Multi-Server Based Data Management For Resource Namespace Service

Leitao Guo
China Mobile Research Institute
guoleitao@chinamobile.com
http://labs.chinamobile.com

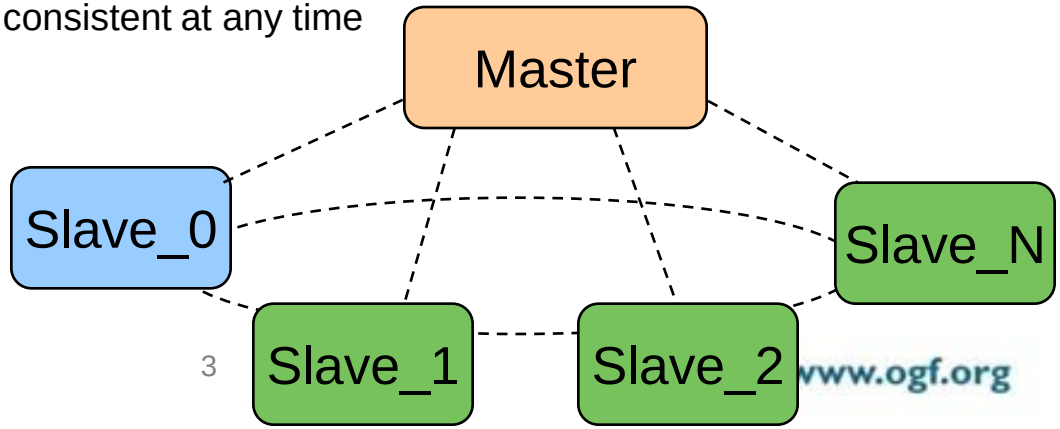
OGF IPR Policies Apply

- "I acknowledge that participation in this meeting is subject to the OGF Intellectual Property Policy."
- Intellectual Property Notices Note Well: All statements related to the activities of the OGF and addressed to the OGF are subject to all provisions of Appendix B of GFD-C.1, which grants to the OGF and its participants certain licenses and rights in such statements. Such statements include verbal statements in OGF meetings, as well as written and electronic communications made at any time or place, which are addressed to:
  - the OGF plenary session,
  - any OGF working group or portion thereof,
  - the OGF Board of Directors, the GFSG, or any member thereof on behalf of the OGF,
  - the ADCOM, or any member thereof on behalf of the ADCOM,
  - any OGF mailing list, including any group list, or any other list functioning under OGF auspices,
  - the OGF Editor or the document authoring and review process
- Statements made outside of a OGF meeting, mailing list or other function, that are clearly not intended to be input to an OGF activity, group or function, are not subject to these provisions.
- Excerpt from Appendix B of GFD-C.1: "Where the OGF knows of rights, or claimed rights, the OGF secretariat shall attempt to obtain from the claimant of such rights, a written assurance that upon approval by the GFSG of the relevant OGF document(s), any party will be able to obtain the right to implement, use and distribute the technology or works when implementing, using or distributing technology based upon the specific specification(s) under openly specified, reasonable, non-discriminatory terms. The working group or research group proposing the use of the technology with respect to which the proprietary rights are claimed may assist the OGF secretariat in this effort. The results of this procedure shall not affect advancement of document, except that the GFSG may defer approval where a delay may facilitate the obtaining of such assurances. The results will, however, be recorded by the OGF Secretariat, and made available. The GFSG may also direct that a summary of the results be included in any GFD published containing the specification."

- OGF Intellectual Property Policies are adapted from the IETF Intellectual Property Policies that support the Internet Standards Process.

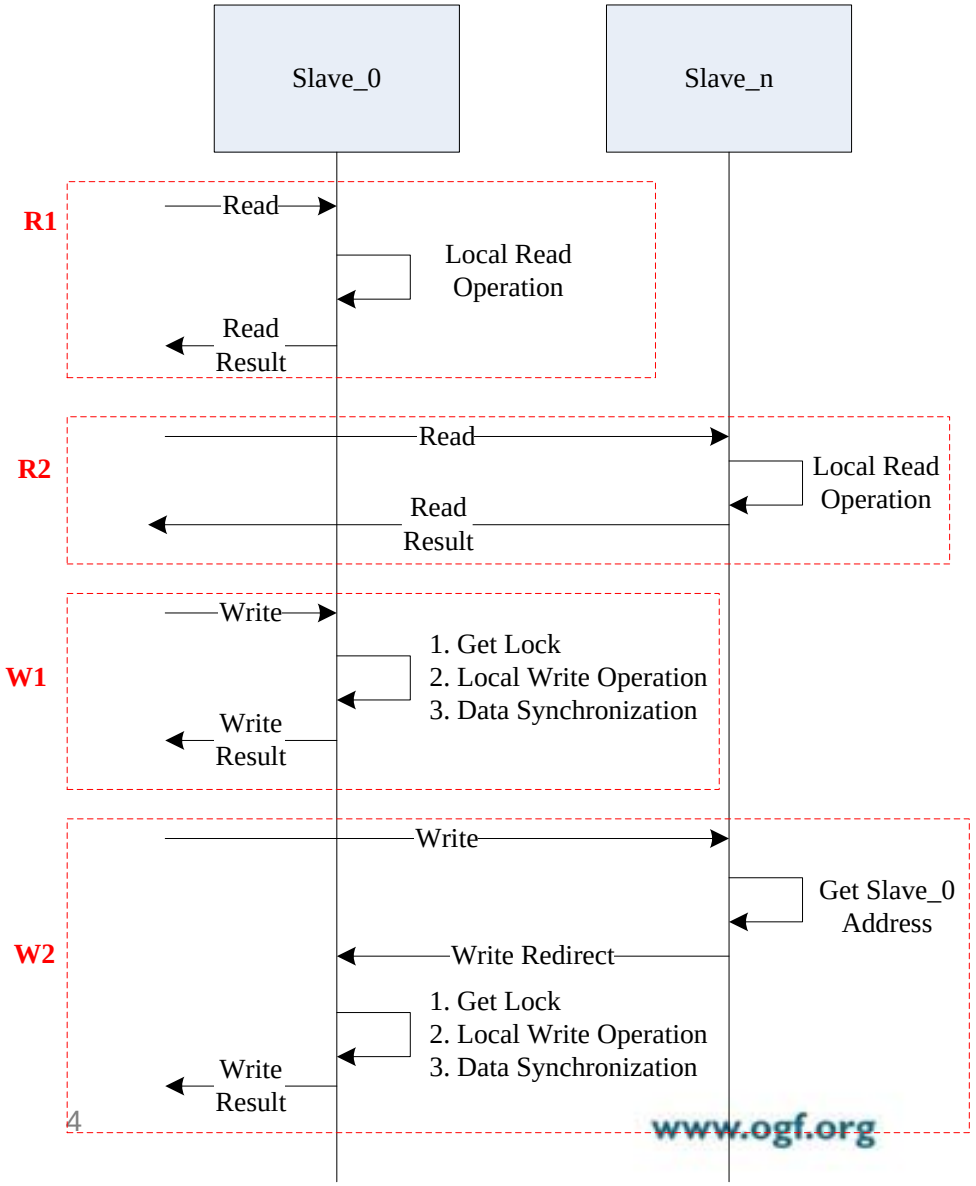
Overview

- Why we need a multi-server based data management for RNS
  - **Scalability**
    - RNS is a critical service in grid and distributed network, almost every operation on data will access this service, single-server based system will cause bottleneck problem
  - **Availability**
    - Namespace data are stored in the dynamic nodes, especially in grid environment, where nodes can come and go easily, so the QoS of RNS can not be guaranteed
- A glance of multi-server based data management
  - Master Node (1)
    - Provide the standard port types to the RNS requester
    - No namespace data are stored in this node
    - Request dispatcher: simply redirect every request to the slave nodes
  - Slave Node (N)
    - 2 kind of slave nodes: slave_0 and slave_* (1,2,3…)
    - Slave_0: response to the **write/read request**, such as *add, update and remove* functions
    - Slave_*: response to the **read request**, such as *list and query* functions
    - Data on each slave are kept consistent at any time



Process Flow of Port Type Provided by RNS

- 4 cases
  - R1: the read request is dispatched to slave_0 node
  - R2: the read request is dispatched to slave_* node
  - W1: the write request is dispatched to slave_0 node
  - W2: the write request is dispatched to slave_*



Join of a new node

- Master assigns a name/id to this new node
- Master asks one slave node to synchronize the namespace data to this newly joined node
- Master enable this node

Node Failure

- Slave_* node failure
  - Not affect the system availability, other slaves can also respond to the read request
- Slave_0 node failure
  - Choose another node to take over the failed slave_0 node
  - System availability will be affected during choosing another slave_0 node, but system will continues after that
- Master node failure
  - System will be unavailability during choosing a new master or restart the master
  - Namespace data will not lost

Conclusion

- Multi-Server based data management can improve the scalability
  - All slave nodes can respond to read request
  - Can dynamically add new slave node to improve the system performance
- Multi-Server based data management can enhance the availability
  - Master node failure will not lost the system data
  - System can also work when slave_* nodes failure
  - Slave_0 nodes failure can only cause a slight effect to the system availability

- What we are doing in CMRI
  - Developing a prototype of this multi-server based data management using open source software
    - Hadoop / Zookeeper
  - Planning to carry out an experiment on scalability and availability in typical scenarios
  - Enhance the system functions with the help of experts from OGF GFS-WG, and contribute our experiences to OGF

Full Copyright Notice

Thank You!

Leitao Guo
Project Manager
Service Support Department
China Mobile Research Institute

guoleitao@chinamobile.com