Quarry Queries

User 1: The Polish

'The Polish' is an active user on Polish Wiki since more than a year and has 125+ queries that giving insights on various trends on Polish Wiki. Following are the three queries that I thought were interesting and covered different SQL concepts.

1. **Query 1:** https://quarry.wmflabs.org/query/1999

   Title: Mobile edits per month

   This query gives us a count of monthly edits done to polish wiki articles using mobile from 2013 to 2016. The query starts with picking up first six characters of the timestamp using command LEFT and naming the column as 'month' to display in the output using 'AS'. To count the iterations, command 'COUNT' is used. The above data is picked up from table revision_userindex. A keyword match of the word '%mobile%' in the column 'ts_tags' of table 'tag_summary' is done to find edits done from a mobile account. Since two separate tables are queried, 'INNER JOIN' is used to join both tables on the basis of matching information in columns 'ts_rev_id' in table tag summary and 'rev_id' from table revision_userindex. These can be called Primary keys using which the matching and counting is performed. Finally, the entire data is grouped by 'month' using command 'GROUP BY'.

2. **Query 2:** https://quarry.wmflabs.org/query/2214

   Title: Links to given multiple articles on plwiki

   This query queries tables for links to articles searched using their titles on plwiki. The user selects column 'page_title' from tables 'page' and 'pagelinks' using keys 'page_id' from table 'page' and 'pl_form' from pagelinks. The string match is done using command 'IN' to column pl_title and multiple string values are passed separated by commas. The output is grouped using command 'GROUP BY' on column 'page_title' and is sorted in ascending order using command 'ORDER BY' page_title 'ASC'. All 63 entries get generated.

3. **Query 3:** https://quarry.wmflabs.org/query/747

   Title: Top 25 most edited articles on plwiki

   This query helps us find the most edited articles on plwiki. However, the response is limited to 25. The user uses table 'recentchanges' and select column 'rc_title' and retrieve the names of the articles and also adds a 'COUNT' command on the number of times it is edited and displays the same as 'num_edits'. The grouping is done using command 'GROUP BY' on column 'rc_title' and it is ordered in descending order of number of edits mentioned earlier. The responses are limited to 25 using command 'LIMIT 25'.

User 2: Fae

'Fae' is an active user on Commons Wiki and has a complex set of queries that gives trends on mass uploads and edits on Commons Wiki. Following are the three queries that I thought were interesting and covered different SQL concepts.

1. **Query 1:** https://quarry.wmflabs.org/query/7110
   Title: Which of my uploads were deleted most recently?
   This user uploads data in bulk and hence it gets difficult to figure out if any deletions go unnoticed. This query helps the user extract data on deletions from the 'File' table. The data extracted might not belong directly to the user but he must have contributed to it in some way or the other in the past since we get the file via this query. We extract the file name using 'fa_name' and display it as 'File' using command 'AS'. We also extract the reason for deletion and display that as a comment. From the 'filearchive' table we extract the time stamp and take only the left 8 characters using command 'LEFT'. All of the above is queried only for this particular user '1086557'. Also, user has provided keywords for exclusion from deleted reason using command and the entire display is formatted in descending order using 'ORDER BY DESC'.
2. **Query 2:** https://quarry.wmflabs.org/query/9540
   Title: Deleted files with Eiffel
   This query queries the table file archive to check for images that were deleted and contained the word Eiffel in either the file name or file deleted reason. This list contains a lot of images around the Paris terror attack dates. The query takes file name 'fa_name' and displays the same as 'File' using command 'AS'. The same approach is used with timestamp and media type. All of this data is extracted from the table 'filearchive' where the variable 'fa_deleted_reason' or 'fa_name' contain '%Eiffel%' in the string. The response is limited to 1000 using command 'LIMIT'
3. **Query 3:** https://quarry.wmflabs.org/query/10695
   Title: Overwrites of ThiefOfBagdad
   This query helps us find all the images that were over ridden by a particular user by the name of 'ThiefOfBagdad'. Apart from the commands used above, the new commands used here are DISTINCT and ON. 'DISTINCT' will give us an output of only unique files with unique names basically. We use the 'ON' command after the 'JOIN' command to let the join happen only if the conditions mentioned after the command 'ON' are true.

User 3: Matěj Suchánek

'Matěj Suchánek' is an active user on Czech Wiki and has a complex set of queries that gives trends in Czech Wiki. Following are the three queries that I thought were interesting and covered different SQL concepts.

1. **Query 1:** https://quarry.wmflabs.org/query/3645
   Title: We miss you!
   This query returns a list of users who have been currently inactive although were really active once upon a time. Their historically higher engagement is shown by number of articles published and their last publish date. MAX and MIN are the new commands used in this query which returns maximum and minimum values of the parameter passed. This query also excludes the users whose names include the string 'BOT' with various combinations to avoid non-human users. The queries are limited to 200 in number.
2. **Query 2:** https://quarry.wmflabs.org/query/2853
   Title: 50 most active users on Wikidata: Project chat
   This query queries from wikidatawiki which contains data about wiki. It gives us the list of 50 most active users on wikidata: Project chat. From 'revision' table 'rev_page=394' corresponds to Project Chat page. More the revisions, more active is the user considered and is grouped by 'rev_user' i.e. users name. The query used 'LIMIT 50' to limit queries to 50.
3. **Query 3:** https://quarry.wmflabs.org/query/2174
   Title: Largest redirects in cswiki
   This query helps us find large pages with redirects. From table 'page' we select columns page_title, page_namespace, page_name and use command 'AS' to name it as Title, namespace and length. We output pages where redirect is on 'page_is_redirect=1' and length of page 'page_len>200'. To display, we order by page_len in descending order.

Word Count: 1058