costs

Cloud Storage Controller



Is Cloud Storage better than on-site storage?

- ❑ Not suitable for certain workloads (e.g. data warehousing)
- ❑ Does unlock new capabilities
 - ❑ Unlimited capacity provisioning
 - ❑ Primary storage with CDP-like benefits
 - ❑ Instantaneous disaster recovery from anywhere
 - ❑ Simplified technology refreshes
- ❑ Makes cloud storage economics compelling
- ❑ Easier management

- ❑ No one storage model works for everybody and every application
- ❑ Cloud storage is better as a means to an end for most use cases – not an end itself
- ❑ Cloud storage controllers let you
 - ❑ Get the benefits of cloud storage
 - ❑ Eliminate integration challenges of cloud storage
 - ❑ Combine the best of both cloud and local storage

Questions?

The Evolution of Cloud Storage - From "Disk Drive in the Sky" to "Storage Array in the Sky"

Allen Samuels

Co-founder & Chief Architect

