NetId1: yaoxiao9 NetId2: xiaoxin2

Group81-Project Report
CS425-MP3

Design

1. UDP-based connection used for memberlist update, node operation(join, leave,) and failure detection; 2. TCP-based connection used for grep command; 3. a simple distributed file system: within the SDFS system, commands “put”, “get”, “ls”, “store” are supported, and the master is responsible for storing all the file distributing information and handle the system logic. The file transformation is based on “ssh” command and the message transfer is based on TCP connection.

Function

1. memberlist update, node operation and failure detection
Using MP2, based on UDP connection.
2. put
When the user inputs a “put” command on a slave node, firstly, this slave node builds a TCP connection with the master; secondly, master returns response about where to put there replicas to the slave node; thirdly, this requester slave node builds TCP connections with all the replicas and transfers the file to them. The put information is along with a file version, so that when a file update is brought up, we could update the file version and that we could ensure to get the latest version of the file.
3. get
When the user inputs a “get” command on a slave node, firstly, this salve node builds a TCP connection with the master; secondly, master gets all the replica nodes and check the version of each of them, if there is any replica that is not up-to-date, master would repair them, and then return the replica nodes to the slave node; thirdly, the slave gets the latest version to its local system
4. ls
When the user inputs a “ls” command on a slave node, firstly, the slave node builds a TCP connection with the master, and the master would return a response about where the target file is stored; secondly, the slave node would list all the locations
5. store
When the user inputs a “ls” command on a slave node, the slave node just looks up its own structure body about which local files it stores and list them
6.delete
When the user inputs a “delete” command on a slave node, firstly, this salve node builds a TCP connection with the master; secondly, the master get where this target file is stored and ask all the replica nodes to delete the target file

Algorithm

1. master re-election
We use bully algorithm, when a master is down, a vote would be initiated, and the node with the highest vote would be selected to be the new master
2. replication strategy
Because we should tolerate up to 3 failures, we should store all the files to 4 replicas so that there is at least one alive node that stores the target file. And when some of the replicas are

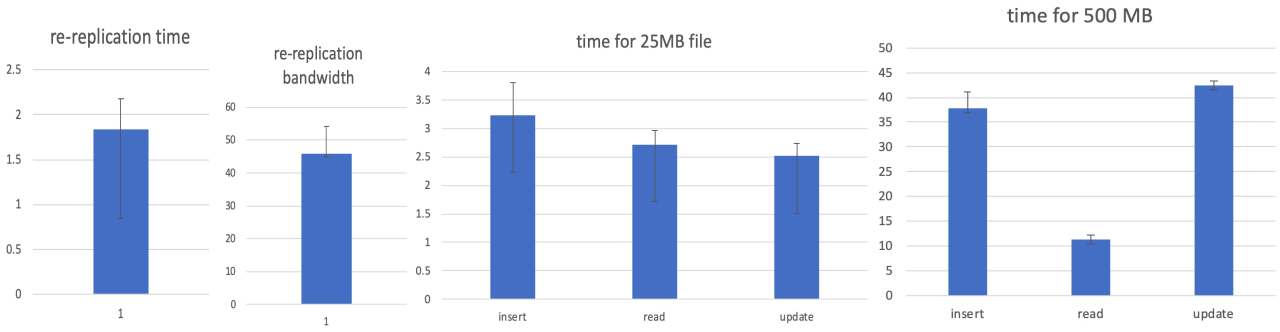
failed, the failure detection system would catch this and update the replica location, then a re-replication and a new replication list would be initiated.
3. We use quorum algorithm to calculate a number to decide when we should return the put/get ACK.

Using of MP1

Once an operation is done, there would be a corresponding log information initiated and written into the log file. So we could use the grep function in MP1 to do the grep pattern matching to see whether the operations meet our expectations.

Measurement

caculated by seconds									
		1	2	3	4	5	6	average	std
re-replication time upon a failure	time	1.57	1.69	2.43	1.56	1.8	2	1.8416667	0.331386
re-replication bandwidth upon a failure	bandwidth	53.3	47.9	30.4	50.9	45.1	48	45.933333	8.1123774
time of 25MB	insert	2.63	2.87	4.21	3.15	3.52	3.01	3.2316667	0.5637168
	read	2.57	2.74	2.81	2.41	3.13	2.62	2.7133333	0.2469548
	update	2.61	2.44	2.47	2.84	2.15	2.56	2.5116667	0.2269288
time of 500MB	insert	33.11	35.65	40.21	37.43	38.34	42.13	37.811667	3.2125779
	read	10.43	11.34	12.64	10.54	10.94	11.76	11.275	0.8329646
	update	43.21	41.23	43.45	42.12	42.65	42.03	42.448333	0.8235634
time to store Wikipedia corpus	4 machines	53.12	52.78	54.28	54.02	53.63	54.37	53.7	0.644267
	8 machines	51.52	52.53	53.03	53.26	52.39	51.64	52.395	0.707863



Analysis

1. Time of insert and update are close, and time of read is slightly less than them and it is increasingly significant when the file size increases. This meets our expect, since when read, we only need to write to 1 file, but when insert and update, we need to write to 4 replicas(ACK by 3 replicas).
2. When file size increases, the operation time increases, which meets our expect.
3. There is no significant difference between the time of 4 machines and 8 machines, which meets our expect, since the time is associated with the number of replicas instead of the number of machines.