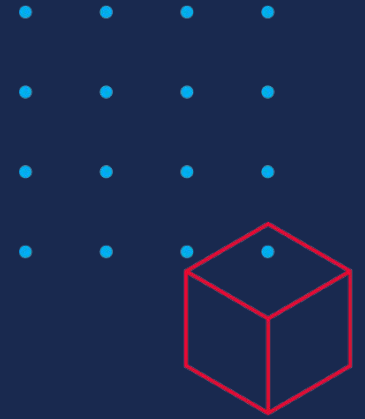
KubeCon | CloudNativeCon

Europe 2023



Bring Elastic and Resilient Multi-Tenancy to TiKV

PingCAP 2023.04

Agenda

KubeCon CloudNativeCon Europe 2023

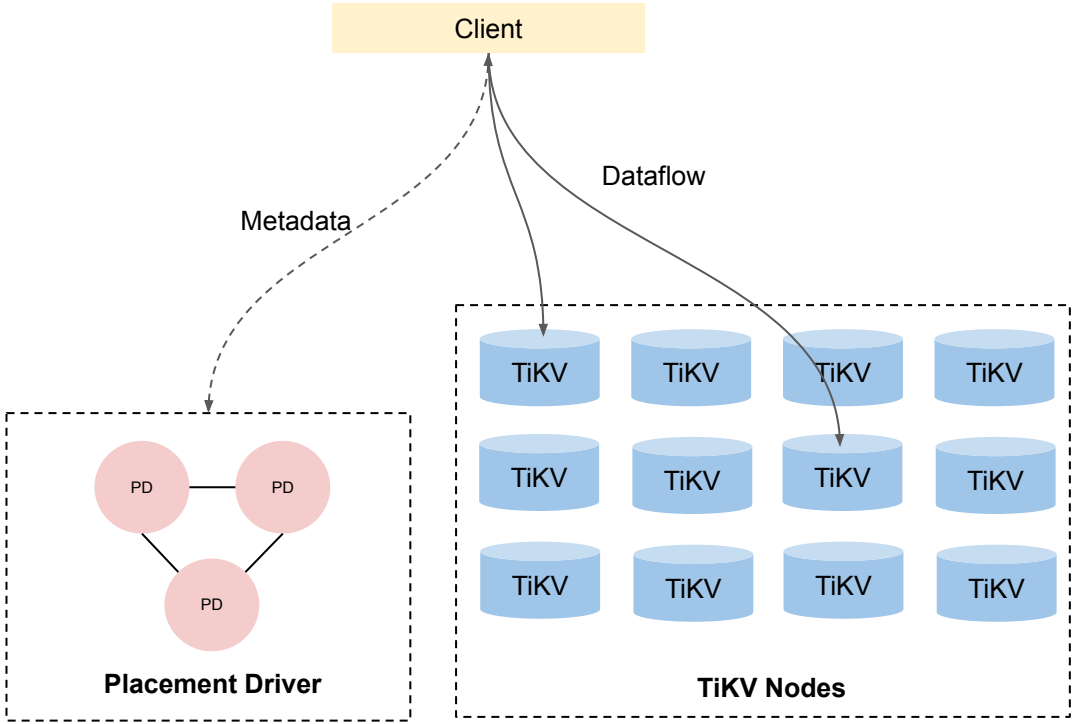
- What is TiKV
- Requirements & Challenges of Multi-Tenancy
- Solutions for Elastic and Resilient Multi-Tenancy
  - Isolating Different Tenants with Keyspace
  - MicroServices in Placement Driver
- Future Outlook

TiKV

What is TiKV

TiKV

- The **storage layer** for TiDB
- Distributed **Key-Value** store
  - Support **ACID Transactions**
  - Replicate logs by **Raft**
  - **Range** partitioning
    - Split / merge **dynamically**
  - Support coprocessor for **SQL operator pushdown**
- Visit https://tikv.org/



TiKV

Requirements & Challenges of Multi-Tenancy

Requirements for Multi-Tenancy

- Isolation & Quality of Service (QoS) for different groups of services
- Share some infrastructure to reduce costs
- Simplify cluster maintenance
- Use Transactional KV & RawKV in the same cluster

Challenges of Multi-Tenancy

KubeCon CloudNativeCon Europe 2023

- Elastic
  - Support data volume from **TBs** to **PBs** in single cluster
  - **Millions** of tenants
- Resilient
  - Isolate & reduce blast radius
  - Recover from fatal errors or disasters

TiKV

Solution for Elastic and Resilient Multi-Tenancy

Isolating Different Tenants with Keyspace

Data Isolation with Keyspace - Key Modes

KubeCon CloudNativeCon Europe 2023

Modes

- TiDB Mode (internal encoding)

  m<<xxx>> // TiDB metadata

  t<<tableID>>_r<<rowID>> => [col1, col2, col3, col4] // table row

  t<<tableID>>_i<<indexID>>_indexedColumnsValue_rowID => nil // index row

- TiKV Mode (internal encoding)

  x<<xxx>> // TxnKV Keys

  r<<xxx>> // RawKV Keys

Data Isolation with Keyspace - Key Encoding

KubeCon CloudNativeCon Europe 2023

Before:

<<mode-prefix>><<user-key>><<timestamp>>

After:

<<mode-prefix>><<keyspace>> <<user-key>><<timestamp>>

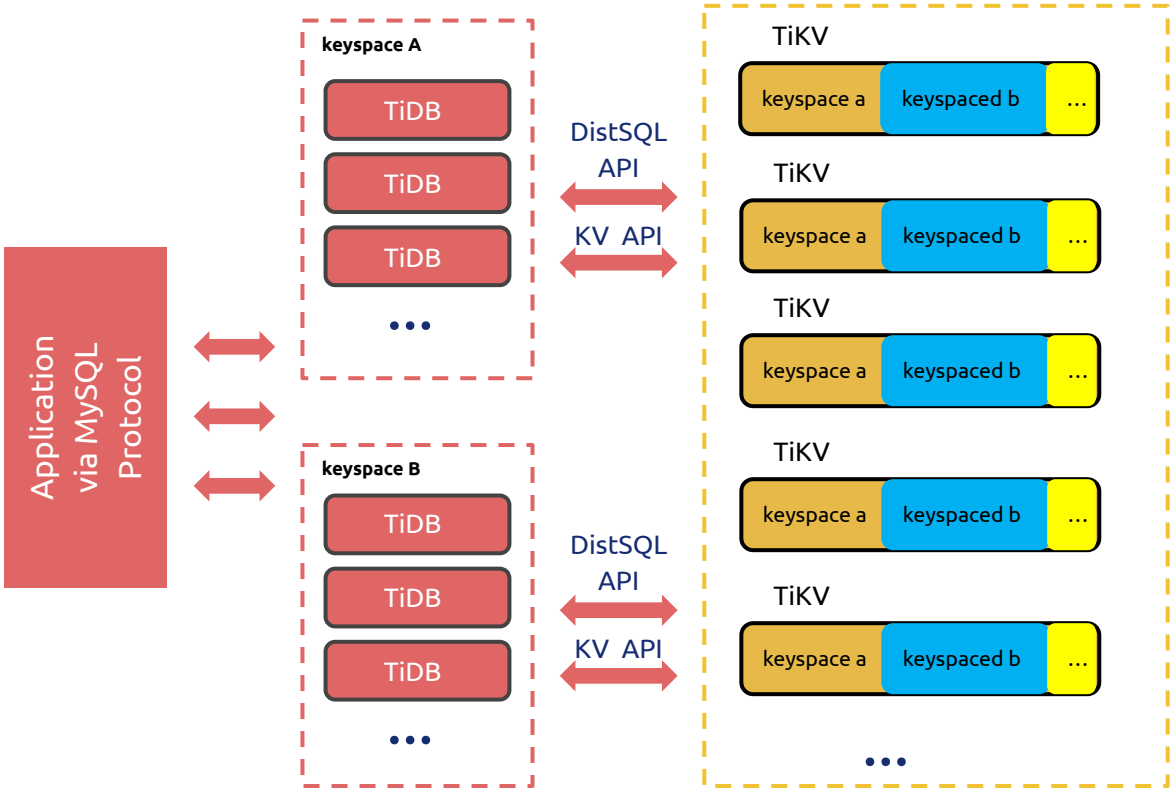
TiKV

Data Isolation with Keyspace

Keyspace management capability in PD:

1. Allocation of keyspace ID by keyspace name
2. Creation of a new keyspace by keyspace meta
3. Updating of keyspace meta and configuration
4. Management of keyspace lifecycle
5. Retrieval of keyspace metadata

Data Isolation with Keyspace - Overview



MicroServices in Placement Driver (PD)

MicroServices in PD

“TiKV can be scaled by scaling the TiKV node, but placement driver is not scalable. :(“

“PD is responsible for managing the cluster metadata, data access routing, TSO allocation, ID allocation, and scheduling of TiKV resources. It has too many responsibilities to handle :(“

“I am concerned about the stability and high availability of PD. It is a crucial component of TiKV, and if it cannot function properly, TiKV will also be affected :(”

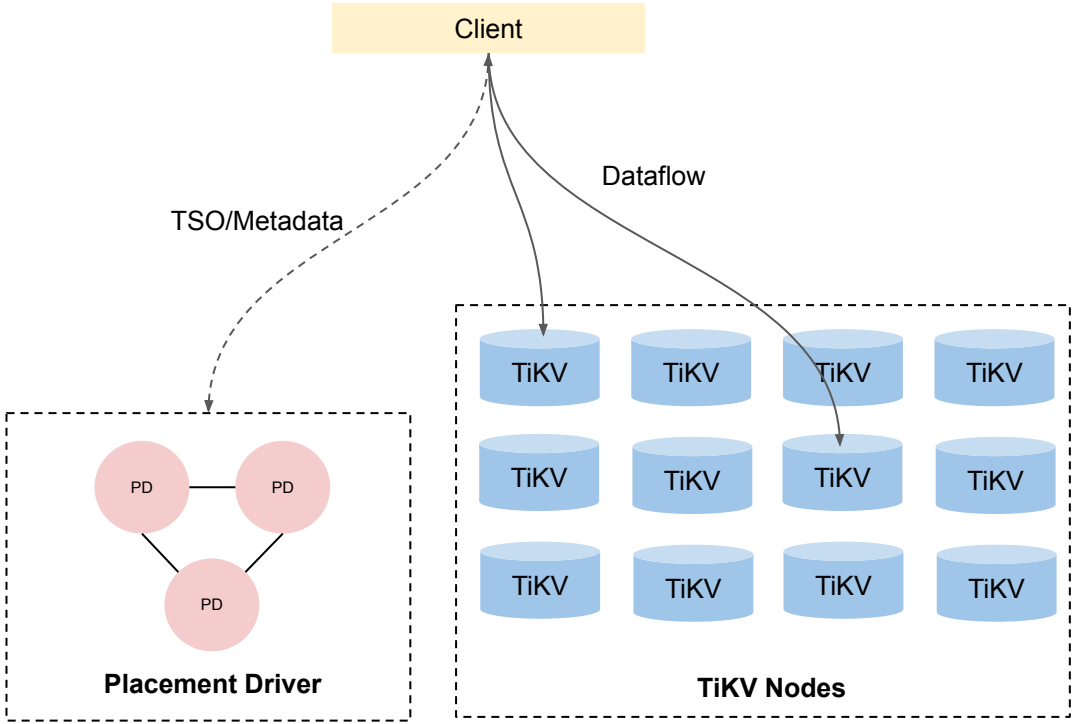
…

MicroServices in PD

KubeCon CloudNativeCon Europe 2023

Placement Driver Provides:

- API(Metadata) service
- TSO Service
- Scheduling Service
- Allocate Service
- …

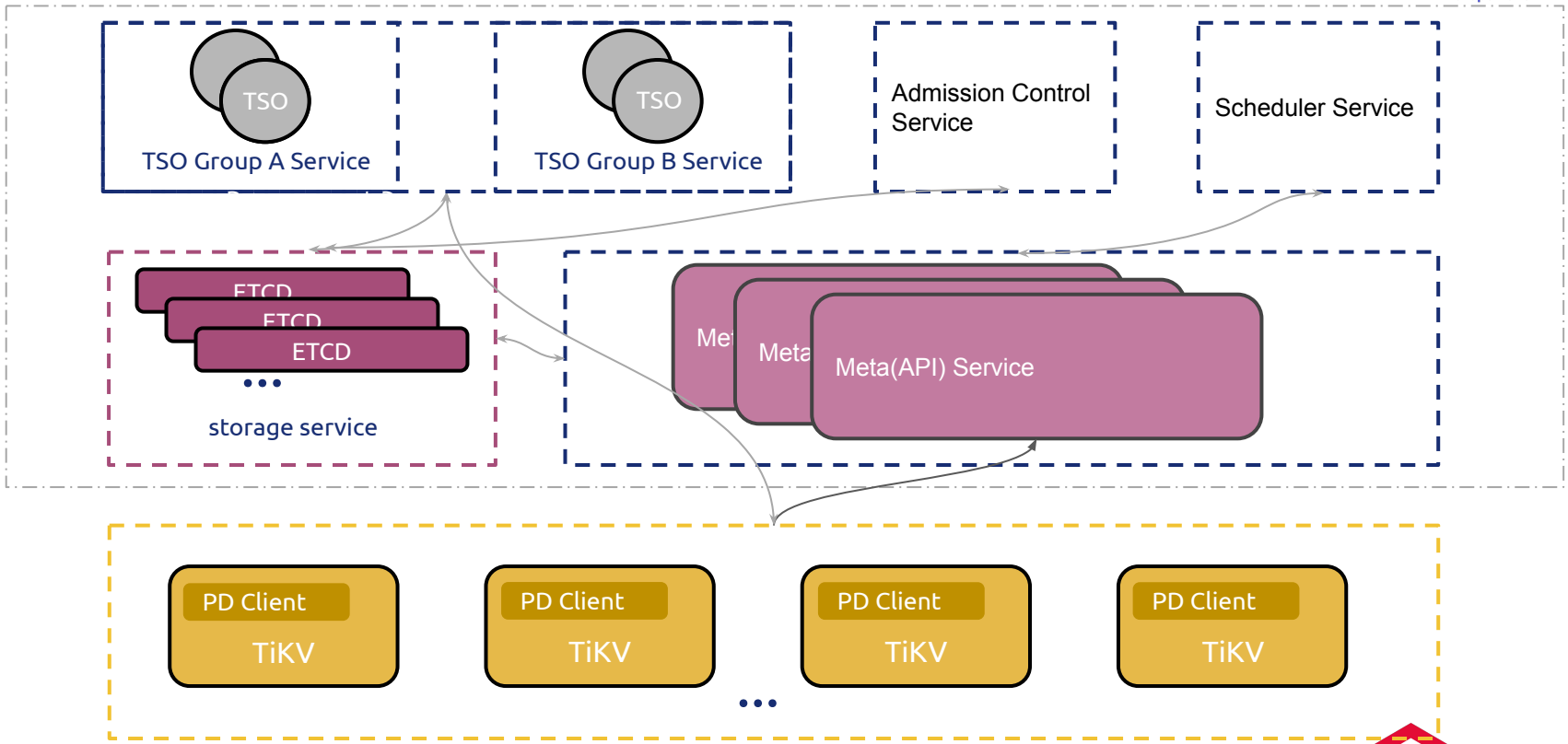


TiKV

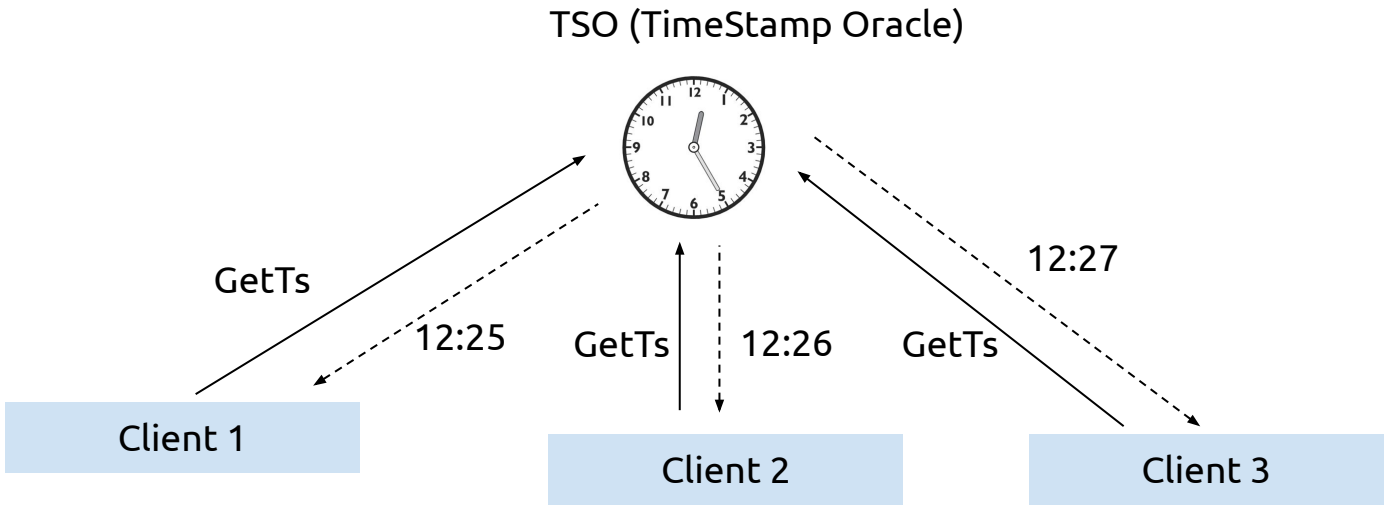
MicroServices in PD

	Level	Resource Required (Data Size)
TSO Service	**Critical**	**High**
Metadata Service	**Critical**	**High**
Allocate Service	Critical	Low
Scheduler Service	Middle	Middle
Others	Low	Low

MicroServices in PD - Overview

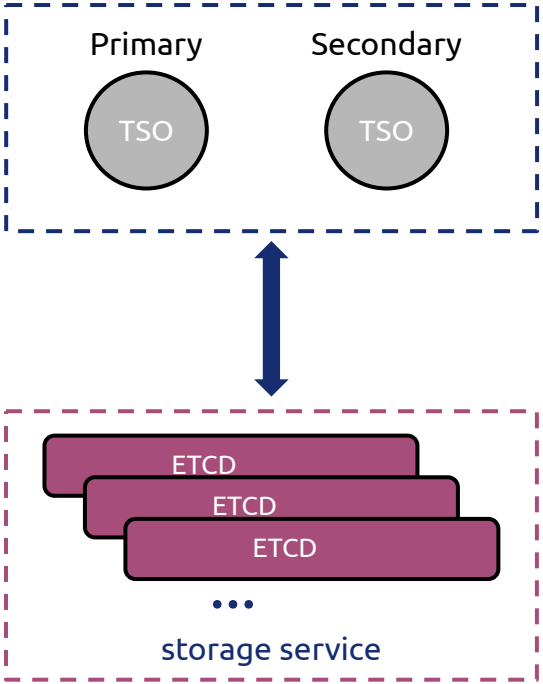


MicroServices in PD - TSO Service

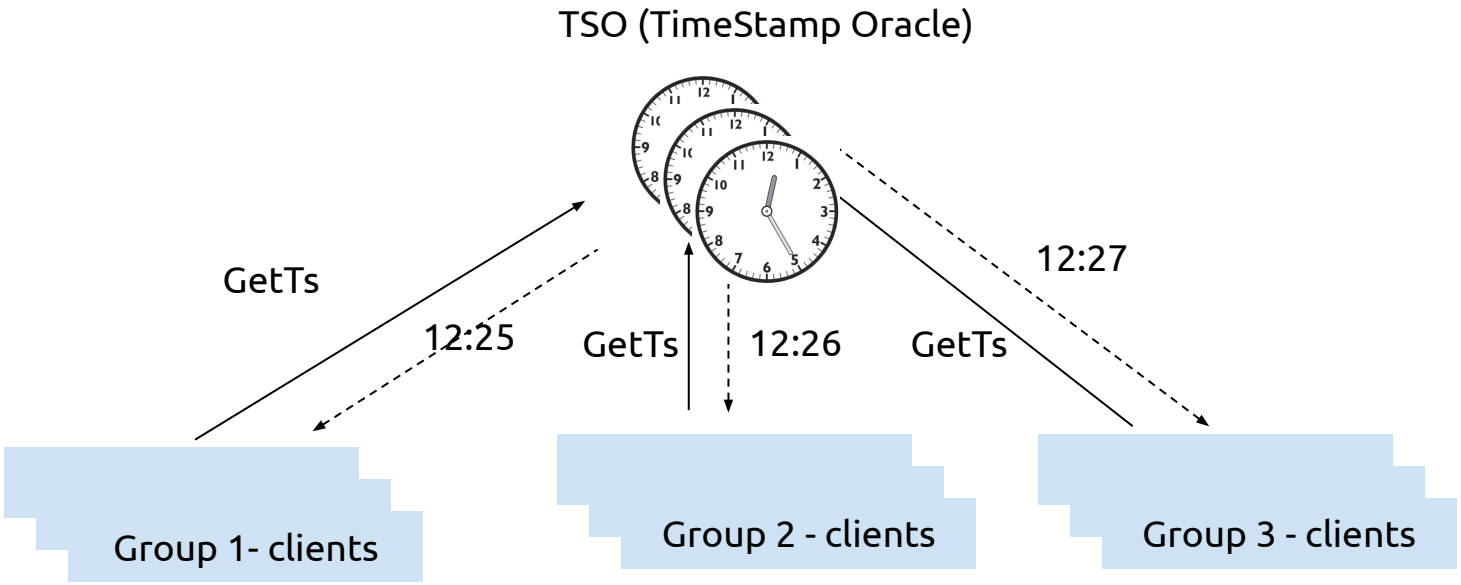


TSO allocates monotonically increasing timestamps

MicroServices in PD - Overview



MicroServices in PD - TSO Service

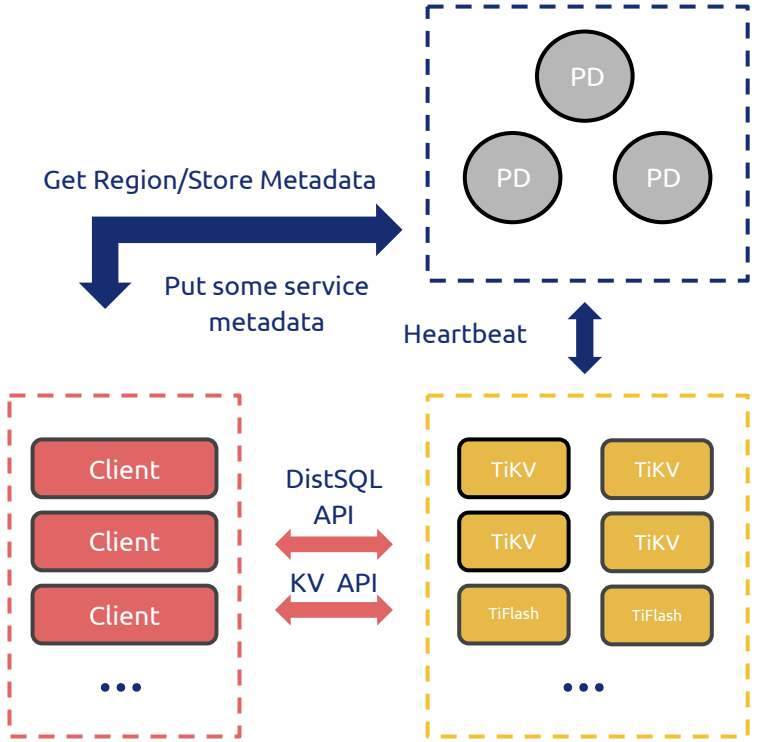


TSO allocates monotonically increasing timestamps

TiKV

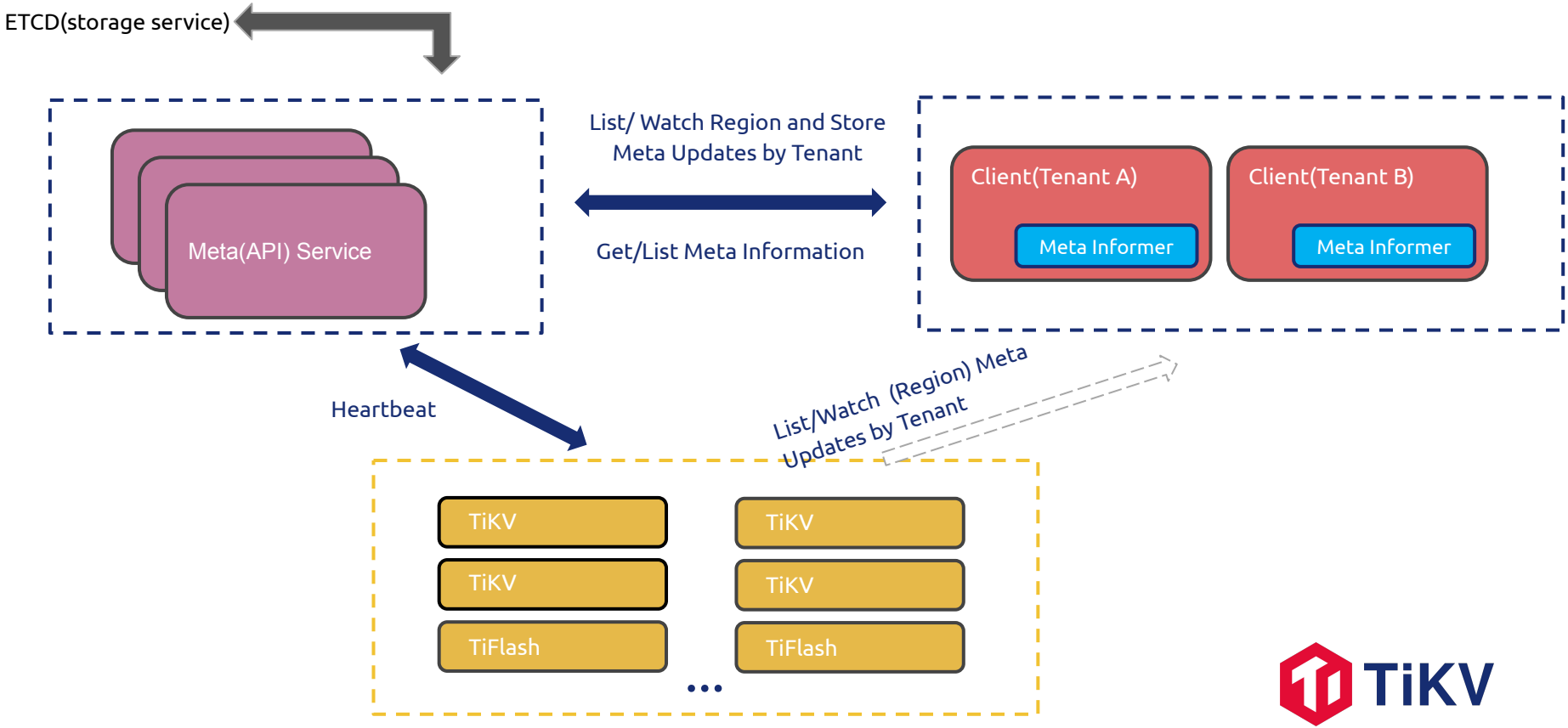
MicroServices in PD - Metadata(API) Service

- PD Collects Metadata Information from TiKV through Heartbeat
- Client can store some metadata to PD (e.g. elections)
- Client get important information about the TiKV cluster (e.g. routes to TiKV nodes)



TiKV

MicroServices in PD - Metadata(API) Service



Future Outlook

Future Outlook

KubeCon CloudNativeCon Europe 2023

- Provide fine-grained control of resource for tenants
  - CPU, memory, disks capacity and throughput
- **Serverless** based on Multi-Tenancy
  - Provide **virtual cluster** for every tenant
  - Start or stop a cluster **in seconds**
  - Scale in and out **automatically**

TiKV

KubeCon CloudNativeCon Europe 2023

Thanks.

TiKV