



June 13-15, 2016

| Marriott San Mateo

| San Mateo, CA

# Scale-out Searchable Storage for the Cloud Age

**Tony Barbagallo**  
**Caringo**

# Expectations of File Storage Have Changed

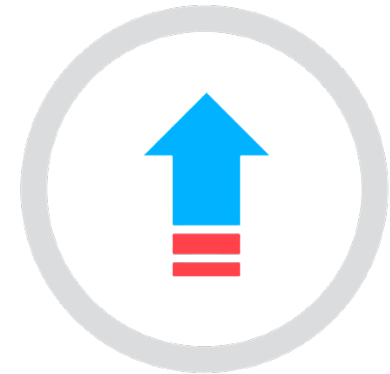
Endless Capacity



Accessible Anywhere

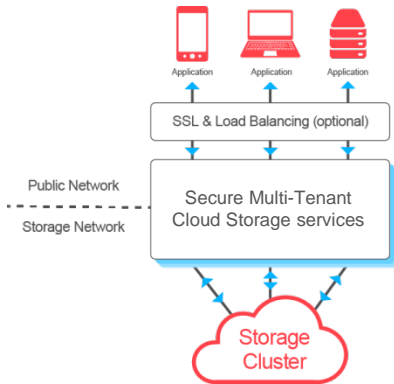


Always Available



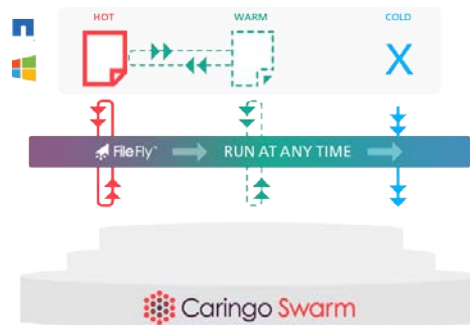
**What I Want – When I Want It!**

# Scale-out Searchable Storage



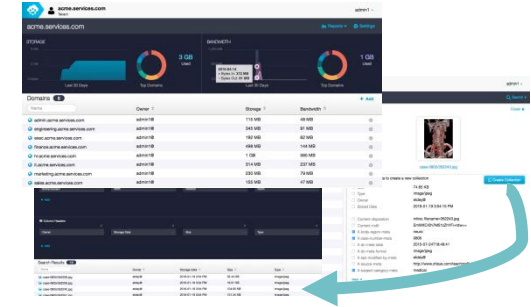
## Data Management Platform

- Multi-tenancy
- Public or Private
- Metering and Quotas
- Access Controls
- Multi-protocol Access



## Offload Legacy Storage

- Transparent File Access
- No Application Changes
- Streamline Backups
- Data is Continuously Protected



## Organization & Analytics at Scale

- Metadata stored with Files
- Integrated Search
- Content Portal
- Collections
- Big Data Enabled
- No File System Limits

# Architecture Overview



Applications | Automated Tiering from Filers



HTTP



Multi-Tenancy • Orchestration • Security • Metering

SCALE-OUT STORAGE CLUSTER

Automated Content Management  
Elastic Content Protection

SWARM  
SEARCH

SWARM PORTAL  
MANAGEMENT

- Configuration
- Monitoring
- Reporting

CONTENT

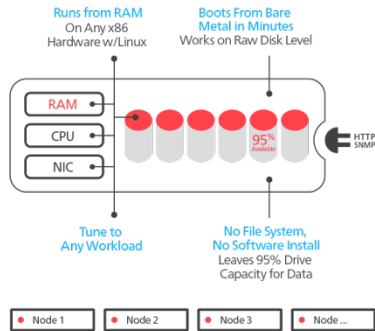
- Content Viewing
- Collections
- Meta Data Entry



Standard x86 Servers, VMs or Microsoft Azure



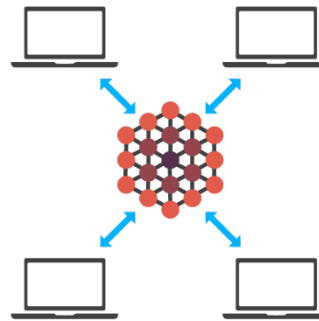
# Scale-out Storage Platform Technical Overview



## Complete Software Appliance

No file system, RAID, LUNs or DBMS required

Runs from RAM;  
95% capacity for files

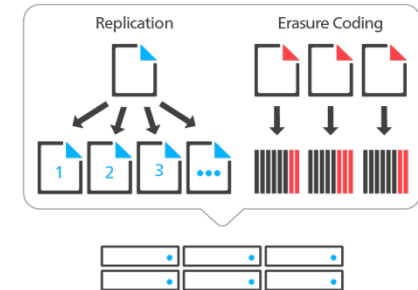


## Flexible Deployment

Storage nodes perform all functions

Fast proactive recovery

Darkive™ adaptive power conservation



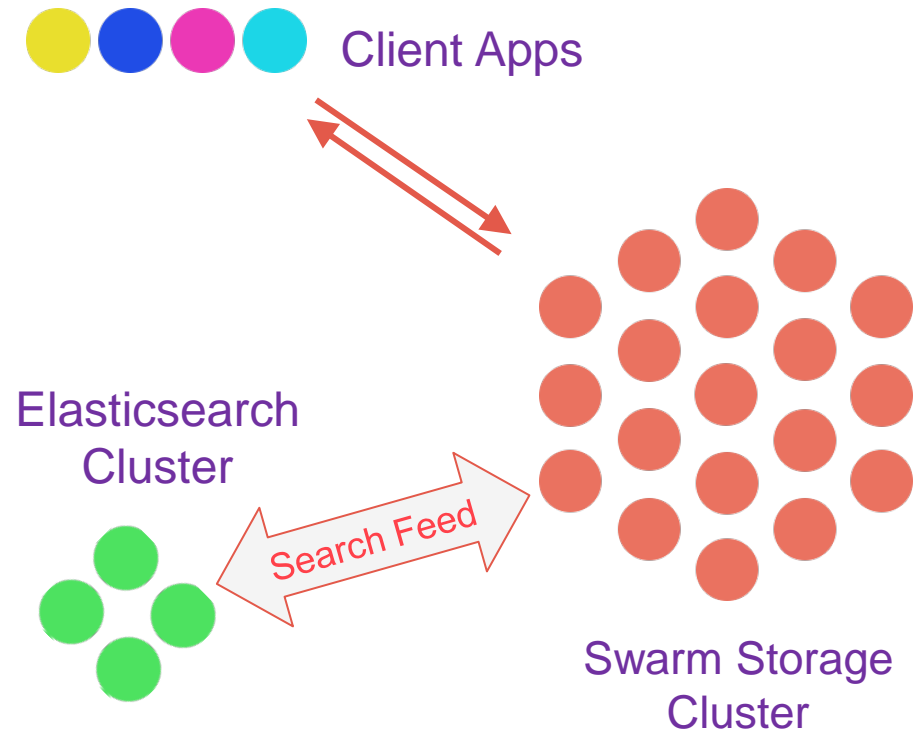
## Continuous Protection

Patented EC & replicated files on the same node

Feeds and subclusters for replication and DR

# Swarm Search Infrastructure

- Client Applications
  - send SCSP+ commands and receive information to/from the Swarm cluster.
- Swarm Cluster
  - Connects to the Elasticsearch(ES) servers through a Search Feed.
- Search Feed
  - Transmits the metadata from the storage cluster to the ES servers and the search queries and replies.
- Elasticsearch servers
  - Index the metadata and service search requests.



# Swarm Search

- Actionable insight
  - Targeted analysis
- Dynamic organization of content
  - Classification, key words, descriptive content
  - Multiple ways to track content
- No separate “Big Data” project required
  - Self-contained within the storage system

List domain objects within a cluster:

```
GET http://CLUSTER/?domains&format=json
```

List buckets within a domain:

```
GET http://CLUSTER/?format=json
```

List objects within a bucket:

```
GET http://CLUSTER/mybucket?format=xml
```

Search for objects with metadata matching:

```
GET http://CLUSTER/?format=json&content  
type=image/jpeg
```

# Collections – Organization and Search at Scale

The screenshot displays a cloud storage interface with three overlapping panels. The top panel shows search results for 'Medical Data' with 113 results. The middle panel shows search settings for a 'New Collection' with criteria for search scope, search criteria, and column headers. The bottom panel shows a detailed view of a search result for 'case-0805/352243.jpg', including metadata and a 'Create Collection' button. A red arrow points from the 'Create Collection' button to the search results table.

**Medical Data Search Results**

Name
case-0805/352238.jpg
case-0805/352239.jpg

**New Collection Search Settings**

Search Scope: Entire Domain

Search Criteria:

- x\_case\_number\_meta: 0805
- x\_subject\_category\_meta: medical
- x\_body\_region\_meta: neuro

Column Headers:

- Owner
- Storage Date
- Size
- Type

**Search Results (18)**

Name	Owner	Storage Date	Size	Type
case-0805/352238.jpg	etdey@	2016-01-19 3:54 PM	52.44 KB	image/jpeg
case-0805/352239.jpg	etdey@	2016-01-19 3:54 PM	16.51 KB	image/jpeg
case-0805/352240.jpg	etdey@	2016-01-19 3:54 PM	104.65 KB	image/jpeg
case-0805/352241.jpg	etdey@	2016-01-19 3:54 PM	101.34 KB	image/jpeg

**Case Details: case-0805/352243.jpg**

Select metadata to create a new collection

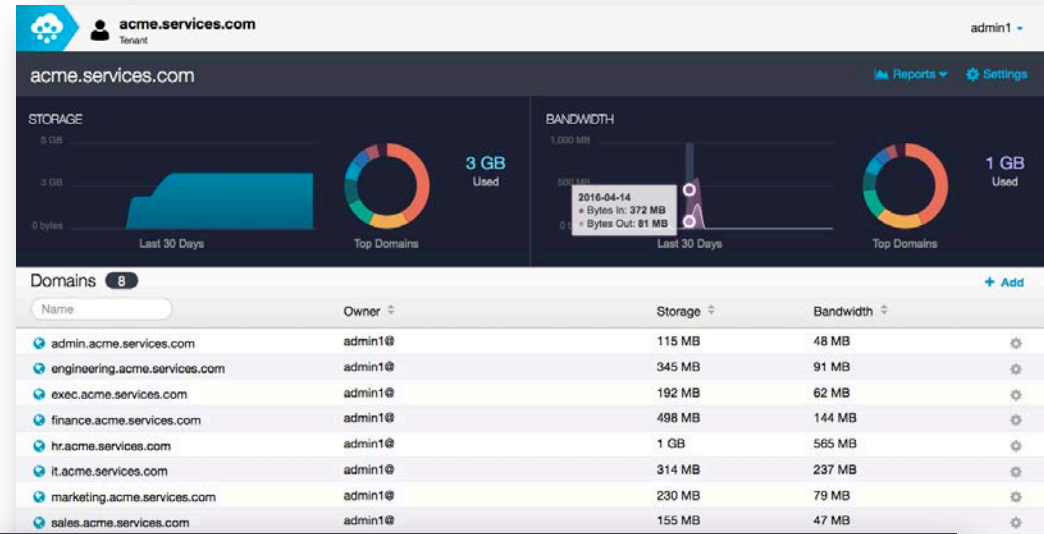
- Size: 74.65 KB
- Type: image/jpeg
- Owner: etdey@
- Stored Date: 2016-01-19 3:54:15 PM
- Content-disposition: inline; filename=352243.jpg
- Content-md5: EmWXC/0h7M51zZHyFi+tdw==
- X-body-region-meta: neuro
- X-case-number-meta: 0805
- X-dc-meta-date: 2015-07-24T18:48:41
- X-dc-meta-format: image/jpeg
- X-last-modified-by-meta: etdey@
- X-source-meta: http://www.ctisus.com/teachingfiles/
- X-subject-category-meta: medical

[Create Collection](#)



# Multi-tenant Management

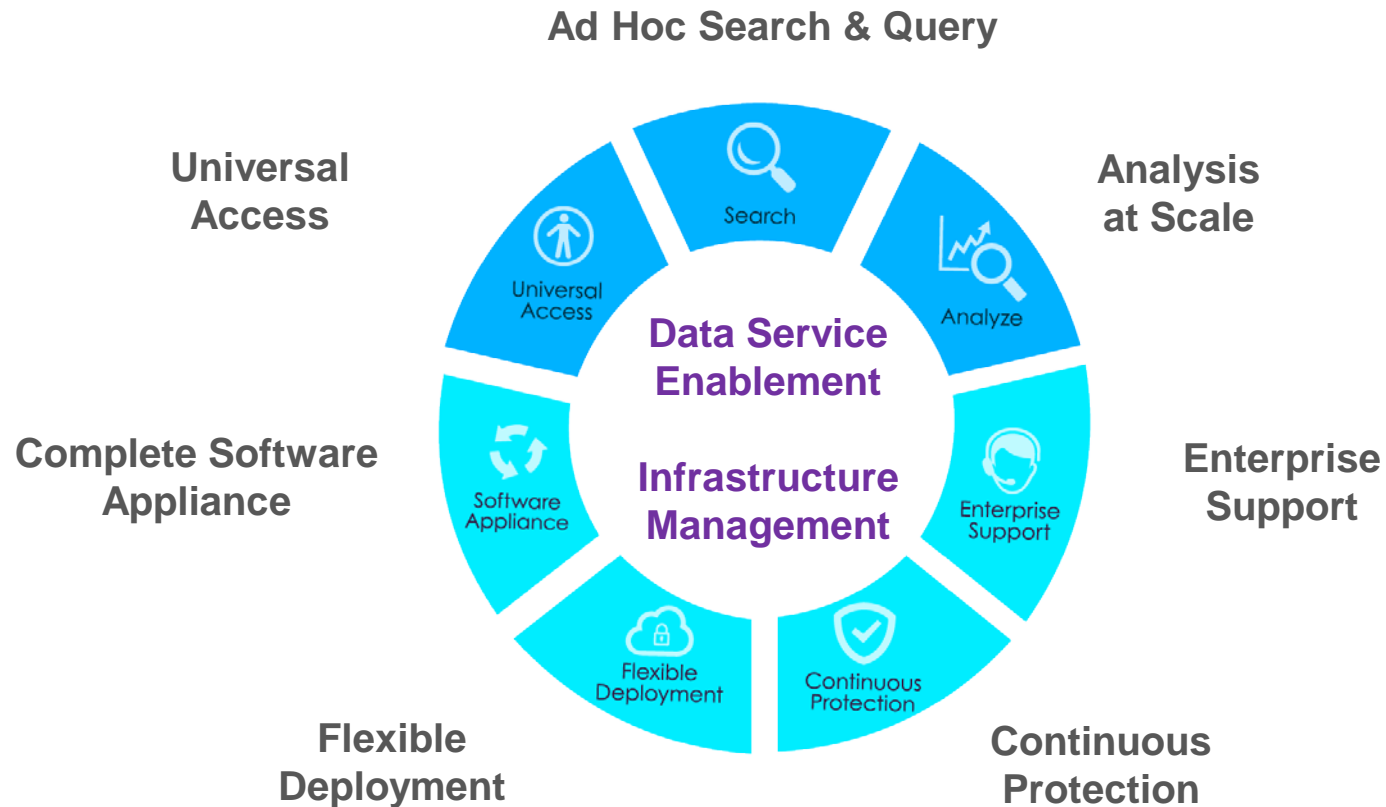
- Secure
  - Connects to LDAP, Active Directory, Linux PAM
  - Token-based authentication
- Robust system administration
  - RESTful API & via web portal
  - Flexible naming scheme, no bucket limits
  - Use your domain
- Interoperable
  - HTTP, S3, HDFS, NFS, SMB



The 'New Token' dialog box contains the following information:

- Owner: admin1
- S3 Secret Key:  S3 Secret Key \* grKO9XgBqB4II2cjPYtyAbuGqT
- Expiration Date: 1 year (dropdown menu with options: 1 year, 3 years, 5 years, Exact date (yyyy-mm-dd))

# Limitless Storage at a Fraction of the Cost



# Thank You

[www.caringo.com](http://www.caringo.com)



[facebook.com/caringostorage](https://facebook.com/caringostorage)



[linkedin.com](https://linkedin.com)



[@caringostorage](https://twitter.com/caringostorage)