Akka Classic Cluster

A high-level tour of how to build application cluster

Silde: https://github.com/xiaozhiliaoo/my-slides/blob/master/akka-classic-cluster.pptx

Code: https://github.com/xiaozhiliaoo/akka-practice/tree/main/akka-classic-cluster-sample

李力

2022-04-13

内容大纲

- 集群核心概念
- 集群功能与模块
- 集群设计与实现
- 集群教务系统案例
- 集群技术其它选择
- Akka与分布式系统泛型
- Akka与应用架构
- 参考资料

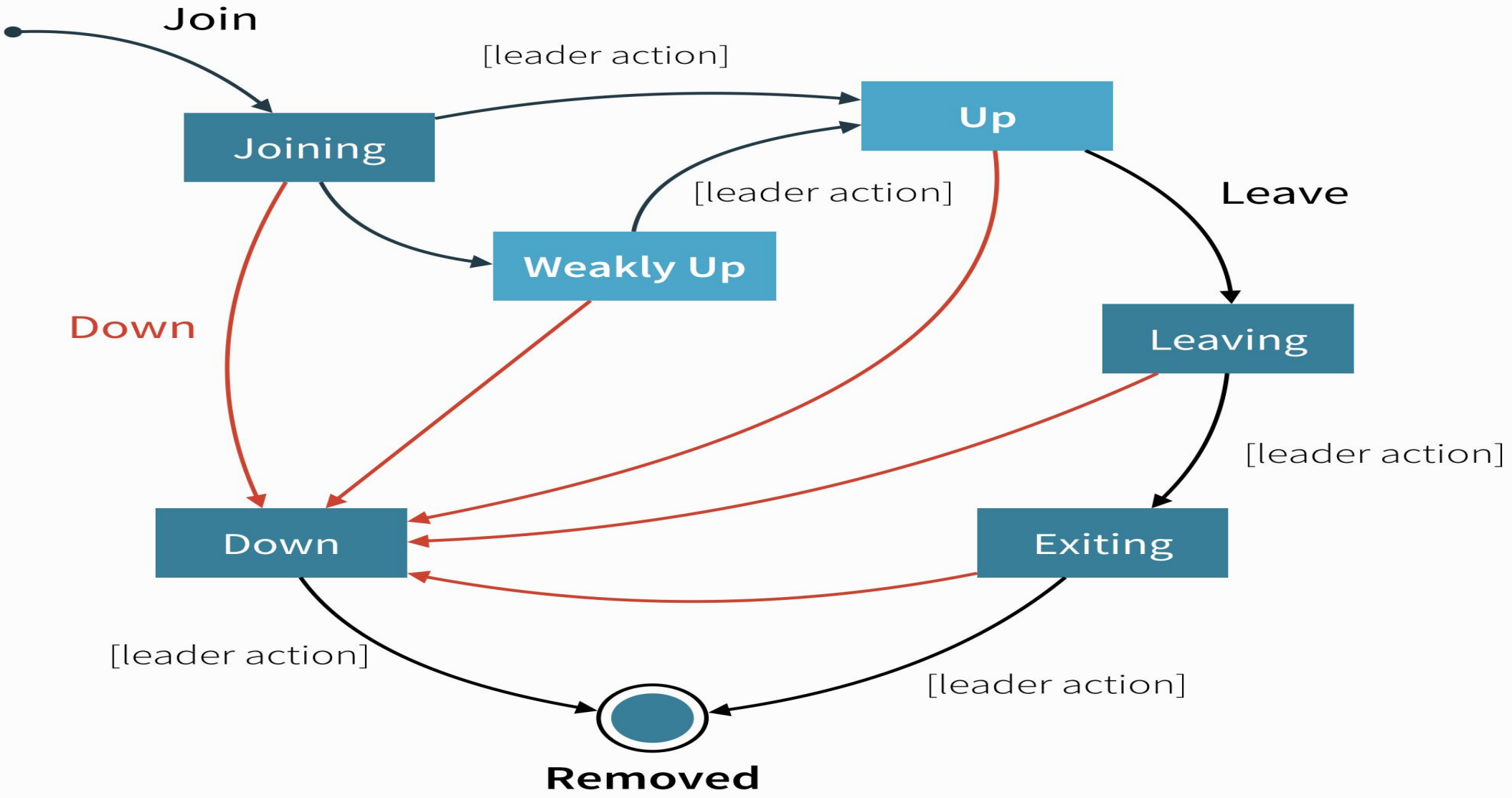
集群核心概念-案例

- 至少两个条件: membership，coordinator，可以满足sharding/partitioning
- application
  - 普通微服务(X)
  - 多节点master-worker(√)
  - 游戏
- middleware
  - Flink/Spark Streaming(√)
  - Redis Cluster(√)
  - Mongo ReplicaSet(√)
  - Hazelcast(√)
- databases
  - MySQL Master-Slave(X)
  - MySQL Cluster(√)
  - TiDB(√)
- web server
  - Tomcat/JBoss Cluster(√)
- framework
  - spring cloud cluster(leader lock)
  - hazelcast
  - **akka**

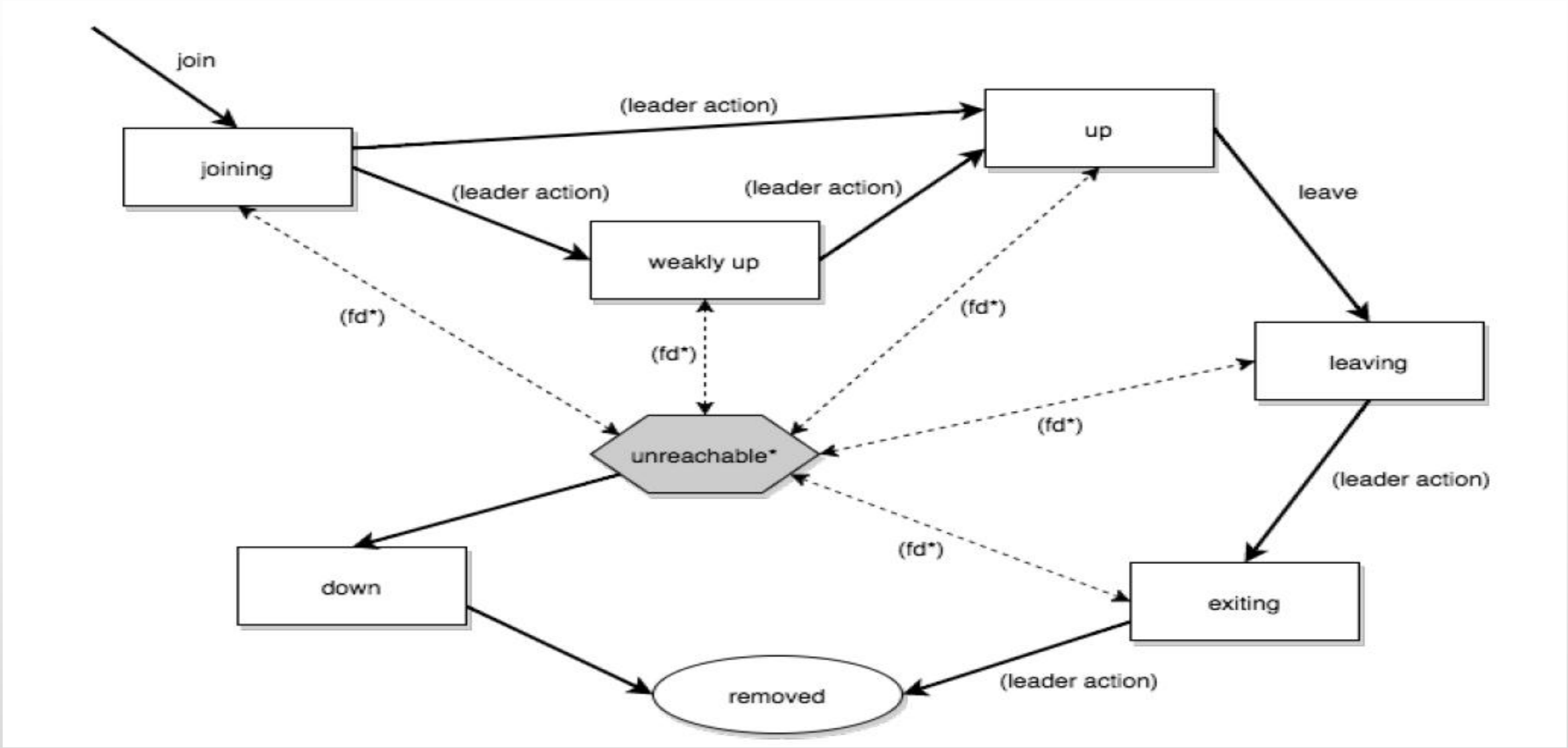
集群核心概念-akka cluster

- Membership(Cluster Membership Service)
  - MemberEvent驱动MemberStates变化
  - MemberStates变化组成MemberLifeCycle
- Coordinator
  - Leader
    - Cluster Convergence
    - MemberState转换
  - Lease

集群核心概念-状态变化


Join
[leader action]
Joining
Up
[leader action]
Leave
Weakly Up
Down
Leaving
[leader action]
Down
Exiting
[leader action]
[leader action]
Removed

集群核心概念-状态变化+不可达检测

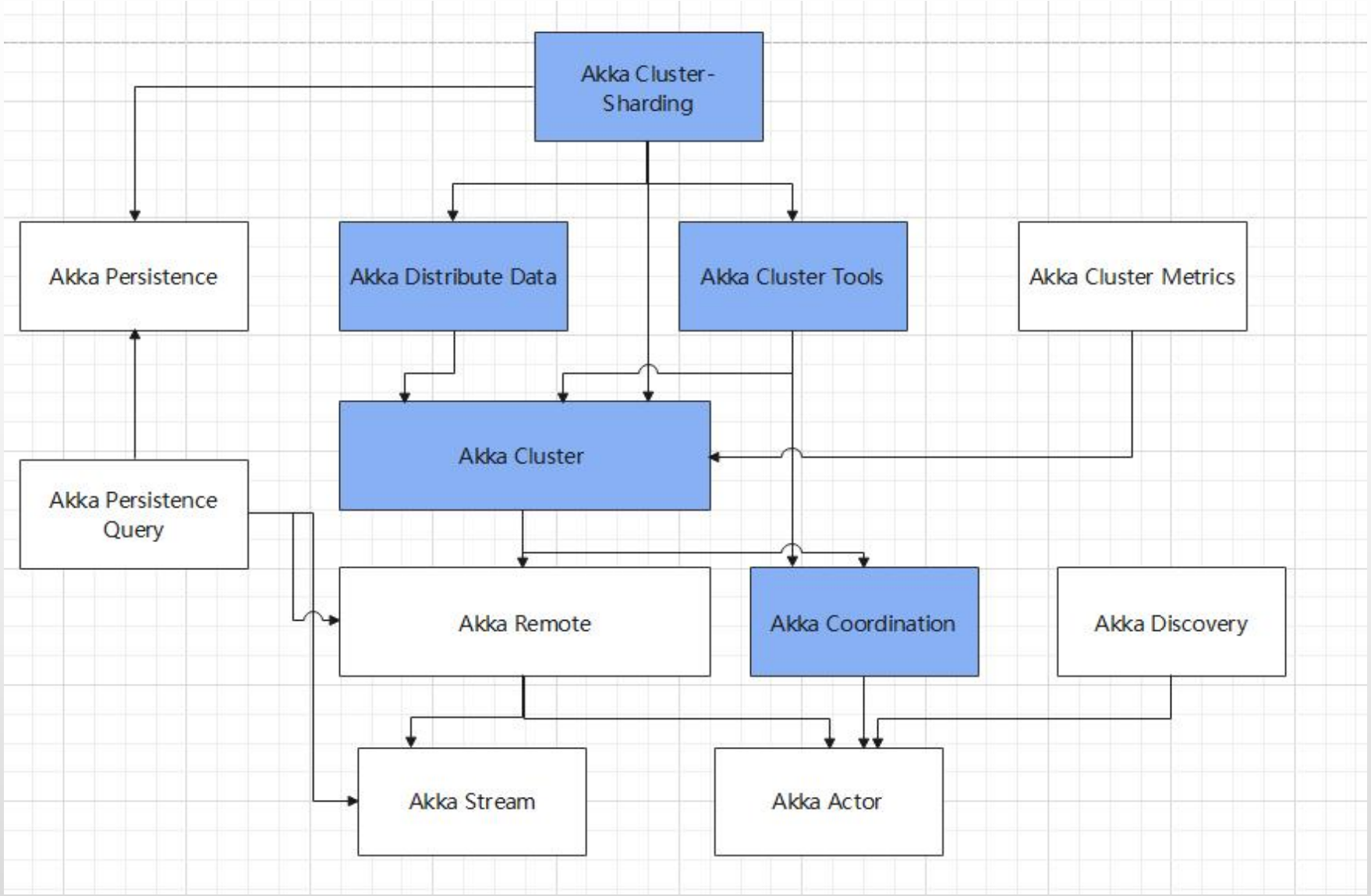


集群核心概念-代码演示

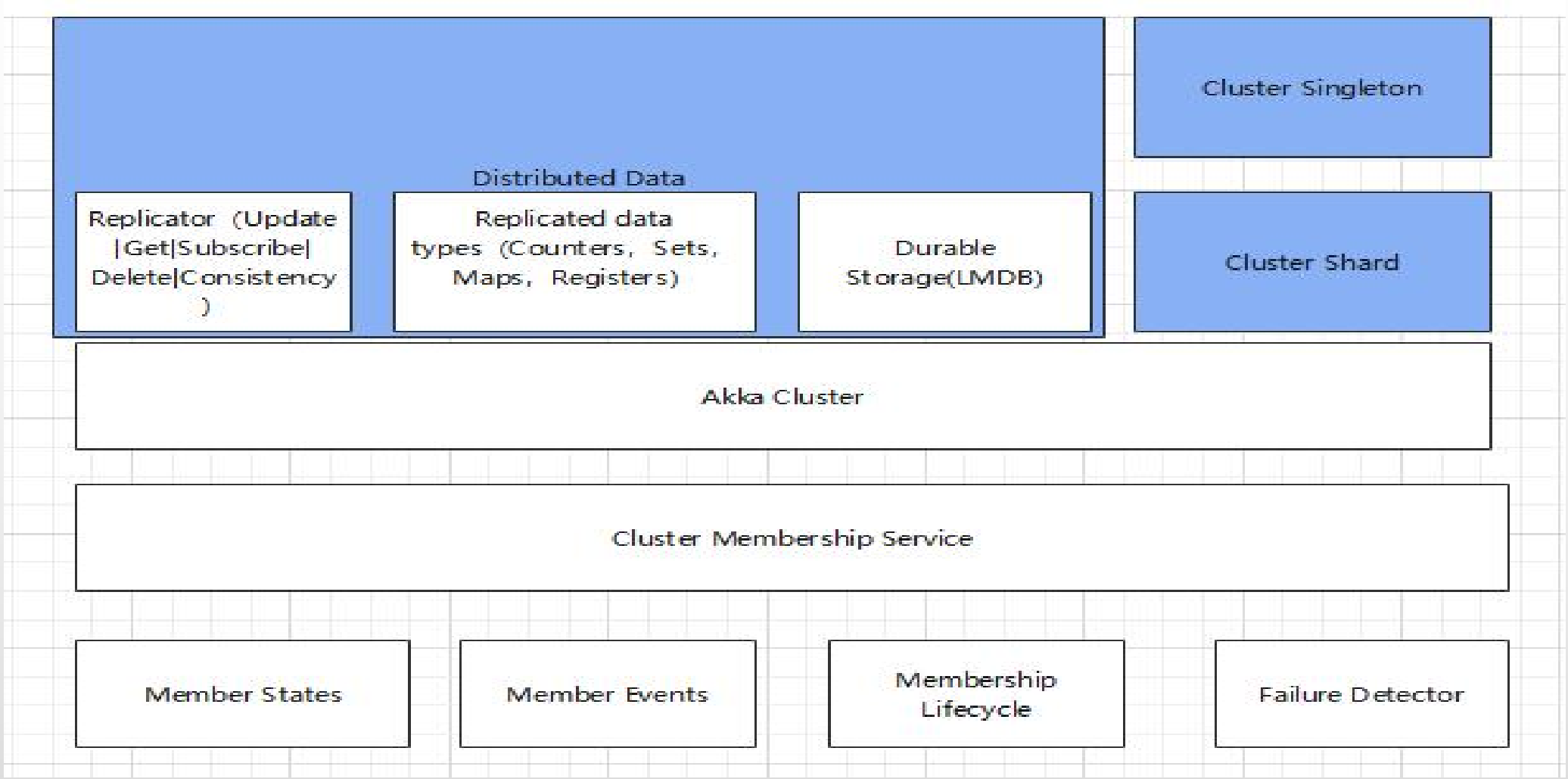
- demo1：动态变化的集群动画
  - 演示集群启动
  - 直接访问dashboard.html
  - **(demo1 code)**

- demo2：集群成员变更事件通知/演示JMX
  - **(demo2 code)**

集群功能和模块-cluster



集群功能和模块-cluster detail



集群功能和模块-Classic Distributed Data

- KV Store
  - K： Unique Indentifier V：CRDTs
- Data Consistency
  - Gossip 和 Direct Replication
  - 必须保证的一致性：Read your write
  - NWR: ReadAll+WriteAll, ReadMajority+WriteMajority,WriteLocal+ReadLocal
- Replicator
  - Update：WriteLocal，WriteToN，WriteMajority，WriteAll
  - Get：ReadLocal，ReadFromN，ReadMajority，ReadAll
  - Delete
  - Subscribe
- WeaklyUp：true

集群功能和模块-Classic Distributed Data

- Data Type
  - Counters: GCCounter, PNCounter
  - Sets: GSet, ORSet
  - Maps: ORMap，ORMultiMap, LWWMap, PNCounterMap
  - Registers: LWWRegister(reverseClock:FWW, defaultClock:LWW), Flag

- Durable Storage
  - LMDB

- **(demo3 code)**

集群功能和模块-Classic Cluster Singleton

- 集群中某一类型的Actor只有一个
  - 系统统一入口或者出口
  - 集群任务的总路由器
  - 运行最久的节点上
  - 单点瓶颈
  - ClusterSingletonManager，ClusterSingletonProxy
  - WeaklyUp：false

- **(demo4 code)**

集群功能和模块-Distributed Publish Subscribe

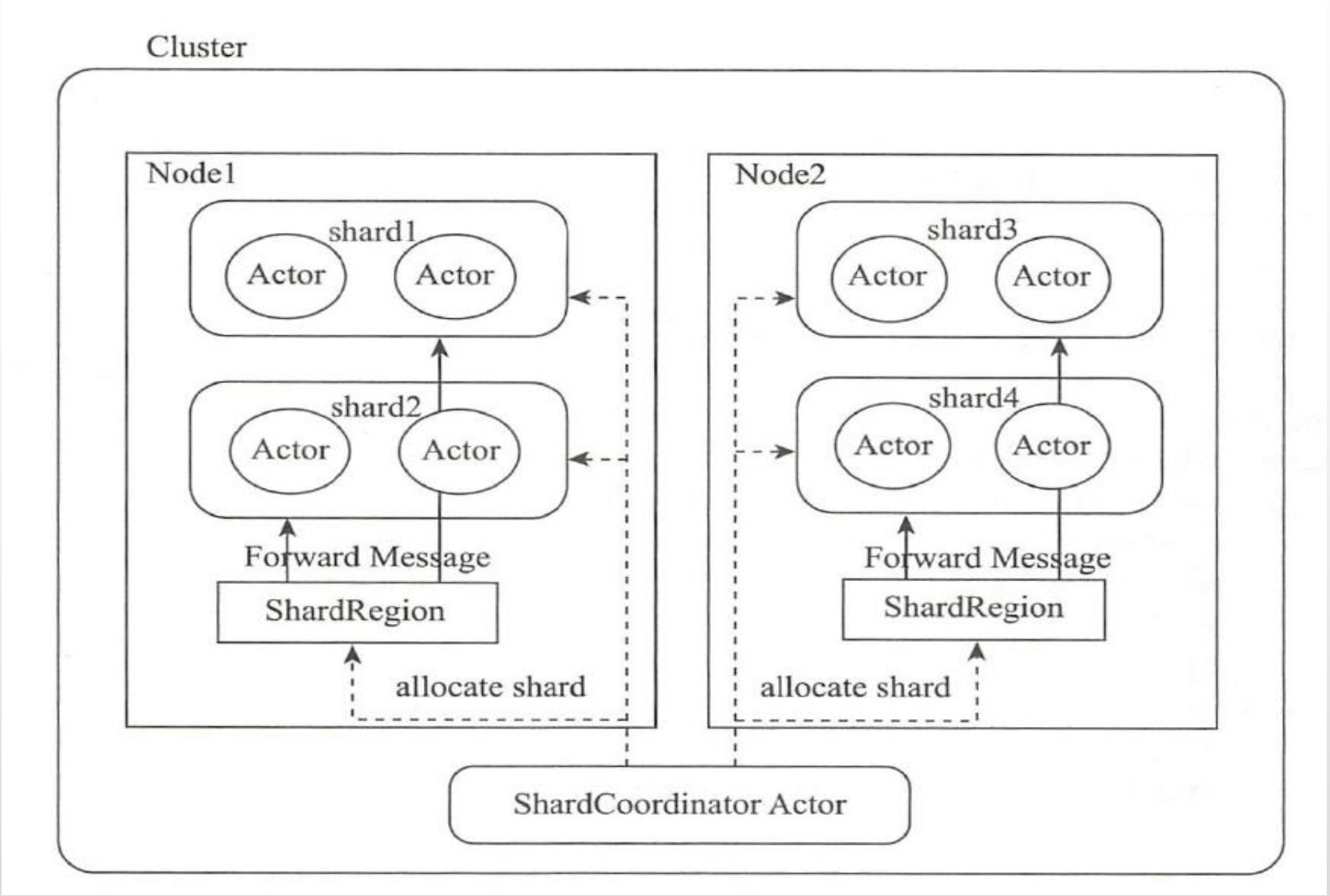
- 发布订阅功能
  - 中介者：DistributedPubSubMediator，管理 Actor 引用的注册表
  - 注册表最终是一致
  - 至多一次传递消息
  - WeaklyUp：true

- (demo5)

集群功能和模块-Classic Cluster Sharding

- 分布式系统常用功能
  - 提高写吞吐
  - 单机内存上限
- 同样的事情，不同的名字
  - ShardRegion:akka
  - partition:hazelcast,kafka
  - shard:MongoDB,ES,Solr
  - region:Hbase,TiKV
  - tablet:Bigtable
  - vnode:Cassandra,Riak
  - vBucket(virtual buckets):Couchbase
  - slot:Redis
- 带来问题
  - 是寻找到分区信息Routing
  - 增减节点时候Rebalance分区

集群功能和模块-Classic Cluster Sharding



集群功能和模块-Classic Cluster Sharding

- Shard，Entity
- ShardCoordinator
  - LeastShardAllocationStrategy(shard allocation,rebalance)
    - **旧：rebalance-threshold=1，max-simultaneous-rebalance=3**
    - **新：rebalance-absolute-limit=0，rebalance-relative-limit=0.1**
  - Singleton
  - 默认持久化：DDataShardCoordinator 和 PersistentShardCoordinator（deprecated 2.6.0）WeaklyUp：false
- ShardRegion
  - MessageExtractor
    - entityId(msg 找到Actor),shardId(msg 找到分片)
    - HashCodeMessageExtractor
- Rebalance
  - 新增/删除/故障情况
  - 迁移Entity到新的节点
- **(demo6 code)**

集群功能和模块-Cluster Split Brain Resolver

- 脑裂
  - 网络分区
  - 机器崩溃
  - 进程长时间没响应（过载、CPU 不足或长时间的垃圾收集暂停）
- 策略
  - keep-majority
  - keep-oldest
  - down-all
  - static-quorum
  - lease-majority

- RabbitMq：ignore、pause_minority、pause_if_all_down、autoheal

集群设计与实现-设计概念

- Akka Cluster provides a fault-tolerant **decentralized peer-to-peer** based Cluster **Membership** Service with no single point of failure or single point of bottleneck. It does this using **gossip** protocols and an automatic **failure detector**.

- Gossip/Gossip Convergence (Scala Code)
  - Consul(Serf)
  - Redis Cluster
  - Cassandra/Dynamod
  - **(demo7 code)**
- Phi Accrual Failure Detector (Scala Code) (Java Code)
  - HazelCast (Code)
  - Consul
  - **(demo8 code)**
- Vector Clocks (Scala Code)
  - Riak
  - Dymanod
- Leader(No Select)

集群设计与实现-源码实现速览

- 集群消息
  - ClusterMessages.proto/序列化
  - routing
  - sbr(split brain resolver)

集群使用场景

- 教务系统
  - common-akka
    - Spring-Managed Actor
    - Actuator Endpint
    - AkkaService,SingletonService
    - Discovery by rancher, seeds
    - 脑裂：quorum
  - 定时任务
  - 结算平台
  - Kafka延时队列

集群其它技术

- 借助单机/分布式存储：etcd/zookeeper/nacos/consul/doozerd/mysql/MFS/NFS
- 借助中间件/框架：Hazelcast，Akka，Serf(Gossip)，JGroups ，Erlang/OTP(非Java)
- 借助协议：raft，gossip，zab，paxos。需要利用开源实现来构建系统。

- 详细内容可见博客：https://xiaozhiliaoo.github.io/2021/12/20/java-application-cluster/

Akka Cluster与分布式系统泛型

- 体系结构：非集中式
- 进程：Actor与线程池
- 通信：消息，多播(Gossip)
- 命名：结构化命名的树状目录
- 同步：向量时钟，Gossip
- 一致性和复制：DistributedData一致性可调，Cluster最终一致性。复制：Leaderless
- 容错性：故障检测，Let it crash，父子级监督机制
- 安全：无

Akka与应用架构

- 反应式架构/分布式领域驱动设计
- Fast Data Architectures for Streaming Applications
- LAMP VS SMACK
- Akka Play Lagom Spray全家桶
- 流和表的融合.

参考资料

- starter: https://developer.lightbend.com/start/
- 《Akka实战》
- 《反应式应用开发》
- 《Akka应用模式》