Fault-Tolerant Storage Servers for the Databases of Redundant Web Servers in a Computing Grid

Minhwan Ok

Fault-Tolerant Storage Servers for the Databases of Redundant Web Servers in a Computing Grid

MinHwan Ok

Korea Railroad Research Institute, Woulam, Uiwang, Gyeonggi, Korea
mhok@krri.re.kr

Abstract. Computing Grid in this paper is a Grid computing environment that supplies applications which run in a local computing site only, without any modification or adaptation for running globally in the Grid computing environment. Each stage of a running application is transcribed at all the management databases coupled with respective Web servers. The consistency is maintained by double-checking of every acknowledgement against a write to all the management databases and a circulated read response from either database. The storage spaces could be integrated into a single one by storage managers within a computing site. The modification of a file is broadcast to the storage managers sharing the storage space and their allocation tables are updated immediately. The system architecture is in a distributed control type, potentially the best match for Cloud computing.

Keywords: Scalable Web service, computing Grid, Fault-tolerant, Storage virtualization.

1 Constructing the Computing Grid

In the system model, the applications are provided to the users by Web service. A client computer connects to the Web server of the coordinator. Coordination Service is composed of user interfaces to log-on the computing Grid and to input parameters with user input/output data transfer. DB Organizer is the manager of information including parameters input, software title selected, and details concerned with user ID. It also selects appropriate computing site for the user. DB Connector is a client of DB Organizer and read/write information/report from/to the management database. Applications are launched, controlled and landed through Application Manager, which would be a kind of RFB Service. In Fig. 1, Application Manager delivers the commands included in the order, which is received via DB Connector, to the Application. When transferring the output data, Application Server should be re-authenticated for the security of user data.

On writing information to the management DB of the originator coordinator, the originator writes the same information to the management DBs of the other coordinator, if one or more coordinator does not respond to the writing, the coordinator is presumed crashed-down. This is broadcast to the remained coordinators. On reading

information from the management DB of the originator coordinator, the originator initiates reading from the management DB of the next coordinator, after reading the information from the its own management DB. The reading is relayed returning to the originator, thus it is named *Circulated Read Request*. If one or more coordinator does not respond to this reading, the coordinator is presumed crashed-down and this is also broadcast to the remained coordinators.

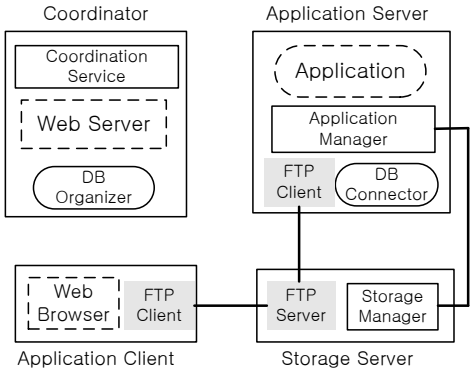

Fig. 1. System architecture of the computing Grid

Fig. 2 illustrates the circulated read request/response, the broadcast write is illustrated in Fig. 3. In both information writing and information circulated reading, the originator creates *Replay Roll* if it detects any crashed-down coordinator. Replay roll is the list of DB transactions from the point the crash-down is detected to the point the crashed-down coordinator broadcasts its restart. Then the coordinator updates its management DB following the replay roll the originator has sent. Keeping up the recorded information identical is the major issue in this follow-up scenario.

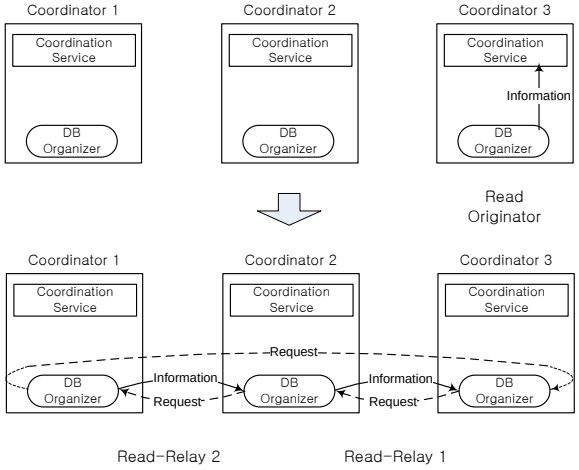

Fig. 2. Information reading from either management DB is relayed, which will be circulated to the originator and the originator responds with the information last.

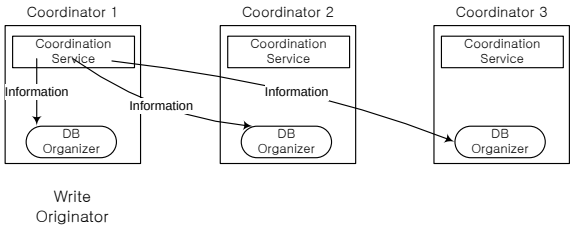

Fig. 3. Information writing to the management DB of the other coordinators are acknowledged later and the coordination service continues without these acknowledgements.

2 Multiple Storage Servers

Ancillary to the selected one among local computing sites the software installed, one of storage servers is designated for sizable storage space to process large quantities of data with the storage farm. The storage space for application running is confined within the application server in the previous work[1]. It is also confined in the application server in this work, except that the whole data is partitioned and the partitions are replaced to be processed in the application server in the manner similar to the virtual memory. The application has restarted in the case of the storage server failure in the previous work, however the application rolls back to the previous phase of the current partition in this work. For active/standby failover, the allocation table is mirrored to the other storage server. A couple of storage servers are assigned to backups of each other, and the storage manager has dual modes between active/standby in normal status and active/active in abnormal status, of the other server. Coupling two storage servers is static and conducted by the administrator.

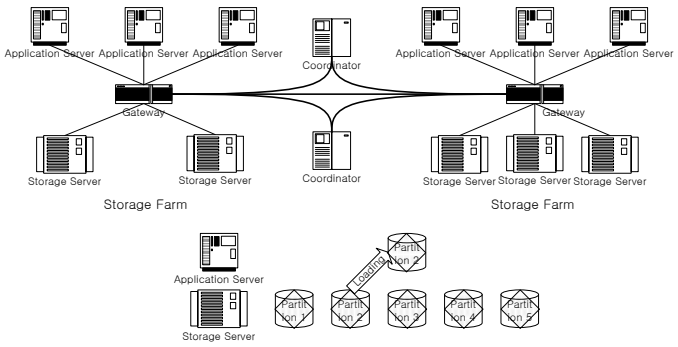

Fig. 4. Computing Grid Organization and storage management on partitioned data

Since multiple storage servers govern the storage farm, the write to a storage device is allowed while the storage manager has the token of the device that the token is traversing storage servers otherwise. Once the storage manager wrote to a storage device, it broadcasts allocation information of the written file to other storage managers so

that storage managers sharing the storage space would update their allocation tables. The storage devices could constitute one single storage space spanned from one device to another, for availability losing an advantage of parallelized access of striping. When one storage device fails, it is broadcast among the storage managers by a storage manager detected the fault. The storage space of the device is marked 'missing' at all the storage managers, analogous to bad blocks, and access to files located at the space is restricted from then. A storage manager should have the token of a device when it has to write to the device and it waits for the token. The investigation protocol is described in Fig. 5 for the case the token of a device is not returned before the failure of a storage manager.

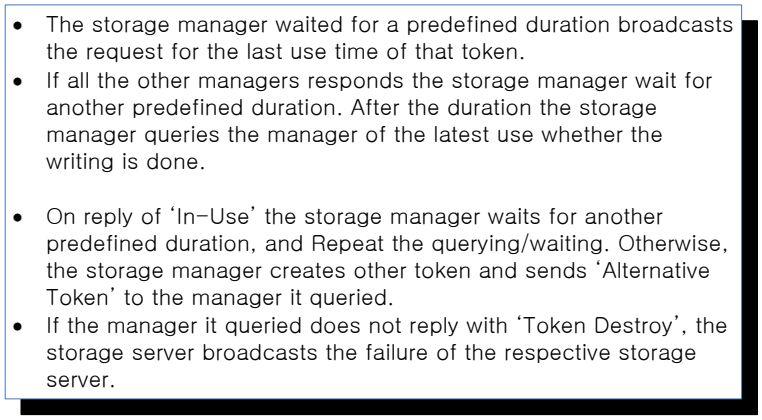

Fig. 5. Investigation into the storage manager which is queried whether its writing is done.

References

1. Ok, M.-H., Lee, K.-s.: A Consolidation Model of Web Application Servers toward a Simplified Computing Grid. International Conference on Multimedia and Ubiquitous Engineering, Seoul Korea, 757-761 (2007)