Linux Clustering Technologies

Mark Spencer

November 8, 2005

Presentation Topics

- Business Drivers
- Clustering Methods
  - High Availability
  - High Performance
- Cluster Filesystems
- Volume Managers

Business Drivers

- Service Continuation
  - example: Oil & Gas databases
    - Oracle RAC
- Uptime = $$
  - example: Online Sales
    - Amazon.com
- High Density Compute Farms
  - example: Medical
    - University of Calgary – Faculty of Medicine

Technologies

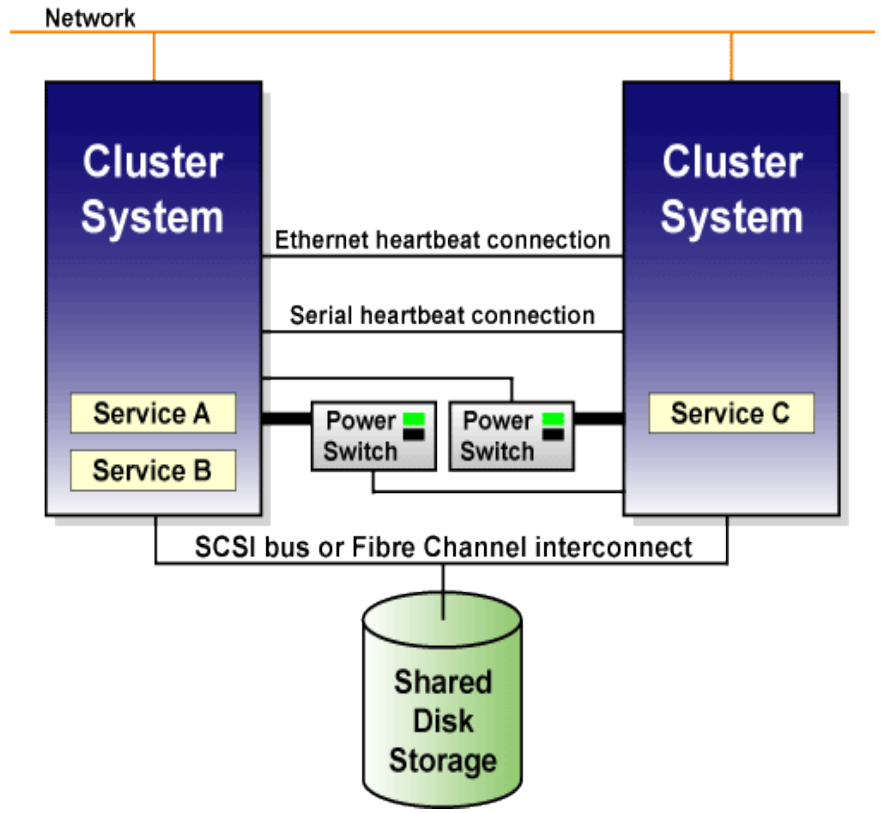
- High Availability
  - Active / Passive
    - one node does all of the processing
    - passive node monitors active node for service failure
    - Pros:
      - Service(s) are up 99.99% of the time
    - Cons:
      - Expensive Solution
        - twice the hardware costs, with no change in performance or scalability
      - There is STILL limited downtime

Technologies

- High Availability (cont.)
  - Active / Active
    - both nodes processing service(s) requests
  - Pros:
    - 100% service availability (unless BOTH nodes fail)
    - load-balancing effect on processing
  - Cons:
    - requires N+1 nodes to achieve
    - at least 1.5 times more expensive than Active / Passive configuration

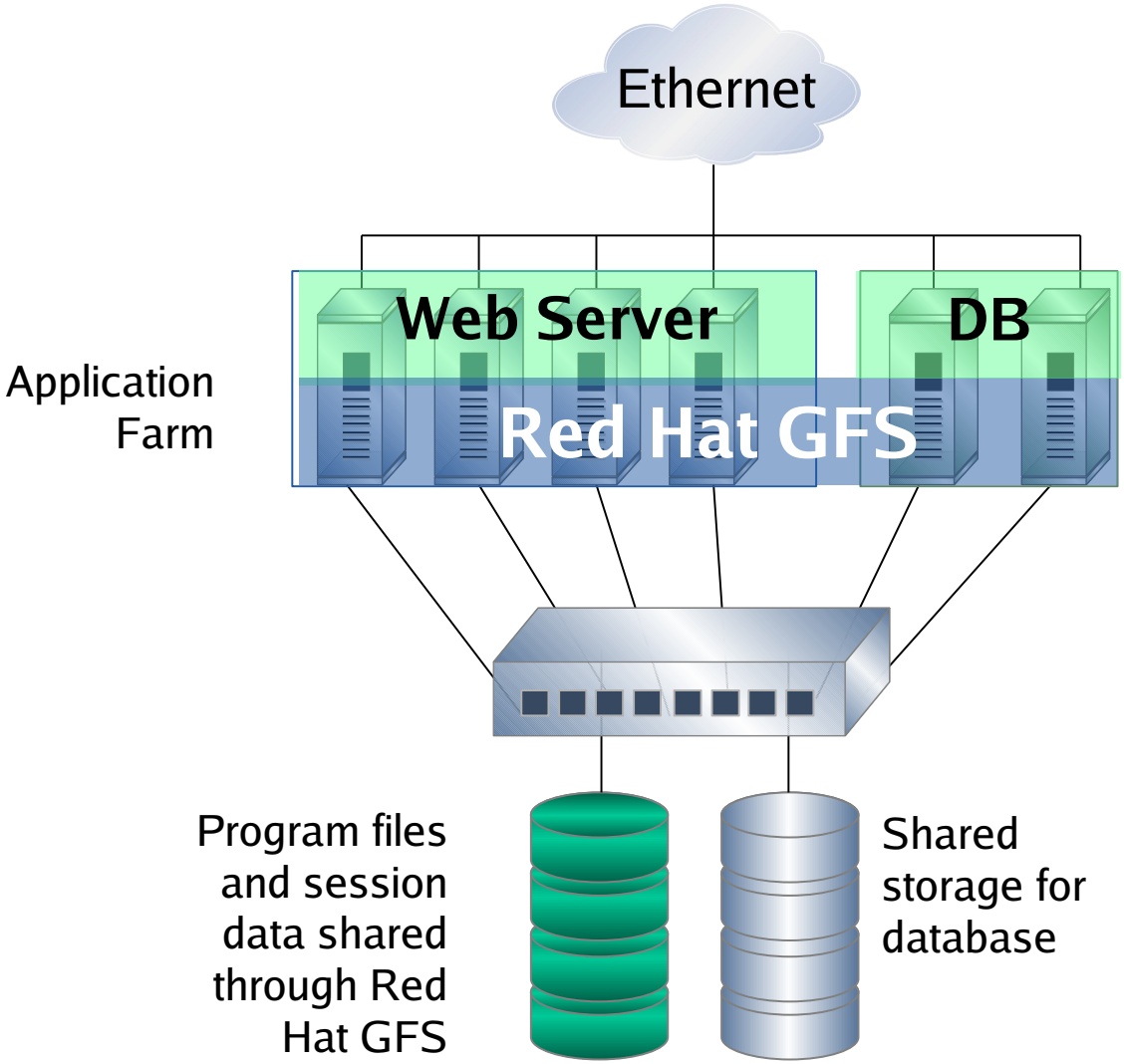
Simple Cluster Configuration

- 2 nodes
  - heartbeat connection
    - allows monitoring
  - shared storage
    - for common configs & data
  - individual network interfaces
    - allows IP Address takeover



redhat

Example of load-balanced service(s)



- Clients connect to load-balanced applications, which use shared storage back-end.

redhat

Technologies

- High Performance Computing
  - many nodes providing processing power
  - Pros:
    - cost-efficient when compared to mainframe technologies
    - balances load amongst many CPUs
    - example: Beowulf, OpenMosix
      - harnesses “idle” CPUs
  - Cons:
    - highly sensitive to network interruptions
    - requires extensive setup / configuration time

Clustering Filesystems

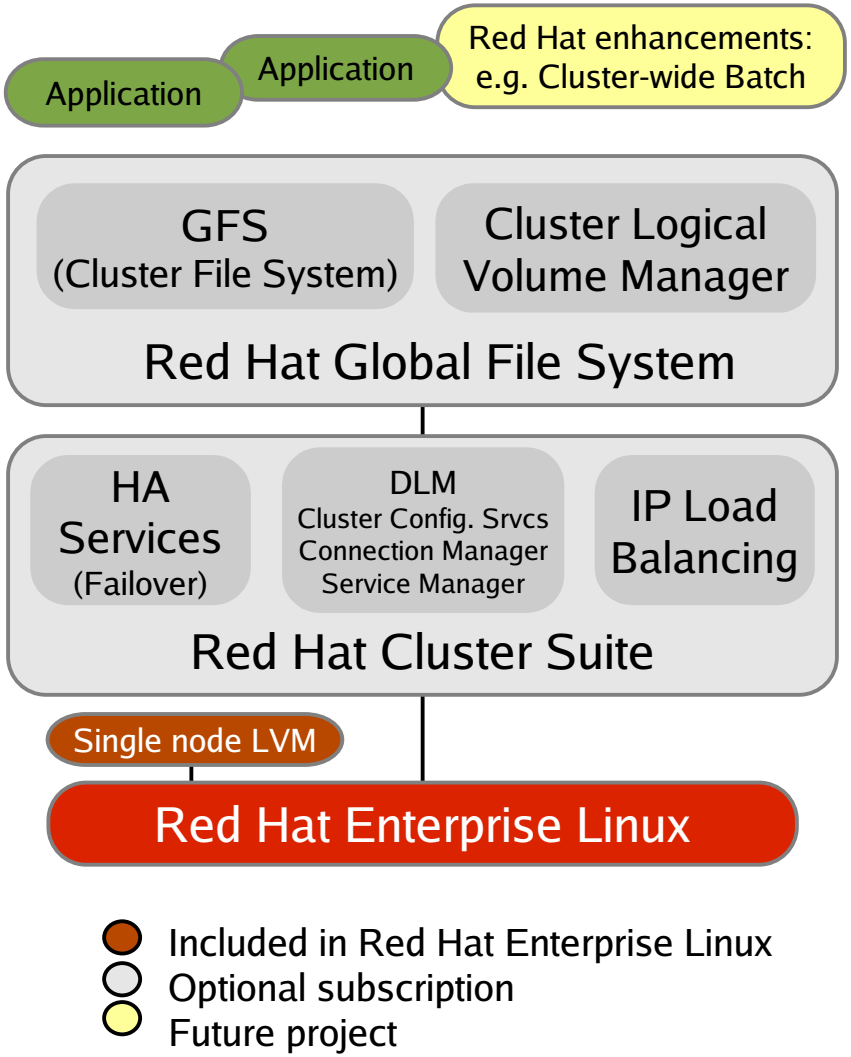
- Many cluster-aware services require concurrent filesystem access
  - Example: load-balanced NFS Servers
    - needs to handle multiple reads & writes to the SAME file
  - How to handle concurrency issues ?

Clustering Filesystems

- Red Hat's GFS
  - uses a lock manager to handle concurrency
- LustreFS
  - object storage device shares objects, not blocks
  - still considered highly experimental
- CODA FS
  - distributed filesystem, using concept of replication amongst nodes
  - allows disconnected access to filesystem resources

Red Hat Clustering Solutions

- Red Hat Global File System provides a fully open-source cluster file system that offers cluster-wide concurrent read-write file system access
  - Improves availability, scalability and performance
  - Recommended for medium-large configurations
  - File-level concurrency is provided by the application
  - Includes cluster logical volume manager
- Red Hat Cluster Suite provides application failover
  - Improves availability
  - Recommended for small configurations
  - Available separately; included in Red Hat



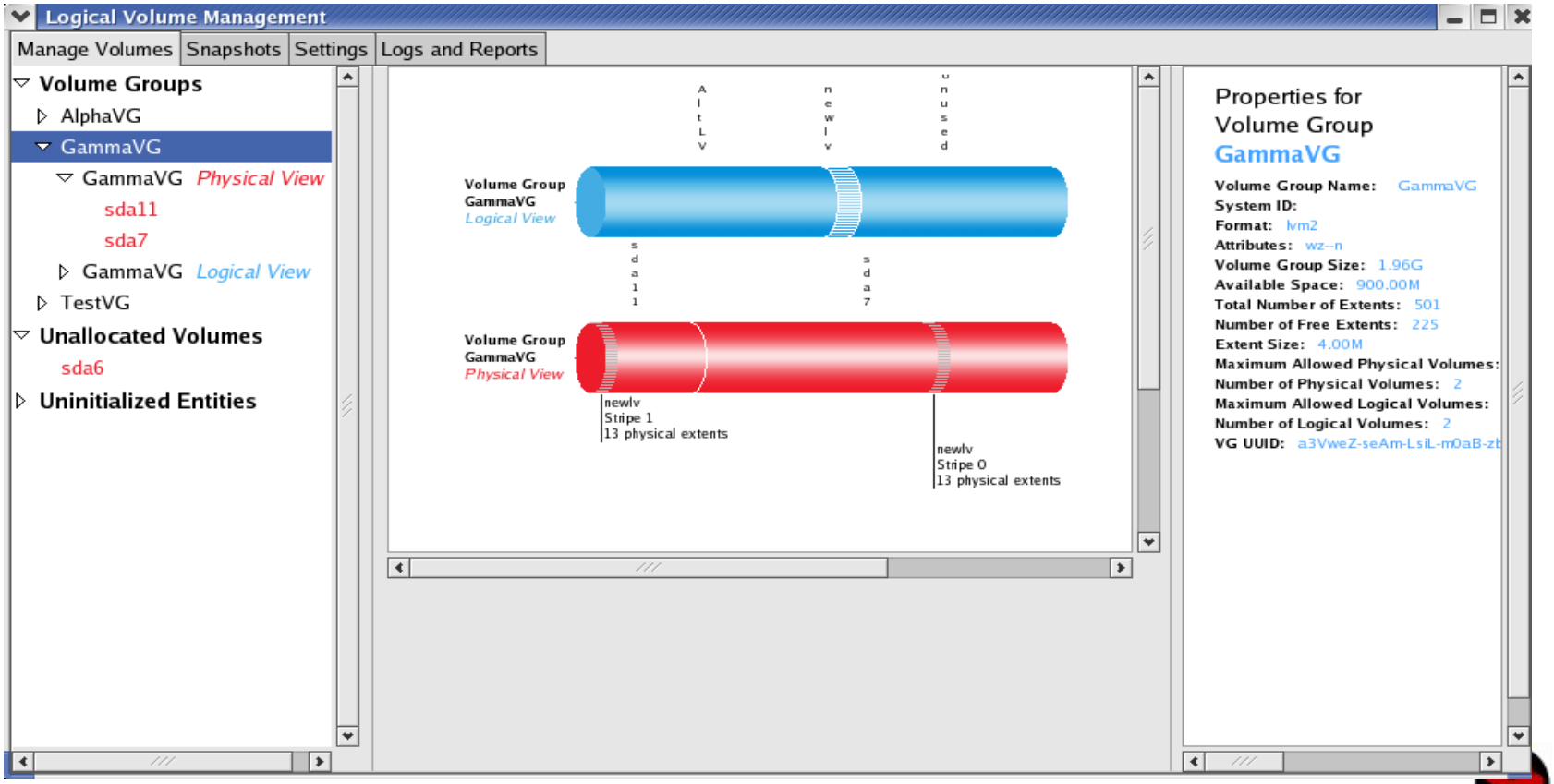
redhat

Volume Managers

- Handle the job of allocating shared storage
  - need to be cluster-aware
  - need to handle volumes for multiple nodes
- Examples include
  - Red Hat's LVM2
    - aka CLVM
  - IBM's EVMS
    - not seen much in Enterprise clients
    - cost a factor ?

Cluster LVM

- CLVM2 brings ease of administration
  - dynamically grow / shrink filesystems



redhat

Red Hat Global File System v.6.1

- New version for Red Hat Enterprise Linux 4
- Provides two major technologies
  - GFS cluster file system – concurrent file system access for database, web serving, NFS file serving, HPC, etc. environments
  - CLVM cluster logical volume manager
- Distributed Lock Manager
- Data and meta-data journaling (per-node journals, clusterwide recovery)
- Maximum filesize & file system size: 16TB with 32-bit systems, 8EB with 64-bit systems
- Supports file system expansion
- Requires shared storage
  - Supports several topologies: SCSI, SAN, iSCSI, GNBD

redhat

Red Hat GFS File Services

- Scalable: Architecture scales to hundreds of servers; supported up to 300 servers
- Robust: 120+ proven production deployments
  - Sectors: Financial Services, Automotive, Oil & Gas, EDA
  - Applications: Oracle 9iRAC, NFS, Web/Application Server, SAP, Custom
- Technology Differentiators:
  - High Availability:
    - Multi-Journaling and Distributed Metadata
    - Multi-Path support
    - Multi-Fence
    - OmniLock Architecture
  - Manageability:
    - Quotas
    - Online re-Sizing
    - POSIX compliance
  - Performance:
    - Direct I/O
    - Data Journaling
    - Deferred Locking

redhat

Questions

redhat