PingCAP

Using Rust to Build a Distributed Transactional Key-Value Database

LiuTang | tl@pingcap.com

PingCAP

About me

- Chief Architect at PingCAP
- TiDB and TiKV
- Open source projects
  - LedisDB
  - go-mysql
  - go-mysql-elasticsearch
  - rust-prometheus
  - ...

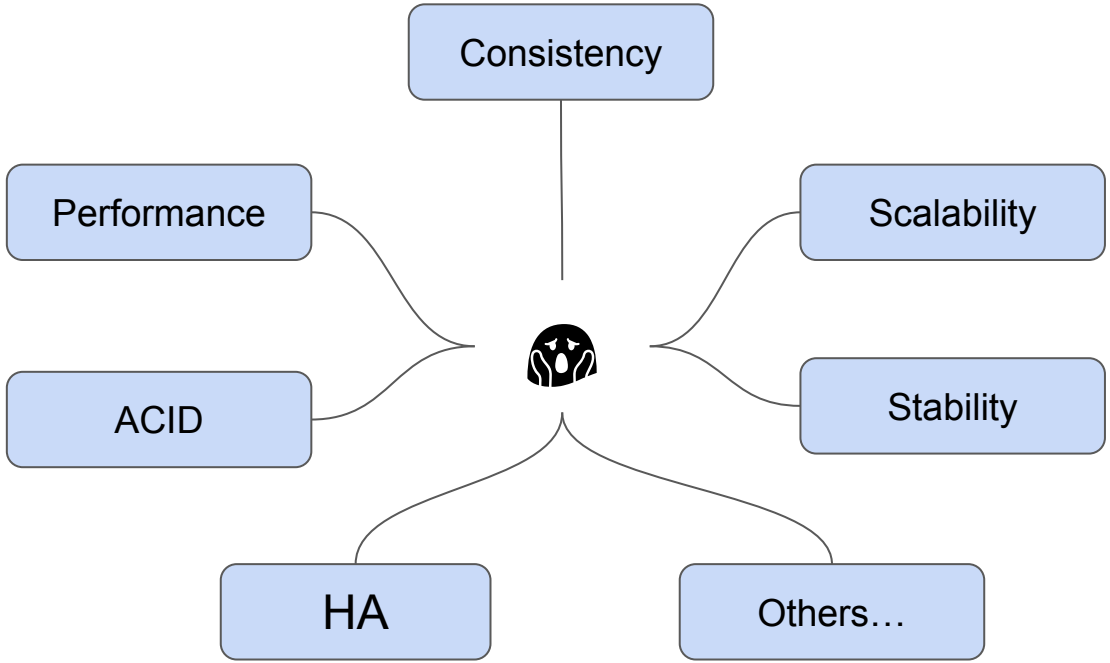
PingCAP

Agenda

- Introduction
- Hierarchy
  - Storage
  - Raft
  - Transaction
  - RPC Framework
  - Monitor
  - Test
- Combine them all

PingCAP

When we want to build a distributed transactional key-value database...


Consistency
Performance
Scalability
ACID
Stability
HA
Others…

PingCAP

A High Building,
A Low Foundation

PingCAP

Language

PingCAP

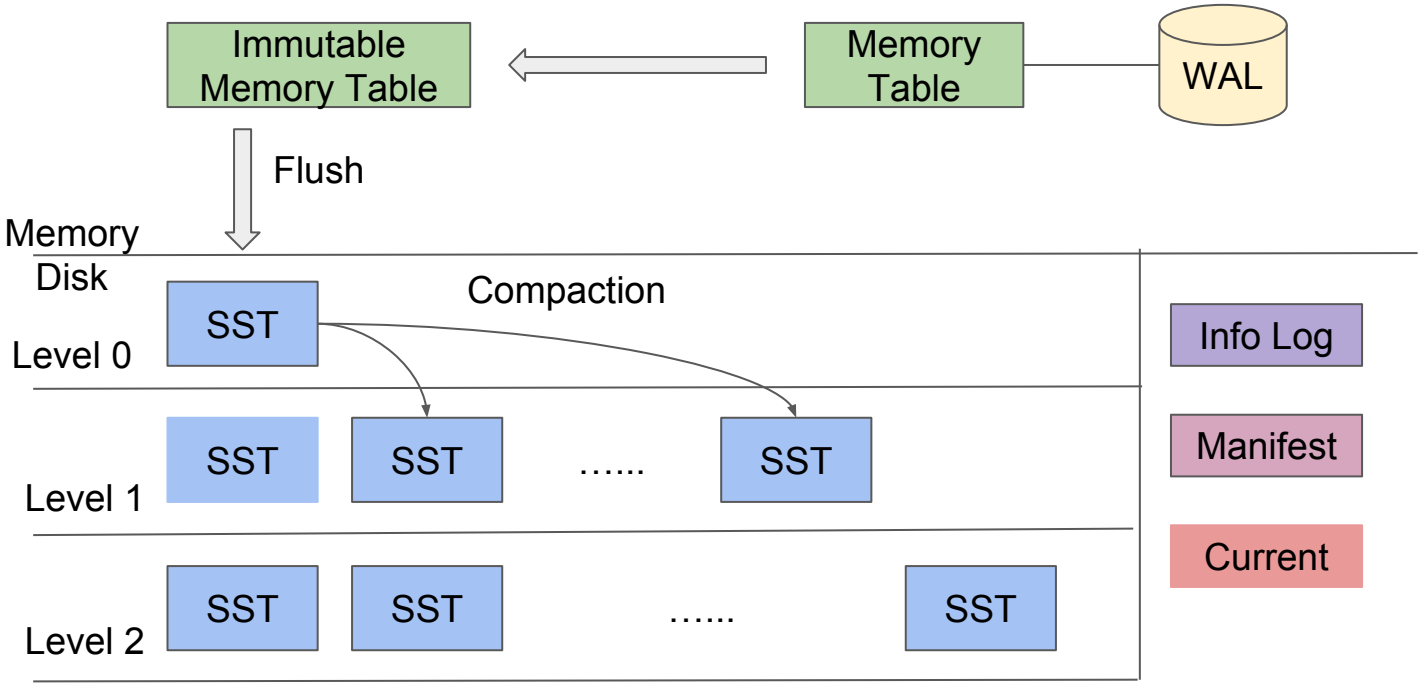
PingCAP

Let’s start from scratch!!!

PingCAP

PingCAP

RocksDB



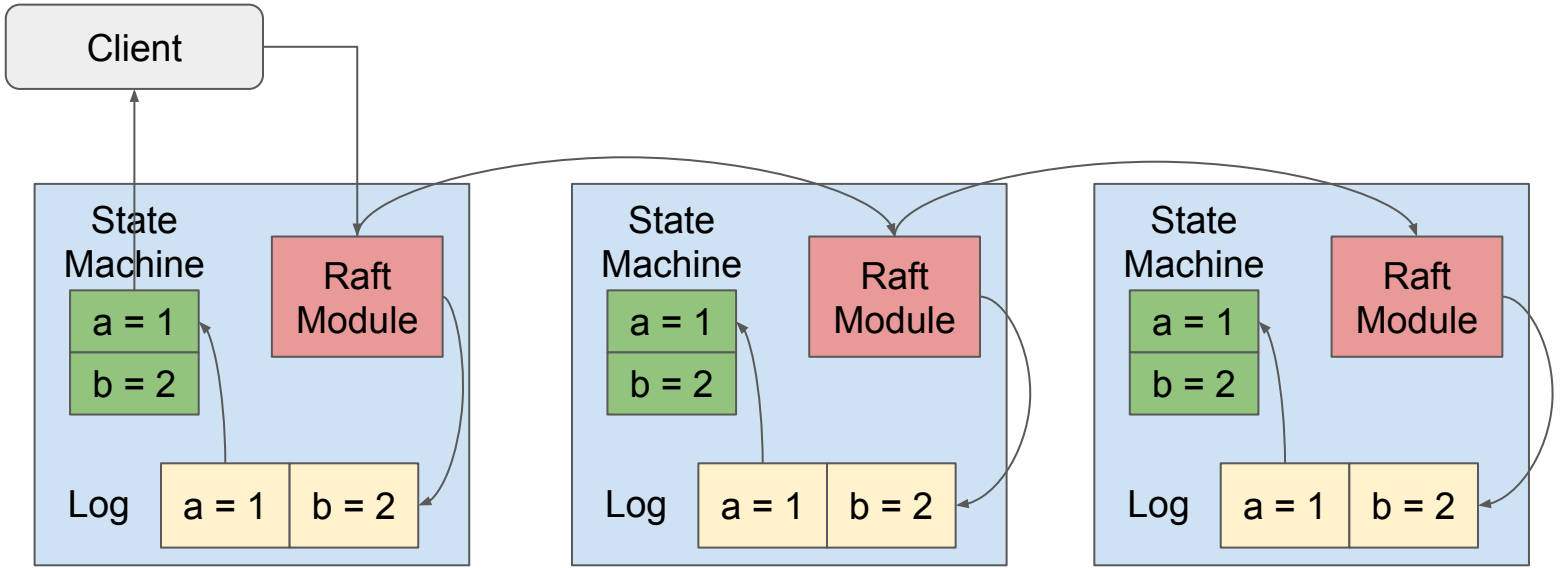
PingCAP

https://github.com/pingcap/rust-rocksdb

PingCAP

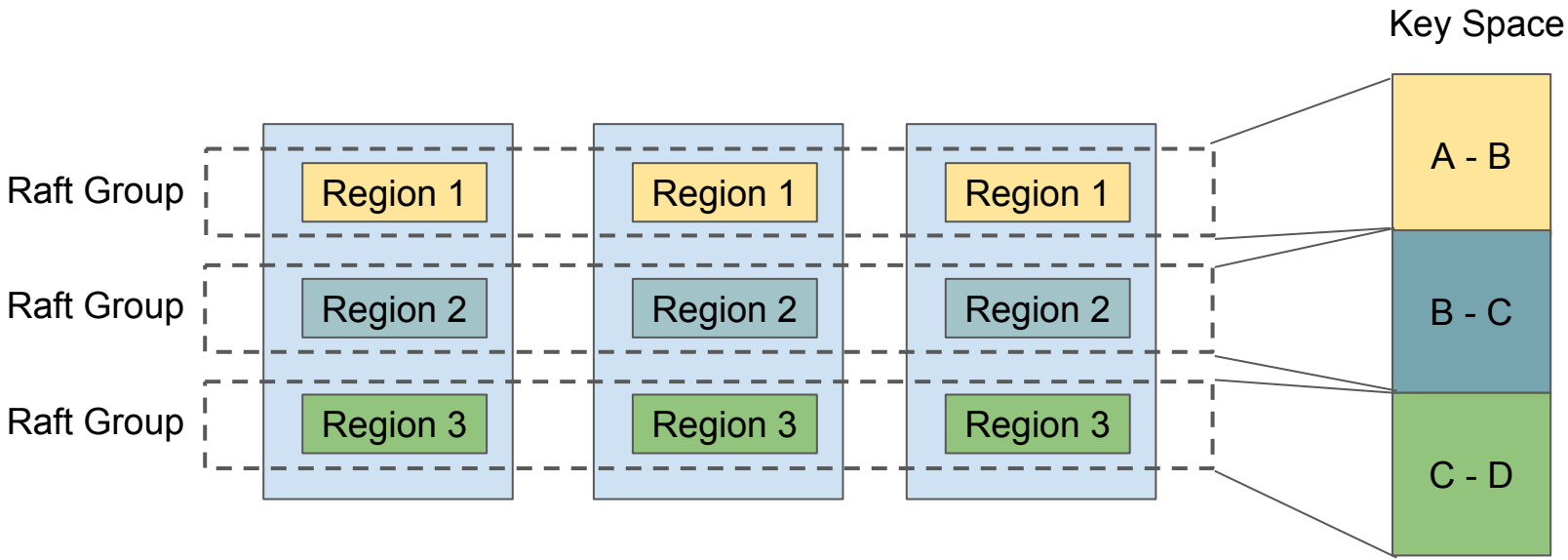
PingCAP

Raft



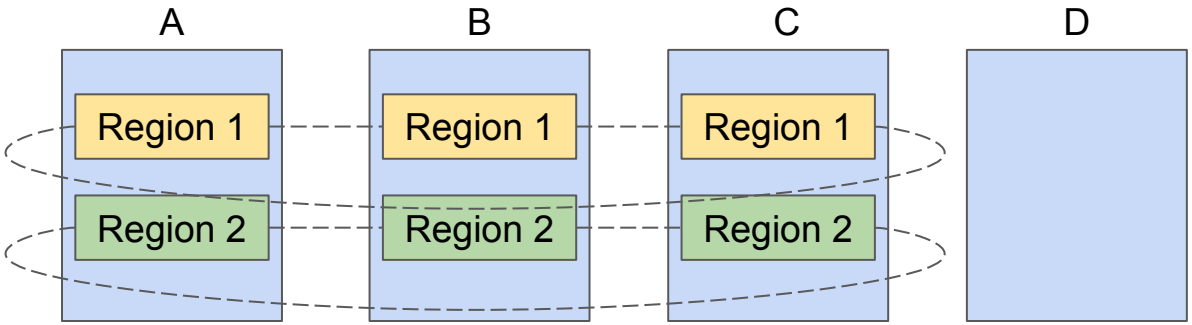
PingCAP

Multi-Raft



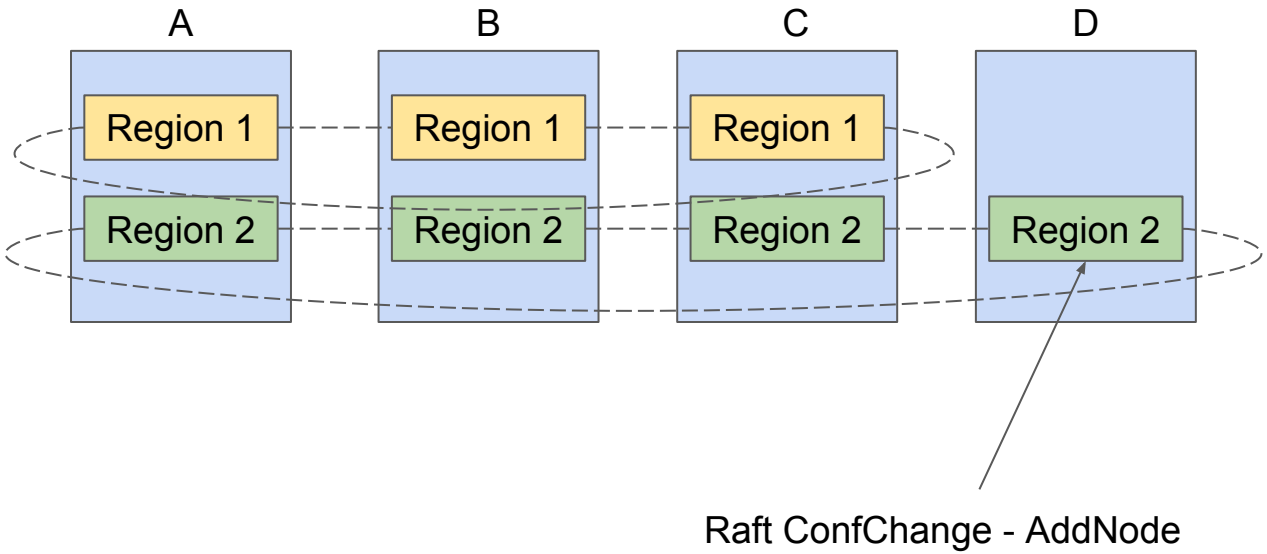
PingCAP

Multi-Raft - Scalability



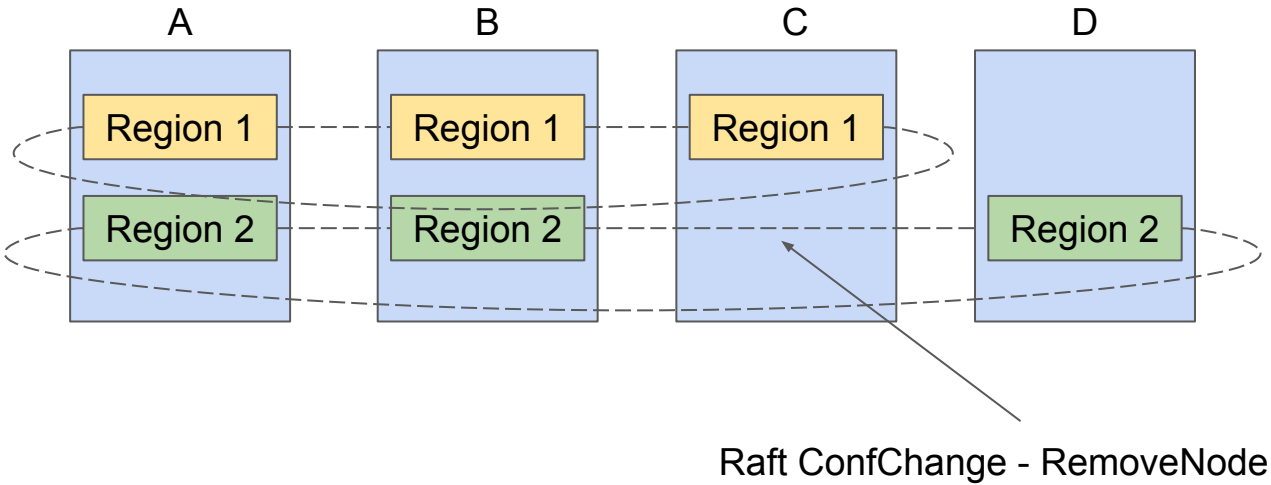
PingCAP

Multi-Raft - Scalability



PingCAP

Multi-Raft - Scalability



PingCAP

https://github.com/pingcap/raft-rs

PingCAP

Transaction

PingCAP

Transaction

```
let mut txn = store.begin()
let value1 = txn.get(region1_key)
let value2 = txn.get(region2_key)
// do something with value
txn.set(region1_key, new_value1)
txn.set(region2_key, new_value2)
txn.commit()
// or txn.rollback()
```

How to keep consistency crossing multi-Raft Groups?

PingCAP

Transaction

1. Inspired by Google Percolator
2. Optimized Two Phase Commit (2 PC)
3. Multiversion Concurrency Control (MVCC)
4. Snapshot Isolation
5. Optimistic Transaction

PingCAP

PingCAP

gRPC

- Mode
  - Unary
  - Client streaming
  - Server streaming
  - Duplex streaming

- Using Futures to wrap the asynchronous C gRPC API

```
let f = unary(service, method, request);
let resp = f.wait();
```

PingCAP

https://github.com/pingcap/grpc-rs

PingCAP

PingCAP

Prometheus

- Type
  - Counter
  - Gauge
  - Histogram

```
lazy_static! {
   static ref HTTP_COUNTER: Counter = register_counter!(
       "http_request_total",
       "Total number of HTTP request."
   ).unwrap();
}

HTTP_COUNTER.inc();
```

PingCAP

https://github.com/pingcap/rust-prometheus

PingCAP

Testing

PingCAP

Testing - Failure Injection

```
// Ingest a failure
fn function_foo() {
    fail_point!("foo");
}


// Run and Trigger the failure
FAILPOINTS=foo=panic cargo run
```

PingCAP

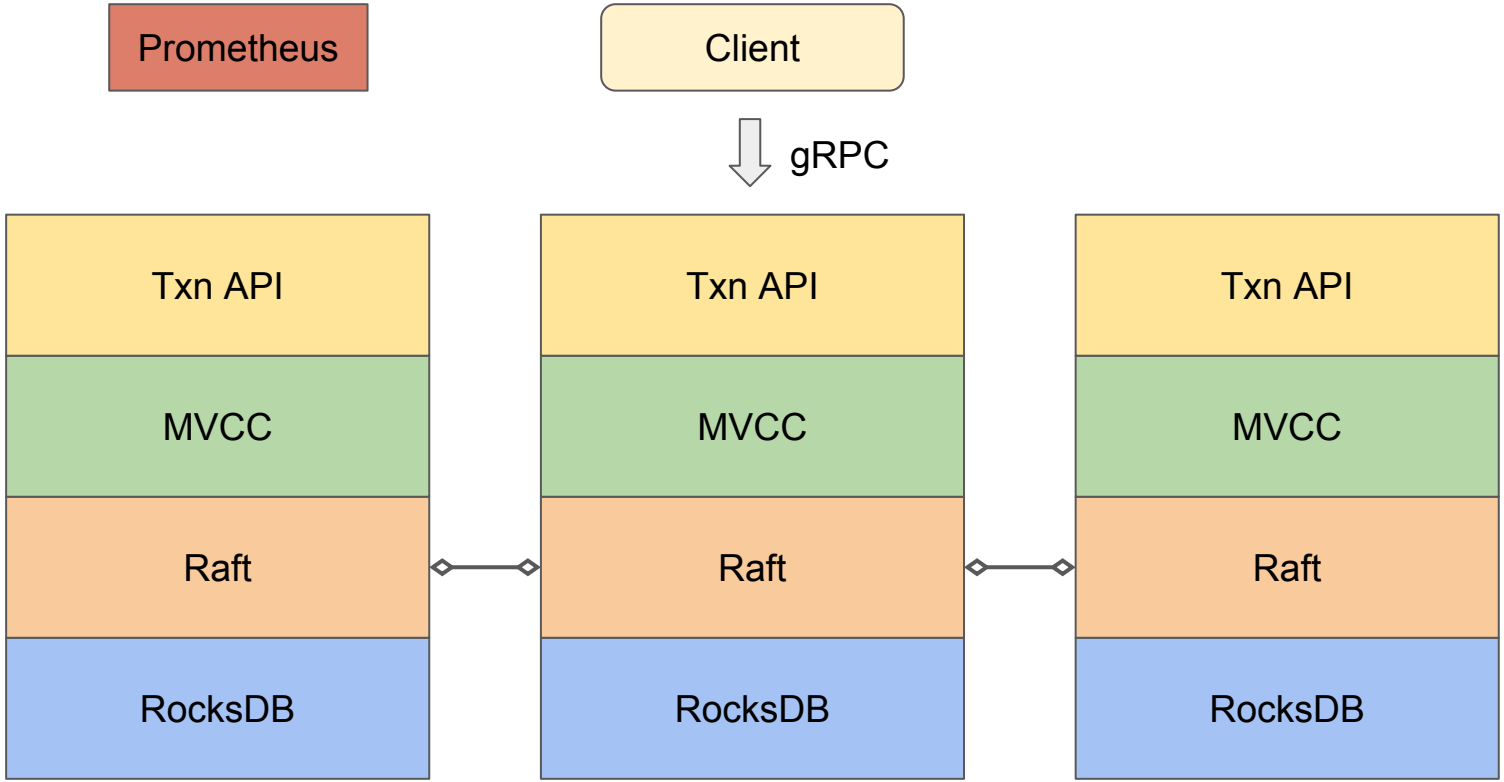
https://github.com/pingcap/fail-rs

PingCAP

Architecture

PingCAP

Architecture



PingCAP

https://github.com/pingcap/tikv

PingCAP

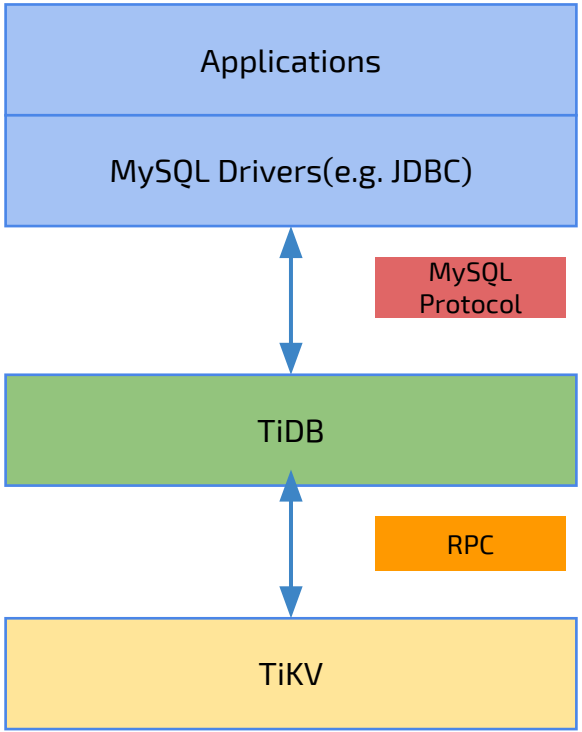
Beyond TiKV

PingCAP

A Distributed Relational Database

PingCAP

TiDB

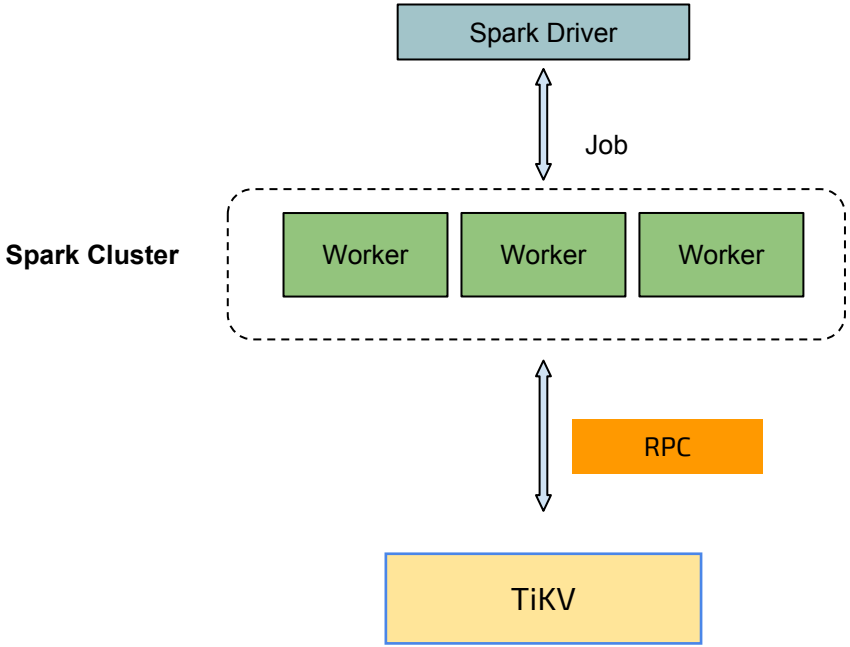


PingCAP

A Distributed Analytical Database

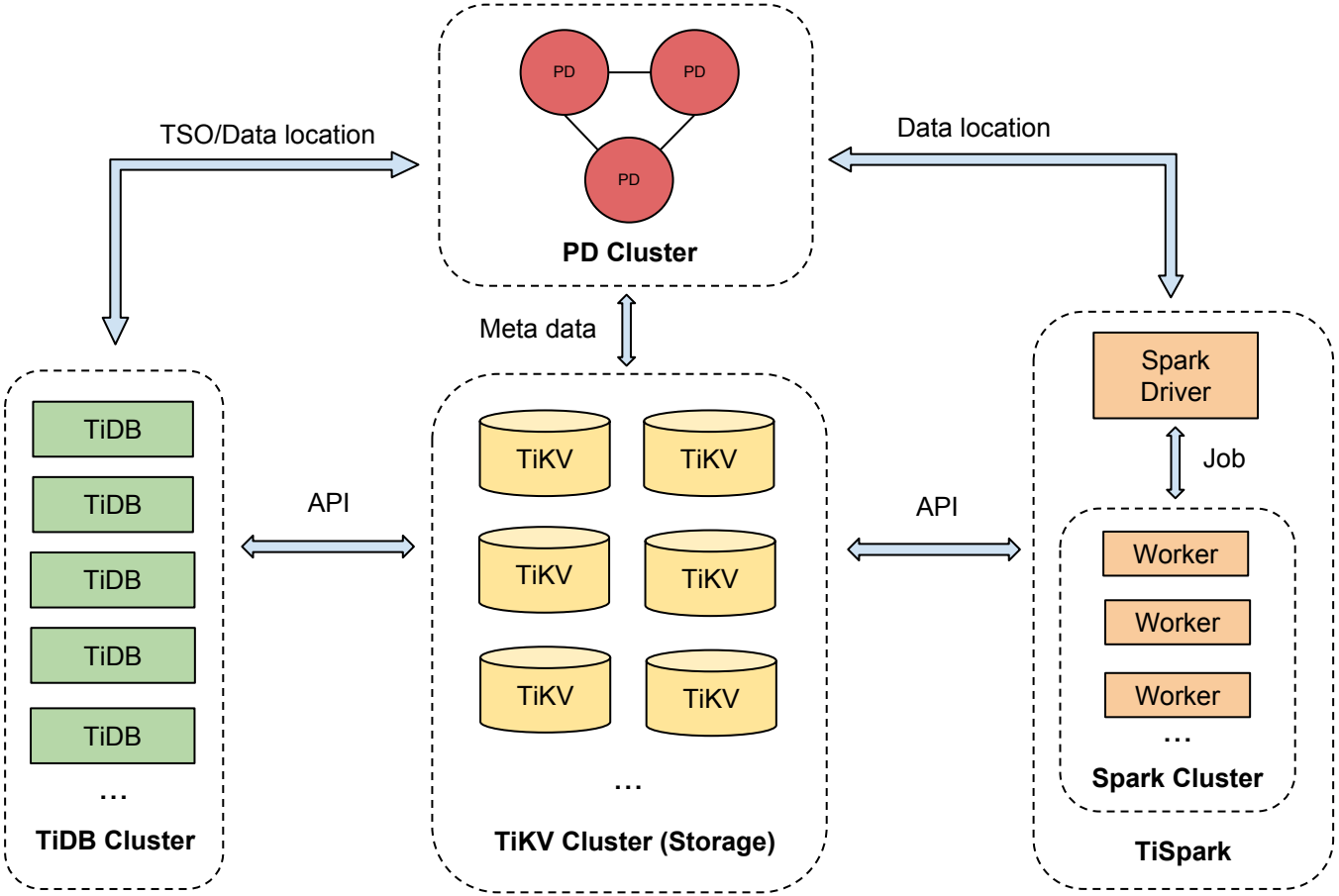
PingCAP

TiSpark



PingCAP

Hybrid Transactional/Analytical Processing Database


PD
PD
PD
PD Cluster
TSO/Data location
Data location
Meta data
TiDB
TiDB
TiDB
TiDB
TiDB
...
TiDB Cluster
API
TiKV
TiKV
TiKV
TiKV
TiKV
TiKV
...
TiKV Cluster (Storage)
API
Spark Driver
Job
Worker
Worker
Worker
...
Spark Cluster
TiSpark

PingCAP

Thank you!

https://github.com/pingcap/tidb

https://github.com/pingcap/tikv

We are hiring…

@China @Silicon Valley @Home