

Taming The Beast:

Managing High-Growth Postgres Databases at CircleCI



Brice Nkengsa

whoami

Brice Nkengsa

Staff Software Engineer @ CircleCI

8+ years of engineering experience

Focus on backend architecture & reliability

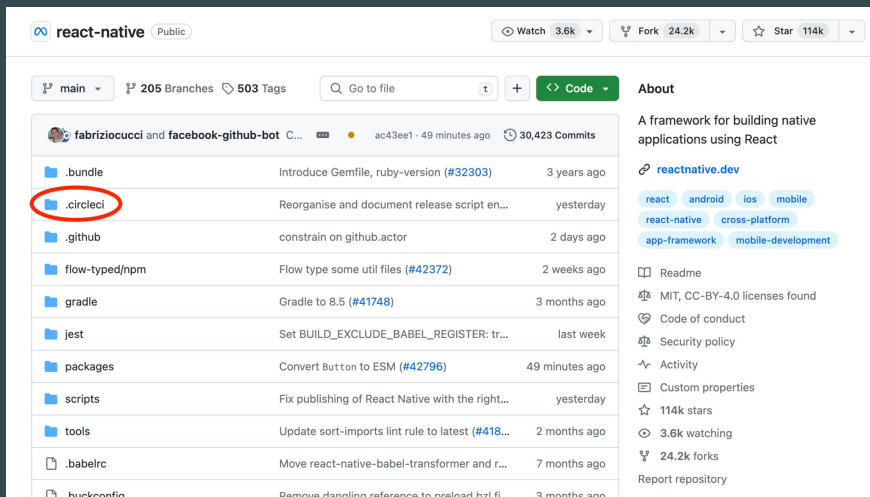
Enthusiastic about developer experience

Amateur golfer and occasional traveller



How The Beast Came To Be

- CircleCI is a global CI/CD platform
- 4M workflows, 20M jobs run per week
- 150+ services, 70+ postgres DBs



react-native (Public)

Watch 3.6k Fork 24.2k Star 114k

main 205 Branches 503 Tags

Go to file

Code

About

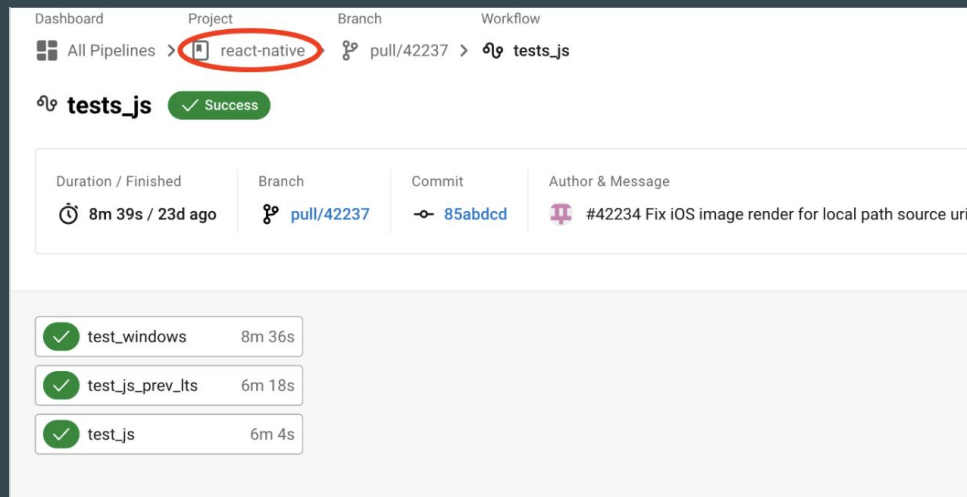
A framework for building native applications using React

reactnative.dev

react android ios mobile
react-native cross-platform
app-framework mobile-development

Readme
MIT, CC-BY-4.0 licenses found
Code of conduct
Security policy
Activity
Custom properties
114k stars
3.6k watching
24.2k forks
Report repository

File	Description	Last Commit
.circleci	Reorganise and document release script en...	yesterday
.bundle	Introduce Gemfile, ruby-version (#32303)	3 years ago
github	constrain on github.actor	2 days ago
flow-typed/npm	Flow type some util files (#42372)	2 weeks ago
gradle	Gradle to 8.5 (#41748)	3 months ago
jest	Set BUILD_EXCLUDE_BABEL_REGISTER: tr...	last week
packages	Convert Button to ESM (#42796)	49 minutes ago
scripts	Fix publishing of React Native with the right...	yesterday
tools	Update sort-imports lint rule to latest (#418...	2 months ago
.babelrc	Move react-native-babel-transformer and r...	7 months ago
.buckconfig	Remove dangling reference to preload.bzl fi...	3 months ago



Dashboard Project Branch Workflow

All Pipelines > react-native > pull/42237 > tests_js

tests_js Success

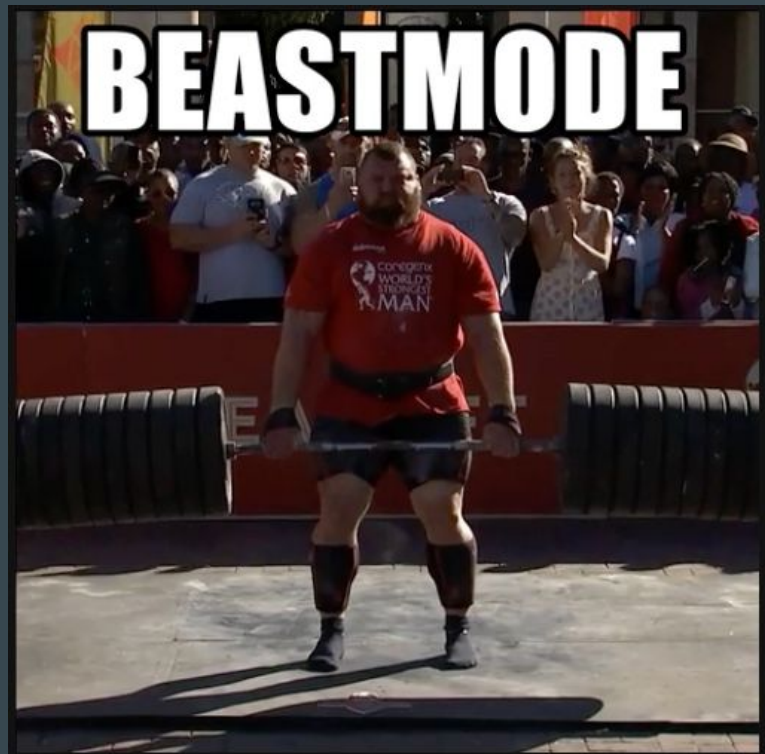
Duration / Finished	Branch	Commit	Author & Message
8m 39s / 23d ago	pull/42237	85abdcd	#42234 Fix iOS image render for local path source uri

test_windows	8m 36s
test_js_prev_lts	6m 18s
test_js	6m 4s

How The Beast Came To Be

High-growth DB:

- 5TB, growing 500GB/quarter
- Poor performance == incidents
- Major version upgrade == significant downtime

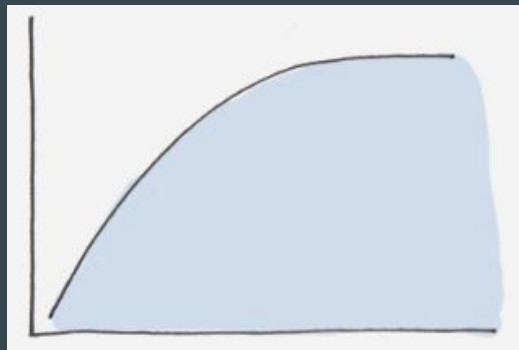


The Journey To Taming The Beast



Storage Reduction

Immediate storage savings by deleting low-hanging fruit



Growth Restriction

Curb the data growth to reasonable levels



Optimization

Establish data systems and strategies for the long term

Reduce Storage

- Drop unused columns, tables, & indexes
- Use **BRIN** indexes wherever possible
- Offload blob data to Object Storage

All Indices

INDEX NAME	TABLE NAME	DATA SIZE ▾
_created_at_index		80.7 GB
_created_at_index		65.7 GB
_branch_name_created_at_index		64.7 GB
_created_at_index		58.2 GB
_created_at_index		42.8 GB
_events_created_at_index		41.5 GB
_created_at_index		39.6 GB
_created_at_index		13.8 GB

Curb The Growth

- Determine data retention policy
- Implement API-level data retention
- Delete obsolete data



Whoops....not so fast

- How are the replicas doing?
- Index bloat
- Reclaim disk space
 - VACUUM vs pg_repack

The screenshot shows the GitHub repository page for `pg_repack`. At the top, it indicates the repository is public and shows 59 watches, 164 forks, and 1.7k stars. Below this, the repository name `pg_repack` is displayed with a dropdown menu for branches, currently set to `master`. It also shows 8 branches and 27 tags. A search bar and buttons for 'Add file' and 'Code' are visible. The main content area lists files and folders with their commit history:

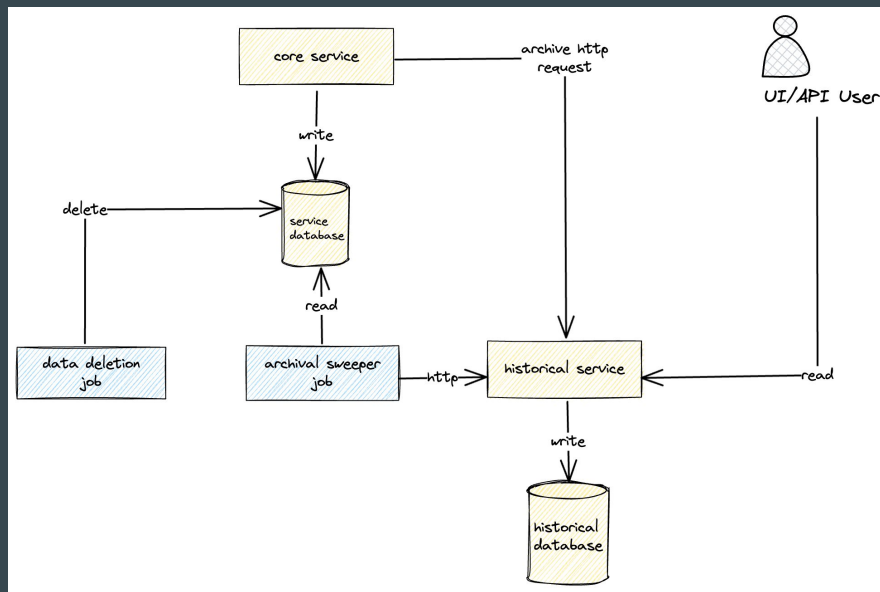
File/Folder	Commit Message	Time Ago
<code>.github/workflows</code>	Issue #338: Remove old EOL PostgreSQL versions (#374)	3 months ago
<code>bin</code>	Move INSERT query of repack, repack_trigger to C (#368)	4 months ago
<code>doc</code>	Issue #338: Remove old EOL PostgreSQL versions (#374)	3 months ago
<code>lib</code>	Fix wrong OID format type in repack_trigger	2 months ago
<code>msvc</code>	Dropped executable flags	12 years ago
<code>regress</code>	Issue #338: Remove old EOL PostgreSQL versions (#374)	3 months ago

On the right side, there is an 'About' section with the following text: 'Reorganize tables in PostgreSQL databases with minimal locks'. Below this, there are links for 'Readme', 'BSD-3-Clause license', 'Activity', 'Custom properties', '1.7k stars', '59 watching', '164 forks', and 'Report repository'.

Optimize

- Single historical data store
- Data archival process
- Partition new tables

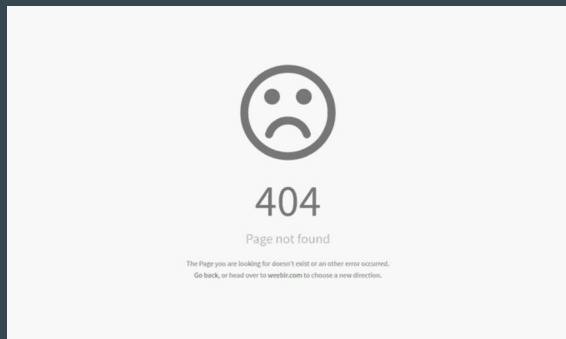
([pg_partman](#))



The screenshot shows the GitHub repository page for **pg_partman**, a PostgreSQL extension for partition management. The repository is public and has 13 branches and 82 tags. The main branch is **master**. The repository is managed by **keithf4** and has 260 commits. The repository description is "Partition management extension for PostgreSQL". The repository includes several tags and a README file. The repository is watched by 55 users, forked by 271 users, and has 1.8k stars. The repository is managed by **pg-partition-manager**.

File	Commit Message	Commit Date
bin/common	Update base documentation	last year
doc	v5.0.1 fix edge case condition when upgrading to version 5	2 months ago
sql	prepare for 5.0.0 final release	5 months ago
src	prepare for 5.0.0 final release	5 months ago
test	prepare for 5.0.0 final release	5 months ago
updates	simplify 5.0.1 update	2 months ago
.gitignore	v3.1.2 Fixed reapply_indexes.py to work properly in PG10	7 years ago
.pre-commit-config.yaml	add procedure unit tests. add pre-commit hooks. add we...	6 months ago

Key Takeaways



Retention

Brief retention policy as early as possible



Rehearsals

Rehearse any major database maintenance



Documentation

Write it down!

THANK YOU!

github.com/briceicle

linkedin.com/in/briceicle

brice@circleci.com