Q & A's from AEM GEM session Jan 20th, 2016 - Oak Lucene Indexes

1. Q: For precise index selection is it advised always use native query?
    * A: if you can't manage to have the query engine select a specific index based on cost tweaks (recommended way), native query is your only alternative

2. Q: Can mixins be used in the place of nodeType?
    * A: yes

3. Q: any plans to support compactVersion =2 to support acl for child node.
    * A: acl are already checked by the query engine so you are not able to access nodes outside of acls, if this is about aggregation, please clarify

4. Q: Does the properties order in index definition match with exact query?
    * A: it doesn't need to, no

5. Q: Oak index config allows for regular expressions for property names. If the properties encompassed by that regex are of different types (like Date, String, Long) are those properties orderable
    * A: yes. However use regex support with caution and create explicit property definition node for properties you are interested in.

    It's provided to simplify global fulltext index definition where you specify regex to index all and have specific property definition to exclude certain properties from getting indexed

6. Q: Why comment is not propertyIndex?
    * A: could you clarify the question? what is 'comment'?

7. Q: Is it true that when you want to query on multiple paths, the query would be run against / and then the result would be filtered for the paths searched ?
    * A: yes, you cannot have a query constraint on multiple paths, you'd need to use node types for example

8. Q: In Chetan's example there were a property called comment' for fulltext?
    * A: Yes that much is sufficient for fulltext clause to be evaluated

9. Q: What's the best way of calculating how long an "ORDER BY" operation will take?
    * A: it depends on the dataset size, other than that, I'm not sure what you'd look for

10. Q: Thanks Chetan & team for making this great session
    * A: :)

11. Q: Regarding compactVersion I was referring to document which says "only the access right of the parent (aggregate) node is checked, and the access right of the child nodes is not checked"
    * A: yes, in the case of aggregation, so you'd have a sort of leak where you'd try to guess-query for a node based on properties of its child nodes that you don't have access to, not a very common case

12. Q: Is Suggestions feature available in AEM 6.1 ? If not is there a plan to port it back? And by when?
    * A: It's made more feature rich in upcoming 6.2 release. Suggestion feature in 6.1 was bit limited.

13. Q: For tools, I've also found the Oak-run console utility to export the index and analyze with Luke lucene tool to get an visualization to be useful
    * A: will be covered later in the presentation

14. Q: For query diagnostic tool enhancement should we log issue with daycare OR with acs?
    * A: As ACS has a faster release cycle you may want to put it on ACS tracker first

15. Q: is there a way to know from jmx the heap used by index ?
    * A: Index files are read as memory mapped files so they would not consume much space on heap

16. Q: Is the lucene index binary stored in the external DS when an external DS is configured?
    * A: Yes

17. Q: For Textextraction mbean is to possible to display how much time it took to extract
    * A: It displays overall time spent in text extraction

18. Q: Is there a way to fasten the reindex process?
    * A: Reindexing process has got lot better and we look into way of improving it more. One approach that has been found to be useful is to make use of 'includedPaths' to provide scope of what all the indexer has to look into. Refer to the docs for more details as it needs to be used with caution

19. Q: Does Luke provide capability to check index consistency and regenerate index for selected content files only?
    * A: No. It's only meant to get insight into how index looks like. With Oak it cannot be used to repair the index files

20. Q: How to match the doc id in luke with AEM asset ?
    * A: Each lucene document has a stored property ':path' which serves as mapping between Lucene view and JCR view of repository

21. Q: Does Luke handle searching indexed Date properties well?
    * A: Luke is meant to diagnose index data. Query is still performed within Oak outside of Luke

22. Q: Does luke respect ACL of files?
    * A: No. Luke need access to raw lucene index files. It's meant as a diagnostic tool

23. Q: Are there real world experiences on how much indexing could be delayed when a system is on heavy load - Seconds, Minutes, Hours? Is there any limit in a way that an indexing is forced after a certain amount of time?
    * A: In most customer setup and our test we have seen Async indexing to take ~15-20 secs under average load. Unless there are lots of new binary files for which text extraction is to be performed it should run fast

24. Q: Regarding reindex process we want to avoid much down time in production. In aem 5x we use to index in other nodes & copy the index file. I was looking more in this term how to fasten the reindex process?
>   * A: Note that while reindexing is in progress that older copy of index is still in use but yes reindexing would put some load. Would be better if you can put up this requirement on oak-dev and we can discuss this there

25. Q: Why is it preferable to create the index node in the path to be limited to versus using includedPaths. For example, why would you create an index for content at /content/oak:index instead of just specifying includedPaths?
>   * A: With 'includedPaths' you need to be careful as it might cause your query to miss out on valid result from excluded path.
>
>   Putting index under path would ensure that it's only used for query which have the path restriction specified and hence more safer to use

26. Q: So, if I take an offline dump of indexes using Luke, all data is open for anyone to have a look into
>   * A: Yes but you need admin access to the system to take the lucene dump!

27. Q: we have seen some performance issues when using lower case queries - is there a good way to craft a lowercase query?
>   * A: It's not yet supported. Would suggest to create an issue so that it can be prioritized

28. Q: Where do you see this checkpoint node? Is it from oak console?
>   * A: Yes. Checkpoint node is hidden and is not visible via crxde

29. Q: AEM Publish: when rendering content based on a query and this content is decached it is quite possible that updated data is not available when the decached content is requested again (with high load). What would be the best practices here? - Using plain property indexes (I understand that this is discouraged)? Is there a way to get notified when a certain index was updated (to decache manually)?
>   * A: What do you mean by 'decached' content

30. Q: Is there a way to skip index node if there is any error & proceed with next content rather than blocked at failed content?
>   * A: no, you'd need to fix the error so the indexer can move on

31. Q: Thanks Chetan and Alex for the insight :)
>   * A: Thanks

32. Q: Great and very useful presentation. Thanks.
>   * A: Thanks!

33. Q: Great Job Alex and Chetan. Thanks
>   * A: Thanks!

34. Q: Are there major differences or advice regarding indexing on AEM 6.0 (versus AEM 6.1)? We are having great difficulty with our 6.0 queries and often get log messages saying we should create indexes, often for fundamental actions like logging in.
        * A: Thanks Sham!

35. Q: can skip index node qualify has enhancement?
        * A: any proposal is welcome! Just make sure you document the use case, based on client scenarios

36. Q: Is there a to ensure that my Query is going to use particular Index only?
        * A: explain tool, if needed you can tweak the cost output by index

37. Q: when you make index configuration changes does it rebuild the entire index?
        * A: not by default. You would need to set reindex=true for reindexing to trigger

38. Q: Question was perhaps about the use of Property "propertyIndex" to true in lucene index definition… how is that different from Property Index definition
        * A: first off, it would be async, so possibly lagging behind. Otherwise it should be not much different query-wise.

39. Q: and when the reindex is completed does it set reindex=false?
        * A: yes, it will reset the reindex flag

40. Q: @Sean: Logging in will cause the notification count to be displayed, so there will be queries for things like items in inbox.
        * A: Answer by Zyg

41. Q: Thanks Chetan and Alex
        * A: Thanks Rohan!

42. Q: If the index is not found for any property or full text, it does a traversal query for entire repository and slows downs the system? Does it by design
        * A: Yes. There are settings which can enable which would cause the query to fail if it reads certain number of nodes while evaluating the query

43. Q: right, so the default lucene index that comes with AEM indexes all text and binary by default?
        * A: Yes