Building a Bitcask-like store in Rust

$ whoami

- Hi! I’m Anirudh Sudhir
- Sophomore at PES University, Bangalore
- Systems enthusiast
- Current interests include distributed systems and databases
- Internethome: https://sudhir.live

Motivation

- Build a Rust-based project
- Explore databases
- Discovered https://github.com/pingcap/talent-plan/ - Distributed systems and databases in Rust
- Started out with the key-value database project - Bitcask-like store in Rust

What is Bitcask

- “A Log-Structured Hash Table for Fast Key/Value Data”
- Initially designed as a storage engine for the Riak distributed database
- Inspired by concepts from log-structured file systems, log compaction in LSM trees

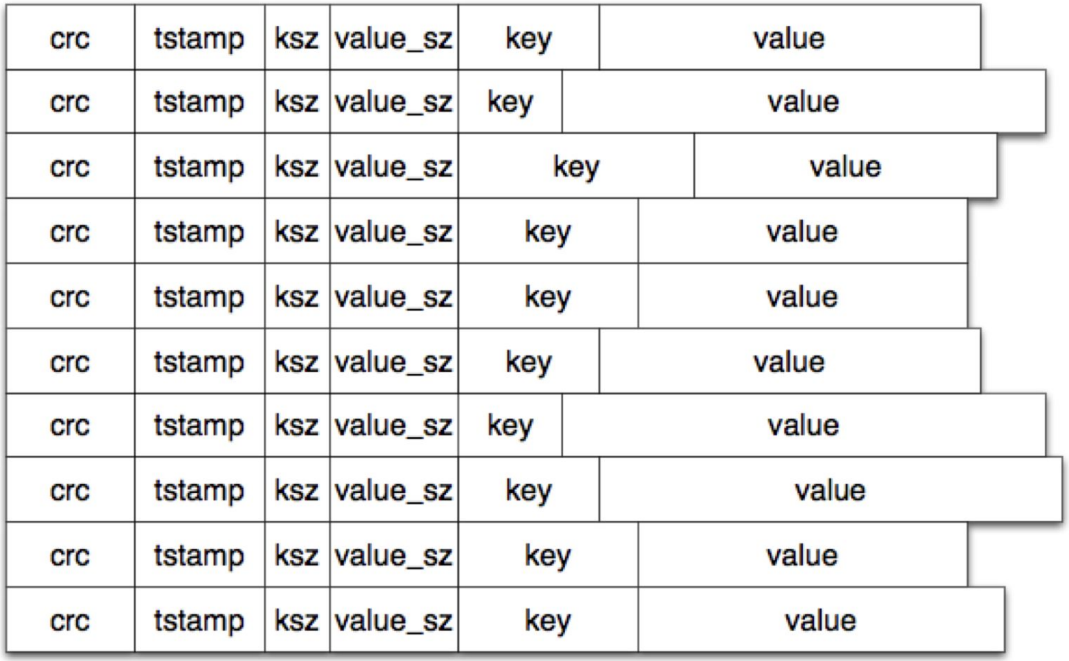
Bitcask architecture - Overview

- Two primary components:
  - Keydir - In-memory index
  - Data files - On-disk log

- A bitcask instance - directory with several log files
  - One active WAL
  - Multiple immutable logs

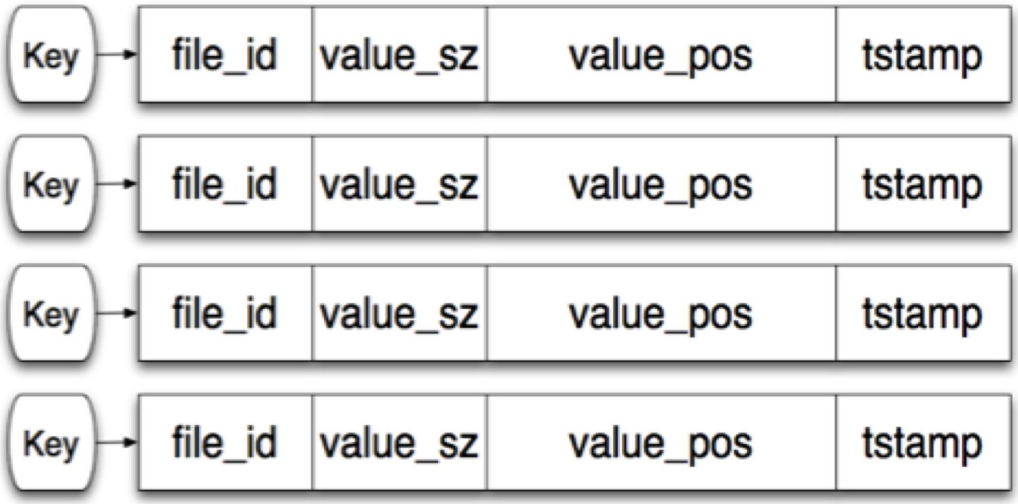
Bitcask architecture - Datafiles

- Key-value entries are appended to the active on-disk log
- High write throughput - single disk seek



Bitcask architecture - Keydir

- A HashTable that maps keys to log pointers (offsets of the entries in the logs) and additional metadata
- Key retrieval - few in-memory operations and a single disk seek



Bitcask architecture - Log compaction

- Essential for efficient disk utilisation - Remove stale entries
- Immutable files are processed
- A new set of data files with only latest live keys
- References
  - The Bitcask paper
  - Arpit Bhayani's blog

Implementation

Implementing Bitcask

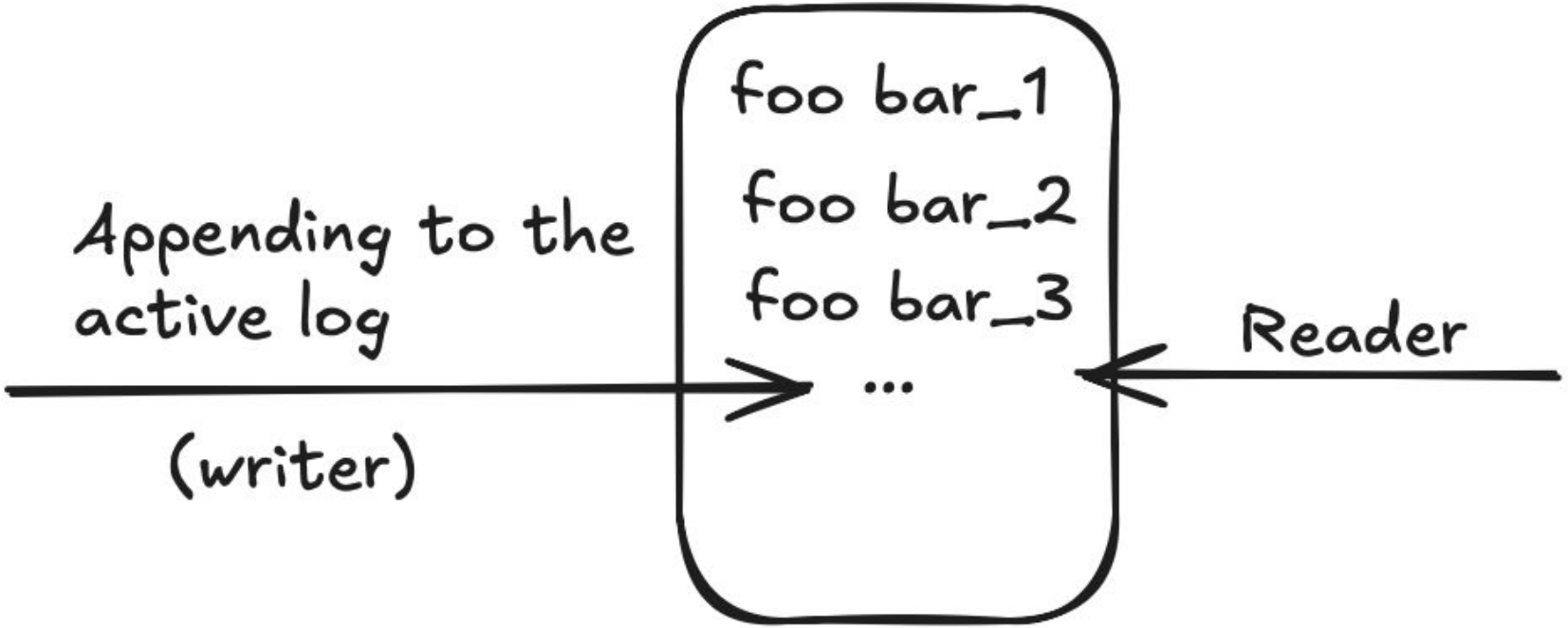
- Initially - in-memory key-value store
- Log entries (key-value pairs) - serialised to MessagePack and flushed to disk
- Index maps keys to log pointers


hobbes(main*)$ ls
CONTRIBUTING.md Cargo.toml README.md hobbes-server target
Cargo.lock LICENSE hobbes src tests
hobbes(main*)$ ./hobbes get foo
Key not found
hobbes(main*)$ ./hobbes set foo bar
hobbes(main*)$ ./hobbes get foo
bar
hobbes(main*)$ ls
CONTRIBUTING.md Cargo.toml README.md hobbes-server src tests
Cargo.lock LICENSE hobbes hobbes-store target
hobbes(main*)$ hobbes-store
hobbes-store(main*)$ ls
1.db
hobbes-store(main*)$ xxd 1.db
00000000: 92a3 666f 6fa3 6261 72 ..foo.bar
hobbes-store(main*)$ ../
hobbes(main*)$
1: ..s/code/hobbes
2: ./hobbes-server

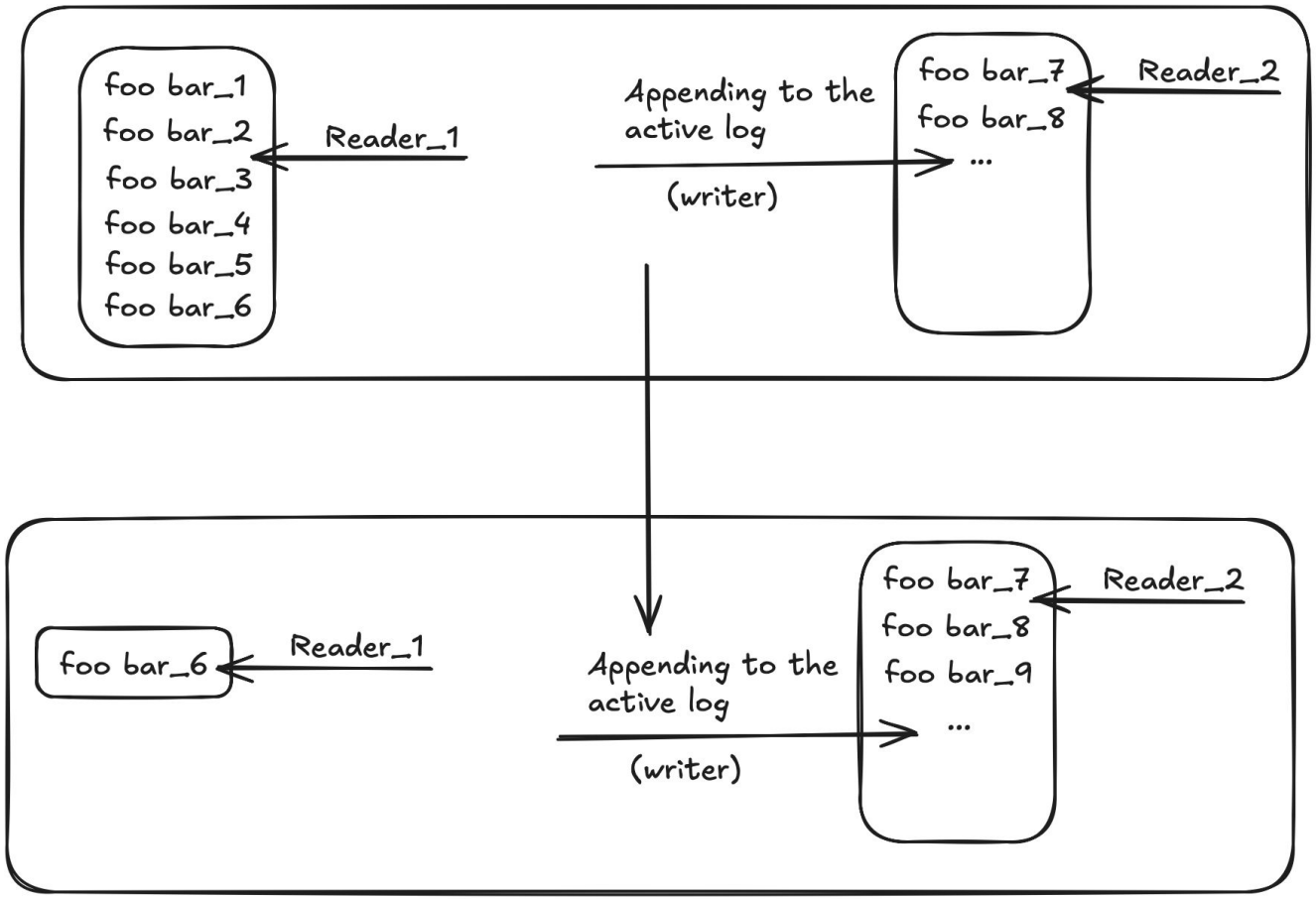
Log compaction

- Once the log hits a file size threshold
  - Append future entries to a new log
  - Set the current log as immutable and compact it
- Compaction - replay the log and store only the latest versions of live keys

Log compaction



Log compaction



```
hobbes  0 hobbes-server -  1 zsh *                                   anirudh
hobbes(main*)$ ls
CONTRIBUTING.md    LICENSE           hobbes           src
Cargo.lock         README.md         hobbes-server    target
Cargo.toml         compaction_demo.sh hobbes-store    tests
hobbes(main*)$ l hobbes-store
total 0
drwxr-xr-x@  3 anirudh  staff    96B Oct 18 10:15 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 10:15 ..
-rw-r--r--@  1 anirudh  staff     0B Oct 18 10:15 1.db
hobbes(main*)$ xxd hobbes-store/1.db
hobbes(main*)$ bat compaction_demo.sh
```

```
       File: compaction_demo.sh
   1   #!/bin/zsh
   2
   3   for ((i = 0; i < 1000; i++)); do
   4     ./hobbes set foo "bar_$i"
   5   done
```

```
hobbes(main*)$ source compaction_demo.sh
hobbes(main*)$ l hobbes-store
total 16
drwxr-xr-x@  4 anirudh  staff   128B Oct 18 10:15 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 10:15 ..
-rw-r--r--@  1 anirudh  staff    13B Oct 18 10:15 1.db
-rw-r--r--@  1 anirudh  staff   2.8K Oct 18 10:15 2.db
hobbes(main*)$
```

```
1: tmux ×  2: ..s/code/hobbes ×  +
```

```
hobbes  0 hobbes-server -  1 zsh *                                    anirudh
total 0
drwxr-xr-x@  3 anirudh  staff    96B Oct 18 10:15 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 10:15 ..
-rw-r--r--@  1 anirudh  staff     0B Oct 18 10:15 1.db
hobbes(main*)$ xxd hobbes-store/1.db
hobbes(main*)$ bat compaction_demo.sh
```

File: **compaction_demo.sh**

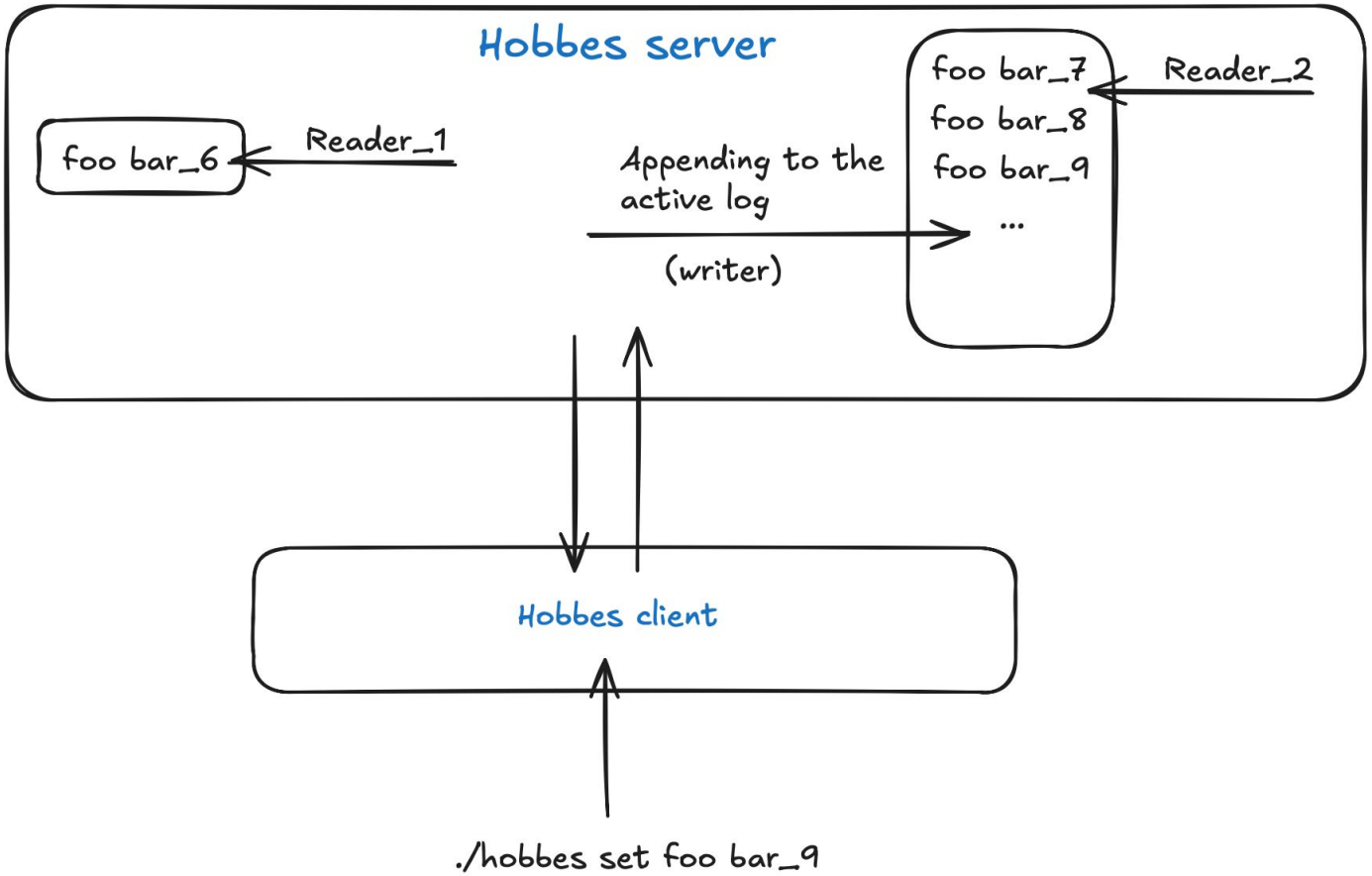
```
1   #!/bin/zsh
2
3   for ((i = 0; i < 1000; i++)); do
4     ./hobbes set foo "bar_$i"
5   done
```

```
hobbes(main*)$ source compaction_demo.sh
hobbes(main*)$ l hobbes-store
total 16
drwxr-xr-x@  4 anirudh  staff   128B Oct 18 10:15 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 10:15 ..
-rw-r--r--@  1 anirudh  staff    13B Oct 18 10:15 1.db
-rw-r--r--@  1 anirudh  staff   2.8K Oct 18 10:15 2.db
hobbes(main*)$ hobbes-store
hobbes-store(main*)$ ls
1.db 2.db
hobbes-store(main*)$ xxd 1.db
00000000: 92a3 666f 6fa7 6261 725f 3737 37         ..foo.bar_777
hobbes-store(main*)$
```

```
1: tmux ×  2: ..s/code/hobbes ×  +
```


hobbes 0 hobbes-server - 1 zsh * anirudh
000009a0: 6172 5f39 3637 92a3 666f 6fa7 6261 725f ar_967..foo.bar_
000009b0: 3936 3892 a366 6f6f a762 6172 5f39 3639 968..foo.bar_969
000009c0: 92a3 666f 6fa7 6261 725f 3937 3092 a366 ..foo.bar_970..f
000009d0: 6f6f a762 6172 5f39 3731 92a3 666f 6fa7 oo.bar_971..foo.
000009e0: 6261 725f 3937 3292 a366 6f6f a762 6172 bar_972..foo.bar
000009f0: 5f39 3733 92a3 666f 6fa7 6261 725f 3937 _973..foo.bar_97
00000a00: 3492 a366 6f6f a762 6172 5f39 3735 92a3 4..foo.bar_975..
00000a10: 666f 6fa7 6261 725f 3937 3692 a366 6f6f foo.bar_976..foo
00000a20: a762 6172 5f39 3737 92a3 666f 6fa7 6261 .bar_977..foo.ba
00000a30: 725f 3937 3892 a366 6f6f a762 6172 5f39 r_978..foo.bar_9
00000a40: 3739 92a3 666f 6fa7 6261 725f 3938 3092 79..foo.bar_980.
00000a50: a366 6f6f a762 6172 5f39 3831 92a3 666f .foo.bar_981..fo
00000a60: 6fa7 6261 725f 3938 3292 a366 6f6f a762 o.bar_982..foo.b
00000a70: 6172 5f39 3833 92a3 666f 6fa7 6261 725f ar_983..foo.bar_
00000a80: 3938 3492 a366 6f6f a762 6172 5f39 3835 984..foo.bar_985
00000a90: 92a3 666f 6fa7 6261 725f 3938 3692 a366 ..foo.bar_986..f
00000aa0: 6f6f a762 6172 5f39 3837 92a3 666f 6fa7 oo.bar_987..foo.
00000ab0: 6261 725f 3938 3892 a366 6f6f a762 6172 bar_988..foo.bar
00000ac0: 5f39 3839 92a3 666f 6fa7 6261 725f 3939 _989..foo.bar_99
00000ad0: 3092 a366 6f6f a762 6172 5f39 3931 92a3 0..foo.bar_991..
00000ae0: 666f 6fa7 6261 725f 3939 3292 a366 6f6f foo.bar_992..foo
00000af0: a762 6172 5f39 3933 92a3 666f 6fa7 6261 .bar_993..foo.ba
00000b00: 725f 3939 3492 a366 6f6f a762 6172 5f39 r_994..foo.bar_9
00000b10: 3935 92a3 666f 6fa7 6261 725f 3939 3692 95..foo.bar_996.
00000b20: a366 6f6f a762 6172 5f39 3937 92a3 666f .foo.bar_997..fo
00000b30: 6fa7 6261 725f 3939 3892 a366 6f6f a762 o.bar_998..foo.b
00000b40: 6172 5f39 3939 ar_999
hobbes-store(main*)$
1: tmux
2: ..s/code/hobbes

Client-server architecture



Pluggable storage engines

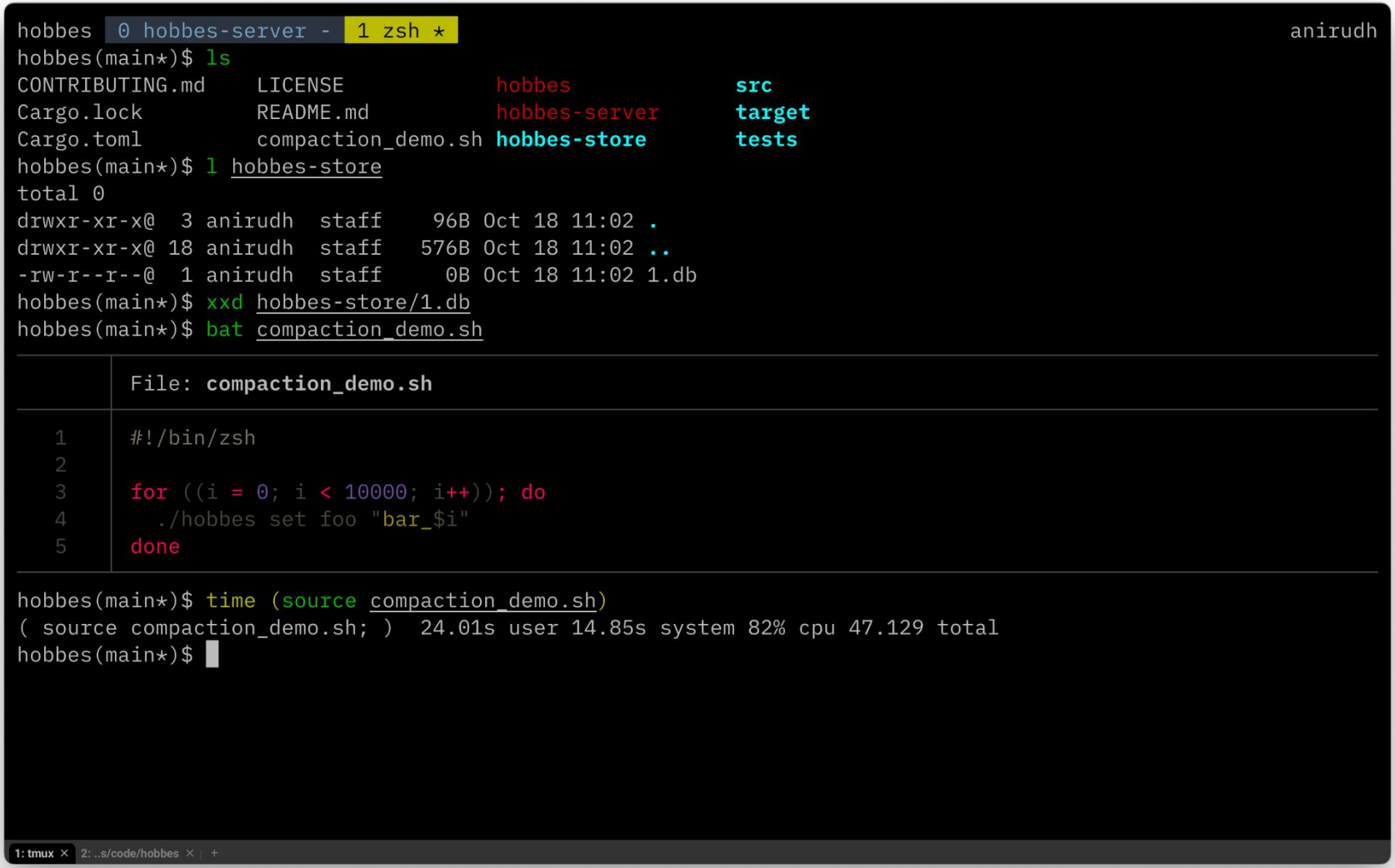
- Hobbes can utilise the
  - hobbes engine - Bitcask architecture
  - sled engine - An alternate embedded database

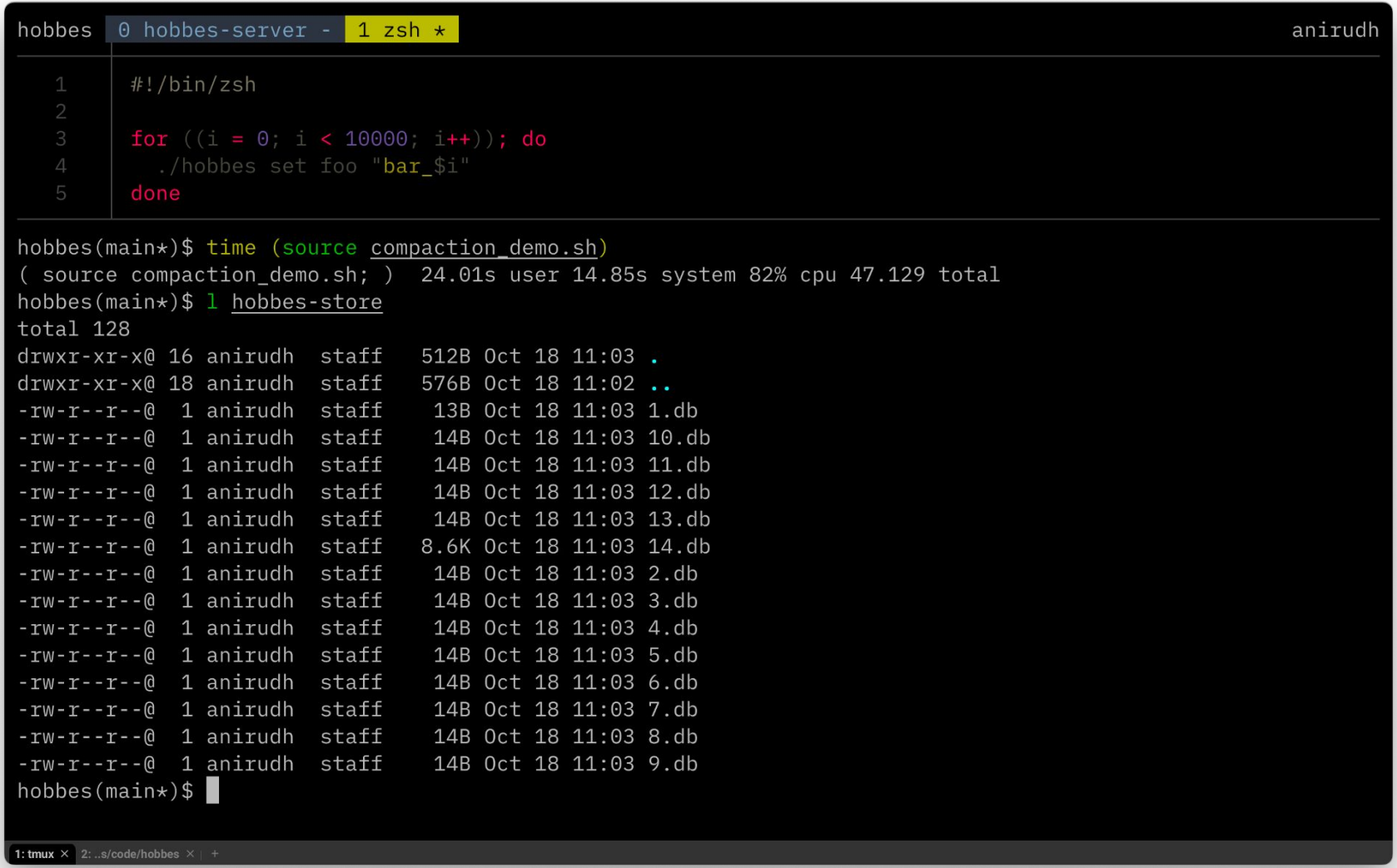
Future plans

- Multi-threaded store, lock-free readers
- Improved log compaction - recursive merging of logs
- Persistent connection between client and server

Early Benchmarks

10,000 key operations - hobbes engine (with log compaction)


hobbes 0 hobbes-server - 1 zsh *
anirudh
hobbes(main*)$ ls
CONTRIBUTING.md LICENSE hobbes src
Cargo.lock README.md hobbes-server target
Cargo.toml compaction_demo.sh hobbes-store tests
hobbes(main*)$ l hobbes-store
total 0
drwxr-xr-x@ 3 anirudh staff 96B Oct 18 11:02 .
drwxr-xr-x@ 18 anirudh staff 576B Oct 18 11:02 ..
-rw-r--r--@ 1 anirudh staff 0B Oct 18 11:02 1.db
hobbes(main*)$ xxd hobbes-store/1.db
hobbes(main*)$ bat compaction_demo.sh
File: compaction_demo.sh
1 #!/bin/zsh
2
3 for ((i = 0; i < 10000; i++)); do
4 ./hobbes set foo "bar_$i"
5 done
hobbes(main*)$ time (source compaction_demo.sh)
(source compaction_demo.sh;) 24.01s user 14.85s system 82% cpu 47.129 total
hobbes(main*)$
1: tmux
2: ..s/code/hobbes
+


hobbes 0 hobbes-server - 1 zsh * anirudh
1 #!/bin/zsh
2
3 for ((i = 0; i < 10000; i++)); do
4 ./hobbes set foo "bar_$i"
5 done
hobbes(main*)$ time (source compaction_demo.sh)
(source compaction_demo.sh;) 24.01s user 14.85s system 82% cpu 47.129 total
hobbes(main*)$ l hobbes-store
total 128
drwxr-xr-x@ 16 anirudh staff 512B Oct 18 11:03 .
drwxr-xr-x@ 18 anirudh staff 576B Oct 18 11:02 ..
-rw-r--r--@ 1 anirudh staff 13B Oct 18 11:03 1.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 10.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 11.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 12.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 13.db
-rw-r--r--@ 1 anirudh staff 8.6K Oct 18 11:03 14.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 2.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 3.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 4.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 5.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 6.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 7.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 8.db
-rw-r--r--@ 1 anirudh staff 14B Oct 18 11:03 9.db
hobbes(main*)$
1: tmux
2: ..s/code/hobbes

```
0  0 hobbes-server -  1 zsh *                                    anirudh
hobbes(main)$ l hobbes-store
total 128
drwxr-xr-x@ 16 anirudh  staff   512B Oct 18 11:03 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 11:31 ..
-rw-r--r--@  1 anirudh  staff    13B Oct 18 11:03 1.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 10.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 11.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 12.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 13.db
-rw-r--r--@  1 anirudh  staff   8.6K Oct 18 11:03 14.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 2.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 3.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 4.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 5.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 6.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 7.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 8.db
-rw-r--r--@  1 anirudh  staff    14B Oct 18 11:03 9.db
hobbes(main)$ xxd ./hobbes-store/6.db
00000000: 92a3 666f 6fa8 6261 725f 3433 3638       ..foo.bar_4368
hobbes(main)$ xxd ./hobbes-store/4.db
00000000: 92a3 666f 6fa8 6261 725f 3239 3338       ..foo.bar_2938
hobbes(main)$ xxd ./hobbes-store/13.db
00000000: 92a3 666f 6fa8 6261 725f 3933 3733       ..foo.bar_9373
hobbes(main)$
```

10,000 key operations - sled engine

```
hobbes  0 zsh *  1 hobbes-server -                                    anirudh
hobbes(main)$ l sled-store
total 16
drwxr-xr-x@  5 anirudh  staff   160B Oct 18 11:17 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 11:17 ..
drwxr-xr-x@  2 anirudh  staff    64B Oct 18 11:17 blobs
-rw-r--r--@  1 anirudh  staff    62B Oct 18 11:17 conf
-rw-r--r--@  1 anirudh  staff    96B Oct 18 11:17 db
hobbes(main)$ xxd ./sled-store/db
00000000: ffba c70f ffff ffff ffff ff7f 0000 0000  ................
00000010: 0000 0080 3f7f d0fb 0600 0000 71a2 c678  ....?.......q..x
00000020: 0501 0001 00eb 2183 a108 0800 0200 0000  ......!.........
00000030: 0000 0000 00f1 2273 6f08 0a00 0300 0000  ......"so.......
00000040: 0000 0001 0100 02cb 82ba 8d06 1100 000f  ................
00000050: 5f5f 736c 6564 5f5f 6465 6661 756c 7403  __sled__default.
hobbes(main)$ time (source compaction_demo.sh)
( source compaction_demo.sh; )  25.08s user 15.19s system 46% cpu 1:25.93 total
hobbes(main)$ l sled-store
total 528
drwxr-xr-x@  5 anirudh  staff   160B Oct 18 11:17 .
drwxr-xr-x@ 18 anirudh  staff   576B Oct 18 11:17 ..
drwxr-xr-x@  2 anirudh  staff    64B Oct 18 11:17 blobs
-rw-r--r--@  1 anirudh  staff    62B Oct 18 11:17 conf
-rw-r--r--@  1 anirudh  staff   221K Oct 18 11:19 db
hobbes(main)$
```

Call to action

- Pingcap TalentPlan course - https://github.com/pingcap/talent-plan/
- Bitcask Paper - https://riak.com/assets/bitcask-intro.pdf
- Github - https://github.com/anirudhsudhir/