

1. Can the tar.bak files be deleted afterwards?
 - a. The .bak files will be deleted automatically if TarMK is able to recover successfully without any data loss. Otherwise, they will be left where they are for further inspection. An occurrence of a failure in the recovery process, though, is very rare.
2. What is the advantage of TarMK over Mongo, which is the more popular NoSQL Database.
 - a. better throughput due to local memory mapped files.
TarMK is for best performance. See also the decision matrix for persistence technology in AEM: <https://docs.adobe.com/docs/en/aem/6-2/deploy/recommended-deploys.html#Microkernels:%20which%20one%20to%20use>
3. What problem does JCR solve in content management space which a popular NoSQL can't?
 - a. JCR offers a hierarchical model that is handy if your data can be abstracted this way. As a practical - and successful - example, see how AEM maps pages, folders and scripts in a comprehensive content structure.
4. what can be done to prevent segment not found exception? and why does it occur
 - a. A SegmentNotFoundException is usually thrown in very exceptional cases. It is difficult to provide a quick-and-dirty recipe to answer your question. Some of the tools described by Michael right now are useful to detect the root cause of these error conditions.
5. I've never seen a .bak file deleted automatically. What process should do this? Also with Oak 1.4, I'm now seeing .bak.ro files in addition to .bak files. What are these?
 - a. there is no auto delete of the .bak files. you are expected to clean those up from time to time. .bak.ro fall into the same category, are created by the read-only tooling like oak-console
6. Every add or update in any node requires entire parent chain to be recreated (which involves creating references to child node in the old tree). What advantage does this approach has? And, how does that compensate for the loss of efficiency (in creating new nodes)?
 - a. MVCC allows a user to have a consistent view on the repository while other users are modifying the content tree. This is an important property offered by data systems and is technically called "repeatable reads". MVCC also has the advantage of coping with concurrent writes in a graceful, non-locking fashion.
7. is oak run check an online or offline function
 - a. I think it's online only on 1.4, but can't say for sure
8. Does oak-run check make any changes (fix inconsistencies)?
 - a. not automatically, it will only point to the good revision. manual intervention is needed to rollback

9. If I have repository ,can I toggle it between standby and primary once in 3 months , Is it recommended to do , will there be any data loss deinf this time ?
 - a. Your question is very much relevant to Cold Standby, but it is not really relevant to this Gem Session. If you have technical questions, address them to oak-dev@jackrabbit.apache.org. Oak's contributors will be happy to help. You can post your question in our help forum: <https://helpx.adobe.com/marketing-cloud/experience-manager.html>
10. with Oak run tool does journal.log read bottom up or top down?
 - a. it will read it bottom up, in chronological order or revisions
11. Before rollout to correct revision, is it possible to check when it was created to know how many changes will be lost?
 - a. you can grep the tar files entries to search for the uuid, and look at the timestamp to figure out the date. latest oak (1.6) has the timestamp info in the journal, but not earlier versions
12. Is there a thumb-rule/best practice to tell how often the offline compaction should be used (eg. once per week)?
 - a. The answer depends a lot on the volume of traffic that you are seeing. More traffic might mean more changes, which means more garbage. You should monitor your instance and act accordingly until you figure out what's your sweet spot.
13. How can I estimate the runtime of Offline Revision GC? And is there a progress bar/percentage done when running GC?
 - a. offline revision GC depends a lot on the avail resources, so you need to establish a baseline on your own. there are logs available to track progress per 10k nodes compacted you might want to enable to have a better idea of progress
14. Is there way to view what is in segment?
 - a. A segment is a very dense binary file. The Oak command line utilities (oak-run) contains a Segment Explorer, a GUI application that helps in exploring your content tree and the content of segments.
15. With MVCC , will not there be locks on any node at all ? or will there be any ? How does it handle two jcr sessions modifying same node at property level, assuming two sessions are modifying the same property and trying to save the node?
 - a. depending on the type of changes, there is a strategy to deal with overlapping commits. one of the session's save call might fail if there are conflicts that cannot be dealt with automatically
16. Does GC actually improve performance of the underlying instance?
 - a. yes. GC reduces the working set, needed memory, improves locality and so on
17. How much free disk space we should have to run offline compaction ?
 - a. As a rule of thumb, you should have enough disk space to contain a copy of your existing repository. The amount of needed disk space might be considerably lower if you use an external BlobStore/DataStore.

18. How to evaluate the resource use required by online tar compaction on a certain AEM instance? How best to detect any excessive resource contention such as on memory and locks by online tar compaction? Suppose thread dumps are useful to some extent in this regards
 - a. The current major release of Oak (1.6 and later) contains safeguards that prevents online compaction from using too many resources like memory and disk space.
19. How do revisions relate to checkpoints? Why does oak-run suggest we remove checkpoints?
 - a. checkpoints are pointers to old revisions. if the checkpoints are very old, revision GC cannot clear those old references so it will not be as effective as it could be, hence the recommendation to remove all unreferenced checkpoints
20. Is it recommended to take segmentore backup regularly or Do we have ample amount of options in oak-run tool for recovery options ?
 - a. backups are mandatory! please schedule those at all times
21. During online compaction, when the previous generation tars are no longer used, are they purged from the memory mapped files as well?
 - a. Memory mapped files are a strange beast on the JVM. From TarMK's point of view, TAR files will be closed and rewritten to remove unused data. The old copy will be removed from the file system, which should be enough to instruct the OS to remove unused data from memory mapping pages.
22. in AMS, we use m4.large instances as default, which is 8 GB RAM
 - a. Regarding resource requirements the general AEM recommendations depending on your actual use case are valid. For example, if you have an Assets heavy scenario, I recommend to look here: <https://docs.adobe.com/docs/en/aem/6-2/deploy/configuring/performance/assets-performance-sizing.html>
23. what is the difference between :async@async and async@fulltext-async?
 - a. this is related to indexing, those are the 2 different indexing lanes. for more info please reach out to the mailing list.
I think you are referring to async indexing, but the TarMK is not about indexing, but storage! Please visit <http://jackrabbit.apache.org/oak/docs/index.html> to access documentation relevant to indexing in general and async indexing in particular.
24. Are the binary tar files on disk locked during access, write, and especially during online compaction?
 - a. Files on disk, once written, are read only. They might be rewritten, but the rewritten files are new files with a different generation. The generation I'm talking about is the letter after the serial number, e.g. the "c" in data0001c.tar.
25. what classes can isolated so a logger can be created for monitoring online tar compaction?
 - a. The class to monitor is org.apache.jackrabbit.oak.segment.file.FileStore for garbage collection and general TarMK operations.

26. how can we assure BlobGC runs after Online Tar compaction and completes before then next Online Tar compaction is triggered the next scheduled time ?
 - a. you need make sure that maintenance windows as configured in the Maintenance Tasks UI (part of Operations Dashboard) do not overlap and have a big enough gap between DSGC and RevGC.

27. What are the best/preferred backup options for AEM repository?
 - a. Please post this question on: <https://helpx.adobe.com/marketing-cloud/experience-manager.html>

28. does the compaction happen on segmentstore or both datastore and segmentstore
 - a. compaction does not touch the datastore. Datastore has a dedicated GC process

29. Which one is recommended in terms of performance for compaction process - single segmentstore having both segments and datastore blobs OR having separate datastore and segmentstore will give better performance?
 - a. the recommended AEM setup is now: segmentstore with an external datastore