

This storage systems  
and database  
functionality to create



**ASSIGN  
BUSTER**

This paper talks about various challenges faced during backup and recovery of NoSQL databases. Since these databases have eventual semantics consistency they cannot use the same backup and recovery methodology used by traditional RDBMS databases. Some of the challenges faced are topology oblivious backup and restore, cluster-wide consistent backup which should result in repair free recovery and removal of redundant copy from backup. This paper presents a solution called BARNS for two NoSQL databases Cassandra and MongoDB.

It addresses those challenges without affecting the performance of NoSQL databases. BARNS is able to achieve the desired results by using light-weight snapshots and clones features of the modern storage systems and database functionality to create cluster consistent backups and most of the recovery related work is done during the backup. On evaluation of BARNS to backup and restore Casandra's four node cluster with data set increasing by 4GB it was observed that first phase took constant time but second phase time increased by 35-40% with every 4GB increase. Recovery time on the other hand was around 60-80 seconds irrespective of data sets. For MongoDB's nine node cluster evaluation it was observed that first phase took around 10 seconds, second phase took around 2.5 minutes and recovery too took around 2.

5 minutes.

The most interesting part of the paper is the technique used for implementing the solution BARNS. Unlike other related solutions which required developers to write custom scripts for backup and restore or copy data outside of the clusters, the BARNS solutions use the built-in database commands and features or shared storage features to solve

<https://assignbuster.com/this-storage-systems-and-database-functionality-to-create/>

above mentioned challenges. Also, since the solution doesn't require copying of data outside the clusters it can also perform backup and recovery on encrypted databases and it's compatible with other NoSQL databases (using their commands and API's).

The second phase (post-processing phase) during Casandra's backup using BARNS is CPU and memory intensive since it performs backup of all the clusters at once. This can be improved if instead of above approach the backup is done using an incremental approach.