

# PocketData Benchmark

Naveen, Sankar, Saravanan, Sathish

# Resolved Challenges from Last Week

Parallelization of Parsing and Extraction of Log Files – User based.

Identification of the application names for 32 million queries – removed *thread\_id* as the mapping criteria.

Added PRAGMA support to latest JSQLParser provided – still unable to parse 2266 queries of the form PRAGMA <name>(<value>) & PRAGMA <name> = <value>.

Extrated few more features of SQL - # of joins (outer, left ...) , # of union...

# Challenges Faced/Facing...

JSQLParser still doesn't parse 4,207,615 queries:

❖ **key (column name):** 1,973,090

*SELECT key, value FROM CalendarCache WHERE key=?*

❖ **(( )) double parenthesis:** 339,575

*SELECT \_id, contact\_last\_updated\_timestamp FROM view\_contacts WHERE ((\_id IN default\_directory))*

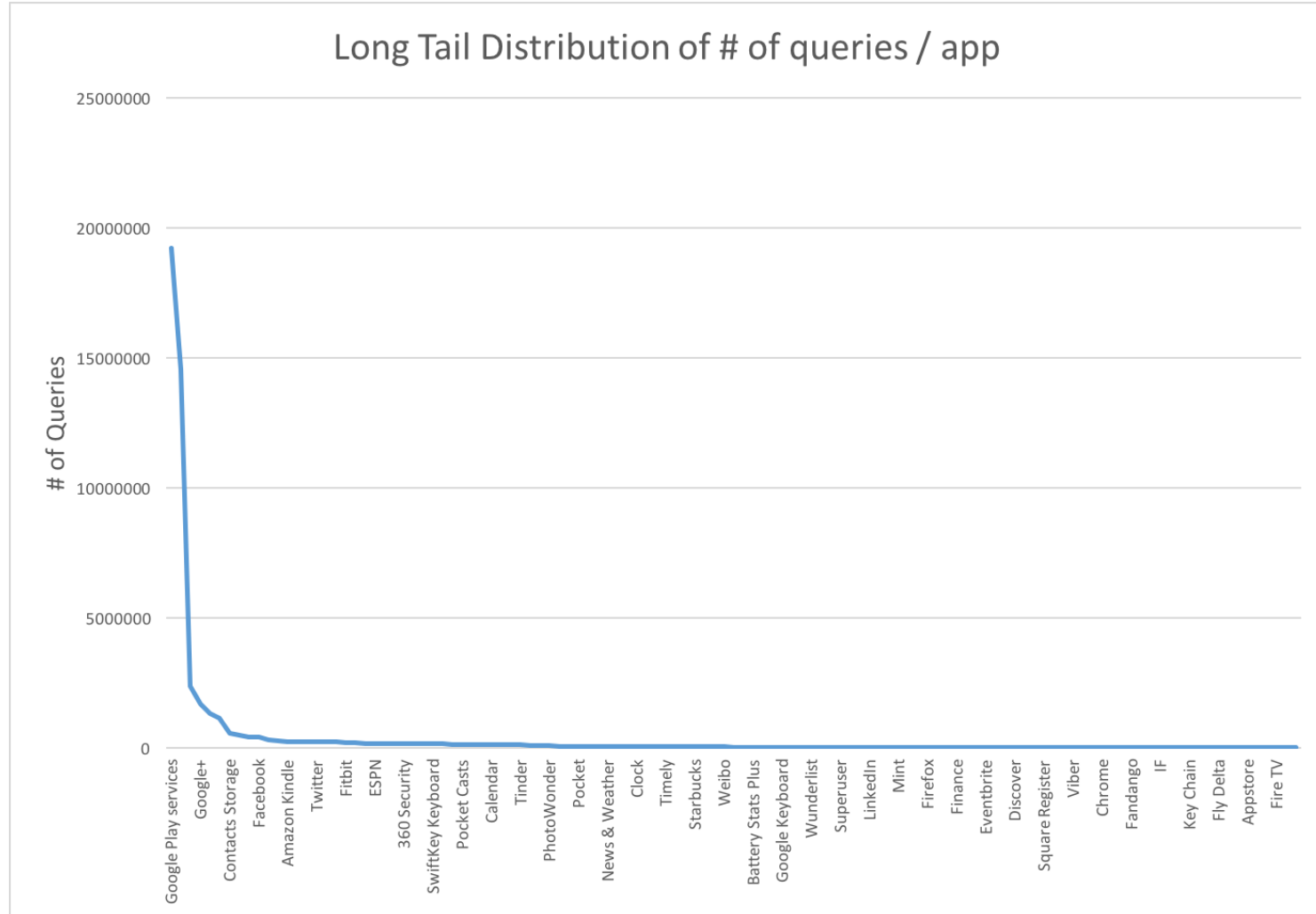
❖ **PRAGMA:** 2,266

*PRAGMA table\_info('nfcTapEvent')*

*PRAGMA secure\_delete = ON;*

More to be analyzed...

# Long Tail distribution - # of queries / app



# Maximum # of queries in 10 ms range

user_id	no_of_queries	avg_response_time (μs)
4	80	57673.125
7	69	56148
10	66	82759.5909
2	60	53961.0167
5	55	54337.0727
9	50	121999
8	47	145619.4681
6	39	98032.8718
11	38	89813.5263
3	37	124932.4324
1	21	87041.381

# Roadmap

- ❖ Parse most of the SQLs.
- ❖ Extract more features.
- ❖ Prepare Report for check-point.
- ❖ Start building the model.