Building a SaaS developer platform using Postgresql

Abhijit Paithankar
ap@crave.io

Introduction

@aptrekr

CRAVE.io

Abhijit Paithankar
ap@crave.io

Founder/CTO at crave.io, Inc

Previously: over 17 years at Nutanix, VMware, Akamai

Agenda

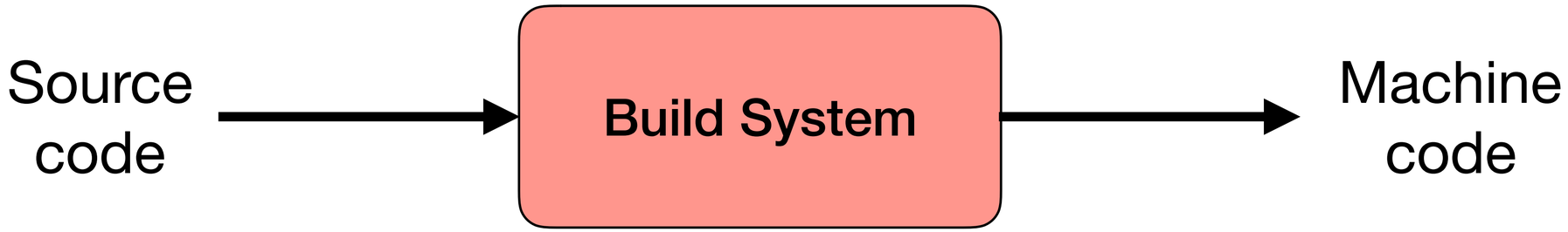
- What is a SaaS developer platform?
- Requirements for a modern developer platform
- How we built one using Postgres
- Demo: Let’s compile Postgres!
- Questions

The story of a developer platform

(or how we built a startup)

- Started as a personal itch: built it as a side project for my side project
- Solved a bunch of interesting problems
- Showed it around
- Got a potential customer very excited: signed a POC!
- Incorporated

What is a build system?



- A build system helps you **build software**
- A structured computation process that converts source code into machine code
- Constructs packages for deployment

What is a build system?

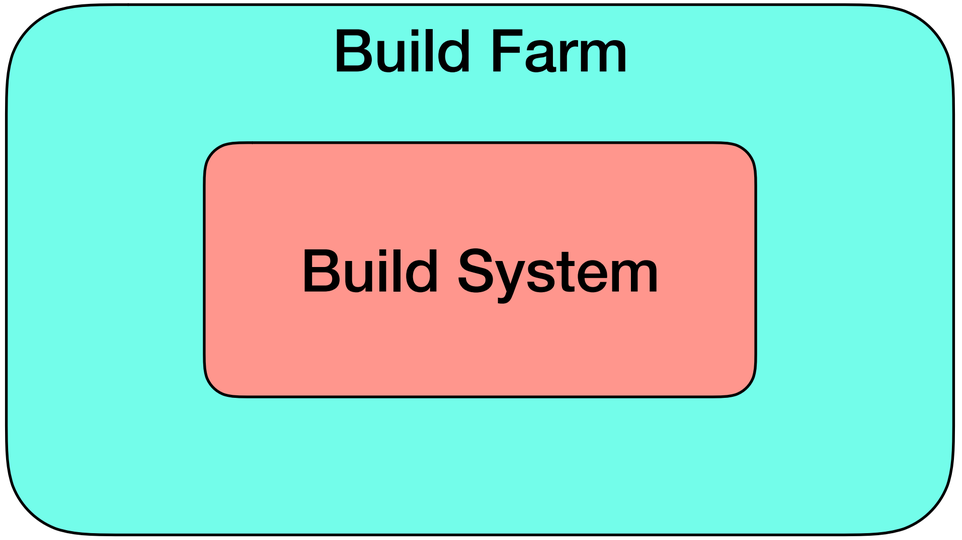
What is a build farm?

A build farm is shared infrastructure for building and testing code.

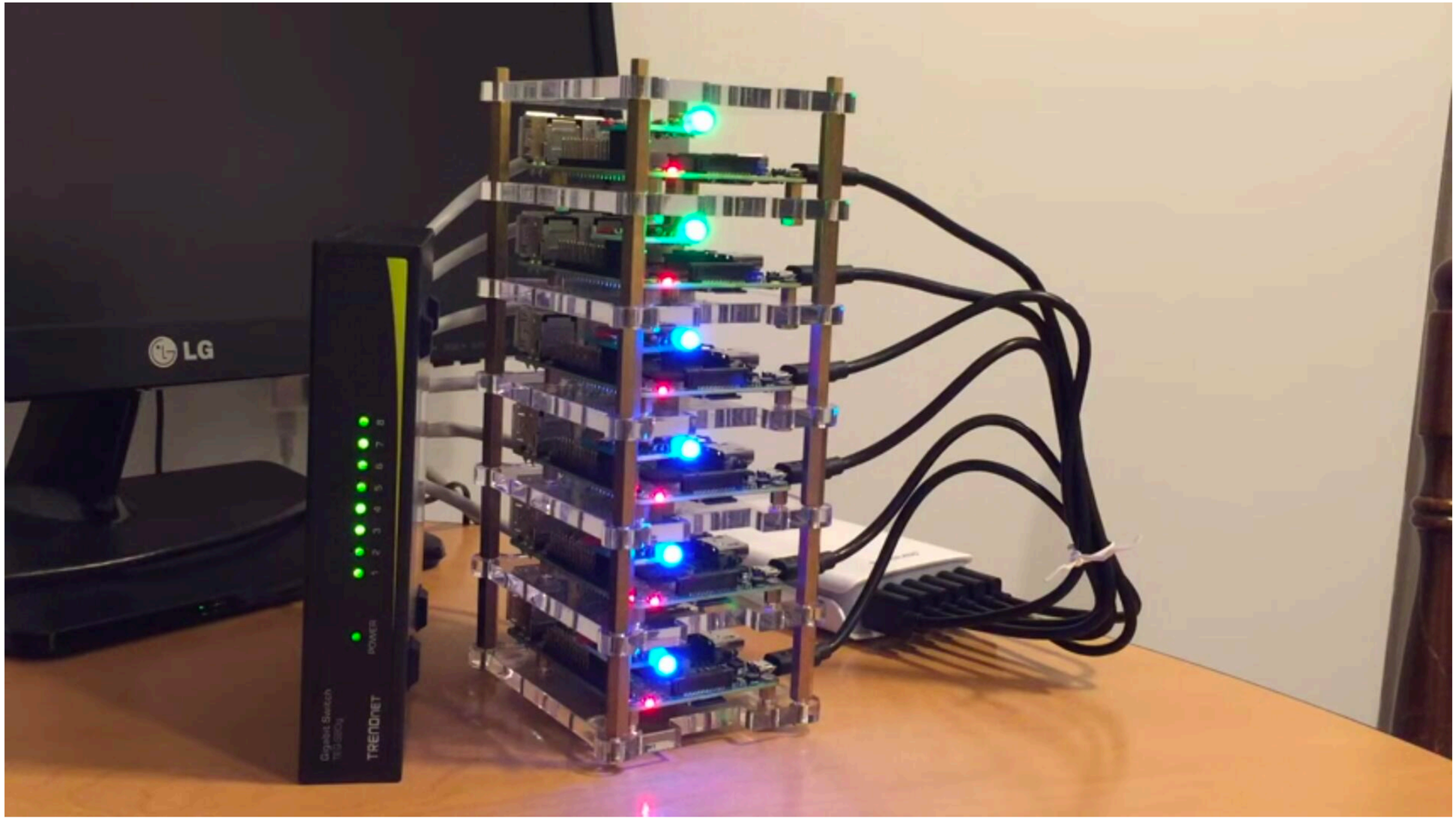
Build farm

- Runs builds and tests at **scale** (1000s to millions per day)
- Support 100s (if not 1000s) of developers
- Provide consistent SLAs

Build farm



Build farm



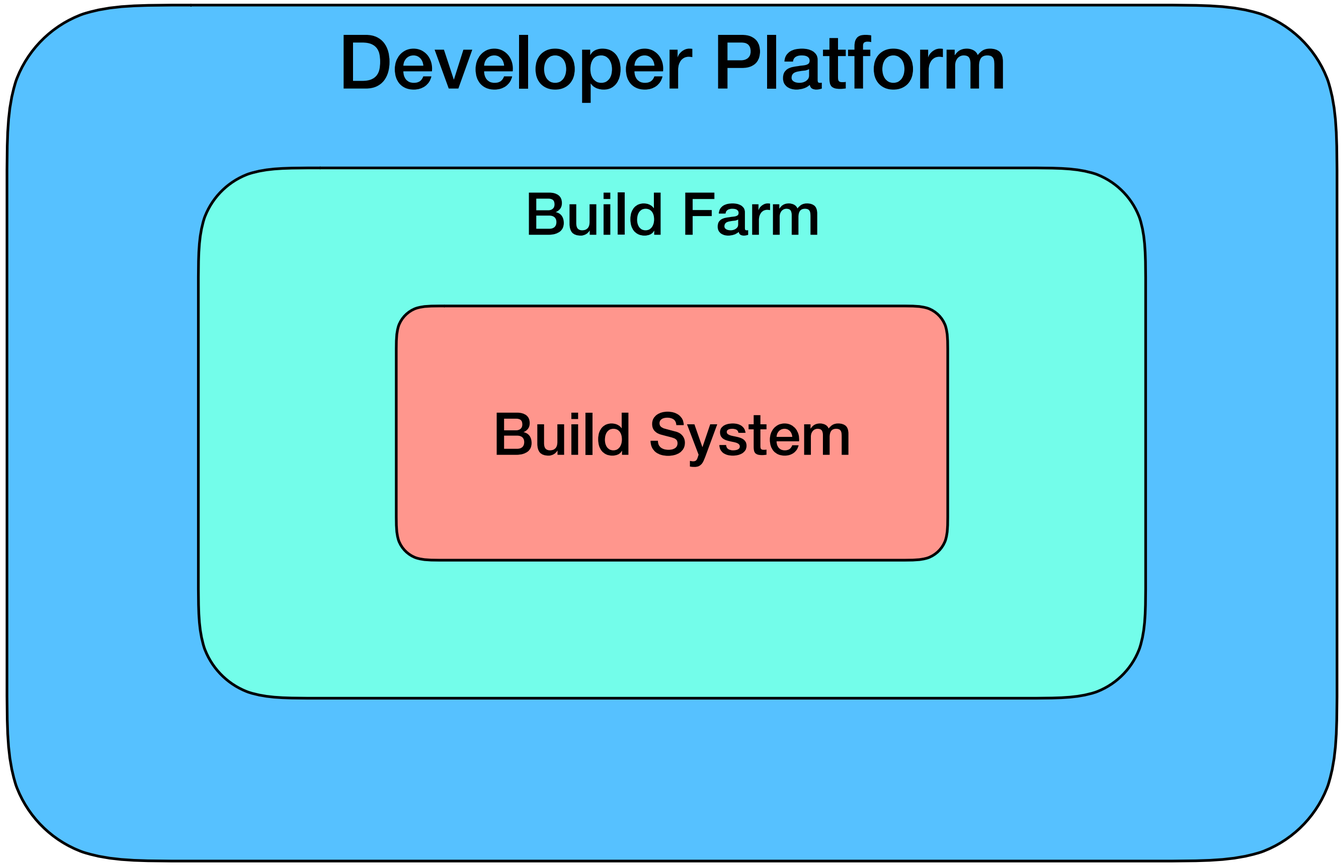
Build farm

What is a Developer Platform?

- Provides key infrastructural for the entire engineering organization to be able to run a build farm at scale.

- Focuses on developer **productivity** and **collaboration**

What is a Developer Platform?



Key goals for developer platforms

- **Speed**: Developers spend lesser and lesser time in waiting for builds, tests and code analysis

- **Manageability**: Make build environments flexible and build infrastructure more efficient and cost effective.

- **Invisible**: Developers should ideally not need to know how the underlying infrastructure or tools work or how they are managed.

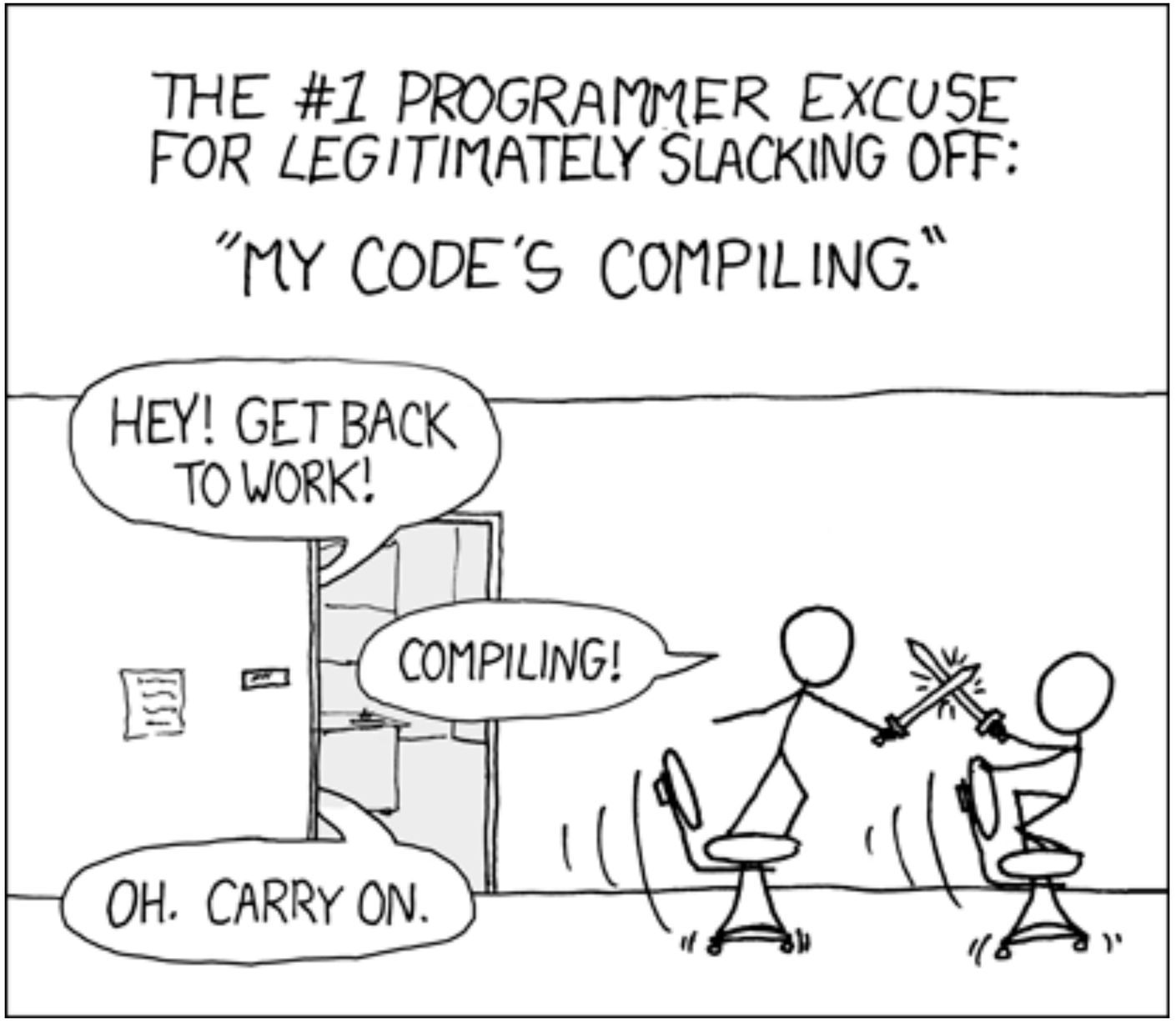
Developer platform

Support **hermetic** build environments for **repeatability**

```
"RepoDigests": [
    "accupara/tannin-build@sha256:3a2415337d626d91cbe334cf2cdfd07c1a451a7a7bb2aa73a7e14e3bf1147180"
],
```

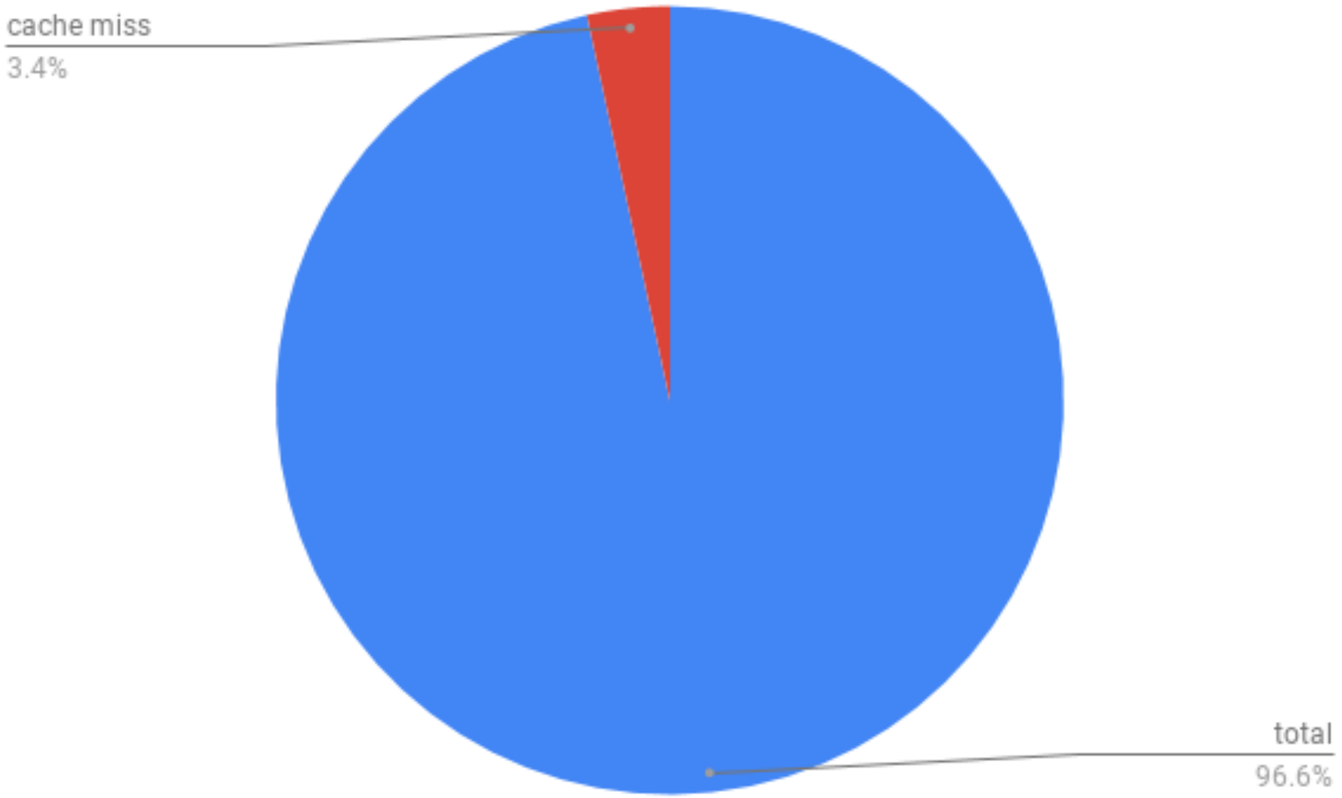
Developer platform

FAST!!


THE #1 PROGRAMMER EXCUSE FOR LEGITIMATELY SLACKING OFF:
"MY CODE'S COMPILING."
HEY! GET BACK TO WORK!
COMPILING!
OH. CARRY ON.

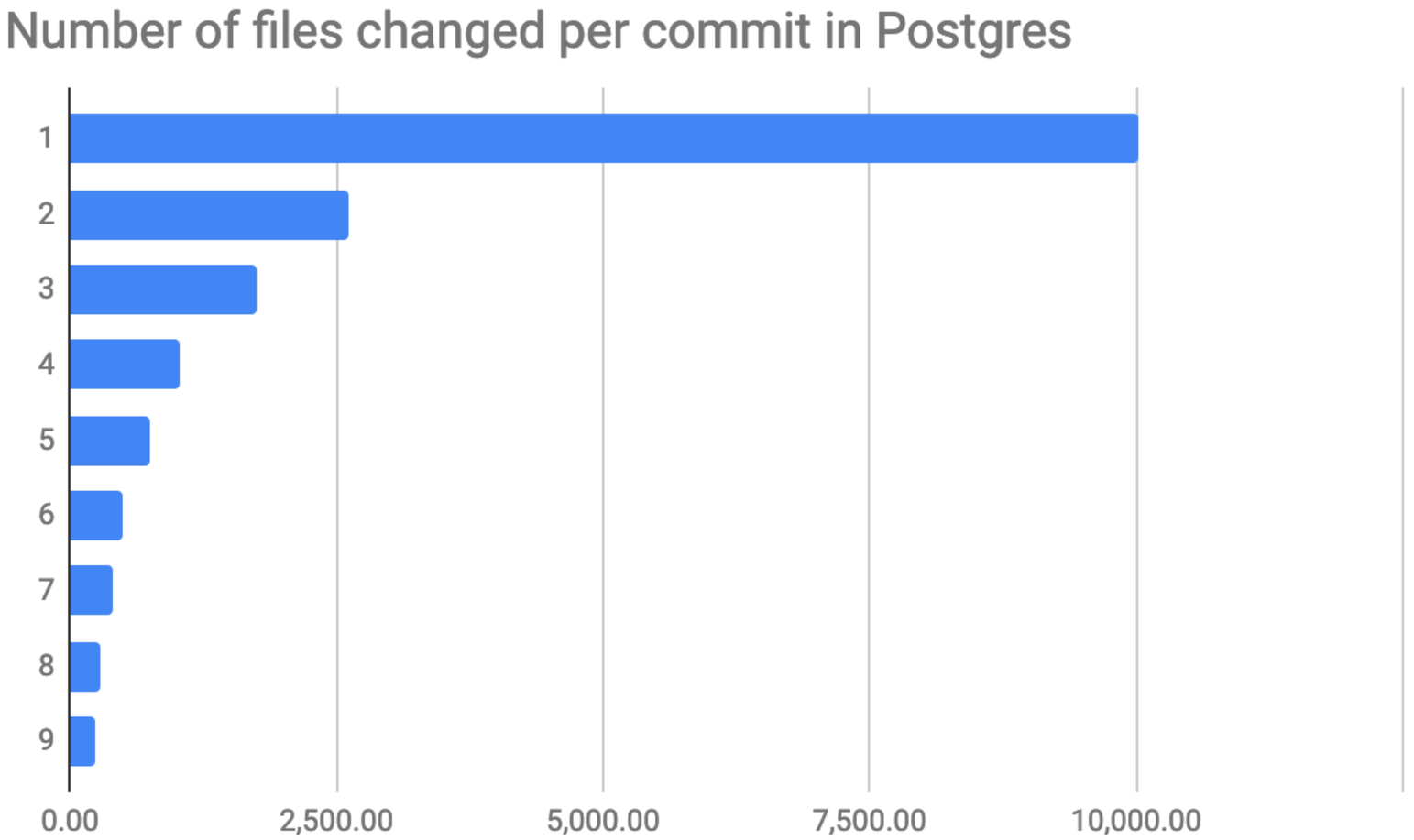
Developer platform

Avoid repeating work



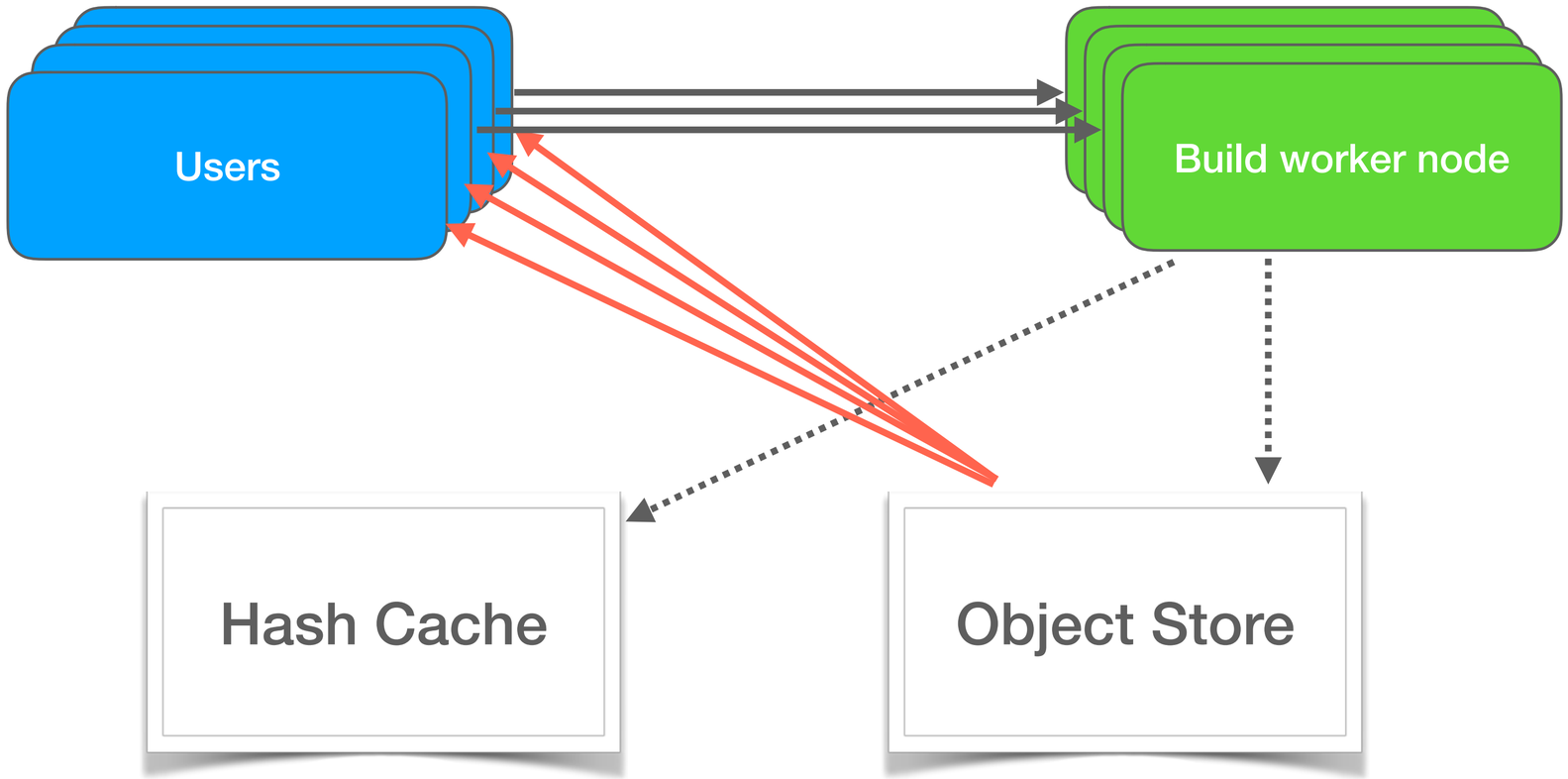
Developer platform

Avoid repeating work



Developer platform

Shared build cache



How we built it using

Postgresql is a great building block for building distributed systems

Where we use Postgres

Job queue

Metrics and analytics

Cluster state management

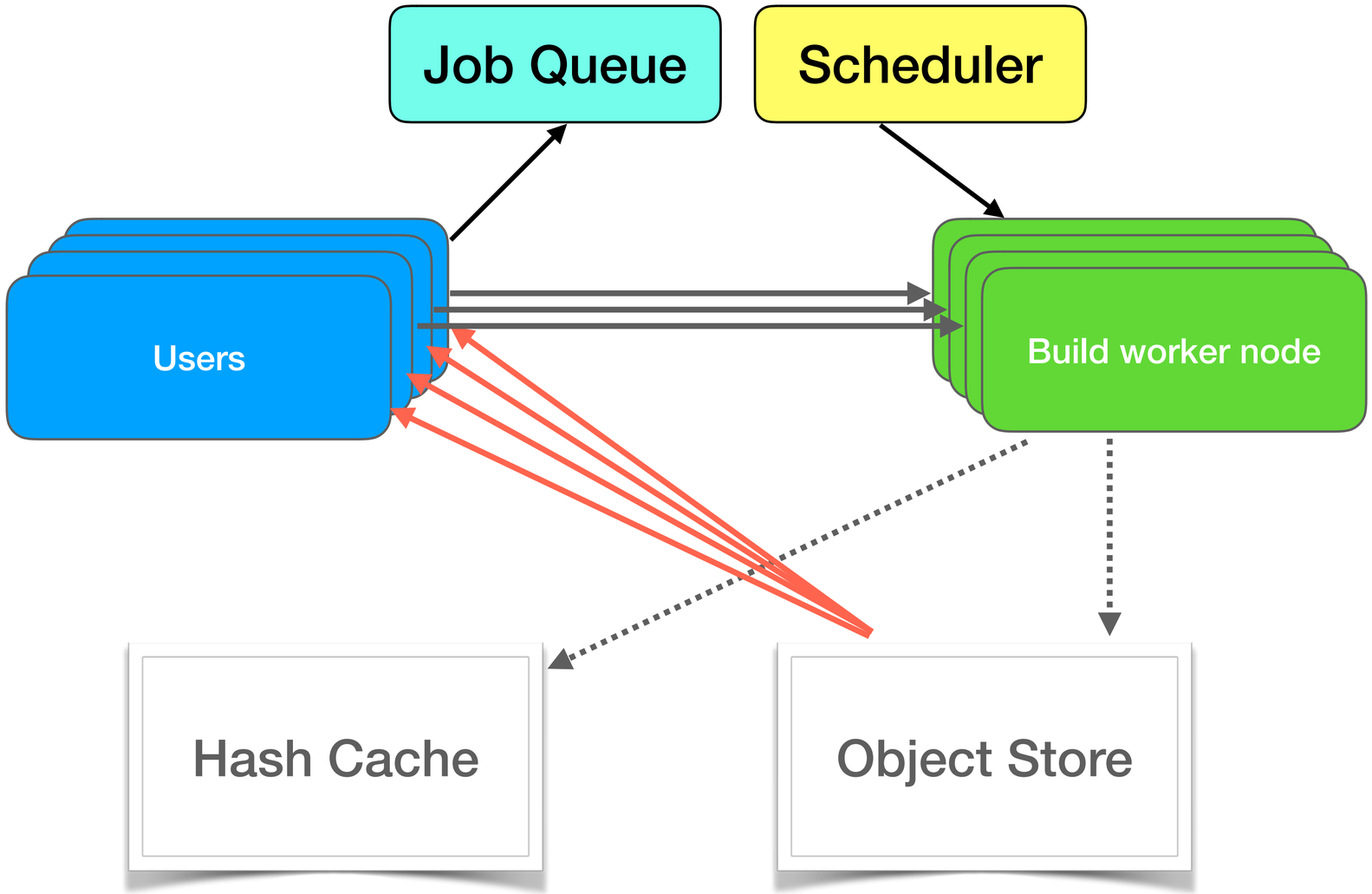
Configuration management

Hash Cache

What is a message queue?

A message queue is a temporary storage for messages between two services

Developer platform



Message queue: Take 1

```
LOCK TABLE build_jobinfo IN ACCESS
EXCLUSIVE MODE;
```

Message queue: Take 2

```
SET TRANSACTION ISOLATION LEVEL
REPEATABLE READ
```

Queue-like workloads are worst case for SERIALIZABLE

Message queue: Take 3

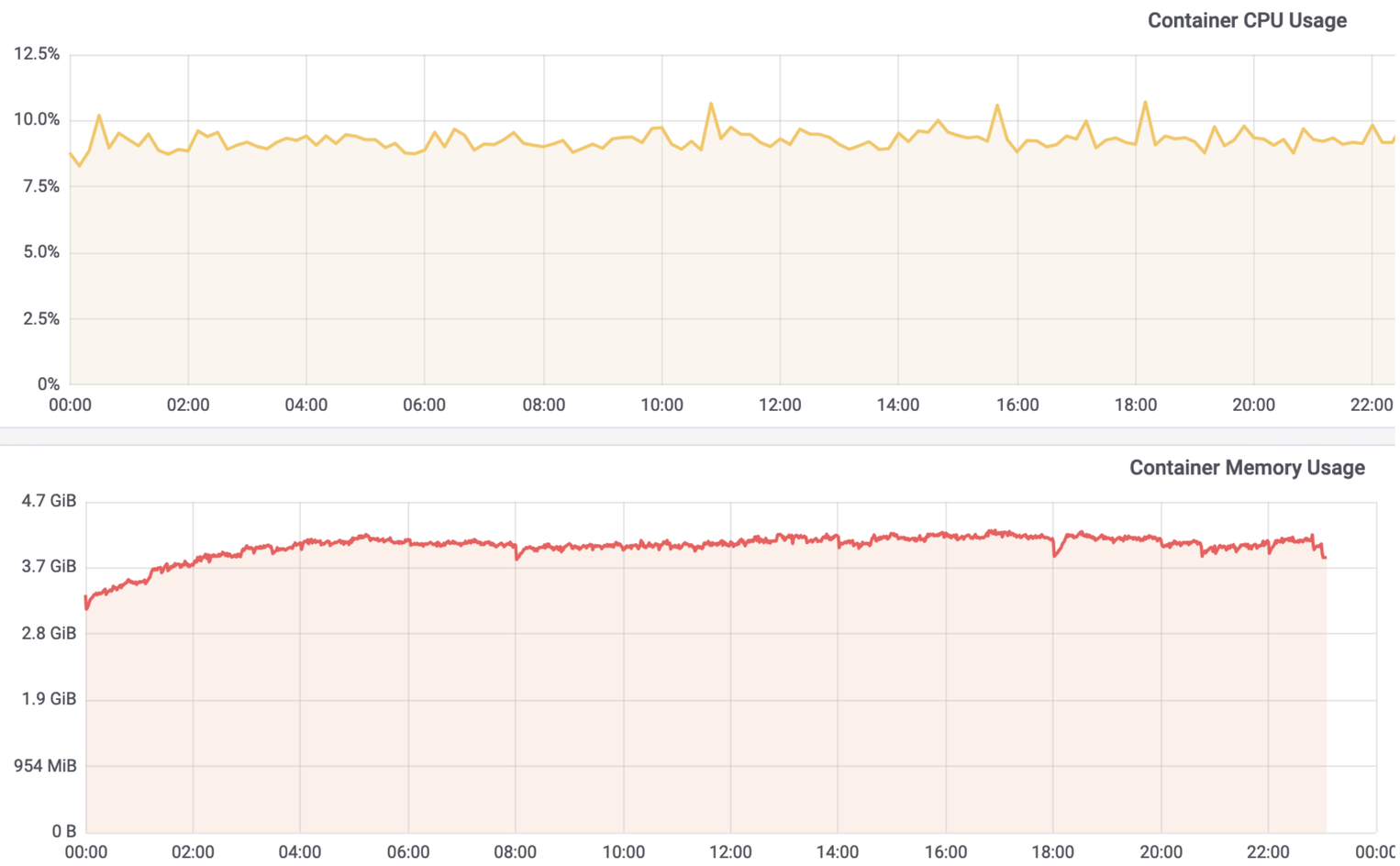
```
UPDATE build_jobinfo SET
“Status” = ‘setup' WHERE
WHERE "id" in
   (SELECT * FROM
    build_jobinfo WHERE "Status"= ‘queued’
    AND "Project_id" IN
       (SELECT "projectId" FROM
        clusterproject WHERE
        "clusterId" = 54)
    ORDER BY \"QueueStartTime\" ASC
    FOR UPDATE
    LIMIT 1)
RETURNING *;
```

Message queue: Take 4

```
UPDATE build_jobinfo SET
“Status” = ‘setup' WHERE
WHERE "id" in
   (SELECT * FROM
    build_jobinfo WHERE "Status"= ‘queued’
    AND "Project_id" IN
       (SELECT "projectId" FROM
        clusterproject WHERE
        "clusterId" = 54)
    ORDER BY \"QueueStartTime\" ASC
    FOR UPDATE SKIP LOCKED
    LIMIT 1)
RETURNING *;
```

Metrics and Analytics

TimescaleDB plugin: out of the box



Metrics and Analytics

- **`timescaledb-tune`**

  ```
  timescaledb-tune --conf-path=/path/to/postgresql.conf
  ```

- **pg_prometheus**

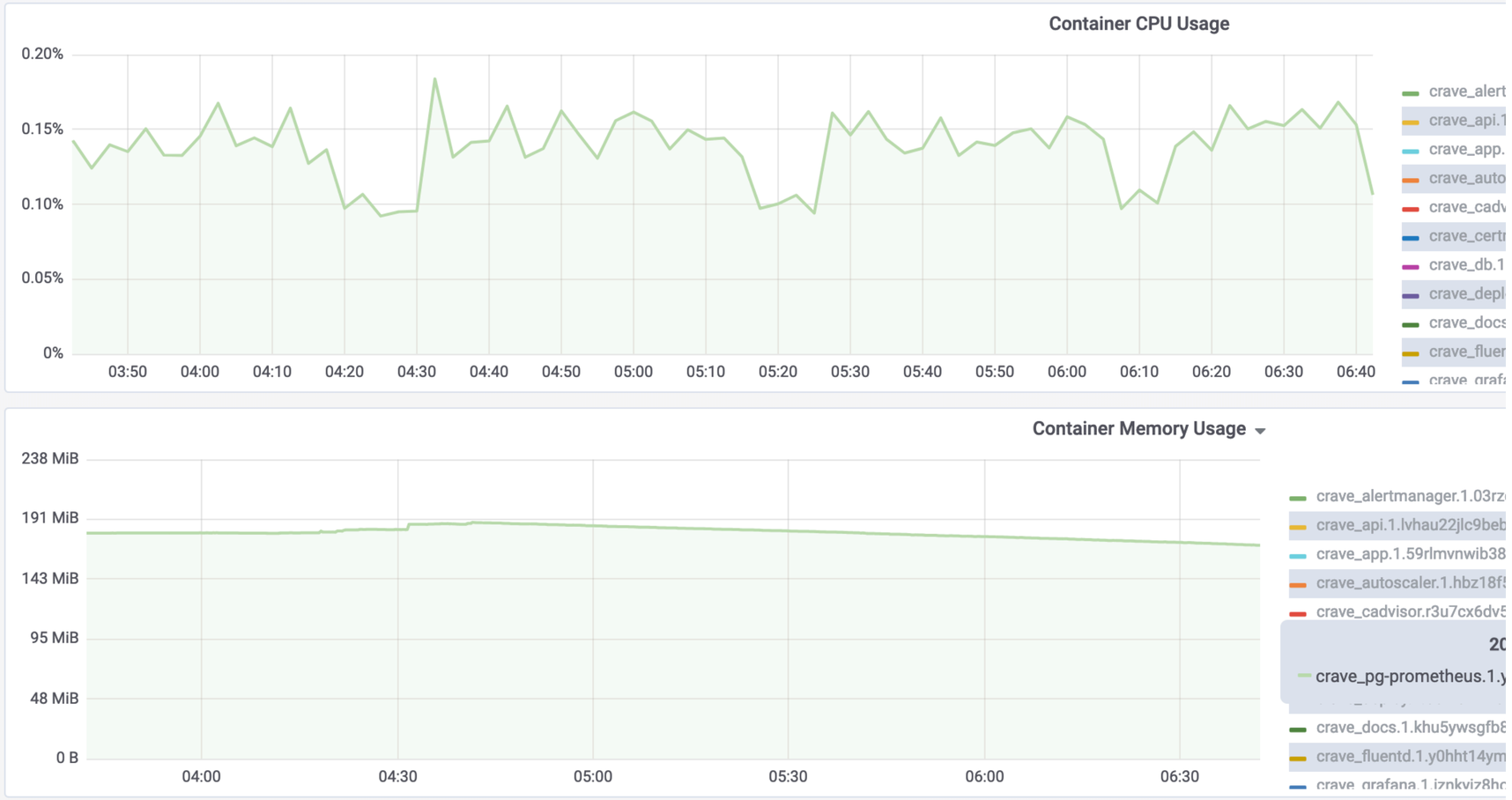
  ```
  synchronous_commit=OFF
  ```

- **prometheus.yml**

```
remote_write:
  - url: "http://prometheus_postgresql_adapter:9201/write"
    queue_config:
      max_samples_per_send: 1000
      batch_send_deadline: 30s
      max_retries: 1
remote_read:
  - url: "http://prometheus_postgresql_adapter:9201/read"
```

Metrics and Analytics

TimescaleDB plugin: after tuning



Cluster state management

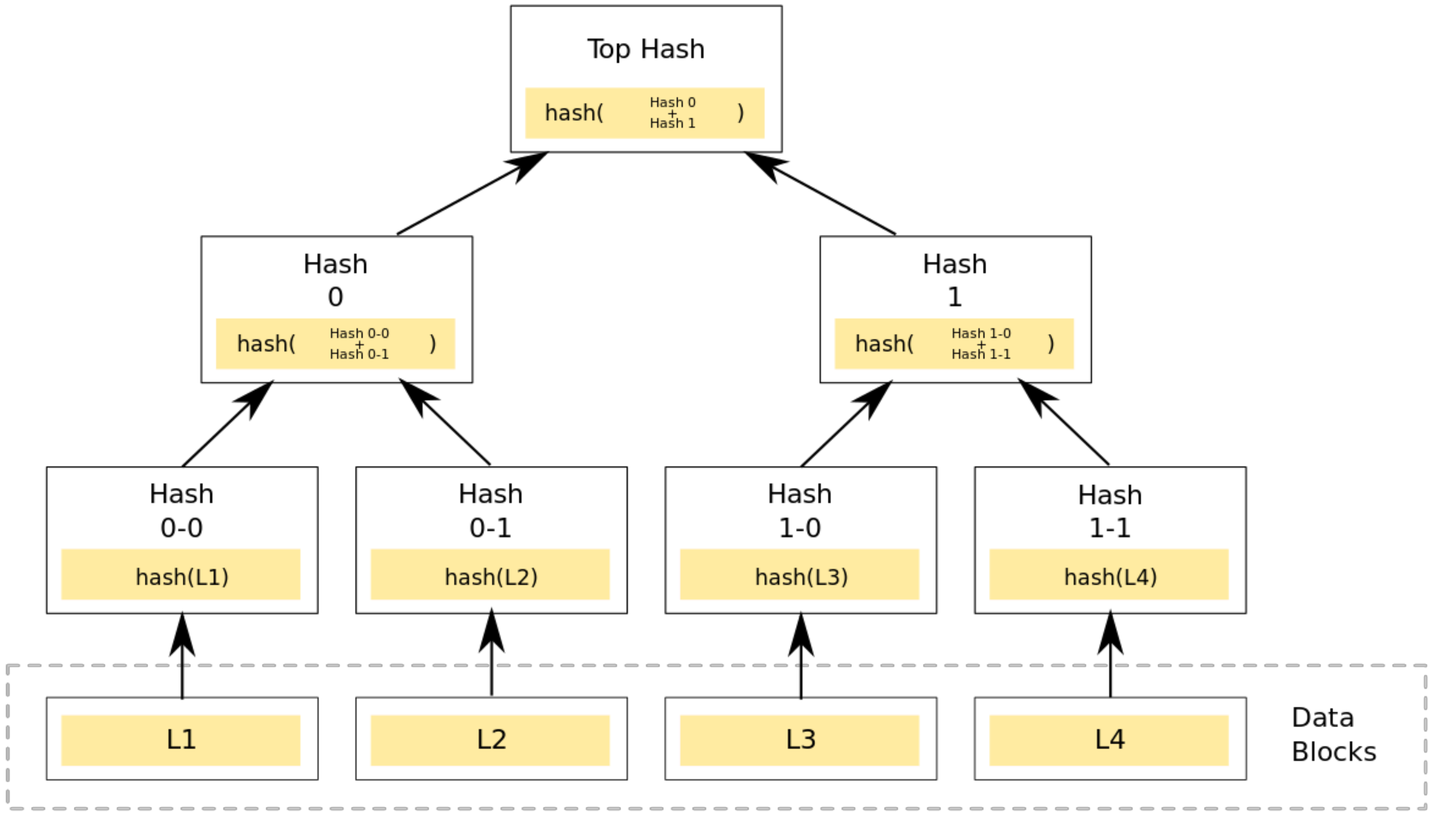
- Cluster state is stored in Postgresql
- Auto scaler, health checker and deployer query cluster state
- Deployer replicates state in the cloud
  - Refresh
  - Plan
  - Apply

Configuration management

- Project configuration is maintained in the DB as **jsonb**
- Auto update of timestamp on update via function
- Clusters poll update timestamp to refresh project configuration state in memory

Hash Cache

hstore: Merkle tree of dependencies and their hashes



Demo

Lets compile

Questions?

ap@crave.io

MANAGED SOFTWARE DEVELOPMENT INFRASTRUCTURE