# Dragon Slayer Consulting Intro

- Marc Staimer - President & CDS
  - 14 years consulting
    - Storage, SANS, SW, networks, server, data centers, MSPs
    - Consults vendors (> 200)
    - Consults end users (> 600)
    - Market and technology analysis
    - Publishes consistently with TechTarget
    - Periodically published for trade magazines
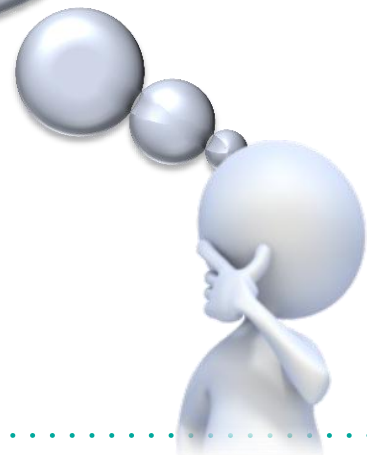 - 32 years industry experience

marcstaimer@comcast.net
503-579-3763

Think First
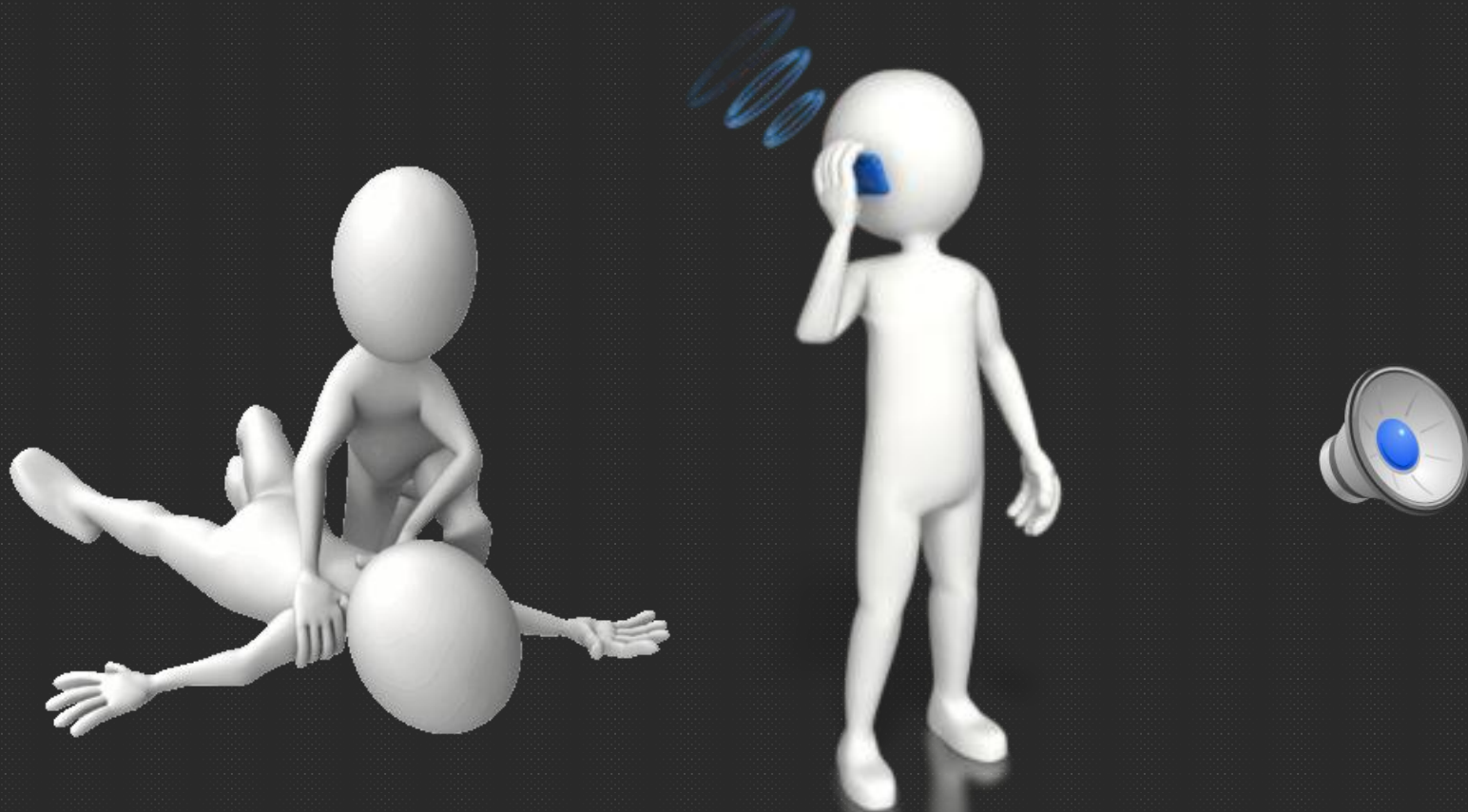
Agenda

- Why cloud storage?
- What's a storage cloud?
- What's a hybrid storage cloud?
- How to cost-effectively build a hybrid storage cloud
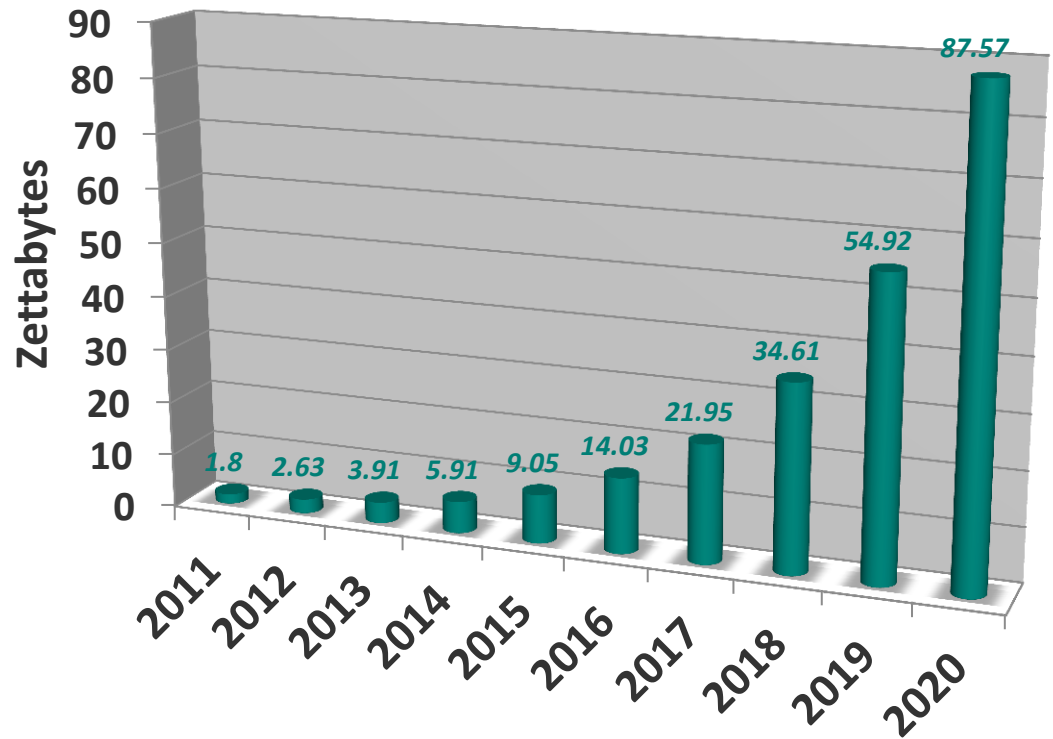- Conclusions & recommendations

Why Cloud Storage?

It Starts with the Problem that Cloud Storage Addresses

How to cost effectively manage the…

- Massive _**EXPONENTIAL**_ digital data growth*
  - Unstructured data
    - Today = ~60%
    - ~62% CAGR
  - Structured data
    - Today = ~40%
    - ~21% CAGR
  - Overall CAGR
    - ~54%
  - 1.8 ZBs 2011

Stored Digital Data Growth Curve

Zettabytes

Year	Value
2011	1.8
2012	2.63
2013	3.91
2014	5.91
2015	9.05
2016	14.03
2017	21.95
2018	34.61
2019	54.92
2020	87.57

*Source: IDC & DSC

Most of That Growth is Unstructured

- Unstructured data
  - 5 to 7X more stored data
  - Growing 3X faster

Structured Data - DBMS

Unstructured Data

Where That Data Growth is Coming From

Human generated data, yes

But, machine generated data is orders of magnitude more

- More info generated outside traditional systems & sources
  - In volume, variety, velocity & data complexity
  - User-generated & machine generated content
    - Orders of magnitude more
  - Aggregation of highly-structured heterogeneous sources
    - Patient records, insurance claims, etc.
    - Raw data, social media, log files, & sensor data

What That Growth Means to Your Storage

- Just 2 PBs today @ 62% CAGR

    - 3 yrs = 8 PBs

    - 6 yrs = 32 PBs

    - 9 yrs = 128 PBs

    - 12 yrs = 512 PBs

- In 13.5 yrs, 2PBs of storage will become 1 EB

- 250TBs becomes 8PBs in 6 yrs.

Problems of Explosive Data Growth

Issues that require new methodologies

- Scaling capacity, performance, and files/objects

  - PBs, GB/s, billions to trillions

- Storage infrastructure

  - Power/cooling, floor space, etc.

- Tech refresh/data migration

- Data resilience

- Data protection

- Data archiving

- Geographic dispersed sharing

Traditional Primary Storage is Really Quite Good

- At serving up storage

  - Fast

  - To lots of apps & clients

  - To different tiers

    - Based on value

  - To different horizontal markets

    - SMB/Mid-tier/enterprise

  - To different vertical markets

  - @ Enterprise scale

Traditional Storage is Not So Good At...

- Massive amounts of passive or archive data

- Tech refresh

- Radically reduced storage cost

- Pay-by-the-drink pricing paradigm

- Reducing storage infrastructure costs

- Data resilience / permanence

  - Required for long-term data archiving

- Geographically dispersed distribution

- Providing DR without multiple data centers
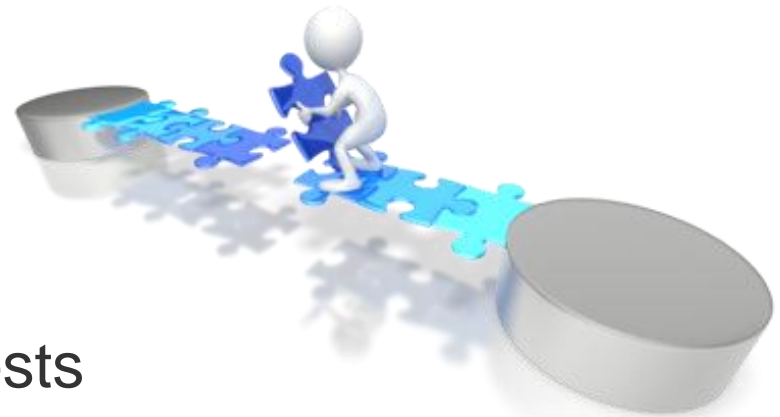
There Are Always Tradeoffs…

There's no such thing as a free lunch!
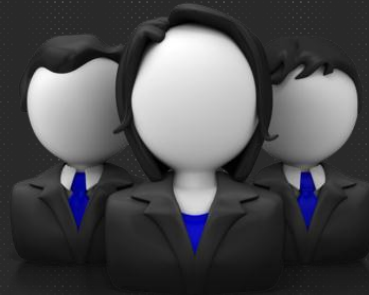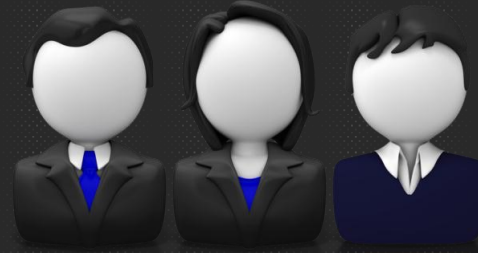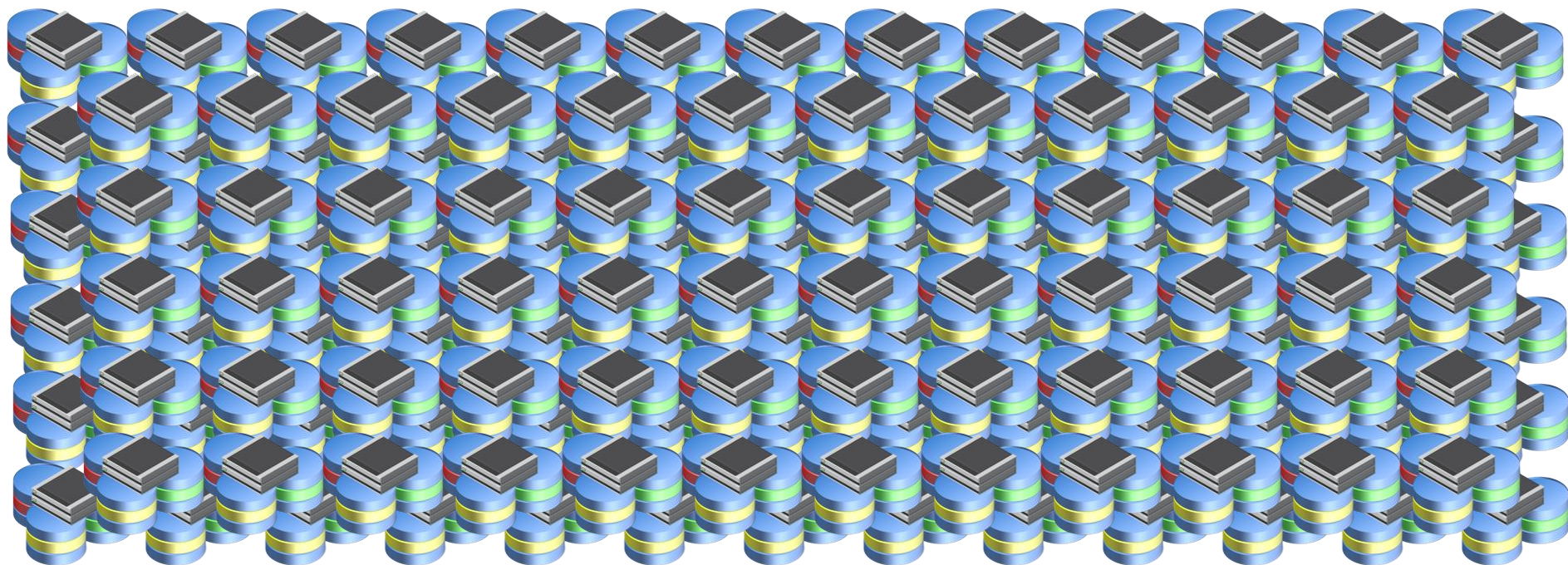
Traditional Storage Systems Have Limited Scale

- Massive passive data amounts ends up as storage system sprawl

- It's human nature…

    - If 1 is good, 2 has to be better, right?

    - What about 10? 20? 100? 200? You get the idea?

Tech Refresh is the Real Killer Because of...

- Excessive data migration, downtime, mgmt & cost

  - Required as storage systems approaches limitations

    - 60% of usable storage utilized

  - Currently averaging 9 to 12 months

    - Because of the manually intensive data migration

      – Data movement only a small part of data migration process
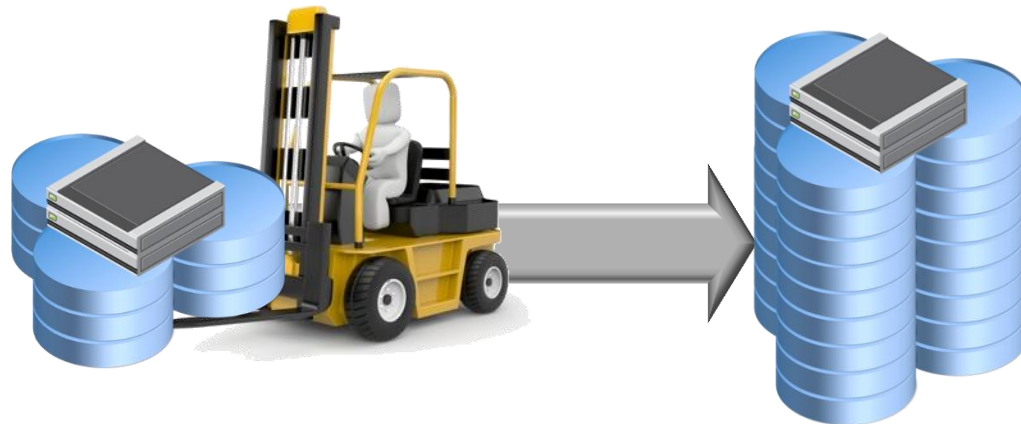
Traditional Storage TCO Way Too High

- Passive data does not have same value as active data

  - But traditional storage is not inexpensive

    - Architecturally aimed at primary data, not passive data

    - Doesn't match pricing to data value very well

    - TCO does not align cost effectively – much too high

Traditional Storage Cost Paradigm: "Pay Me Now"

- Upfront payment

  - Based on forecasted capacity, performance, & file requirements

    - Risk on buyer, not seller of storage

    - Unused storage/performance/software licensing, etc. still paid for

  - Not designed for "pay-for-use" in arrears cost paradigm

Traditional Storage Infrastructure is a Huge Cost

- Typical 3-year OpEx ranges from 5 to 8 x Capex
    - Floor space
    - Rack space
    - Storage networking switches
    - Cables
    - Conduit
    - Transceivers
    - Power
    - Cooling

- OpEx costs too high
    - Must match value of archive data
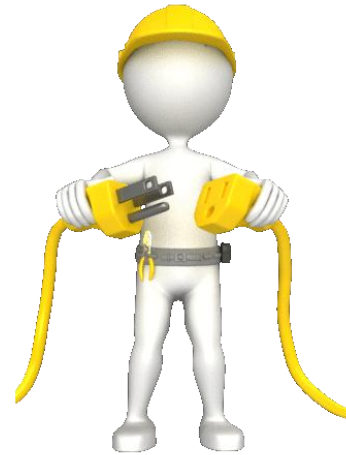
Traditional Storage Has Limited Data Resilience

- Media corrupts digital data

  - Tape/disk/flash/optical

    - Bit rot, corrosion, oxidation, background radiation
    - HDDs: Silent data corruption
      - Torn, partial, misplaced writes
    - Flash has other corruptions
      - Electron leaks, write wear cycle


Photo of actual data corruption on disk; in this case, a result of a failed data recovery from a HDD

- Best practice rewrite data

  - Disk: ~ 2-3 yrs
  - Tape: ~ 2-3 yrs
  - Flash: ~ 2-3 yrs
  - Optical: ~ 5 yrs

  - Primary corruption defense
    - RAID
    - Snapshot
    - Replication
    - Backup

- Fails to meet data

  - Permanence requirements
    - And subsequently, compliance

Traditional Storage Requires Another DC for DR

- Plus, more storage to protect against data disasters

- Essentially DR requires duplicating storage ecosystems

WAN
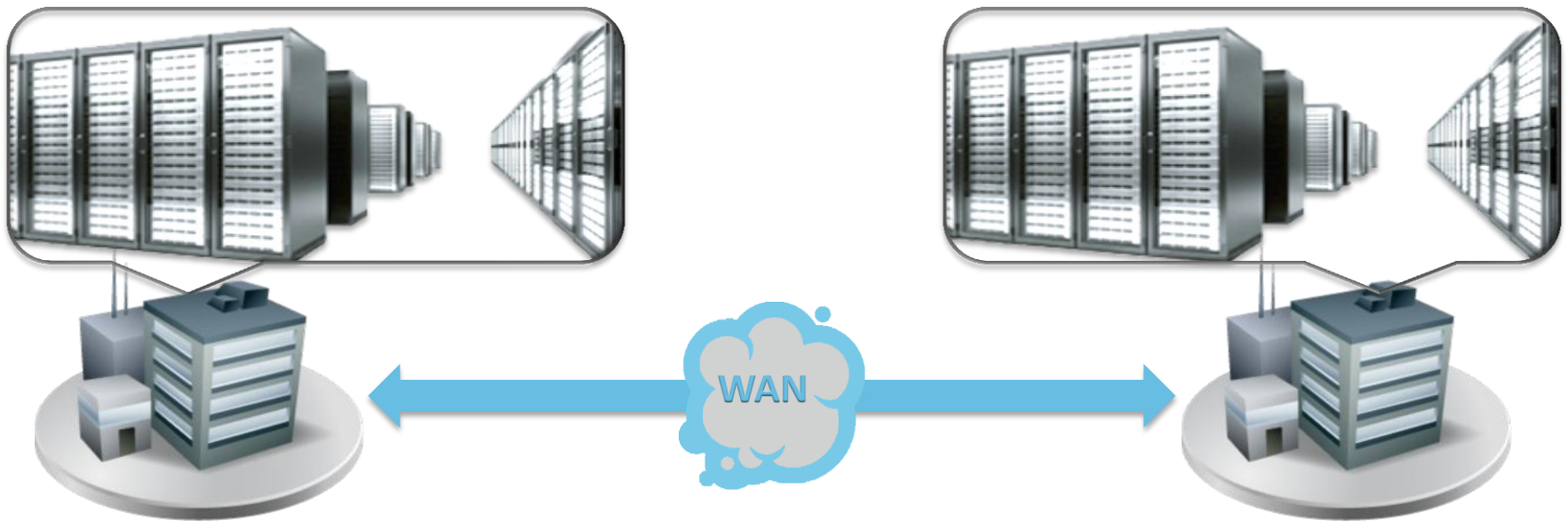
Traditional Storage Has Difficulty with Geographic Dispersed Data Sharing

- Data sharing for:
  - Work flow
  - Content distribution
  - Business analytics
  - Development
  - Partners
  - Safety

- Traditional storage methods are costly & time consuming
  - Snapshot & replicate
  - Mirror

Traditional Storage (800LB Gorilla) Shortcomings Review

- Not so good at:
  - Massive amounts of passive or archive data
  - Tech refresh
  - Radically reduced storage cost
  - Pay-by-the-drink pricing paradigm
  - Reducing storage infrastructure costs
  - Data resilience / permanence
    - Required for long term data archiving

  - Geographically dispersed distribution
  - Providing DR without multiple data centers

Cloud Storage Architected to Solve Those Issues

- From the ground up for archive/passive data/collaboration
  - It's not so good at being primary storage*

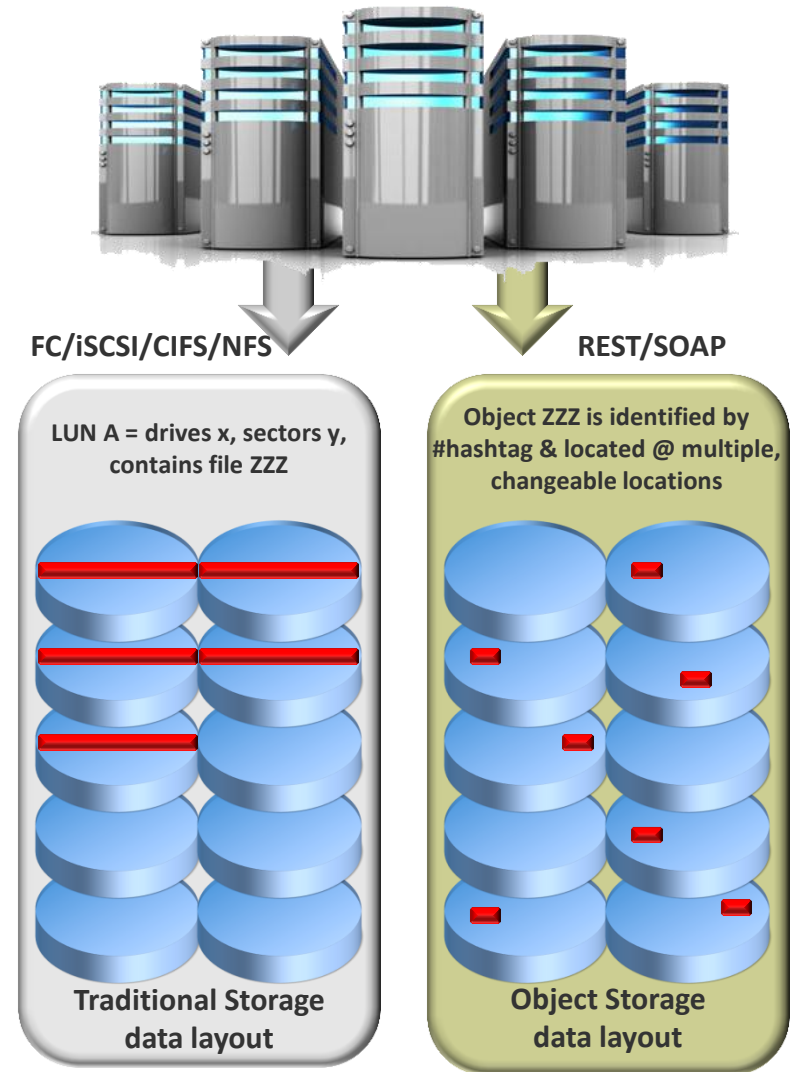*One Cloud Storage exception is Scality Organic Ring*

What is Cloud Storage?

- Cloud storage is
    - Massively scalable in capacity, performance, & objects/files
    - Constructed of many distributed resources acting as one
    - Extremely durable data persistent
    - Inherently fault tolerant
    - Automated data protection w/autonomic healing
    - Application & data non-disruptive tech refresh
    - On-demand resource allocation
    - Bullet-proof security
    - Multi-tenant/self service/detailed billing or chargeback
    - Geographically dispersed & aware
    - Usable storage efficient
    - Multi-protocol access
    - Paradigm shift lower storage TCO

Cloud Storage is Based on Object Storage

- Object storage stores data differently
  - Loosely federated data
  - Vs. consistent storage system
    - Across all resident data
  - E.g. No requirements for
    - Cache consistency
    - Nodal awareness of objects owned by other nodes
    - Single aggregated namespace
  - Data scales based on rules
  - Rules about the data itself
  - Rather than about the system
  - Can scale nearly indefinitely
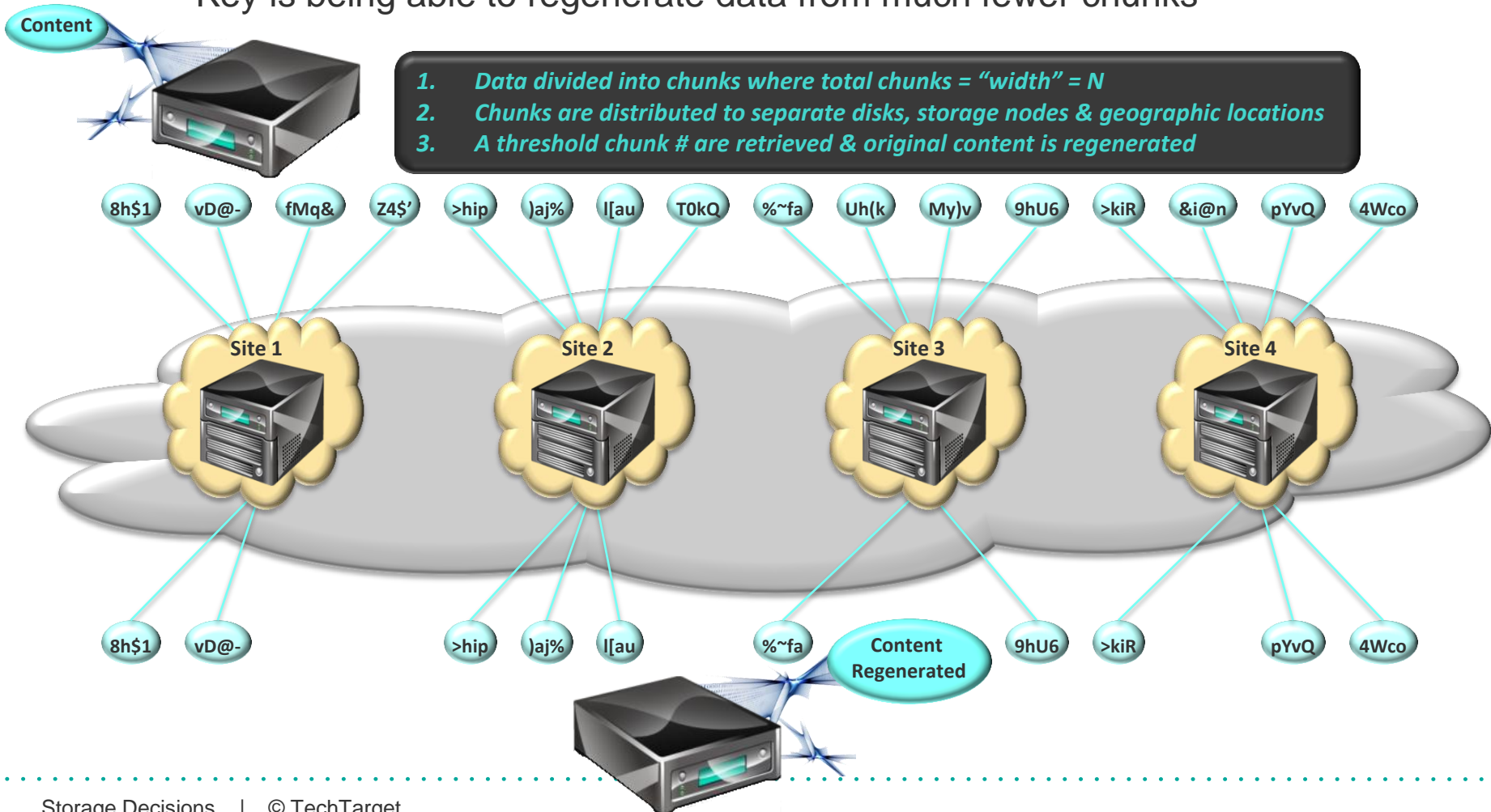
FC/iSCSI/CIFS/NFS

REST/SOAP

LUN A = drives x, sectors y, contains file ZZZ

Object ZZZ is identified by #hashtag & located @ multiple, changeable locations

Traditional Storage data layout

Object Storage data layout

Cloud (Object) Storage Has a Lot More Metadata

- More customized *control* over the data

  - Vs. file system w/fixed amount of metadata

    - File type, creation date, & last-accessed date

  - Vs. SAN storage which typically has virtually none

  - Object storage increases # of possible metadata fields

    - Customizable for specific business and system functions

      – Allows system to manipulate data based on policy triggers

      – Data scales based on rules

  - Rules that also *automate* many traditional manual tasks

   - Tiering, security, migration, redundancy, and deletion

How Erasure Codes Work

- Erasure coding is new kid on the block (pun intended)
  - Breaks objects up into smaller chunks storing them in different locations
  - Key is being able to regenerate data from much fewer chunks

Content

1. *Data divided into chunks where total chunks = "width" = N*
2. *Chunks are distributed to separate disks, storage nodes & geographic locations*
3. *A threshold chunk # are retrieved & original content is regenerated*

8h$1 vD@- fMq& Z4$' >hip)aj% l[au T0kQ %~fa Uh(k My)v 9hU6 >kiR &i@n pYvQ 4Wco

Site 1   Site 2   Site 3   Site 4

8h$1 vD@-   >hip)aj% l[au   %~fa   Content Regenerated   9hU6 >kiR   pYvQ 4Wco
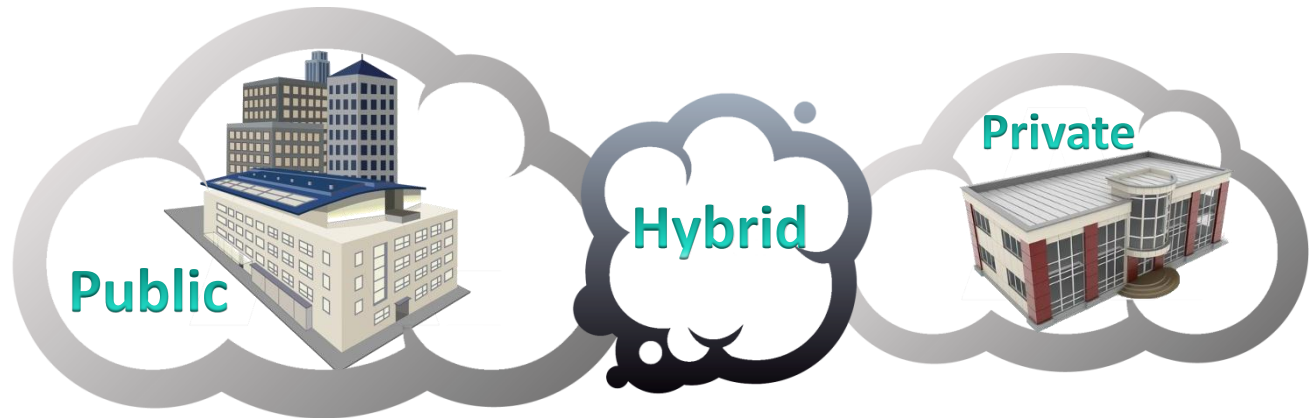
How Multi-Copy Mirroring Works

- Multiple copies generated when written based on policy
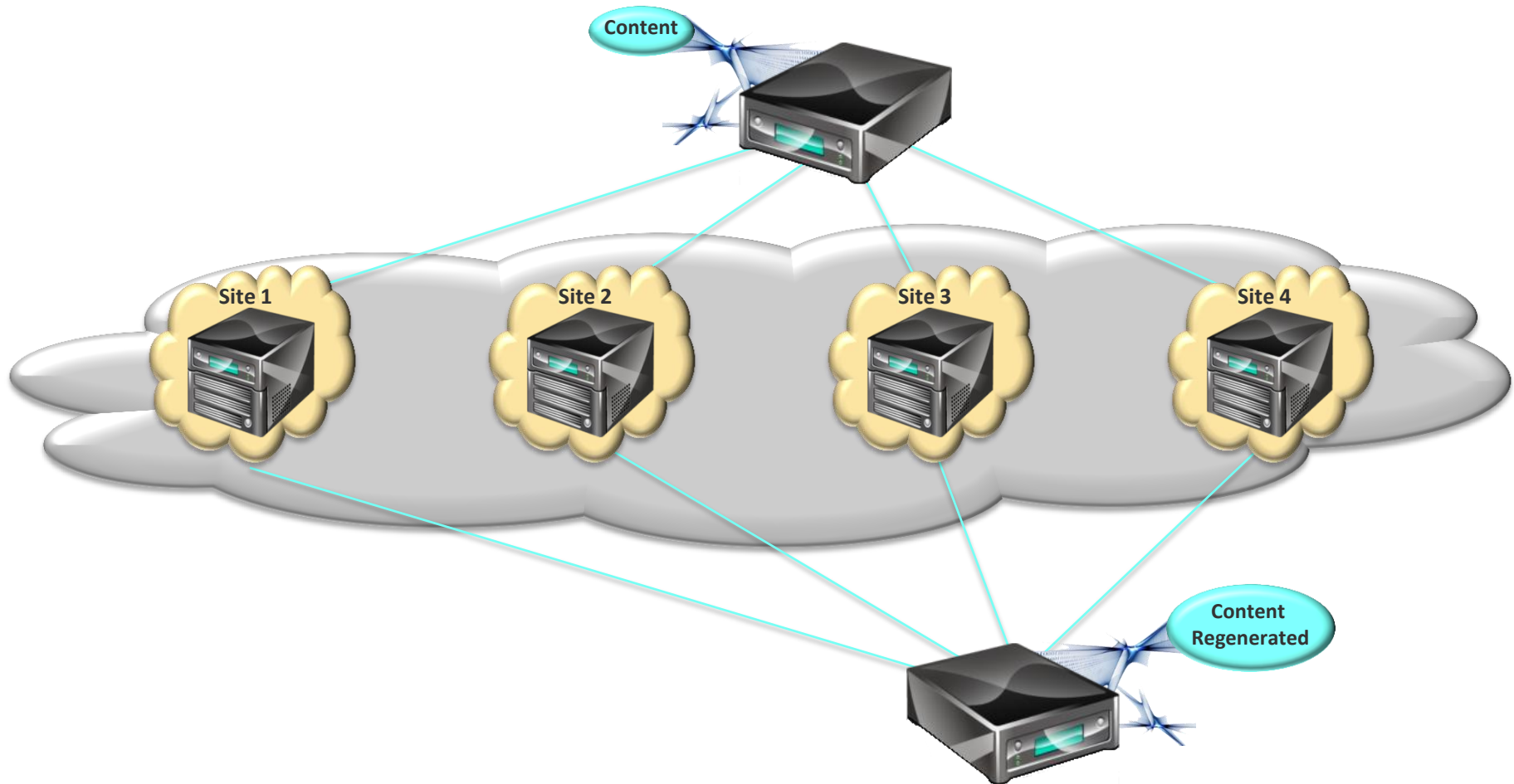
Types of Cloud Storage

- Public

- Private

- Hybrid

Public Cloud Storage

- Storage as a service (STaaS)

  - Over the Internet or VPNs

  - Pay-by-the-drink & only for what's actually being used

  - Accessed via REST and/or SOAP

    - As well as hardware & software gateways

  - Typically has multiple data centers

    - Geographically separated by regions

  - 3 classes

   - Consumer, industrial, enterprise

Private Cloud Storage

- IT owns/operates/manages their cloud storage

  - On customer's premises or co-lo

  - Leveraging cloud storage technology

  - On VLAN and/or VPN

  - Can charge-back to departments if desired

  - Still accessed primarily via REST and/or SOAP

    - As well as hardware & software gateways

  - Provides many public cloud advantages

    - Just requires own or co-lo data centers

Hybrid Cloud Storage

- Solves variety of public cloud problems

  - Long latency

  - Bandwidth

  - Control

  - Local data copies of some data

  - Uploading time to public cloud
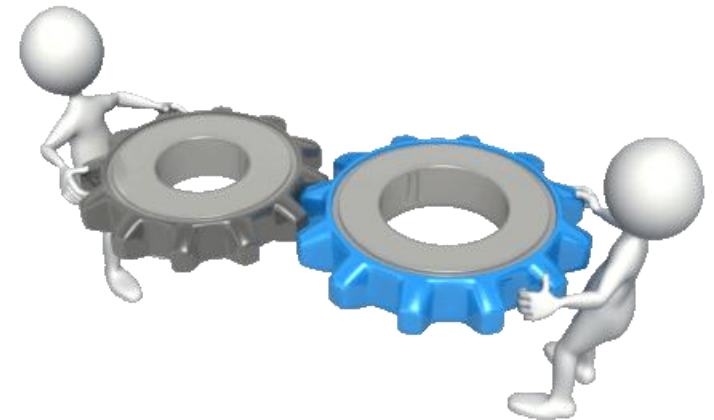
# How to Build a Cost Effective Hybrid Storage Cloud

3 Ways to Cost Effectively Build a Hybrid Cloud

- Local private cloud to public cloud or clouds

- Private storage integrated with public cloud or clouds

- Public cloud extended to private data centers

Local Private Cloud to Public Cloud or Clouds

- Local private cloud that is federated with a public cloud
  - Local control
  - Utilizes public cloud as
    - Archive
    - Backup
    - Work-flow sharing
    - Distribution
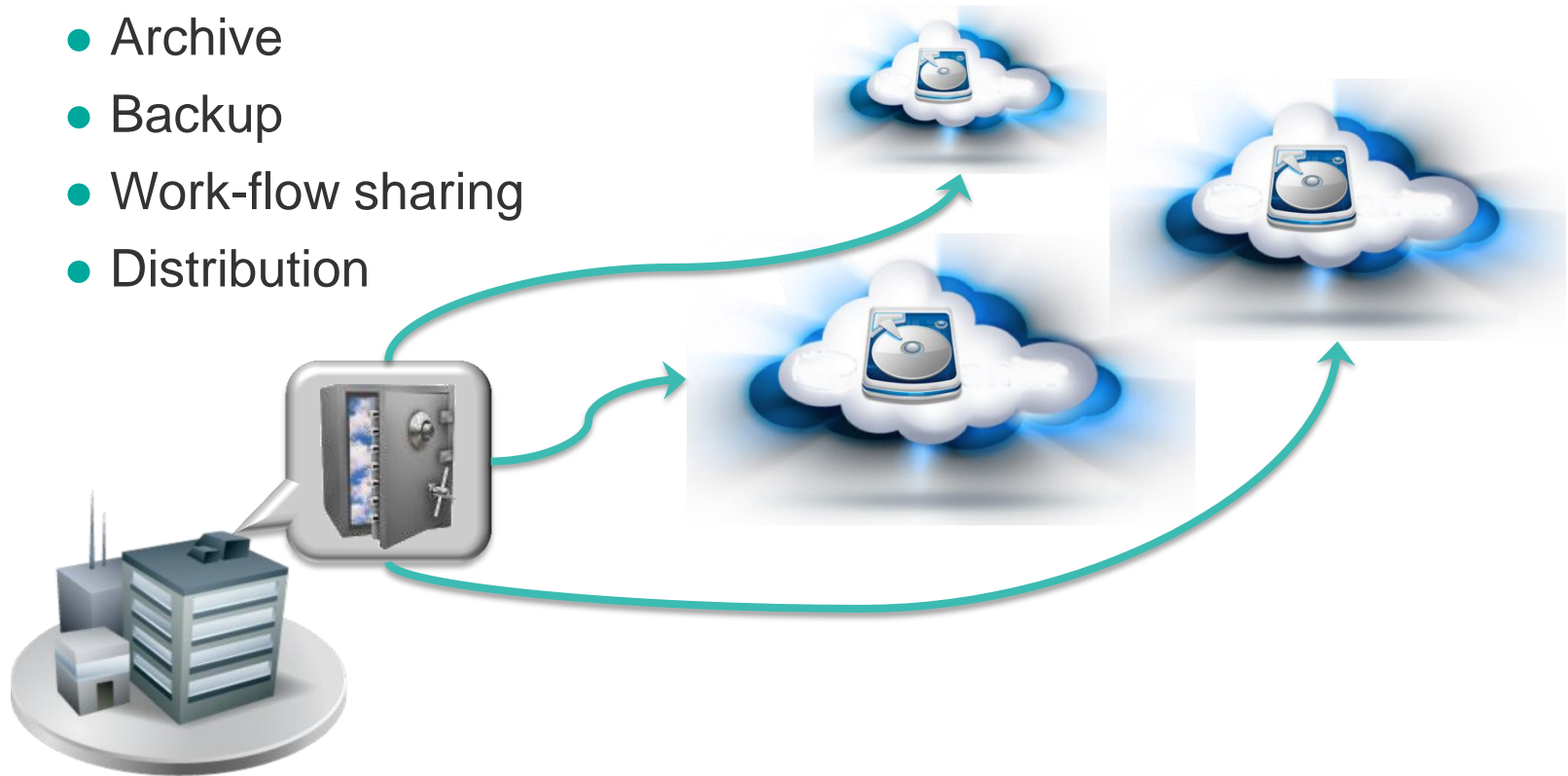
Storage Cloud to Storage Cloud Communication

- Private storage cloud to public storage cloud

  - Convenience of local storage w/storage cloud functionality

    - Lower latency faster response times

    - Local control

  - Takes advantage of public storage cloud

    - For archive, DR, backup, workflow collaboration, etc.

    - Less onsite storage to manage

    - Less local storage infrastructure
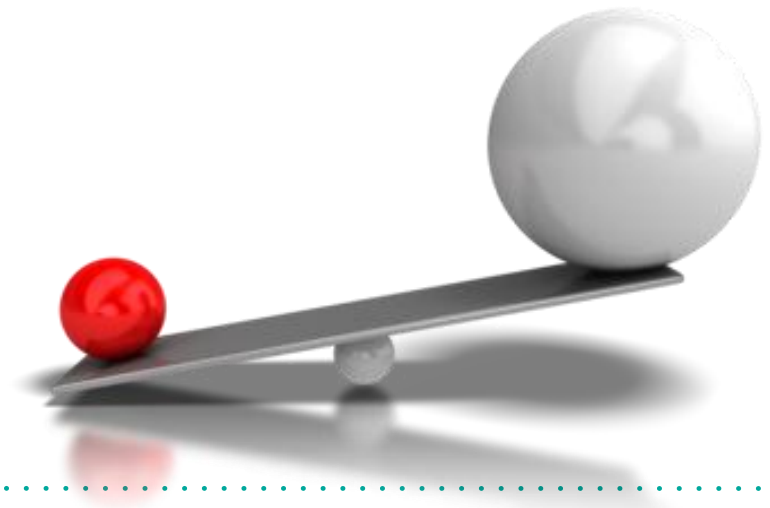
Private/Public Integration Pros & Cons

Pros

- Lower local storage costs
- Better response times
- Enhanced DR
- Heightened collaboration
  - Geographically distributed

Cons

- Compatibility issues
  - Same software on both sides
    - Some cases even same versions
    - Will change with CDMI
- Limited choices
  - EMC Atmos MSPs
    - Atmos versions must be same
 - Nirvanix & their OEMs

Technical Requirements

- Private cloud software

- x86 server hardware

  - Purchased separately

    - Or bundled w/software from vendor

  - Embedded HDDs &/or SSDs

  - Some front end SAN storage

  - 1/10/40 Gb/s TCP/IP Ethernet std. NICs

- 1/10/40 Gb/s Ethernet switch/routers

- Internet access

- Interoperability w/public storage cloud

Private Storage Integrated with Public Cloud or Clouds

- 3 variants: Converts NFS, CIFS, or iSCSI to REST &/or SOAP
  - Primary storage w/cloud integration
  - Secondary storage w/cloud gateways
  - Software gateway to public storage cloud
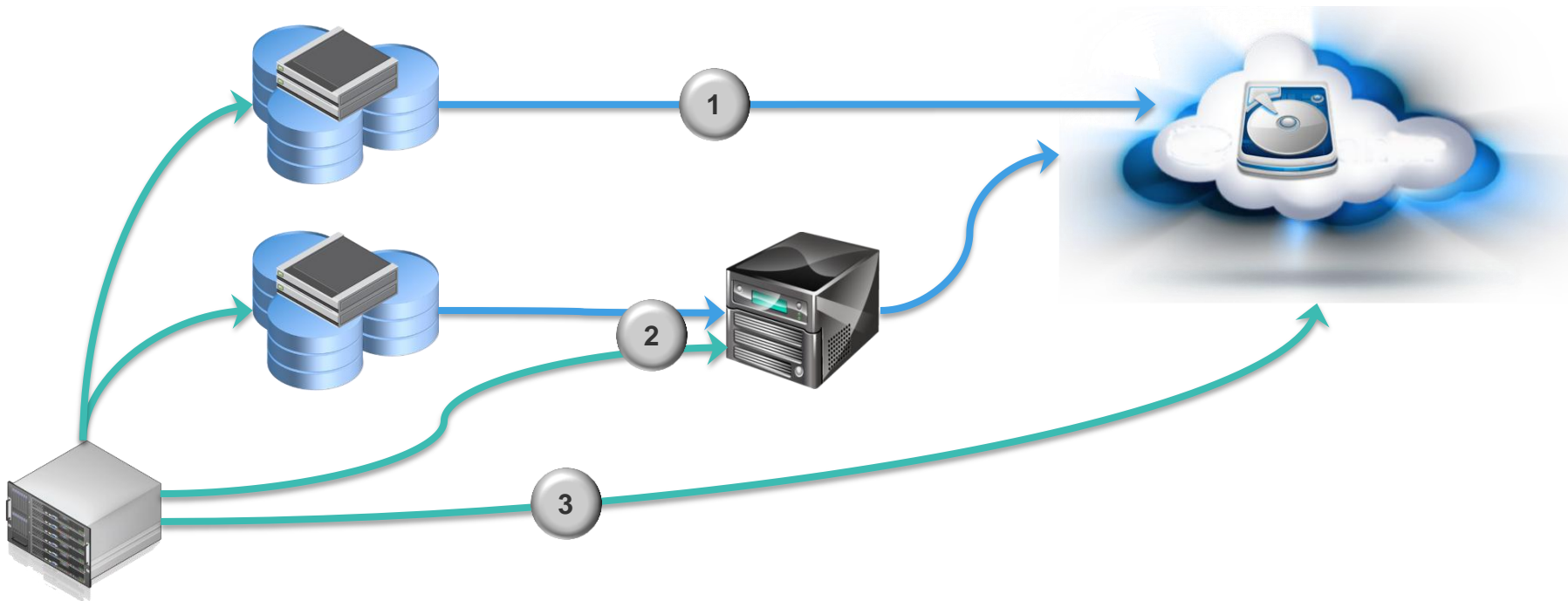
Each Works Differently

- Primary storage w/cloud integration (NAS or iSCSI storage)
  - Provides an unlimited backend to primary storage for
    - Passive data, archive, data protection, collaboration, etc.

- Secondary storage w/cloud GWs (NAS or iSCSI)
  - Requires data be transferred to/from primary storage or apps
    - A data mover of some type must move the data to it
    - Still provides an unlimited backend to primary storage
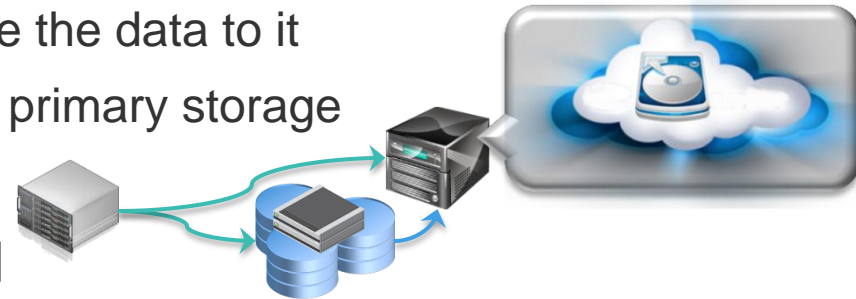
- App integration to public storage cloud
  - Apps/servers directly place data in public cloud
  - App/server controlled & centric
    - Managed like any other storage target

Primary Storage/Public Cloud Integration Pros & Cons

Pros

- App transparency
  - Unlimited cloud tier
- Lower TCO
  - Lower public cloud costs
    - Deduped & compressed
- Simpler work-flow collaboration
  - Specifically with file
- Automatic DP (snaps) offsite
  - Public cloud for DR
    - No duplicate DC

Cons

- Catalog control
  - Requires same vendor storage
    - At all locations w/data access
    - For workflow collaboration
    - To read or write the data
    - Not designed for mobile data
    - Funnel point
- Doesn't avoid tech refresh
  - Although simpler & less of it

Secondary Storage/Public Cloud Integration Pros & Cons

Pros

- Lower TCO on secondary data
  - Lower public cloud costs
    - Deduped & compressed
- Good for archiving or DP target
  - Once data on system
    - Moves data to public cloud (DR)
    - Caches recent data locally
  - Requires other software
    - To move the data

Cons

- In reality a gateway
  - Caches secondary data to cloud
  - Transitory market
- Requires data mover to system
  - Data must be migrated
- Catalog control issue
  - Same as primary storage solution
- Doesn't avoid tech refresh
  - Although simpler & less of it

# App or Server/Public Cloud Integration Pros & Cons

Pros

- App transparency
  - Unlimited cloud tier
- Lower TCO
  - No intermediary storage system
- App/server admin direct control
  - Much simpler work-flow sharing
  - Mobile usable
  - No catalog control issue
- Most DP SW natively supports
  - Public cloud storage

Cons

- Apps must work w/cloud store API
  - REST or SOAP
    - CDMI later
    - Or software gateways
      - NAS or iSCSI
- Higher public cloud costs
  - No auto data reduction tech
  - On primary data put in cloud
   - Unless the app provides it

Technical Requirements

- Storage system or Gateway system (SAN or NAS)
  - With a REST or SOAP backend
  - Compatible w/storage cloud of choice
- Applications writing directly to storage cloud
  - REST or SOAP API compatible w/storage cloud of choice
- Both require Internet access
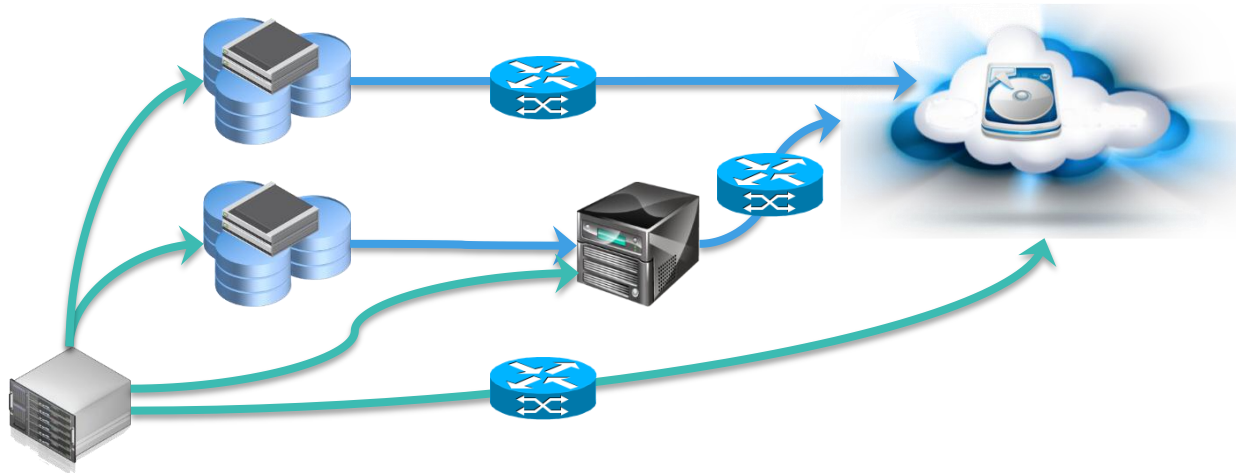
Public Cloud Extended to Private Data Centers

- Instead of moving data to the public cloud

  - With all of the bandwidth issues & latency that involves

  - Move the public cloud to the data

    - Leveraging cloud storage federation
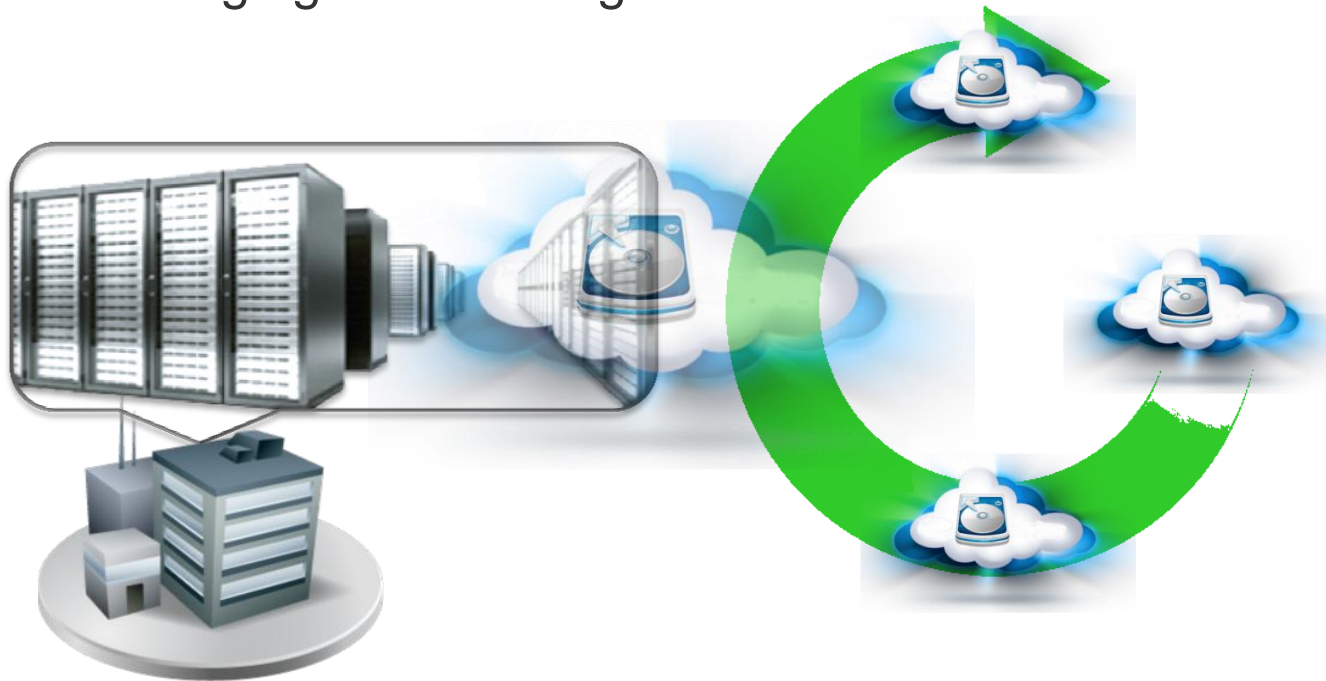
Public Cloud Extension into Private Data Center

- Managed service
  - Handled entirely by MSP in your data center
    - No employed admins required
    - In some cases, repurposes onsite traditional storage for local cloud
  - Convenience of local storage with storage cloud functionality
    - Lower latency faster response times
    - Local control
  - Takes advantage of public storage cloud
    - For archive, DR, backup, workflow collaboration, etc.
    - Less onsite storage to manage
    - Less local storage infrastructure
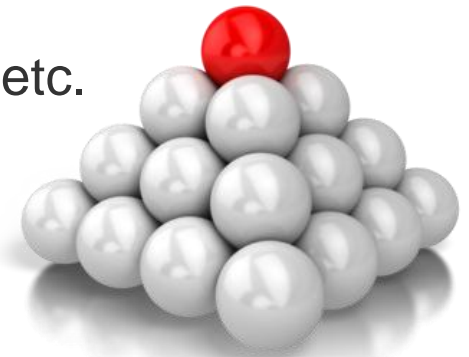
App or Server/Public Cloud Integration Pros & Cons

Pros

- Excellent cloud performance
 - Same unlimited cloud tier
- Lower TCO
  - No intermediary storage systems
  - Lower mgmt costs
- Everything appears local
  - Provides all storage cloud benefits
  - GWs provided by MSP
    - Managed by MSP as well

Cons

- Apps must work w/cloud store API
  - REST or SOAP
    - CDMI later
    - Or SW or HW gateways
      - NAS or iSCSI
- Higher public cloud costs
  - No auto data reduction tech

Technical Requirements

- Storage cloud vendor must extend their cloud
  - To your premise
  - They provide SW & HW to do so
    - Some can repurpose your SAN, DAS, or NAS storage
- You provide Internet connection
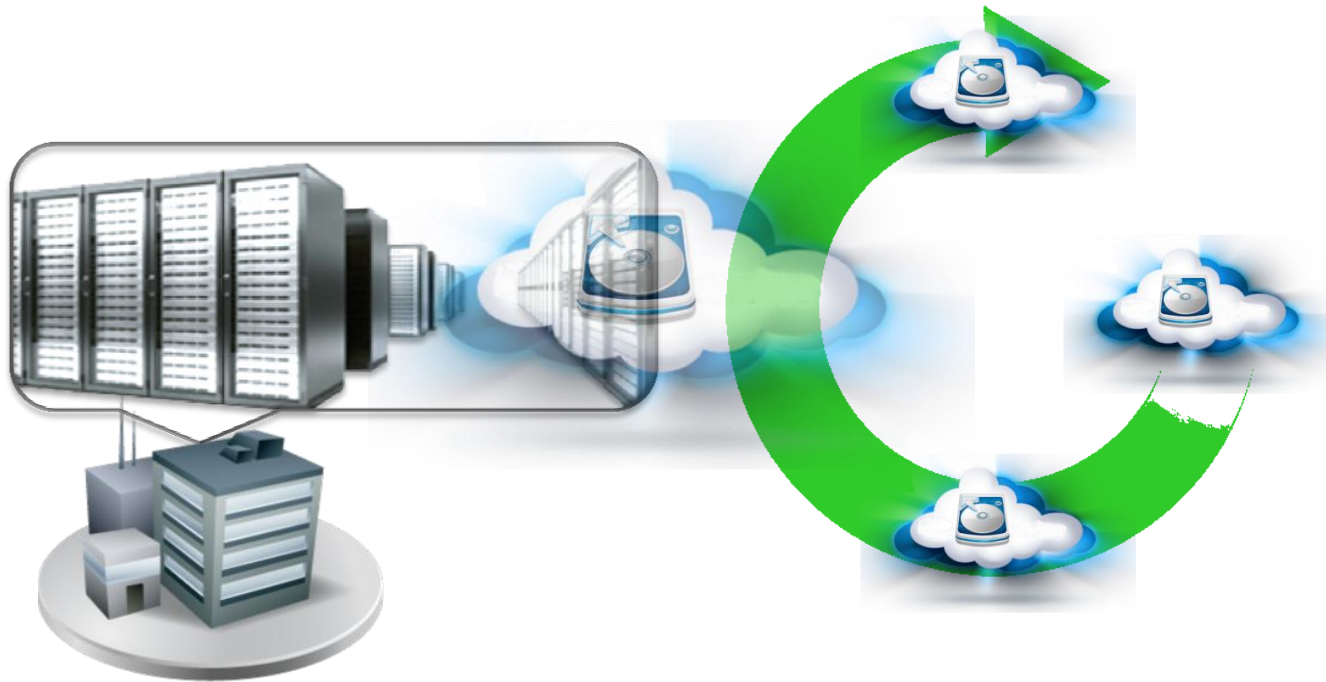
Public Cloud Storage Service Providers

Public Cloud Storage Service Providers			
Vendor	**Headquarters**	**Software**	**Service**
Amazon	Seattle, WA	Amazon Object Storage	S3
ATT	Dallas, TX	EMC Atmos & OpenStack	Synaptic Storage
Cerner	Kansas City, MO	Nirvanix Object Storage	Skybox Cloud Storage
DELL	Austin, TX	Nirvanix Object Storage	DELL Cloud Storage Service
DreamHost	Brea, CA	DreamHost Object Storrage	Dream Objects
DRFortress	Honolulu, HI	Nirvanix Object Storage	Cloud Storage
Google	Santa Clara, CA	Google Object Storage	Google Cloud Storage
HP	Palo Alto, CA	OpenStack Swift	Hp Cloud Object Storage
Hosted Solutions	Raleigh, NC	EMC Atmos	Cloud Storage
IBM Global Services	Armonk, NY	Nirvanix Object Storage	Smart Business Storage Cloud
Microsoft	Redmond, WA	Microsoft BLOB	Azure
Nirvanix	San Diego, CA	Nirvanix Object Storage	Cloud Storage Network
PEER 1 Hosting	Vancouver, BC Canada	EMC Atmos	CloudOne
Rackspace	San Antonio, TX	OpenStaack Swift	CloudFiles
Swisscom	Worblaufen, Switzerland	Nirvanix Object Storage	Swisscom Cloud Services
USC	LA, CA	Nirvanix Object Storage	Digital Repository
Unisys	Blue Bell, PA	OpenStack Swift	Unisys Secure Cloud

Private Cloud Storage Vendors

| Private Cloud Storage Vendors			
Vendor ▼	**HQ** ▼	**Product** ▼	**Appliance and/or SW** ▼
Amplidata	Antwerpe, Belgium	Amplistor	Both
Caringo	KC, MO	CAStor	Software
Citrix	Santa Clara, CA	CloudStack	Software
Cleversafe	Chicago, IL	SliceStor	Both
Compuverde	Karlskrona, Sweden	Object Store	Software
DELL	Austin, TX	DX Object Store	Appliance
DDN	San Diego, CA	Web Object Scaler	Applliance
EMC	Hopington, MA	Atmos	Appliance
HDS	Santa Clara, CA	HDS Object Storage	Both
HP	Palo Alto, CA	Open Stck	Appliance
InkTank	LA, CA	CEPH	Software
NetApp	San Jose, CA	StorageGRID	Appliance
Nirvanix	San Diego, CA	Cloud Storage Network	Software
Quantum	San Jose, CA	StorNext Object Storage	Both
Red Hat	Raleigh, NC	GlusterFS	Software

Conclusions & Recommendations

- Building a hybrid storage cloud makes enormous sense
    - One of the best ways to leverage public storage clouds

- How to build it depends on your specific requirements
    - One size does not fit all
    - There are multiple choices
    - Best fit will depend
    - Building cost effectively is quite doable
    - And you have a lot of different choices

Always Remember…

- Whenever there are 2 or more vendors…
  - Each will tell you the others have it completely wrong
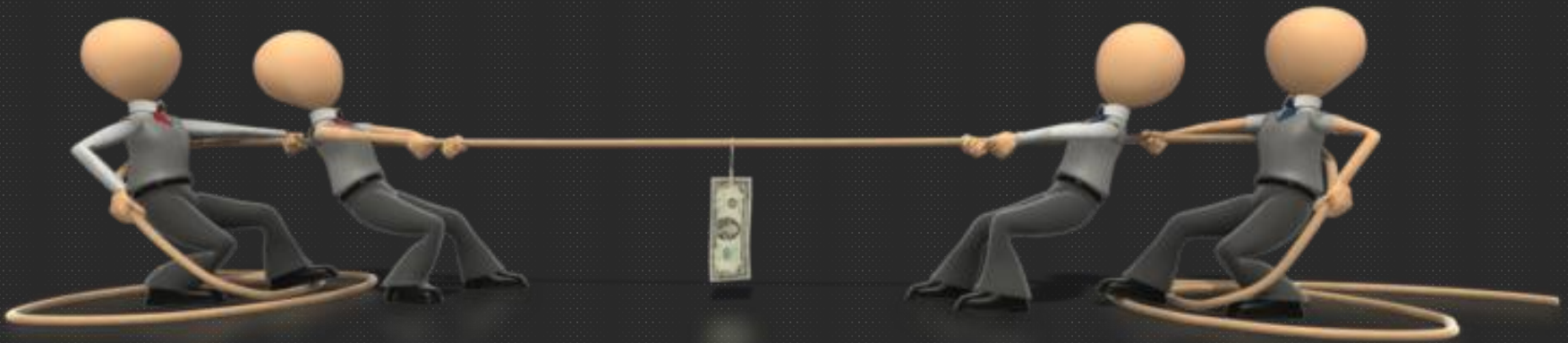
Questions? Thank You!

- Marc Staimer, President & CDS
- Dragon Slayer Consulting
- marcstaimer@comcast.net
- 503-579-3763

TechTarget
Editorial
Speaker Bureau
2012

#SDSF2012