RethinkDB

RethinkDB - Jepsen Test

Evaluating consistency semantics of RethinkDB - an upcoming open-source database for the real-time web

What is RethinkDB?

- An open source database for the “real-time” web
- Started by two CS students of the Stony Brook University in 2009
- Design Goal: Make building realtime applications easier
- Features
  - *Document Oriented*
  - *ReQL as a query language*
  - *Dynamic Schemas*
  - *Indices + Joins!*
  - *Clustered. Easy-to-scale. Use MapReduce under the hood*
- Built from scratch in C++ and completely open-source on Github

ReQL

```sql
SELECT * FROM users
WHERE name = "Peter"
AND age = 30
```

```
r.table("users").filter({
    "name": "Peter",
    "age": 30
})
```

```sql
SELECT * FROM users
WHERE name LIKE "P%"
```

```
r.table("users").filter(
    r.row['name'].match("^P")
)
```

```sql
SELECT * FROM users
ORDER BY name ASC
```

```
r.table("users").order_by("name")
```

```sql
SELECT * FROM users
ORDER BY name DESC
```

```
r.table("users").order_by(
    r.desc("name")
)
```

Pull Architecture

"Instead of polling for changes, the developer can tell RethinkDB to continuously push updated query results to applications in real-time"

Use-cases

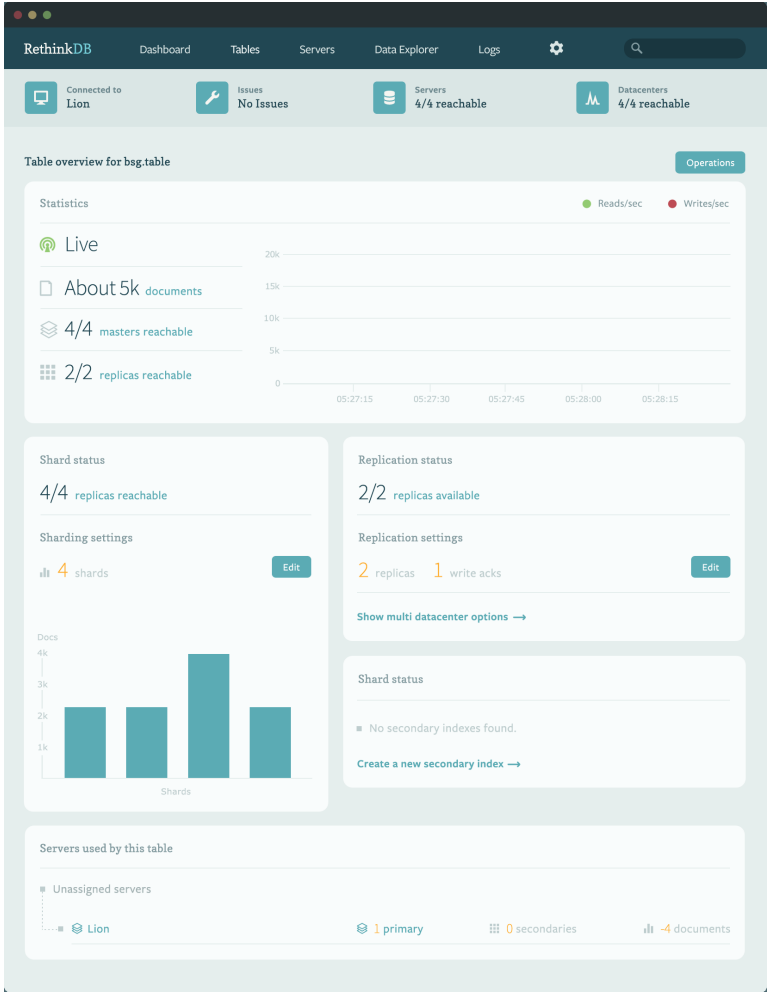
- ➔ Multiplayer games
- ➔ Streaming Analytics
- ➔ Interactive Marketplaces
- ➔ Collaborative web and mobile apps

1-command scaling

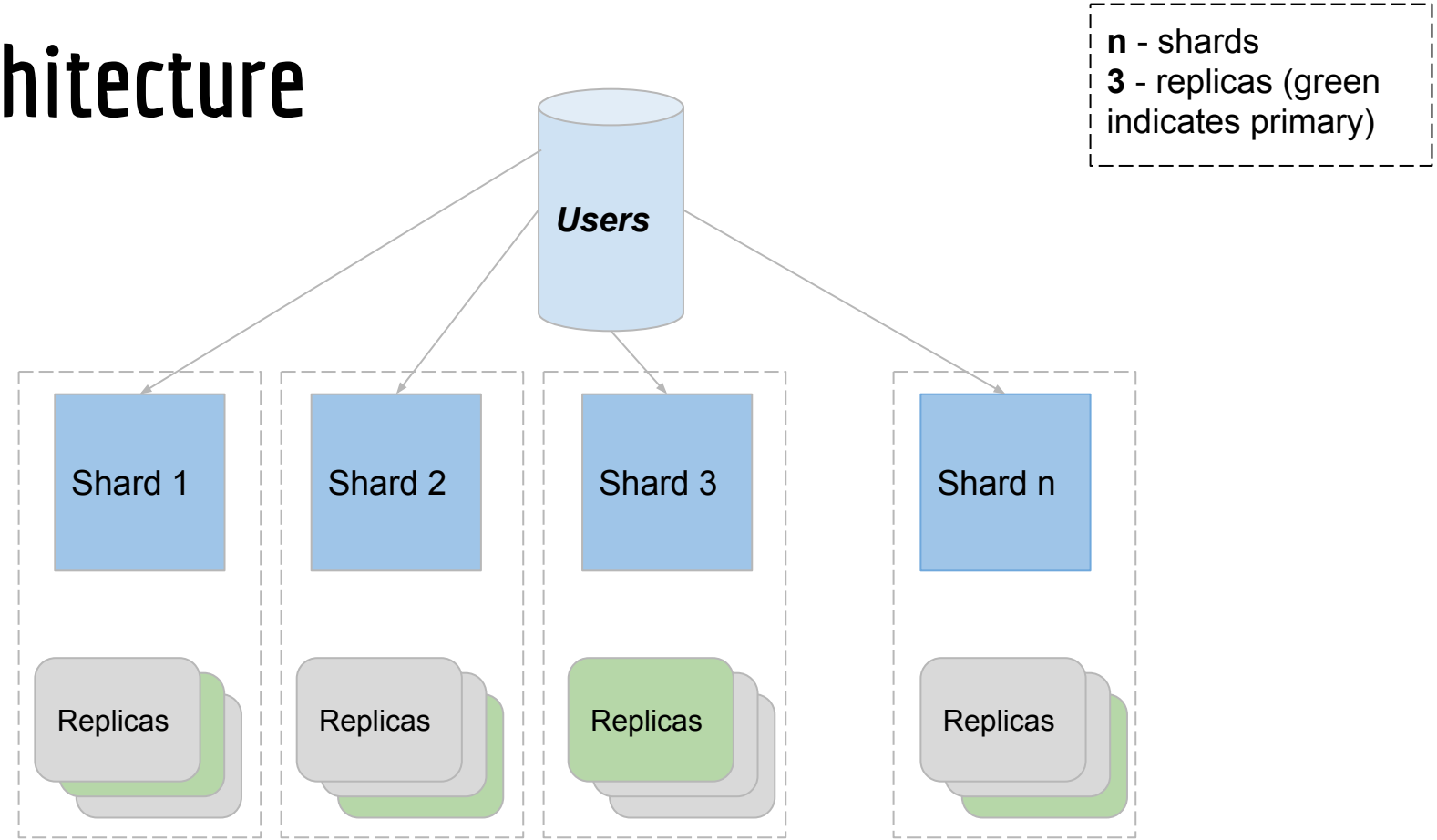
```
// set 5 shards each with 3 replicas
// for the ‘games’ table
r.table(‘games’).reconfigure(shards=5,
replicas=3)
```

1-command pub-sub

```
// listen to the jobs table for changes
r.db(‘rethinkdb’).table(‘jobs’).changes()
```

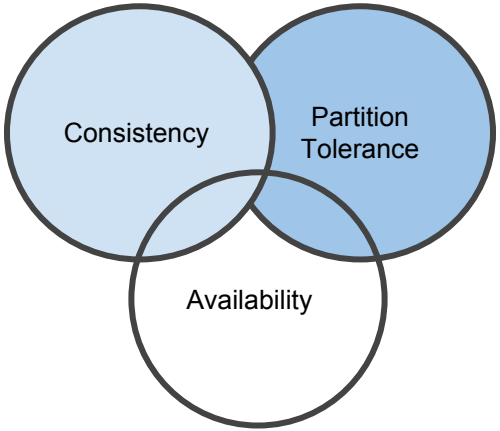


Architecture



Consistency Guarantees

- Architecture is **immediately consistent**
- In essence, it is a **CP (consistent - partition tolerant)** system.



RethinkDB Semantics

- Write acknowledgements
  - *How does RethinkDB confirm that a write was successful?*
  - **Majority** - Majority of replicas confirm successful writes
  - **Single** - Single replica confirms a write.

} Network Failures

- Durability
  - *How does each node in the cluster claim successful writes?*
  - **Hard** - Data committed to disk
  - **Soft** - Data stored in memory

} Disk Failures

- Read mode
  - *Where to read the data from?*
  - **Single** - Return values from memory in primary replica
  - **Majority** - Return values that are safely committed on disk on a majority of replicas
  - **Outdated** - Return values from memory of an arbitrarily selected replica

} Network + Disk Failures

Project Proposal

Goals

- Analyze the consistency properties of RethinkDB.
- Evaluate API semantics around consistency and availability guarantees.

Methodology

- Setup a RethinkDB cluster with one shard, multiple replicas.
- Design test-cases for validating each semantic.
- Use the Jepsen clojure library to run test-cases, mimic network partitions and model failures etc.
- Validate results of operations and derive conclusions.

Questions?

- Prakhar Srivastav (ps2894)
- Ayush Jain (aj2672)

} Team Members