Autonomic Storage Systems

Dilip D. Kandlur
Department Head, Storage Systems
IBM Almaden Research Center
Disk storage trends

Disk Storage Terabyte Shipment and $/MB Forecast 2002-2007

PB (petabyte) shipments

$/MB

Source: IDC 2002
Storage Management Challenges

Storage-related expenditures*, as a % or IT budgets, is also growing rapidly

Transition from Direct attached to Network Storage underway
The Customer Storage Hierarchy

Storage Software/Services
Virtualization & Systems Management

Storage Systems

SCM
HDD
Optical
Tape

Storage Device
HDD $1GB/yr, ↓40%/yr

$10-35/GB for system
Acquisition, ↓35%/yr

$50/GB, (5 yr device life)
Key challenges for managing storage

**Application Availability**
- Non-disruptive data migration
- Eliminate application failures from out-of-space conditions
- Reduce or eliminate planned and unplanned downtime
- Protection of data from disasters, human and system errors

**Personnel Productivity**
- Easier provisioning of new storage for applications
- Manage data from a centralized point
- Easier storage and server consolidations

**System Optimization**
- Data movement from high-cost to low-cost storage based on policies
- Reduce duplicate data through more effective data sharing
- Achieve desired disk utilization through better storage management
Autonomic Vision

“Intelligent” open systems that:
- Manage complexity
- Know themselves
- Continuously tune themselves
- Adapt to unpredictable conditions
- Prevent and recover from failures
- Provide a safe environment
Autonomic Computing

Increase Responsiveness
Adapt to dynamically changing environments

Operational Efficiency
Tune resources and balance workloads to maximize use of IT resources

Business Resiliency
Discover, diagnose, and act to prevent disruptions

Secure Information and Resources
Anticipate, detect, identify, and protect against attacks
Autonomic Storage Systems

- Unified storage infrastructure and common management system
- Storage System manages autonomically to application requirements
  - goal driven, policy-based
  - placement, replication and migration of data to meet performance goals
  - backup/restore/replication to meet availability goals
  - capacity growth and performance growth predictions
  - automated provisioning of storage networks and storage servers
- Minimize complexity of storage hardware
Collective Intelligent Bricks - Concept

- Encapsulate complexity - hardware is packed into Lego-like bricks. A storage system can be built with one type of brick.

- Systems are piles of bricks (the IceCube). Bricks communicate with all neighbors, creating a 3-D mesh.

- Bricks contain only commodity parts, leading to low cost. Bricks fail at a rate typical for commodity systems.

- Reliability is achieved by distributing functions and redundancy across the entire cube. Admins may add bricks, but should not remove bad bricks.

- Object software (Kybos) for storage and management.
Collective Intelligent Bricks advantages

**Monolithic**
- Scaling is extremely coarse
- High entry cost
- Very robust components
- Failure disruptions can be major
- Failed components repaired

**Modular**
- Scaling is fine grain
- Low Management costs
- Low entry cost
- Moderately robust components
- Failure disruptions are small
- Failed components not repaired
Simple Storage Brick

- **Start with a simple brick**
  - No redundancy
  - No especially robust components
- **Distribute Software across bricks**
  - Services can run on any brick
- **Distribute data across bricks**
  - Data Redundancy across bricks

**Components:**
- Processor
- Memory
- Network
- Disks
Brick-based storage system features

- Resources scale in concert
- Self-managing, fail-in place
- Incremental growth by plugging in brick
- Add data without specifying where it goes – system manages resources
Basic Building Block (‘Unit-brick’)

No connectors, cables, fans, fibers....

Slot for ‘thermal bus’

10 Gb/s coupler
IceCube System 'Piles'

- Bricks are stacked on vertical columns (not shown) for power insertion and heat removal
- Bricks communicate with neighbors in a 3D mesh

Almaden prototype - 3x3x3 bricks, 324 disks, 26 TB, 10" on a side, hold the library of congress, 2002
1PetaByte: Conventional vs IceCube
(w/ 250-GB, 18-W, 3.5" disks in either system)

Today...
- 250 kW
- 75+ db Noise
- Waterpipes required

Future...
- 190 kW
- quiet
- Waterpipes required
- 14% Floorspace
CIB Hardware: IceCube

- System is scalable to thousands of bricks (Petabytes)
  - very high physical density, innovative cooling

- 3D-mesh supports parallelism through high speed data exchange
  - Terabits/s in a larger cube, but at low cost due to new ‘capacitive couplers’

- This performance is a key to solving storage reliability problems
  - because it allows fast rebuild after a disk failure (during which data are at risk) and enables many different network-RAID schemes
CIB Software: Unified Storage System Structure

- NAS (NFS, CIFS) Interface
- SAN Filesys++ File Interface
- Virtualized OSD Interface
- Base OSD Interface

System Management

- SAN-Filesystem++ for CIB

Collective Intelligent Bricks

Level 1

Level 2
Capabilities:

- performance and semantics like SAN
- sharing like NAS
- policy-based, centralized storage management (better than both)
Kybos: Storage system software for bricks

- Self-configuration software for bricks
- Object access layer provides virtualized object interface to clients: load balancing, fault recovery
- Network RAID across bricks for high availability, support fail-in-place
STEPS: Policy based Storage Management

- Policy based data life cycle management with StorageTank
- Policies for:
  - File allocation
  - File migration, HSM
- File movement based on administrator specified criteria, including:
  - file age, inactivity, name patterns
  - file sets, pool utilization, and migration time.
- Scalability to billions of files
  - avoid brute-force scans
  - leverage SAN FS architecture
Current Status

- IceCube brick hardware initial debug complete (3x3x3 bricks, 26 Terabytes)

- Anticipate fully populated IceCube in September and operational in Q4
  First demo will be based on minimal effort storage software
  Plan to get Kybos/Storage Tank/OSD software running on IceCube by end-of-year
The Customer Storage Hierarchy

Application data & storage mgmt
Information Lifecycle Management

Storage Software
Virtualization & Systems Management

Storage Systems

SCM
HDD
Optical
Tape

Storage Device
HDD $1/GB, \downarrow 40\%/yr

$50/GB, (5 yr device life)

$10-35/GB for system acquisition, \downarrow 40\%/yr
Application data & storage management
Expand a file system to allocate/extend a database table

1. DBA: Notice tablespace full
   - Or -
   2. Contact Sys Admin

3. Check for space on filesystem; see more needed
   - Or -
   4. Contact Storage Admin

5. Log into host; run commands to see new volumes

6. Add primary volume to volume group, expand group

7. Setup mirroring with remote volume using volume mgr

8. Mount primary volume to filesystem; extend

9. Log into DB; create datafile
   - Or -
   10. Add datafile to tablespace and expand

11. Log into System
   - Or -
12. Install application

13. Repeat previous 4 steps for volume at remote DR site

14. Setup mirroring with remote volume using volume mgr

15. Contact DBA
   - Or -
   16. Installer

17. Installer: Determine space needed
   - Or -
18. Contact Storage Admin

19. Find LUN(s) with right characteristics

20. Make sure LUN(s) are bound to right port

21. Ensure HBA, port are in same zone

22. Mask LUN to allow HBA access

23. Contact Sys Admin

DBA - or -
App Installer
Autonomic Computing Architecture Overview

An autonomic element contains a continuous control loop that monitors activities and takes actions to adjust the system to meet business objectives.

Autonomic elements learn from past experience to build action plans.

Managed elements need to be instrumented consistently.

Policy and symptom service databases are examples of knowledge databases.
Towards Autonomic Systems

System Elements
(Intra-element self-management)

Groups of Elements
(Inter-element self-management)

Business Solutions
(Business Policies, Processes, Contracts)

Customer Relationship Management

Enterprise Resource Planning

Server Farm

Enterprise Network

Storage Pool

Servers  Storage  Network Devices  Middleware  Database  Applications
Thank you