Centera: Fixed Content Objects

David L. Black
Senior Technologist
black_david@emc.com

IEEE Mass Storage Conference
San Diego, CA
April, 2003
Content Data Explosion: New Approach Needed

- **Transactional Data**
  - Continuously updated
  - Many small read and writes
  - Low latency, high performance
  - Examples: Databases, OLAP, OLTP, ERP Systems, Accounting/Finance applications
  - Best addressed by SAN, NAS, or Direct Attached systems

- **Fixed Content**
  - Unchanged after creation
  - Work unit is information object
  - Larger unstructured BLOB sizes
  - Critical processes: storing, retrieving, location independence, global access/sharing
  - Examples: audio, video, pictures, documents, X-rays, etc.
  - Requires flexible content aware file systems, high scalability, reasonable access latency
**Centera At 30,000’**

- **Fixed Content Storage**: Content cannot be updated once stored – non-repudiation
- **No Possibility of Data Loss**: even if a component fails
- **Continuous Availability**: objects are mirrored – No Single Point of Failure
- **Low Maintenance**: Non-disruptive self-healing
- **Off-The-Shelf**: Standard hardware components
- **Scalable**: Easy to add capacity non-disruptively

- **Not**: A general purpose storage array or fileserver
The RAIN Scalable Storage Architecture

Redundant Array of Independent Nodes

Access Nodes provide external API access
Storage Nodes store and protect information

16, 24 or 32 nodes/cab

Automatic node, drive & network configuration
Clustered operation with load balancing and self healing
Node/rack mechanicals designed for easy servicing

850MHz PIII w/256MB RAM
4ea. 160GB EIDE
3ea. 10/100BT ports
How Centera Works

1. Object is created or used by an application.

2. Object is sent to Centera via Centera API over IP.

3. Centera generates unique Content Address.

4. Content Address returned to application.

5. Content Address is retained for future reference.
Why is Content Addressing Important?

- **Content authenticity**
  - Unique “fingerprint” is generated from the content itself
  - Content is validated on delivery
  - Content integrity is continuously validated in background

- **Content Address is location independent**
  - Address is globally unique
  - Not a place in a hierarchy (file system)
  - Not a place in a disk array (logical volume)

- **Identical objects are only stored once**

- **Intrinsic access security**
  - Must have unique Content Address to retrieve an object – can’t browse
**Centera: Fixed Content Objects are Different**

- No changes to data objects after creation
  - “RAID” without writes: No small write algorithm, etc.
  - Replication is easy, no write-ordering issues
  - Coordination in the face of failures is much simpler
    - Avoid RAID intricacies when a single drive has failed

- Naming based on object contents
  - Object name/handle: secure hash “fingerprint”
  - Can bind arbitrary metadata to each object
    - No filesystem limits on metadata structure or size
  - Result: “fileserver” without directories or path names
    - Data catalog: Content management application (e.g., data grid)
    - Complements filesystems for active computation

- Petabyte scale in current product via multiple clusters
  - Simple architecture: better scaling than large filesystem
Customer Requirements & Centera

- Information lasts a long time (10’s of years)
  - Must survive technology migrations
  - Must ensure information integrity

- Information grows to huge scales
  - 100’s of TB to PB’s
  - Add capacity non-disruptively
  - Self-Management is key

- Information lives in multiple places
  - For both access & disaster recovery

- Competitive TCO is required

- Centera objects keep 128-bit address forever
  - Multiple technologies can co-exist;
  - Content authenticity revalidated

- Centera scales linearly from 5TB to PBs
  - Address of object never changes
  - Cost is linear
  - Capacity is “plug-n-play”

- Self Managing, Self-Configuring, and Self-Healing

- Centera can replicate multiple places in an active-active configuration

- Centera uses Best-Of-Breed standard parts