Estimation of availability and reliability in CurveBS

CurveBS uses the RAFT protocol to maintain consistency of stored data. It generally takes the form of 3 replicas of data. If one replica fails, the system can read and write data successfully on the other two replicas. When two copies fail at the same time, the system cannot determine the failure reason because only minor number replica is written successfully. Therefore, manual intervention is required to handle the failure according to the actual situation of the system.

Estimation of availability and reliability in the three-replicas case

Assume that the total number of disks in Curve system is N, the number of replicas is R, and the data recovery Time in the case of failed disks is T. The Annual Failure Rate of disks is AFR, and the average running Time of disks is MTBF (Mean Time Before Failure).

$$AFR = \frac{1}{MTBF/\ (24*365)} * 100$$

In CurveBS, the data on one disk is distributed in about 50 copysets, which means when a disk fails, up to 50 other disks will restore the data on that disk at the same time. According to this, the data recovery time T can be estimated.

$$FIT = \frac{AFR}{24*365}$$

The probability of failure of disk within a year t is:

$$P_1 = 1 - \ e^{-FIT*N*t}$$

Curve uses zones to isolate faulty areas to prevent data loss caused by server problems. In addition, each disk of Curve is distributed on an average of 50 Copysets. Therefore, the probability of failure of another disk related to data on the previous disk is as follows:

$$P_2 = \left(1 - \ e^{-FIT*(N-1)*tr}\right) * \frac{1}{3} * \frac{50}{\frac{N}{3}}$$

The probability of the failure of the third disk in tr during the recovery time is as follows

$$P_3 = \left(1 - \ e^{-FIT*(N\ -\ 2)*tr}\right) * \frac{1}{3} * \frac{50 * 3}{N}$$

The annual probability of cluster data loss is as follows:

$$P = (P_1 * \ P_2 * \ P_3)$$

The probability of no data loss in a cluster is as follows:

$$1 - P$$

Assume that in a CurveBS cluster consisting of 1200 disks, MTBF of each disk is 1.2 million hours, and the data recovery time is 5 minutes, i.e. 0.083 hours. Therefore, the probability of no data loss in the annual of the CurveBS cluster is P = 0.999999781