Taming The Beast:

Managing High-Growth Postgres Databases at CircleCI

Brice Nkengsa

whoami

Brice Nkengsa

Staff Software Engineer @ CircleCI

8+ years of engineering experience

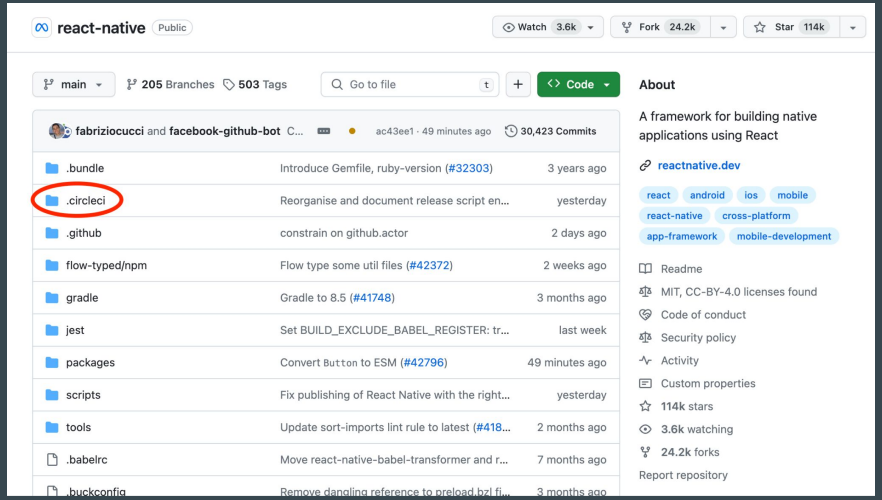
Focus on backend architecture & reliability

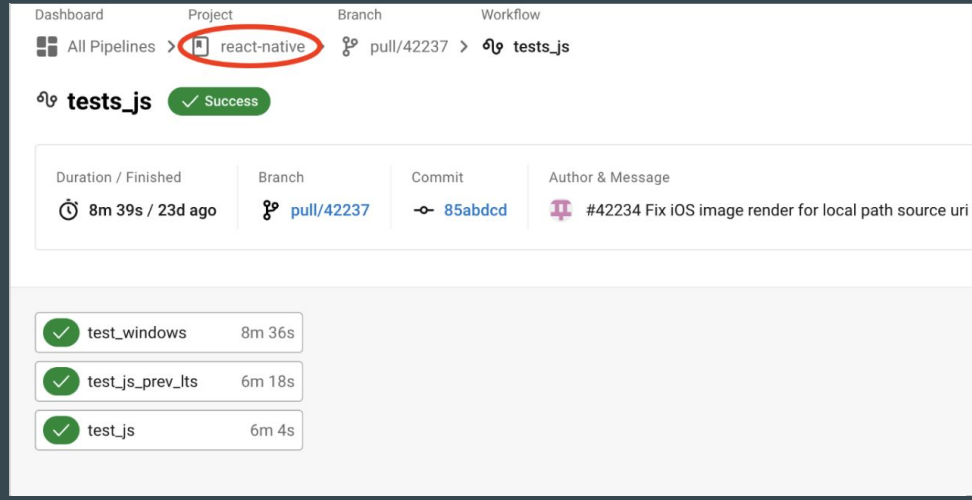
Enthusiastic about developer experience

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



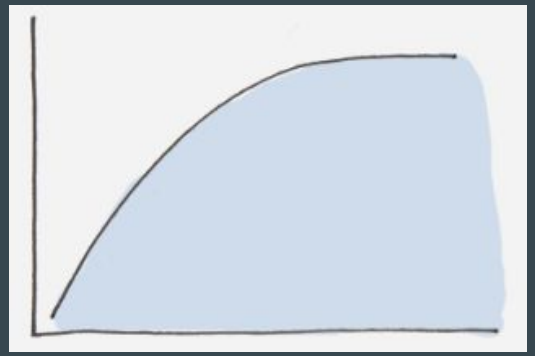


How The Beast Came To Be

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

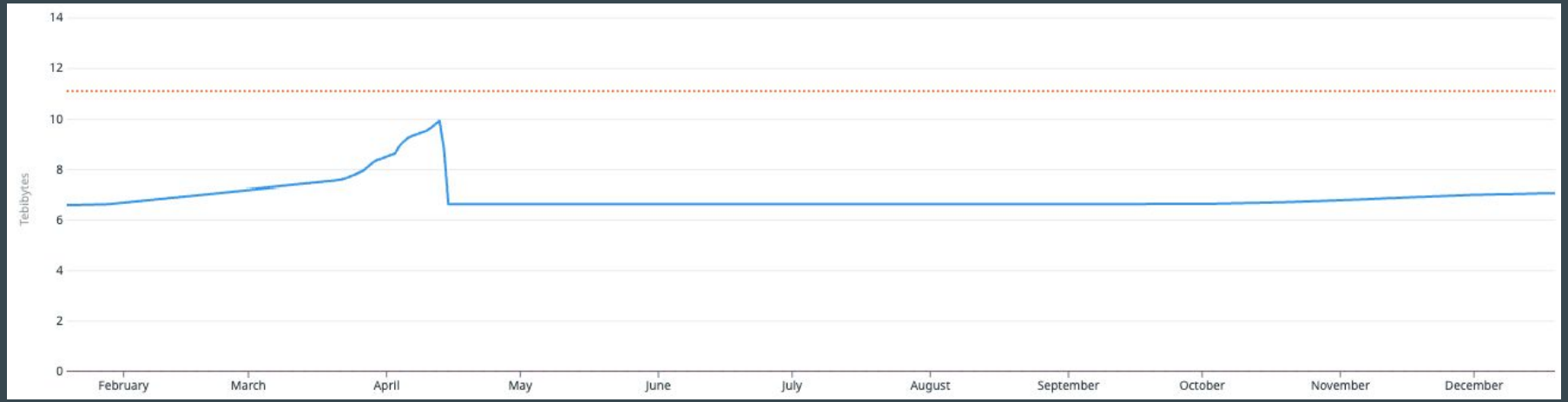
Establish data systems and strategies for the long term

Reduce Storage

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

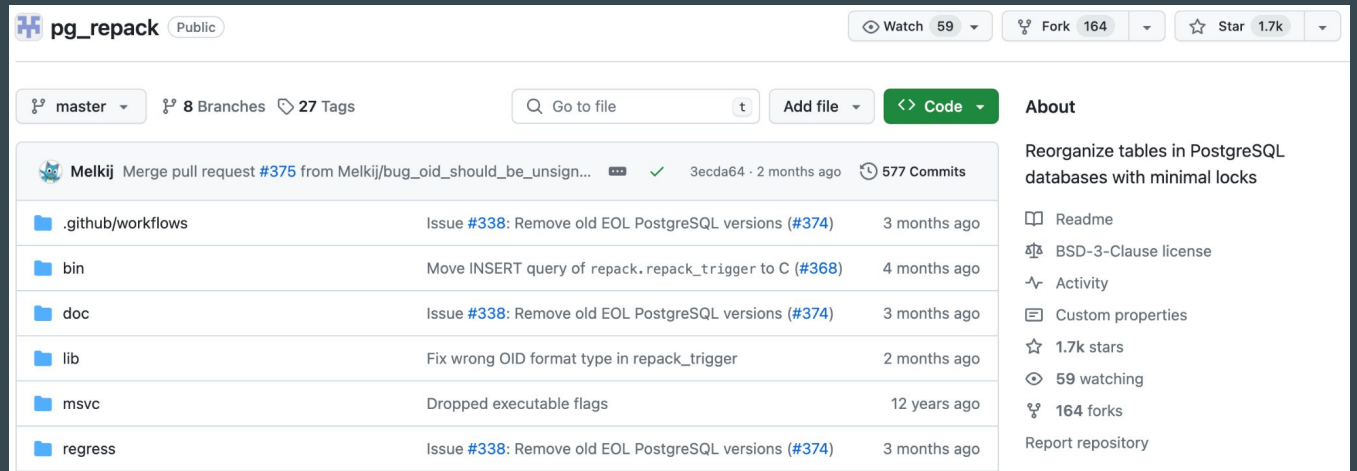
Curb The Growth

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



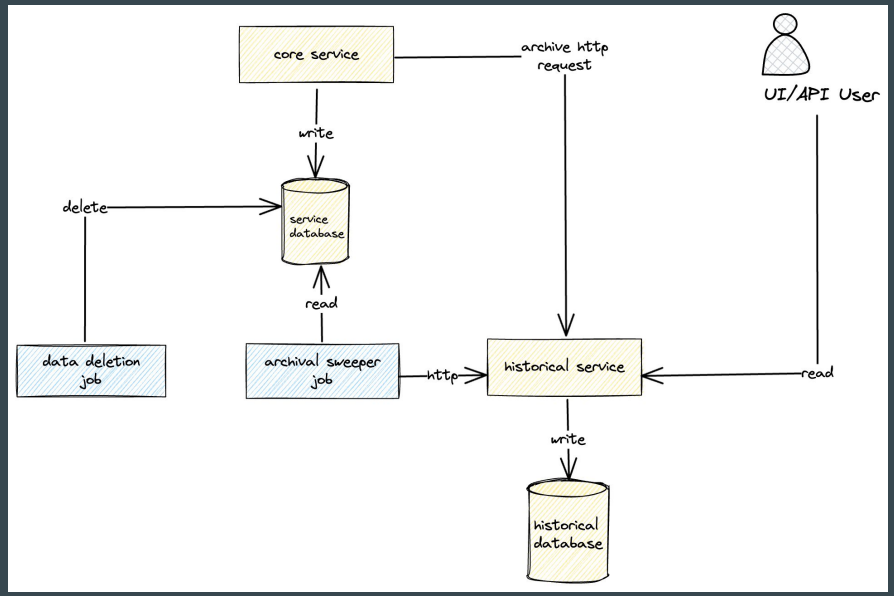
Whoops....not so fast

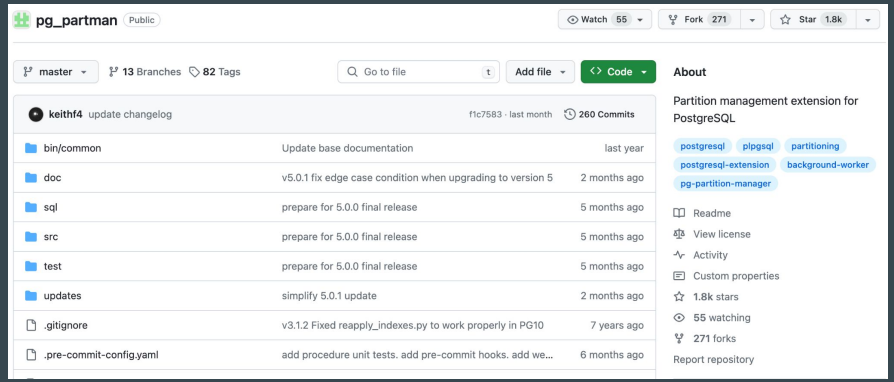
- How are the replicas doing?
- Index bloat
- Reclaim disk space
  - VACUUM vs pg_repack



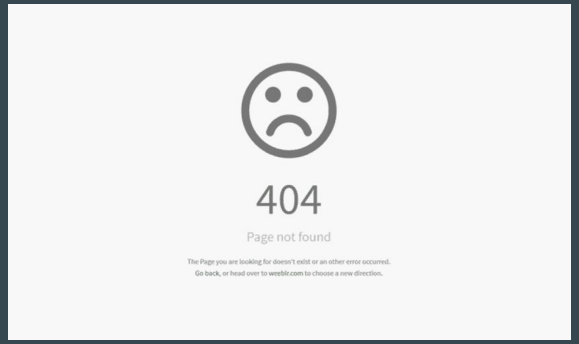
Optimize

- Single historical data store
- Data archival process
- Partition new tables (pg_partman)





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