Enabling Efficient and Reliable Transition from Replication to Erasure Coding for Clustered File Systems

ABSTRACT

Clustered file systems (CFSes) ensure data availability by striping data with redundancy across different nodes in different racks. Two redundancy schemes are commonly used: (i) *replication*, which creates identical replicas for each data block, and (ii) *erasure coding*, which transforms original data blocks into an expanded set of encoded blocks, such that any subset with a sufficient number of encoded blocks can reconstruct the original data blocks. Replication improves read performance as it can load-balance read requests across multiple replicas. On the other hand, erasure coding improves storage efficiency with much less storage redundancy, while it achieves the same or even higher fault tolerance than replication based on reliability analysis.

EXISTING SYSTEM

In Existing System, Asynchronous encoding maintains high read performance for new data via replication, while minimizing storage over head for old data via erasure coding. It also simplifies deployment and error handling, and hides performance degradation. To balance performance and storage efficiency, modern clustered file systems often first store data with replication, followed by encoding the replicated data with erasure coding. We argue that the commonly used random replication does not take into account erasure coding in its design, thereby raising both performance and availability issues in the subsequent encoding operation.

DIS ADVANTAGES

- It does not avoid relocation of encoded blocks after the encoding operation.
- It does not protect against either a two-node failure or a single-rack failure.

PROPOSED SYSTEM

In Proposed System, Encoding-Aware replication, which carefully places the replicas so as to

(i) Eliminate cross-rack downloads of data blocks during the encoding operation.

(ii) Preserve availability without data relocation after the encoding operation.

(iii) Maintain load balancing across replicas as in random replication before the encoding operation.

We conduct extensive HDFS-based test bed experiments and discrete-event simulations, and demonstrate the performance gains of encoding-aware replication over random replication.

ADVANTAGES

- It improves storage efficiency with much less storage redundancy.
- Simplifies deployment and error handling, and hides performance degradation.

SYSTEM REQUIREMENTS

H/W System Configuration:-

Processor	-	Pentium –III
RAM	-	256 MB (min)
Hard Disk	-	20 GB
Key Board	-	Standard Windows Keyboard
Mouse	-	Two or Three Button Mouse
Monitor	-	SVGA

S/W System Configuration:-

Operating System	:	Windows95/98/2000/XP
Application Server	:	Tomcat5.0/6.X
Front End	:	HTML, Jsp
Scripts	:	JavaScript.
Server side Script	:	Java Server Pages.
Database	:	MySQL 5.0

Database Connectivity : JDBC