Scaling Stack Trace Fingerprinting

Author: Mounir El Kirafi, SNE
Supervisor: Luc Gommans, X41 D-Sec

X41 D-Sec

Introduction

- Stack traces often used for debugging, showing active stack frames
- Contains useful information - function name, file name, line and column #
- Leaks information, useful for fingerprinting
  - Framework name
  - Version
  - CVEs

Introduction

- Existing Java stack trace fingerprinting tool: X41 Beanstack
- Extend to JavaScript
- Large number of libraries and data entries
- Java tool 49 million entries for 36 frameworks
    - Scalability issues
    - Need for fast querying
    - Adequate search and storage solution required

Research question

- ***How can the current X41 Beanstack stack trace fingerprinting database be improved for a more efficient storage and querying system?***

- What is the necessary information from JavaScript libraries to populate the database and how can this be extracted?
- What are more optimal database storage solutions for the storage and querying of stack traces?
- How can these solutions be adjusted for better performance in the target use case?

Related work

- Java tool – X41 Beanstack [1]
  - Extract data from .class files
  - Classes, functions, line numbers, function calls
  - input into MariaDB
- Java PoC – tracefp [2]
  - No focus on storage structure
- “generic” database performance research

JavaScript stack trace

- Different formats due to different engines
  - V8 (Chromium), SpiderMonkey (Firefox), JavaScriptCore (Safari), Chakra (IE), Node.JS (server-side JavaScript)
- General syntax:

```
Error type: Error message
  function name (file name:line number:column number)
```

- Filter necessary information out with regex
  - Function name
  - File name
  - Line number
  - Column number
- Extract data from frameworks using source maps

Current Beanstack implementation

Stack trace:

```
java.lang.IllegalStateException: On-the-fly migration has not been activated for this thread. Check servlet filters.
    at com.continental.coremedia.migration.OnTheFlyMigrationController.resolveBean(OnTheFlyMigrationController.java:45)
    at com.coremedia.objectserver.web.AbstractViewController.handleRequestInternal(AbstractViewController.java:169)
    at org.springframework.web.servlet.mvc.AbstractController.handleRequest(AbstractController.java:153)
    at org.springframework.web.servlet.mvc.SimpleControllerHandlerAdapter.handle(SimpleControllerHandlerAdapter.java:48)
    at org.springframework.web.servlet.DispatcherServlet.doDispatch(DispatcherServlet.java:875)
    at com.coremedia.objectserver.web.DispatcherServlet.doDispatch(DispatcherServlet.java:56)
    at org.springframework.web.servlet.DispatcherServlet.doService(DispatcherServlet.java:807)
    at org.springframework.web.servlet.FrameworkServlet.processRequest(FrameworkServlet.java:571)
    at org.springframework.web.servlet.FrameworkServlet.doGet(FrameworkServlet.java:501)
    at javax.servlet.http.HttpServlet.service(HttpServlet.java:617)
    at javax.servlet.http.HttpServlet.service(HttpServlet.java:717)
    at org.apache.catalina.core.ApplicationFilterChain.internalDoFilter(ApplicationFilterChain.java:290)
    at org.apache.catalina.core.ApplicationFilterChain.doFilter(ApplicationFilterChain.java:206)
```

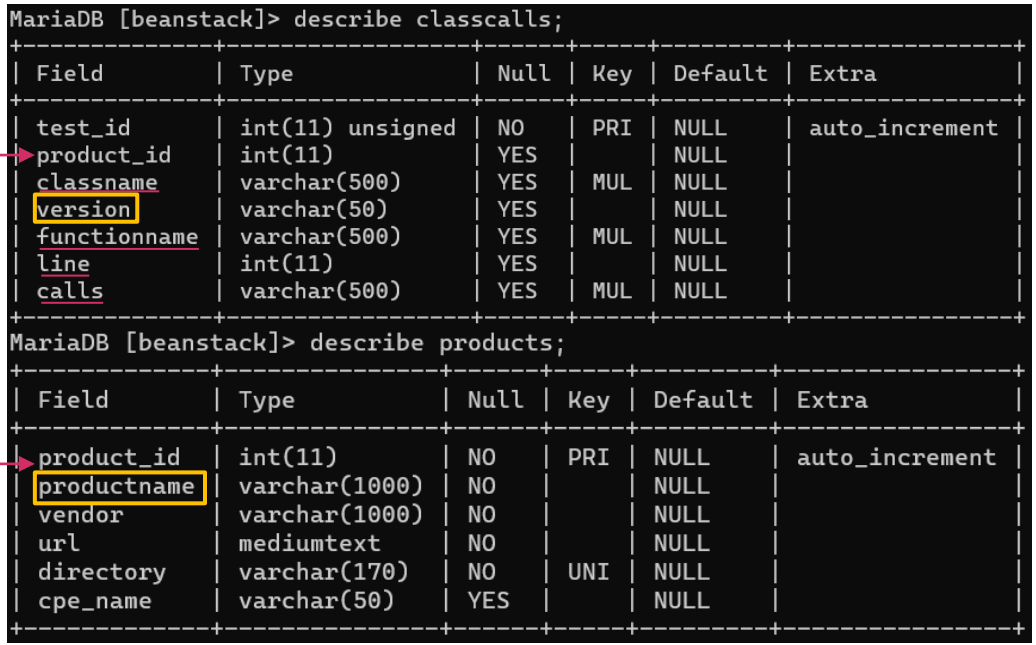
Submit API Request

API key (optional): 2wiEUvnVU-lJhbgJz5B_OdJE

Product	Matched versions	CVEs
SpringFramework	3.2.0, 3.2.1, 3.2.10, 3.2.11, 3.2.12, 3.2.13, 3.2.14, 3.2.15, 3.2.16, 3.2.17, 3.2.18, 3.2.2, 3.2.3, 3.2.4, 3.2.5, 3.2.6, 3.2.7, 3.2.8, 3.2.9	CVE-2018-1270* (9.8) CVE-2014-0225 (8.8) CVE-2015-5211 (8.6) *See all*
tomcat	6.0.24, 6.0.26	CVE-2016-8735 (9.8) CVE-2016-0714 (8.8) CVE-2016-5388 (8.1) *See all*
UrlRewriteFilter	3.2.0	-
JavaMelody	1.36.0, 1.37.0	-

[1] beanstack.io

Current Beanstack implementation



```
MariaDB [beanstack]> describe classcalls;
+--------------+------------------+------+-----+---------+----------------+
| Field        | Type             | Null | Key | Default | Extra          |
+--------------+------------------+------+-----+---------+----------------+
| test_id      | int(11) unsigned | NO   | PRI | NULL    | auto_increment |
| product_id   | int(11)          | YES  |     | NULL    |                |
| classname    | varchar(500)     | YES  | MUL | NULL    |                |
| version      | varchar(50)      | YES  |     | NULL    |                |
| functionname | varchar(500)     | YES  | MUL | NULL    |                |
| line         | int(11)          | YES  |     | NULL    |                |
| calls        | varchar(500)     | YES  | MUL | NULL    |                |
+--------------+------------------+------+-----+---------+----------------+
MariaDB [beanstack]> describe products;
+-------------+---------------+------+-----+---------+----------------+
| Field       | Type          | Null | Key | Default | Extra          |
+-------------+---------------+------+-----+---------+----------------+
| product_id  | int(11)       | NO   | PRI | NULL    | auto_increment |
| productname | varchar(1000) | NO   |     | NULL    |                |
| vendor      | varchar(1000) | NO   |     | NULL    |                |
| url         | mediumtext    | NO   |     | NULL    |                |
| directory   | varchar(170)  | NO   | UNI | NULL    |                |
| cpe_name    | varchar(50)   | YES  |     | NULL    |                |
+-------------+---------------+------+-----+---------+----------------+
```

Current Beanstack implementation

```
MariaDB [beanstack]> show index from classcalls;
+------------+------------+----------------------+--------------+-------------+-----------+-------------+----------+--------+------+------------+
| Table      | Non_unique | Key_name             | Seq_in_index | Column_name | Collation | Cardinality | Sub_part | Packed | Null | Index_type |
+------------+------------+----------------------+--------------+-------------+-----------+-------------+----------+--------+------+------------+
| classcalls |          0 | PRIMARY              |            1 | test_id     | A         |    49657577 |     NULL | NULL   |      | BTREE      |
| classcalls |          1 | idx_test_classname   |            1 | classname   | A         |      506709 |      191 | NULL   | YES  | BTREE      |
| classcalls |          1 | idx_test_functionname |            1 | functionname | A         |      287038 |      191 | NULL   | YES  | BTREE      |
| classcalls |          1 | idx_test_calls       |            1 | calls       | A         |      752387 |      191 | NULL   | YES  | BTREE      |
+------------+------------+----------------------+--------------+-------------+-----------+-------------+----------+--------+------+------------+
MariaDB [beanstack]> show index from products;
+----------+------------+------------------+--------------+-------------+-----------+-------------+----------+--------+------+------------+
| Table    | Non_unique | Key_name         | Seq_in_index | Column_name | Collation | Cardinality | Sub_part | Packed | Null | Index_type |
+----------+------------+------------------+--------------+-------------+-----------+-------------+----------+--------+------+------------+
| products |          0 | PRIMARY          |            1 | product_id  | A         |          31 |     NULL | NULL   |      | BTREE      |
| products |          0 | directory_UNIQUE |            1 | directory   | A         |          31 |     NULL | NULL   |      | BTREE      |
+----------+------------+------------------+--------------+-------------+-----------+-------------+----------+--------+------+------------+
```

Database types

- Relational vs non-relational
  - SQL vs NoSQL
- Relational
  - Structured data with relationships
  - Oracle, PostgreSQL, MySQL, MariaDB
- Non-relational
  - Collections of data with no strict structure
  - Wide Column stores
    - HBase, Cassandra
  - Document Stores
    - MongoDB, Couchbase
  - Key-Value stores
    - Couchbase, Redis, Aerospike
  - Graph databases, search engines, object oriented

Specific use case

- Search function name, file name, line # and column #
- Always known key
- Key-value store most optimal
- Hash lookup, similar to hash index

Possibilities

- Relational database with B-tree index
  - O(log n) lookup
- Relational database with hash index
  - Theoretical O(1) lookup, I/O limited
  - Most implementations only in memory
- NoSQL database with key-value store
  - Built for use case

Data model adaptation

- No need for multiple tables and multiple columns
- Hash required values and use as singular lookup column
  - Xxhash, fast and non-cryptographic hash with low collision rate
  - Hash file name, column # and row #
  - Function name may or may not be known, fuzzy search

```
MariaDB [testdb]> describe beanstack_js;
+----------------+-------------+------+-----+---------+-------+
| Field          | Type        | Null | Key | Default | Extra |
+----------------+-------------+------+-----+---------+-------+
| digest         | varchar(16) | YES  | MUL | NULL    |       |
| function_name  | varchar(10) | YES  |     | NULL    |       |
| framework_name | varchar(10) | YES  |     | NULL    |       |
| version_number | varchar(8)  | YES  |     | NULL    |       |
| cves           | varchar(54) | YES  |     | NULL    |       |
+----------------+-------------+------+-----+---------+-------+
```

Testing

- MariaDB
  - Original data model

```
MariaDB [beanstack_js]> describe filecalls;
+---------------+-------------+------+-----+---------+----------------+
| Field         | Type        | Null | Key | Default | Extra          |
+---------------+-------------+------+-----+---------+----------------+
| call_id       | int(11)     | NO   | PRI | NULL    | auto_increment |
| product_id    | smallint(6) | NO   |     | NULL    |                |
| function_name | varchar(10) | NO   | MUL | NULL    |                |
| file_name     | varchar(15) | YES  | MUL | NULL    |                |
| col           | smallint(6) | NO   | MUL | NULL    |                |
| line          | smallint(6) | NO   | MUL | NULL    |                |
+---------------+-------------+------+-----+---------+----------------+
MariaDB [beanstack_js]> describe products;
+--------------+-------------+------+-----+---------+-------+
| Field        | Type        | Null | Key | Default | Extra |
+--------------+-------------+------+-----+---------+-------+
| product_id   | smallint(6) | NO   | PRI | NULL    |       |
| product_name | varchar(10) | NO   |     | NULL    |       |
+--------------+-------------+------+-----+---------+-------+
MariaDB [beanstack_js]> describe cves;
+----------------+-------------+------+-----+---------+-------+
| Field          | Type        | Null | Key | Default | Extra |
+----------------+-------------+------+-----+---------+-------+
| product_id     | smallint(6) | NO   | PRI | NULL    |       |
| version_number | varchar(10) | NO   |     | NULL    |       |
| cve            | varchar(55) | NO   |     | NULL    |       |
+----------------+-------------+------+-----+---------+-------+
```

Testing

- MariaDB
  - New data model

```
MariaDB [testdb]> describe beanstack_js;
+----------------+-------------+------+-----+---------+-------+
| Field          | Type        | Null | Key | Default | Extra |
+----------------+-------------+------+-----+---------+-------+
| digest         | varchar(16) | YES  | MUL | NULL    |       |
| function_name  | varchar(10) | YES  |     | NULL    |       |
| framework_name | varchar(10) | YES  |     | NULL    |       |
| version_number | varchar(8)  | YES  |     | NULL    |       |
| cves           | varchar(54) | YES  |     | NULL    |       |
+----------------+-------------+------+-----+---------+-------+
```

Testing

- PostgreSQL
  - New data model
  - Hash index
    - Memory caching, in disk storage

```
testdb=# \d beanstack_js
                    Table "public.beanstack_js"
     Column     |         Type          | Collation | Nullable | Default
----------------+-----------------------+-----------+----------+---------
 digest         | character varying(16) |           |          |
 function_name  | character varying(10) |           |          |
 framework_name | character varying(10) |           |          |
 version_number | character varying(8)  |           |          |
 cves           | character varying(54) |           |          |
Indexes:
    "digest_idx" hash (digest)
```

Testing

- Couchbase
  - NoSQL
  - Key-value/JSON store

id	
000000158be3a4ea	{"cves":"uccdbpugfu-oguqrlsnlw-exhqeljmce-melcvewlnf-yinnaeupiy","digest":"000000158be3a4ea","frameworkname":"fdsrbdarze","functionname":"xmcycfvvlj","versionnumber":"10.16.13"}
000000a8a558078f	{"cves":"hoekldnpqz-eqhqaiozbv-xqjhwoskwj-elbezouiju-crgvntiyeb","digest":"000000a8a558078f","frameworkname":"mdjfkgblwp","functionname":"bfpktzqgmk","versionnumber":"15.15.18"}
0000014d070e628b	{"cves":"wackimokzf-teexrtoriv-moejnpxjcl-qkznshztei-pzhhtrxfjg","digest":"0000014d070e628b","frameworkname":"bcwxnvahaj","functionname":"dpxaoddrce","versionnumber":"2.1.19"}
000001748f513be2	{"cves":"ghwecnwjgc-fxisuzacyq-fcaxnfybvl-nuvwqzdqos-dmsstfsddk","digest":"000001748f513be2","frameworkname":"igoarieyod","functionname":"lrgvhfbuds","versionnumber":"19.2.10"}
000001d84ab4a844	{"cves":"ligouczkuc-cttzktvjol-fmshtoitpz-ktyyojyxff-mifqejeder","digest":"000001d84ab4a844","frameworkname":"znsdxpgngs","functionname":"rtccwddyyv","versionnumber":"17.17.12"}
000002398b7cf221	{"cves":"nporvlumlh-adcuyuavks-alpdlojtcn-yimvxkanrd-fqpdljduuh","digest":"000002398b7cf221","frameworkname":"daqfezsnzm","functionname":"wfiabgbmqw","versionnumber":"11.6.17"}

Testing

- Difference between data models
- PostgreSQL hash index speedup
- NoSQL DB speedup
- Speedup querying multiple keys at once vs multiple queries single key
  - `select * from beanstack_js where digest='testdigest'` x2

    vs
  - `select * from beanstack_js where digest='testdigest' or digest='testdigest2'`
  - Minimize query processing time, shift more to database lookup performance

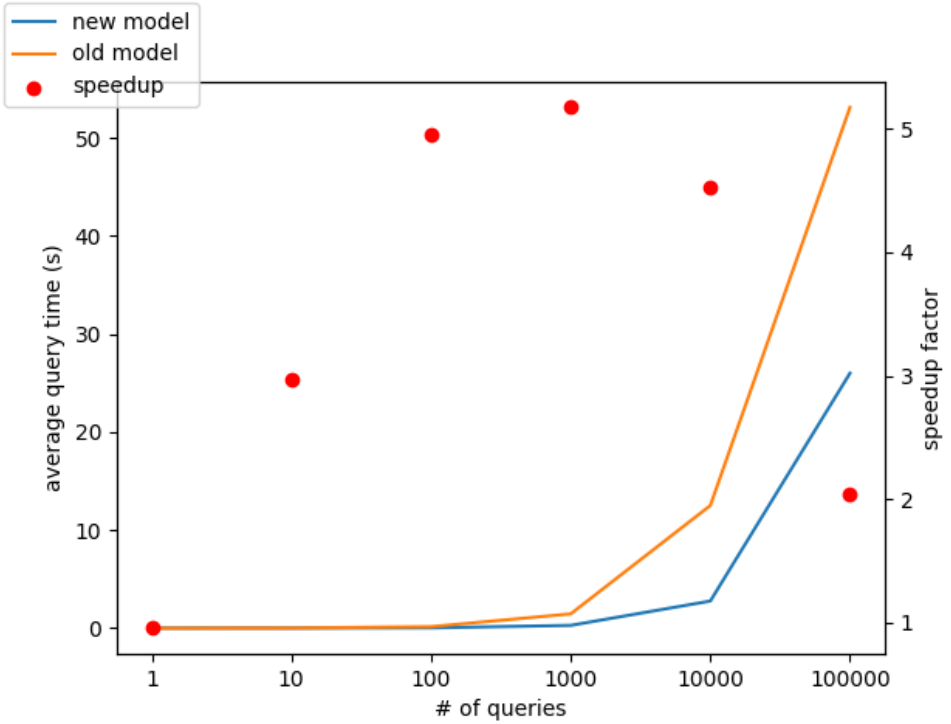
Testing

- Self generated database with 100 million entries
- Randomized strings to represent data
- 10GB file size

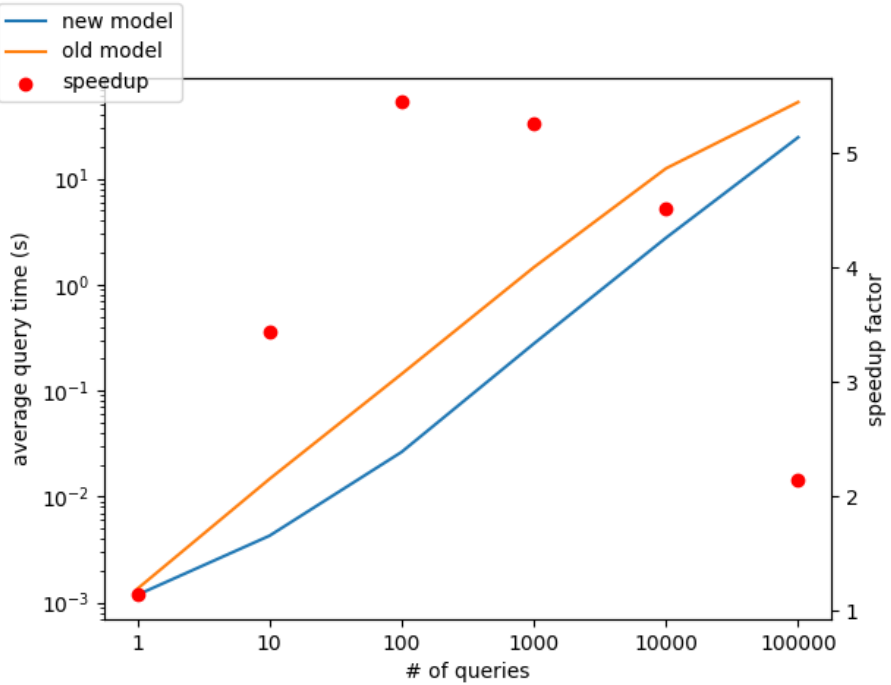
```
testdb=# select * from beanstack_js limit 10;
      digest      | function_name | framework_name | version_number |                                                  cves
------------------+---------------+----------------+----------------+-------------------------------------------------------
 5fdde6b0b8b6da66 | abohwagcoq    | lmvigyqprs     | 9.5.0          | wacaontvoq,fllsbdisyw,pxqjtrufxx,gsiazgphkk,xzlruwmkah
 e3a2ff658fa3cf79 | bdsbvalsft    | yxwglszpgj     | 12.12.9        | tcuozveiyq,cczuiwpgfd,gjhqzricyu,dsbfoiasgs,qnuslxhevc
 e6ff75a2505b36b2 | lrcaqkacmc    | pfhvsvzhjk     | 6.19.11        | spsnxgnfce,pwknqmdgra,dmskrsyzxd,ptbtluhtaj,tsminpsxiu
 16af9f1503c0bfd6 | fvbcnifpey    | gsqwphdror     | 6.3.5          | pimeduhkin,qukkrocmqe,lxyhyazkcr,pvomkapdxh,whfaqfshta
 0ddd3206ab514107 | oldypnnjdr    | mmjfjqrrcd     | 6.18.10        | pekxpfllsv,youpgxnyys,licviftcvj,kaxkcvharf,jluxwoehag
 9c58254150545151 | bvogalxojd    | yptzqqbxdd     | 7.12.13        | uihtqlheql,kcntrwazds,ufrtfxkhkk,lcggbchfop,ihopckndcq
 94e1e2abdbf8b24c | gitxnrbogv    | zzgzdvftso     | 19.11.2        | sctuqpfvwo,pfqqpwatwc,pdyskovszw,izdnceglla,ahtbhxoqkx
 f97d1e634508188f | yqaqppdaji    | ixuamnbskx     | 11.4.11        | ntcmxbkmwr,pdriwpbvgd,ragfuvtjir,itvueilcoh,qsuopxkgcy
 cceb90765c5e0922 | kfkpeqzlkt    | zhwefthpxp     | 7.9.17         | lomwbgashe,hzugeccdtk,xlkxfbmbuw,axfjuiptry,mtccpxelzt
 55e6b382d084bce4 | raetcljywv    | pzbgyvfbkq     | 2.17.18        | thstczemgt,awetzrexeu,nodbifklqr,gucuczholq,uewqqatrgc
(10 rows)
```

Results

- MariaDB “old” vs “new” data model
- New model vs old model speedup factor of 2-5x
- Speedup less significant as # of queries increases

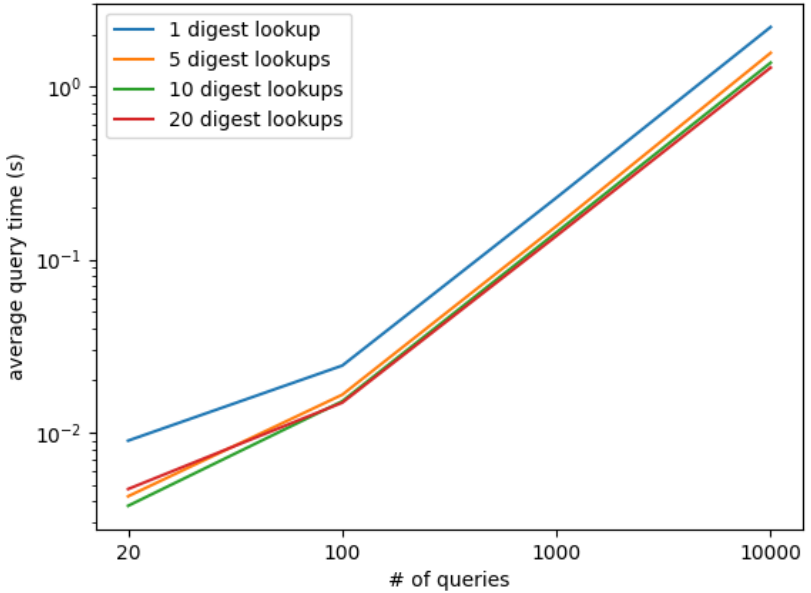


Results


new model
old model
speedup
average query time (s)
speedup factor
# of queries
1
10
100
1000
10000
100000

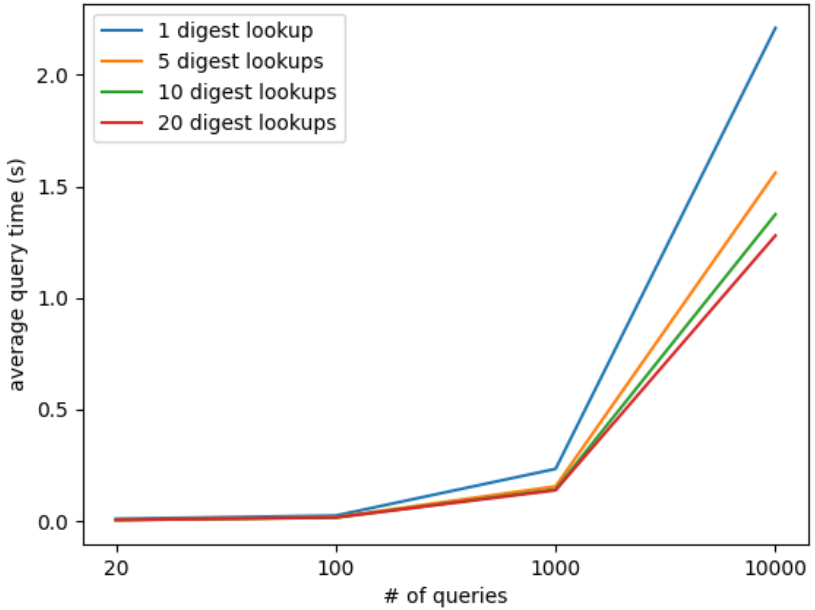
Results

- PostgreSQL speedup using multiple digests in single query
- Due to better IO utilization
  - 1 digest: ~35-37 MB/s disk read
  - 5 digests: ~56-57 MB/s disk read
  - 10 digests: ~67 MB/s disk read
  - 20 digests: ~75 MB/s disk read



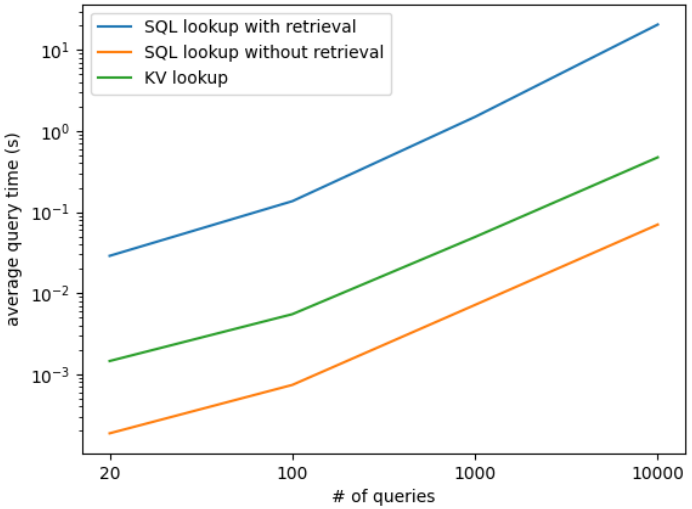
Results

- From 2.2s to 1.3s for 10000 queries
- MariaDB new model takes 2.7s



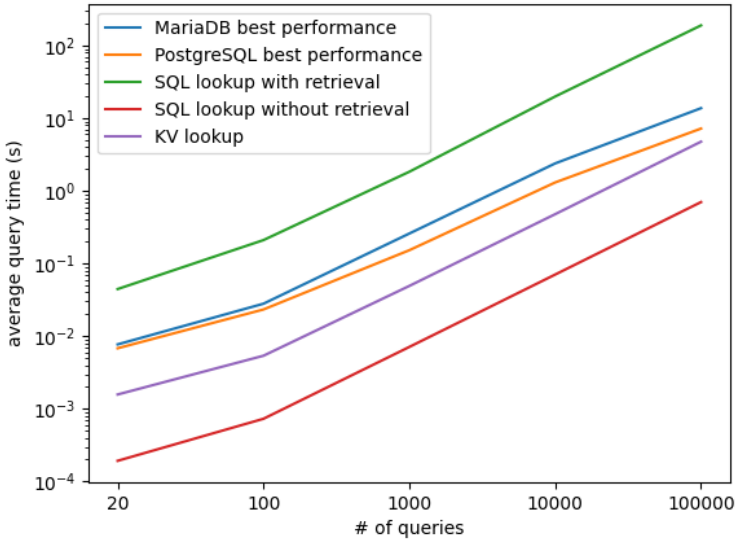
Results

- Couchbase lookup
  - Direct SQL lookup
  - Lookup through Key-Value API
- Direct SQL retrieval very fast
  - Unpacking returned Python iterable very slow
- Key-value lookup in between



Results

- Comparing
  - MariaDB best performance (new model)
  - PostgreSQL best performance (new model, hash indexing, multiple digests per query)
  - Couchbase different lookups
- KV lookup fastest



Conclusion

- Database can be improved through
  - Better data model
  - Better query design
  - Different storage system
- Future work
  - Speedup through distribution over multiple system (sharding)
  - Speedup through parallelism
  - Research effect of hardware optimization

References

- [1] - https://beanstack.io/
- [2] - https://github.com/Skyr/tracefp, https://media.ccc.de/v/EH2014_-_5633_-_de_-_degerloch_-_201404201345_-_java_stacktrace_fingerprinting_-_skyr

Questions?