

Complete a Short Survey



Responses for informational purposes only

Index Strategy Guide



Greg Dostatni DBA @ Command Prompt, Inc.
Postgres Conference 2025

Introduction

**COMMAND
PROMPT, INC.**



Goals

- This is a guide for making decisions about indexes
- Everything is about trade-offs and balance
- Questions are encouraged



But wait.. there is more!

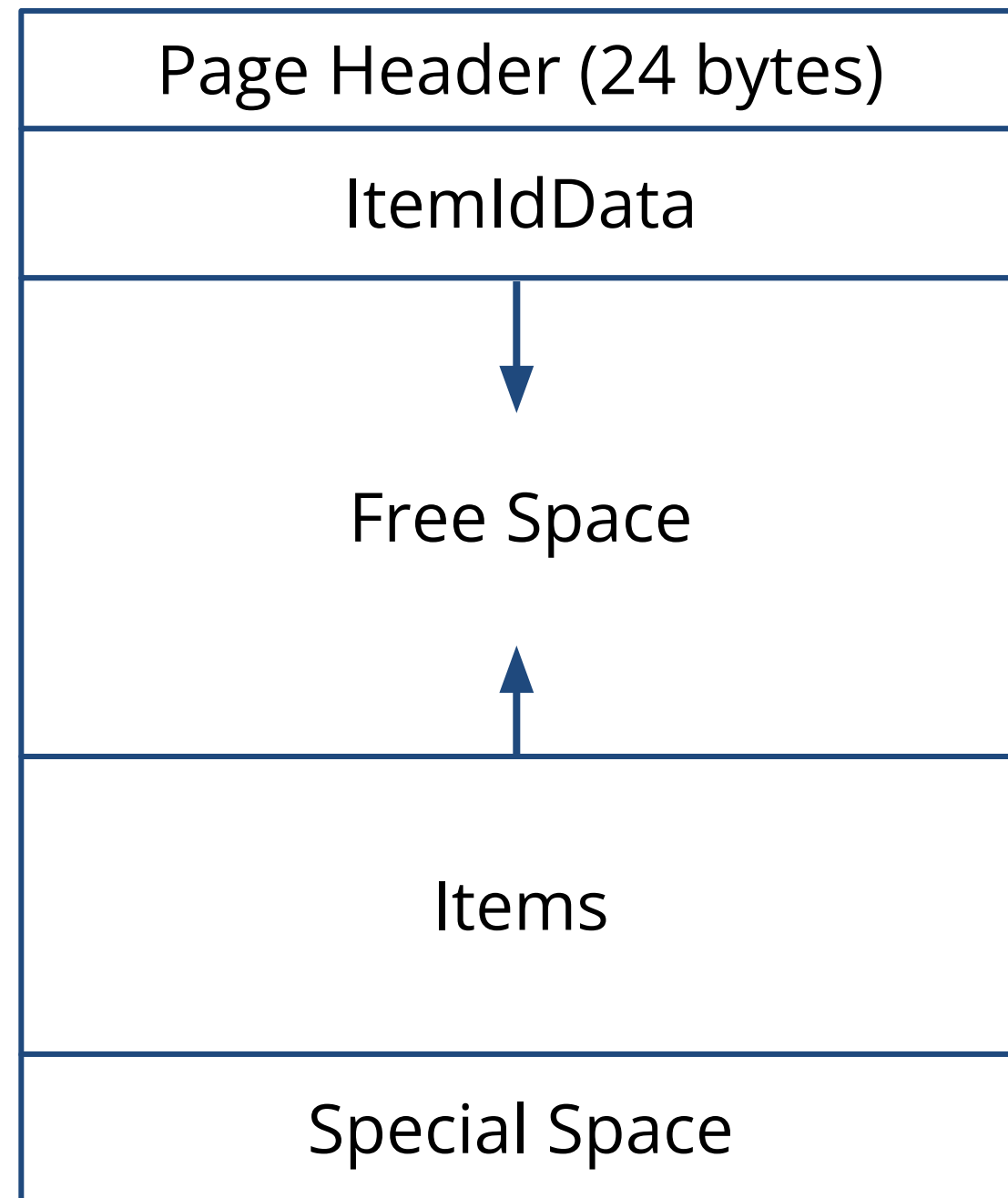
- Data driven - experiments are included
- Data driven, even if data is wrong (5%, 50%, 500%)
- Presentation notes include commands, and setup hints



Indexes are a tradeoff

- Indexes trade disk space and increased IO during changes for reduced IO when accessing existing records.
- PostgreSQL needs to decide to use an index
- Use explain on a query to check what PostgreSQL thinks is the best plan.

Data Page layout



SELECT relname, relkind,

reltuples / relpages AS
avg_tuples_per_page

FROM pg_class

WHERE

relpages>0 and reltuples>100;

Statistics

- pg_class -> reltuples /relpages
- pg_stat_all_tables -> last_autoanalyze, last_analyze, table usage
- pg_stats -> column frequent values and frequencies, avg width

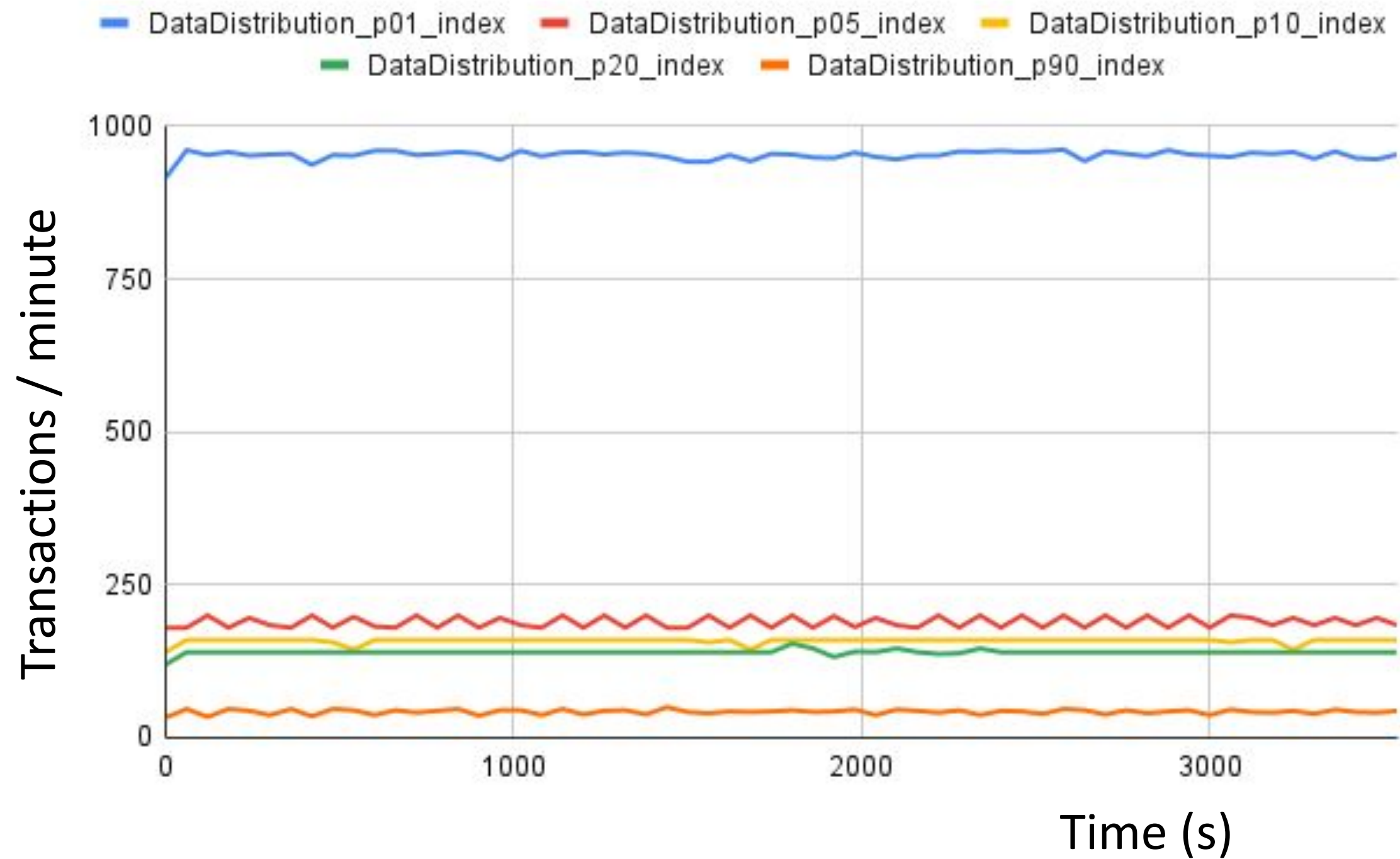
```
SELECT tablename, attname, avg_width,  
               most_common_vals, most_common_freqs  
FROM pg_stats  
WHERE tablename='test' ;
```


Trade-offs

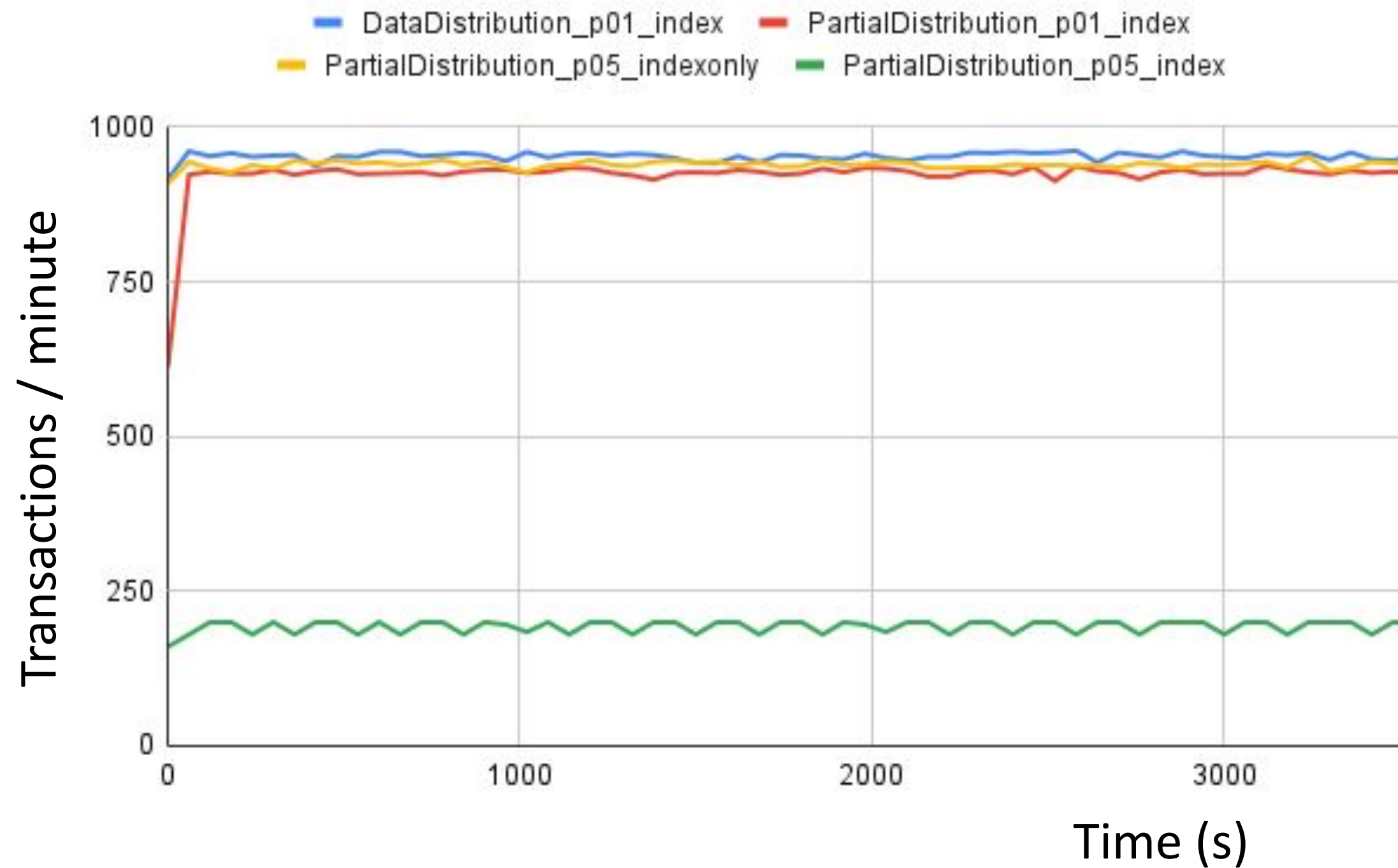
COMMAND
PROMPT, INC.



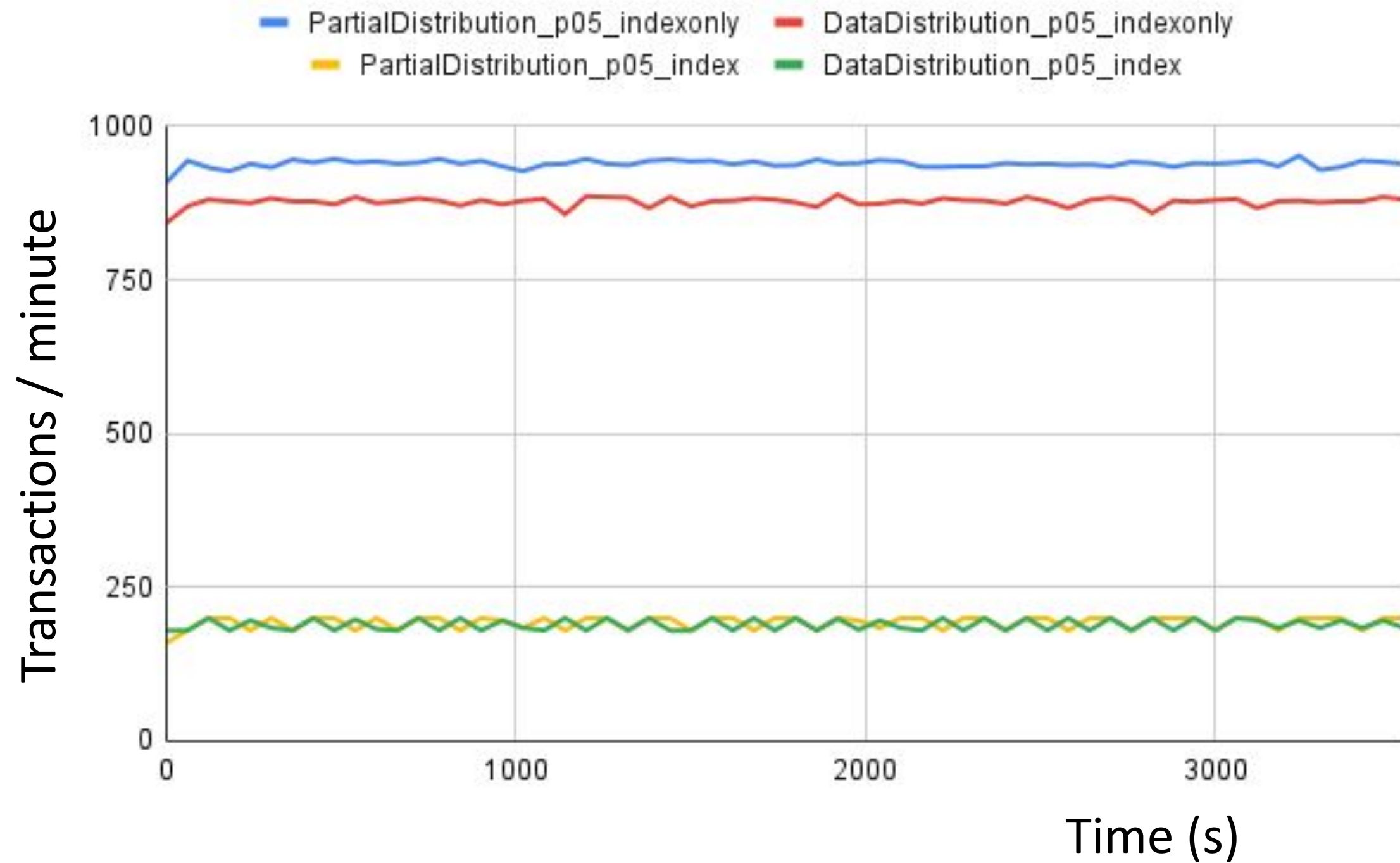
Data distribution



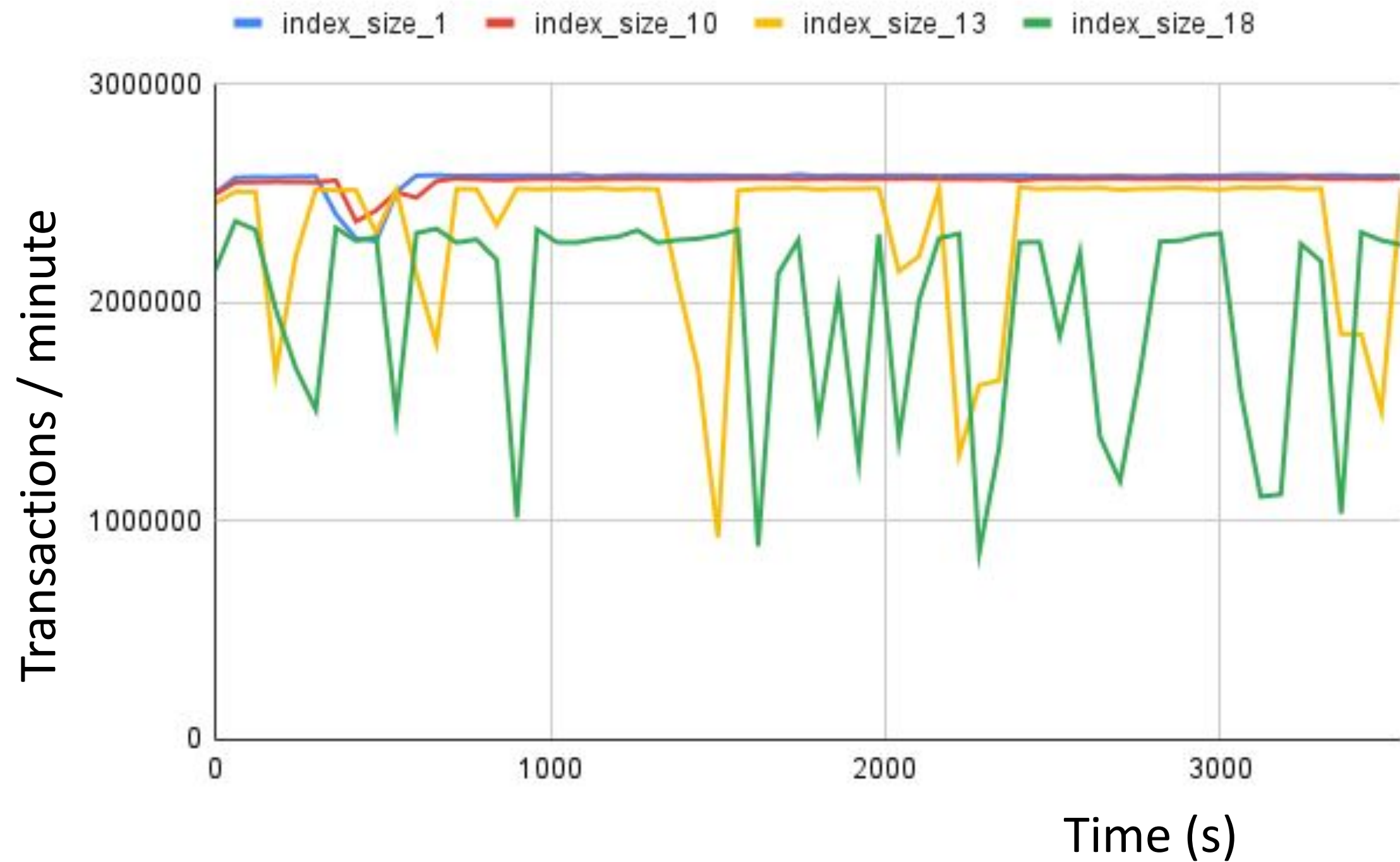
Data distribution - partial indexes



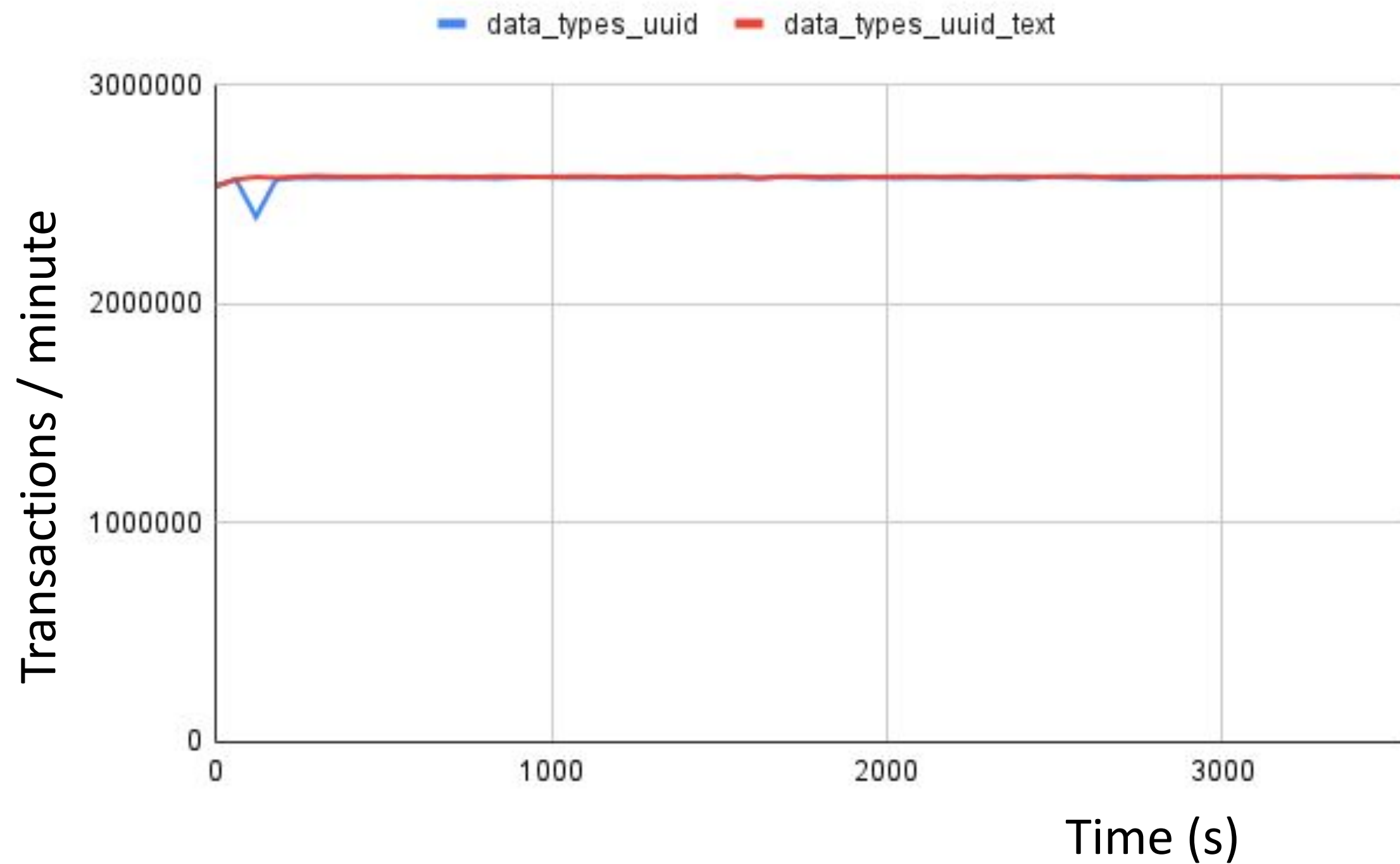
Index only access



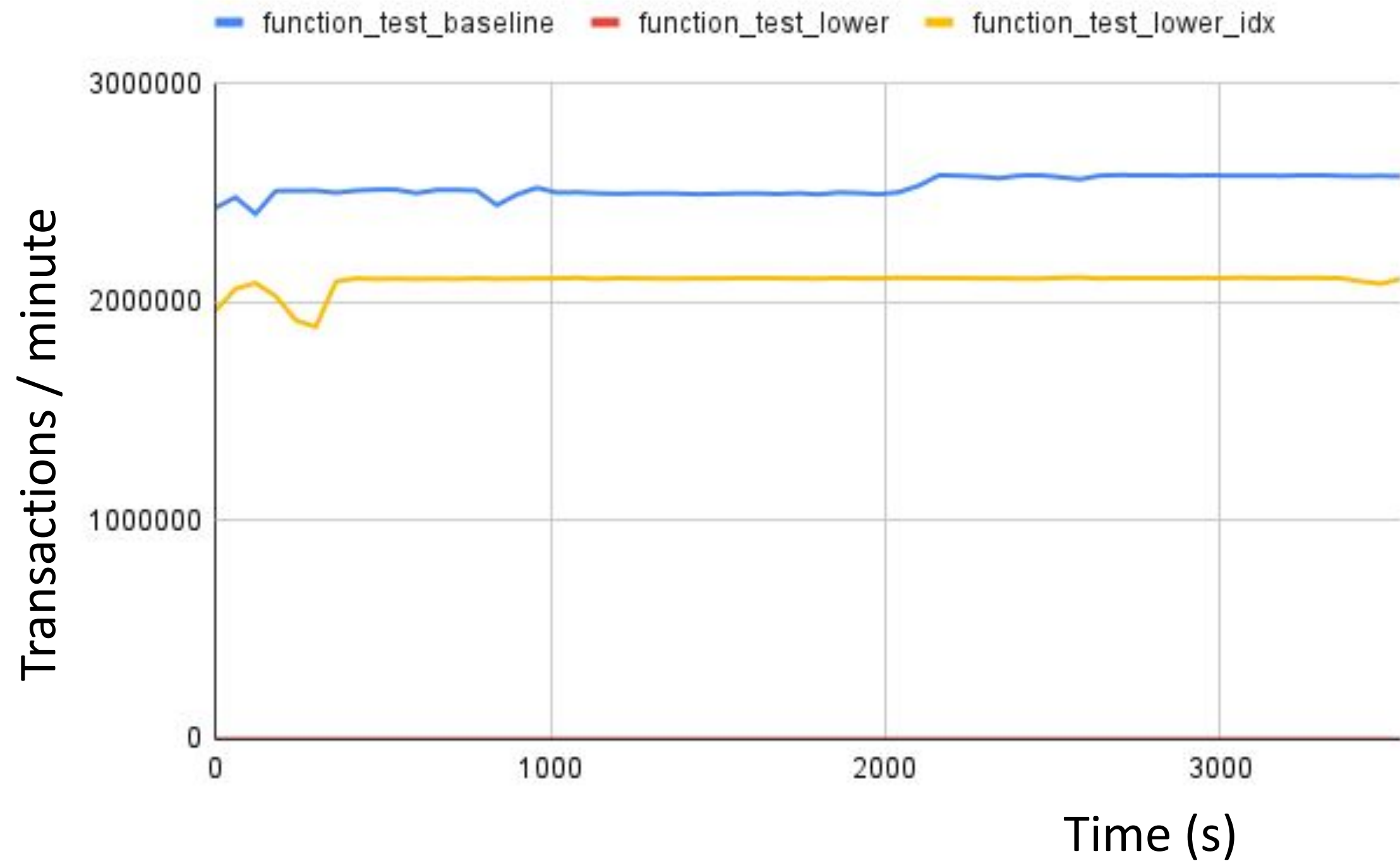
Index size



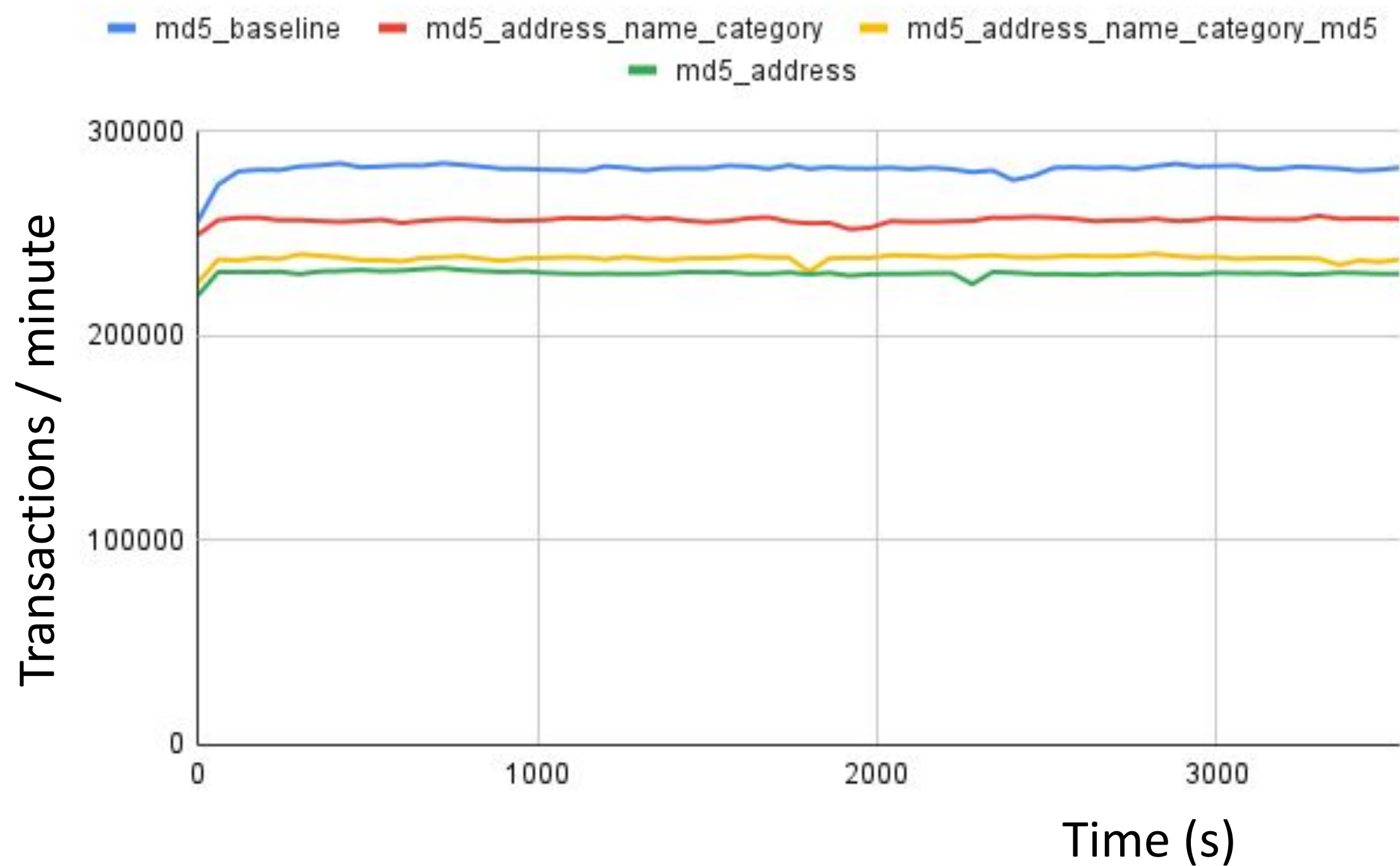
Data types



Functions are bad!

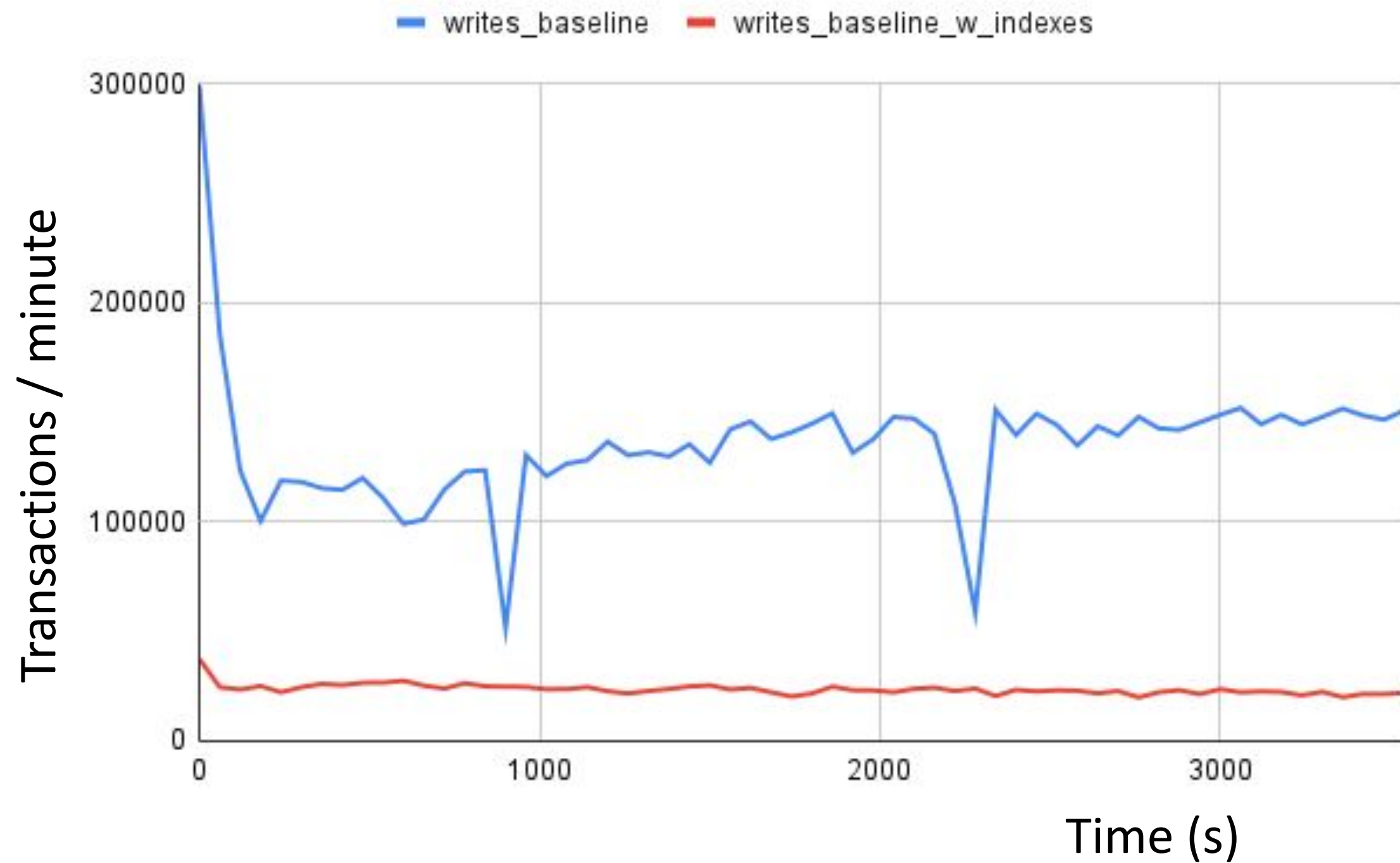


Functions are great!

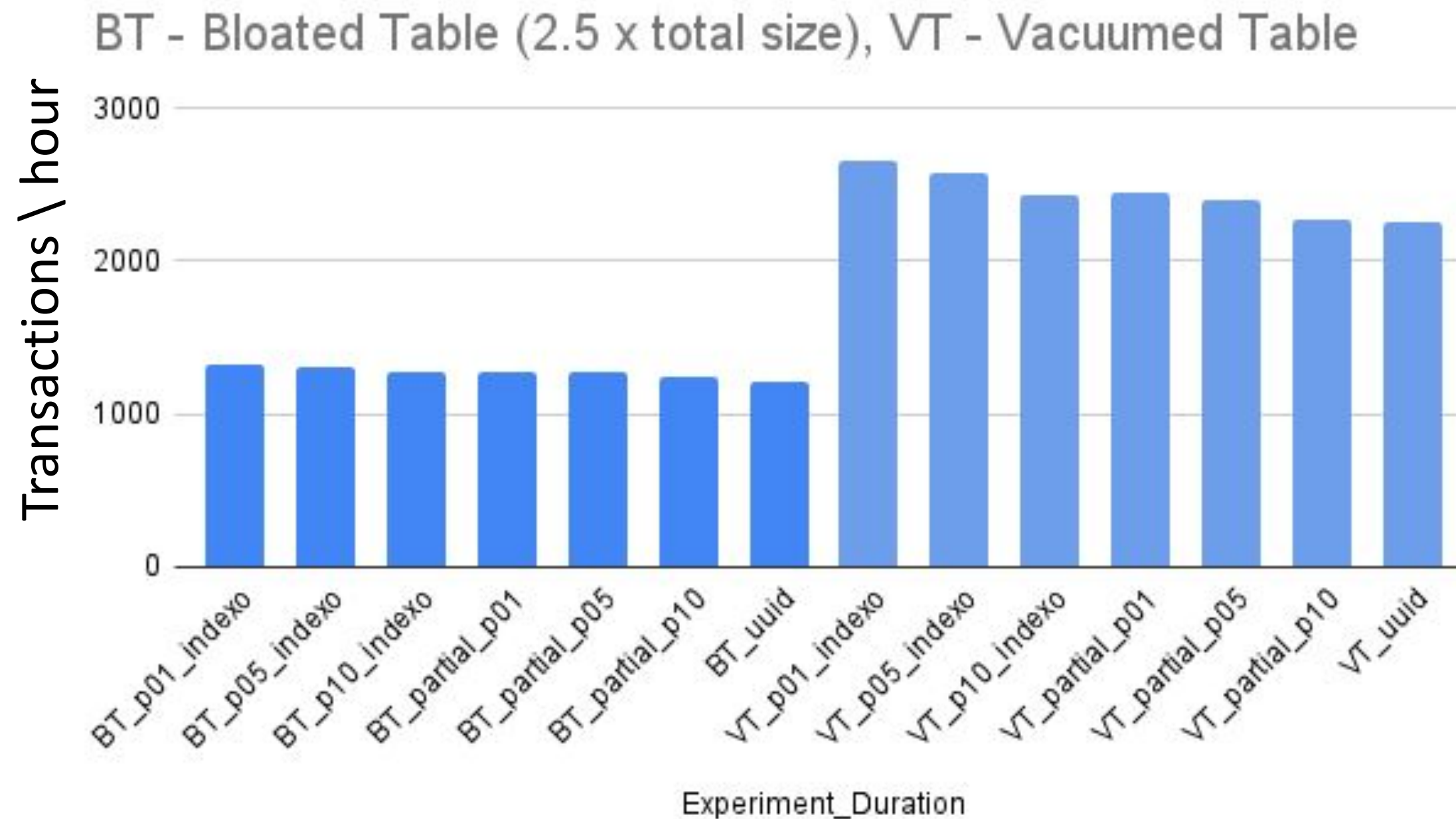


```
create index idx_md5 on
test(
  md5(
    address || '!' ||
    name || '!' ||
    category::text)::uuid
);
```


Cost of writing



Maintenance / bloat



Resources

**COMMAND
PROMPT, INC.**



Test Table

```
CREATE TABLE test (  
  id SERIAL PRIMARY KEY,  
  text_identifier text,  
  subcategory int,  
  p_05 int,  
  p_15 int,  
  p_25 int,  
  p_75 int,  
  name VARCHAR(50) NOT NULL,  
  -- Other user-related columns  
  bio TEXT,  
  address VARCHAR(200),  
  public_key TEXT );
```

```
identifier uuid,  
category int,  
p_01 int,  
p_10 int,  
p_20 int,  
p_50 int,  
p_90 int,  
email VARCHAR(100) NOT NULL,  
  
phone_number VARCHAR(20),  
website_url VARCHAR(200),
```


Use the Index, Luke!

<http://use-the-index-luke.com>



SUMMARY

**COMMAND
PROMPT, INC.**



Recap

- Everything is about trade-offs
- There are many factors including:
 - are statistics up to date?
 - shared_buffers / file cache
 - How many IO operations to access required data?



Questions?

**COMMAND
PROMPT, INC.**



COMMAND PROMPT, INC.

+1 503 667 4564

www.commandprompt.com/contact-us



EXPERTS IN POSTGRES AND OPEN SOURCE INFRASTRUCTURE