Paper review [1]

Yu SU 55323707

1. The problem

The paper proposed a design and implementation of Dynamo, which is a storage system of Amazon. The problem is this system needs to maintain reliability and scalability to provide an always-on service for Amazon. In detail, Amazon has so many customers, servers and data and new data and nodes are imported constantly, so there are strict operational requirements of performance, reliability, efficiency and scalablility on Dynamo.

2. Challenge

There are some challenges of the system to solve the problem. First, it is hard for get both the good consistency and high data availability simultaneously because the data replication algorithm will make the data unavailable when the failure happens. Even if using the optimistic replication techniques, there will be conflicting changes which must be detected and resolved. Second challenge is how to make sure when to perform the process of resolving update conflicts. To make the system is always available, it pushed the complexity of conflict resolution to the reads to make writes will not be rejected. Third, the choices are limited if conflict resolution is done by data store, so we need to figure out a more flexible policy.

3. key insight

Because the system needs to be scalable, the system provides a dynamic partitioning algorithm, which use a variant of consistent hashing with virtual node. The system could be both scalable and available under this algorithm.

Like other file system, Dynamo also use replication to achieve high availability and durability. What's more, for consistency, the system needs to handle the node asynchronously, so the system uses immutable version of data to handle the node in different state.

To guarantee the availability, the system has many new approaches to the failures. On the one hand, the system uses hinted handoff to make sure the read and write operations will not fail because the temporary node or network failure. On other hand, the system uses the ring membership to initiate the addition and removal of nodes from a Dynamo ring. Besides, it also has a decentralized failure detection protocol with a simple gossip-style protocol to make each node know the other nodes come or go.

4. Limitation

I think the main limitation of the system is the performance and flexibility problem. There are too much computation and communication in this system, like hashing, merkle tree calculation, failure detection and gossip protocol, So the system is not flexible, which needs some computation power and network bandwidth to support. And to achieve high scalability, with a lot of new node and data is imported, there will be a lot of workload of computation, which will maybe cause the performance problem.

5. Future work

The paper pays all the attention on the consistent availability of the system, which means it missed or sacrificed some other important properties. In the future, the system should also focus on the performance problem and security problem to make the system is more perfect.

[1] Decandia, G. Hastorun, D. Jampani, M. Kakulapati, G. Lakshman, (2007). Dynamo. *ACM SIGOPS Operating Systems Review, 41*(6), 205. doi:10.1145/1323293.1294281