架构迎接未来变化
IAS2017 • NANJING

How to build high available high scalable distributed application in BigData ecosystem

Junkai Xue
Senior software Engineer / Apache Helix PMC & Committer

架构迎接未来变化
IAS2017 • NANJING

Agenda

- About me
- What is Helix?
- Online Ecosystem
- Offline Ecosystem
- Advantage of Rebalance Strategy Algorithms
- Helix & Containers
- Conclusion
- Q&A

Senior Software Engineer @ LinkedIn Data Infra

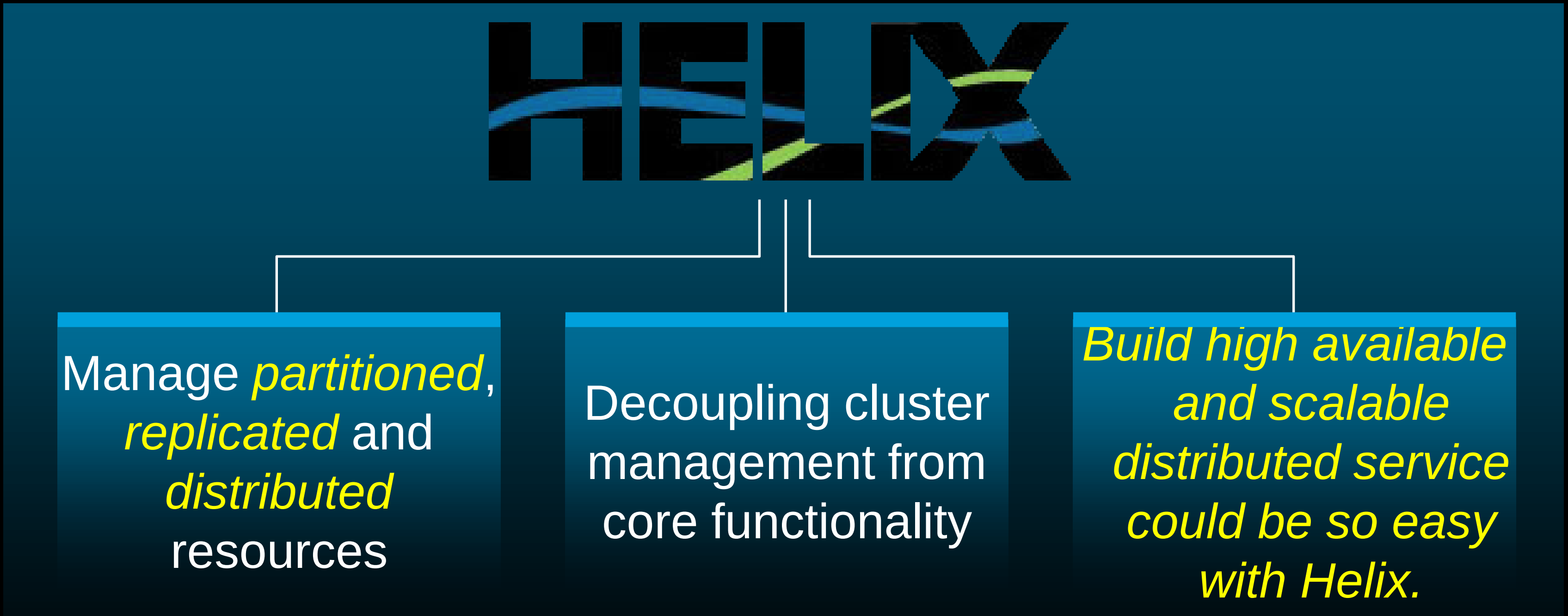
Ph.D from Duke University

PMC & Committer @ Apache Helix

Worked at Teradata Aster, VMware, LinkedIn

What is Helix?

What is Helix

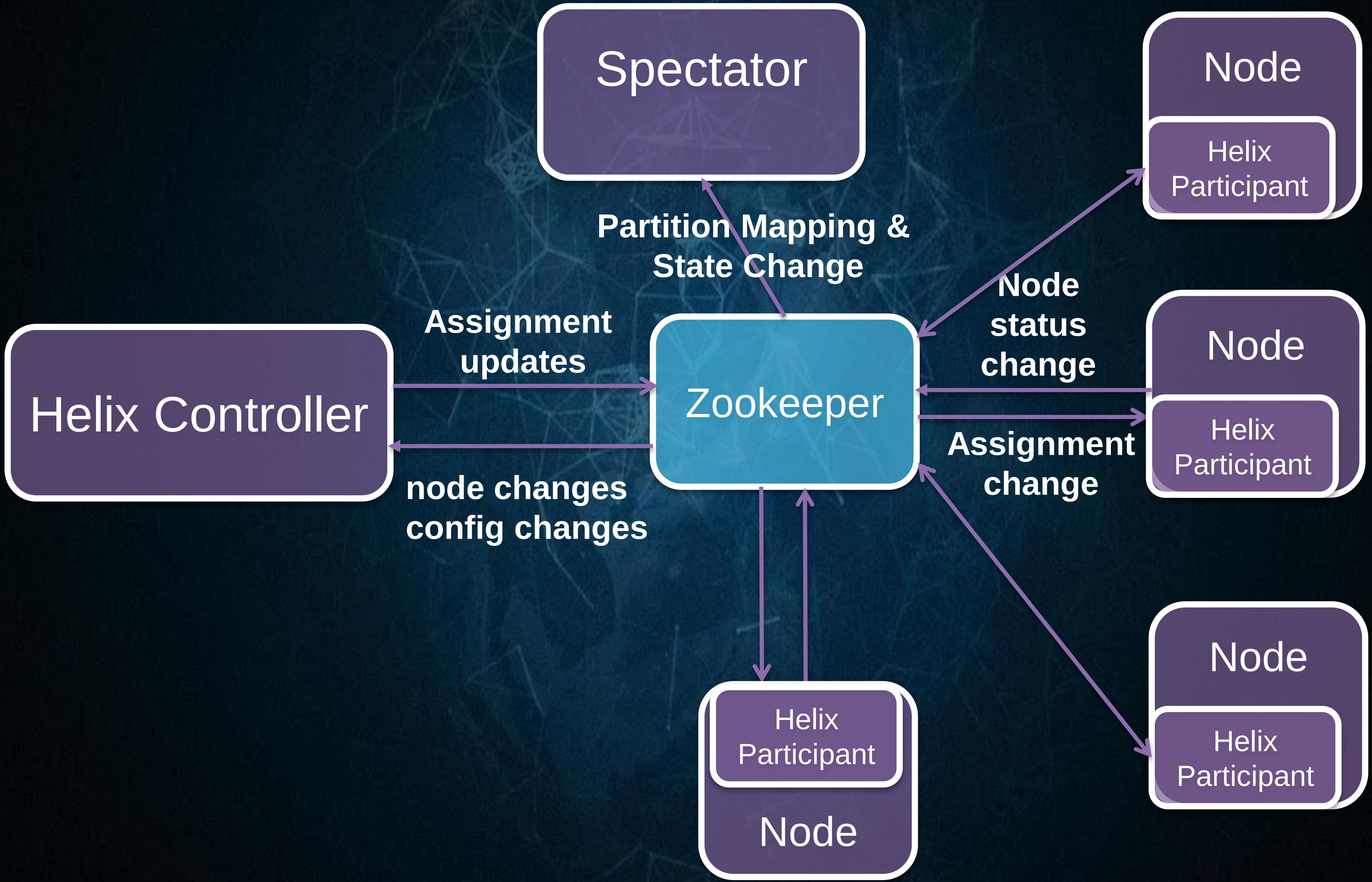


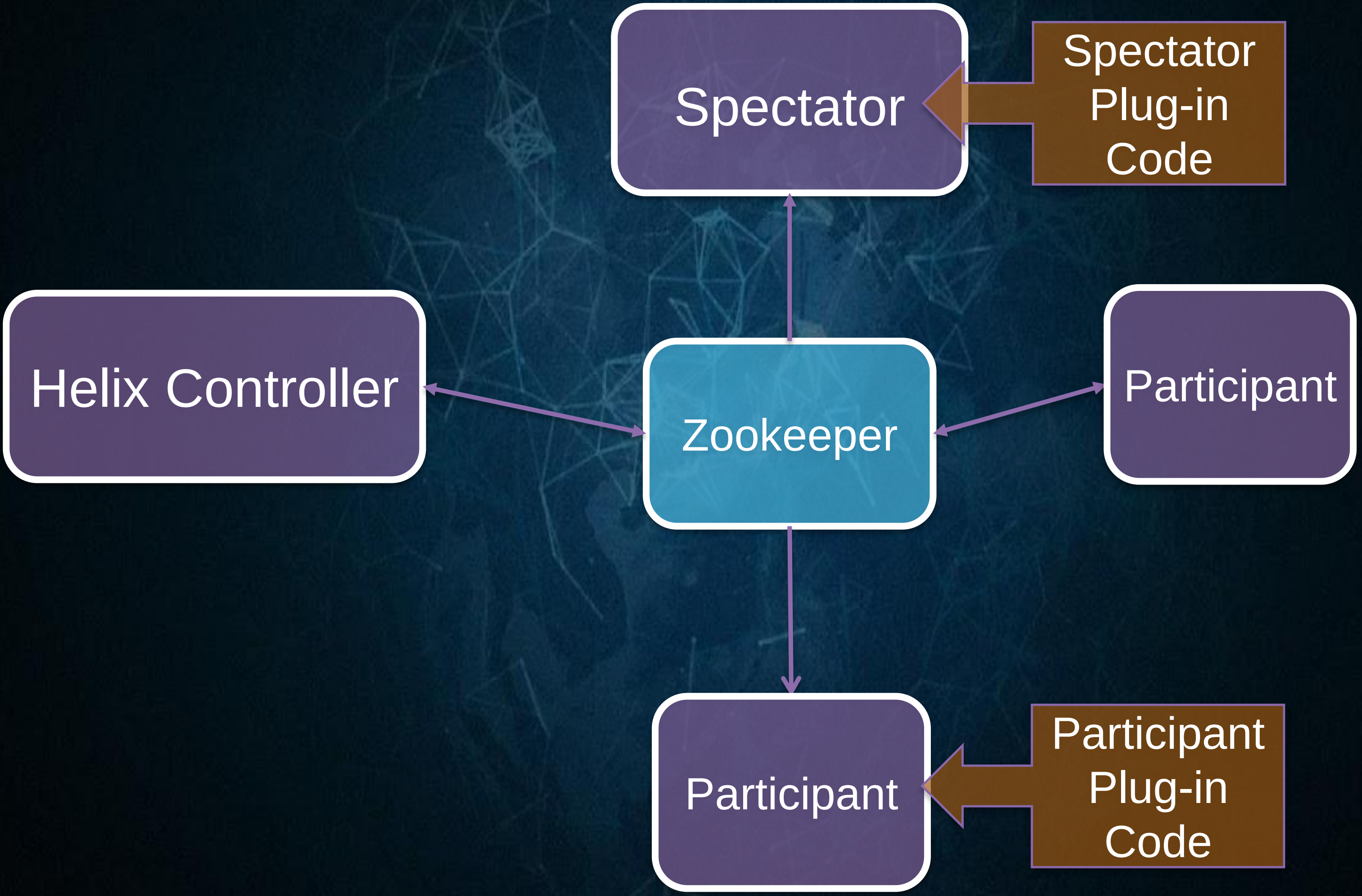
Manage *partitioned*, *replicated* and *distributed* resources

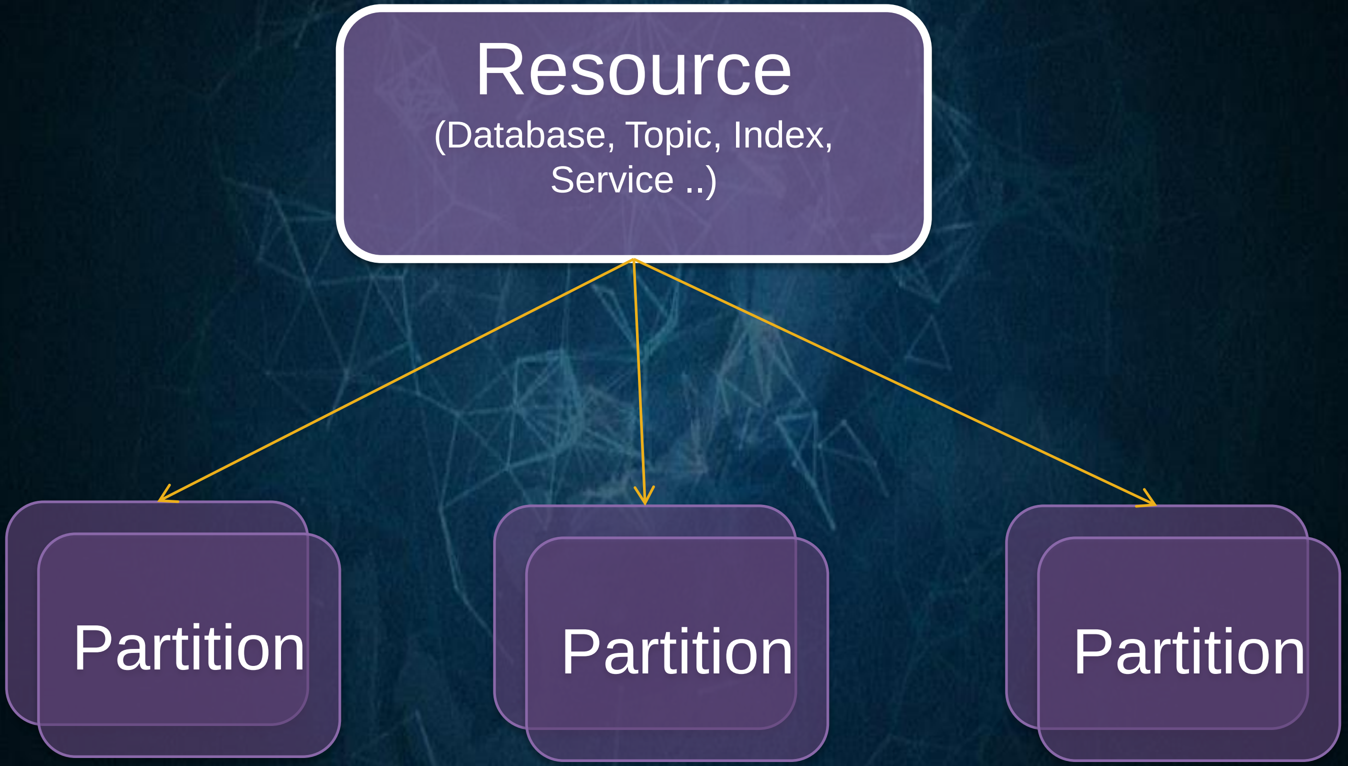
Decoupling cluster management from core functionality

Build high available and scalable distributed service could be so easy with Helix.

A generic cluster management and distributed task framework


Spectator
Node
Helix Participant
Partition Mapping & State Change
Node status change
Node
Helix Participant
Assignment updates
Helix Controller
Zookeeper
Assignment change
node changes config changes
Node
Helix Participant
Helix Participant
Node


Spectator
Spectator Plug-in Code
Helix Controller
Zookeeper
Participant
Participant
Participant Plug-in Code



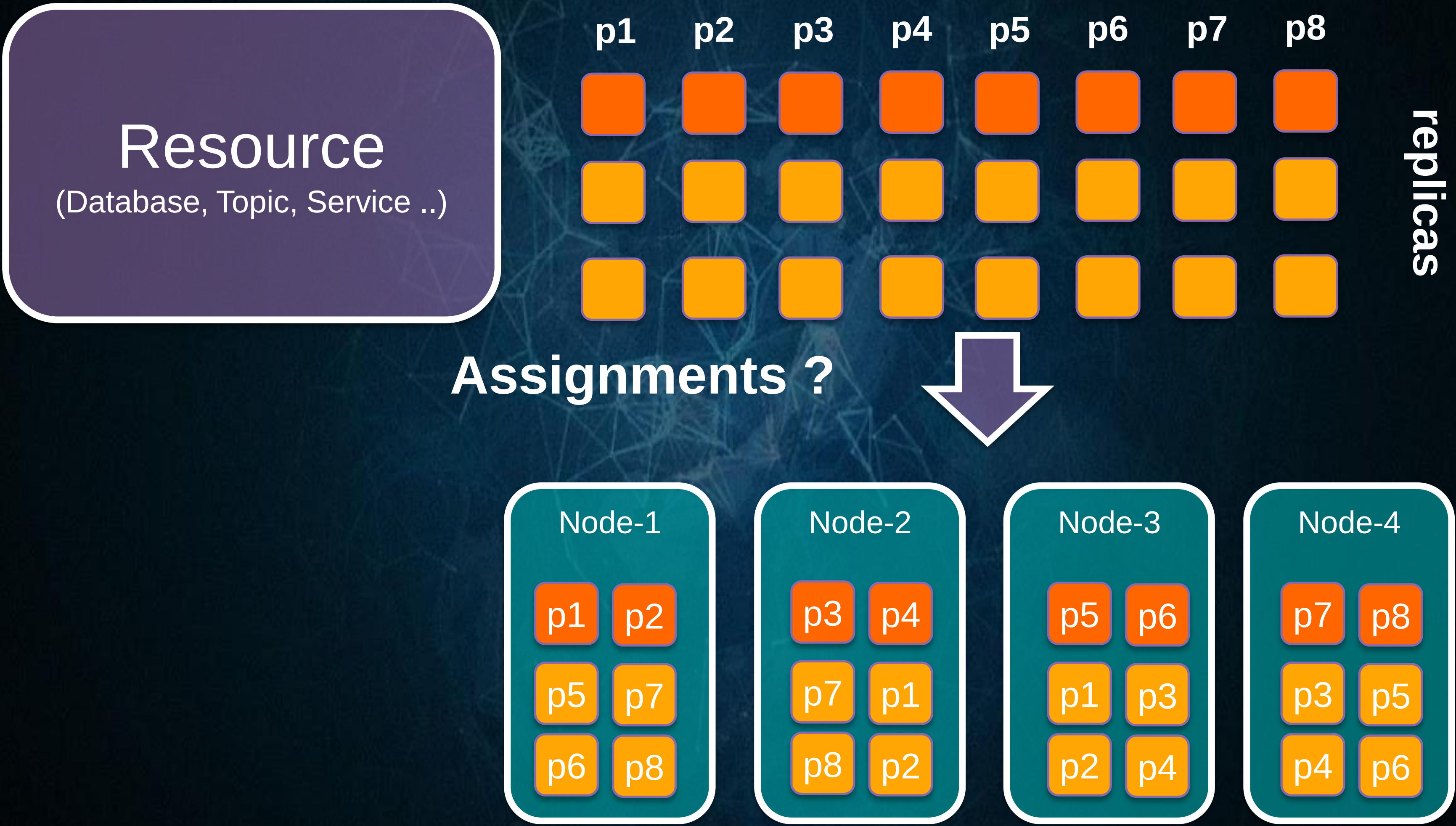
Resource

(Database, Topic, Index, Service ..)

Partition

Partition

Partition



Resource

(Database, Topic, Service ..)

p1 p2 p3 p4 p5 p6 p7 p8

replicas

Assignments ?

Node-1

p1 p2
p5 p7
p6 p8

Node-2

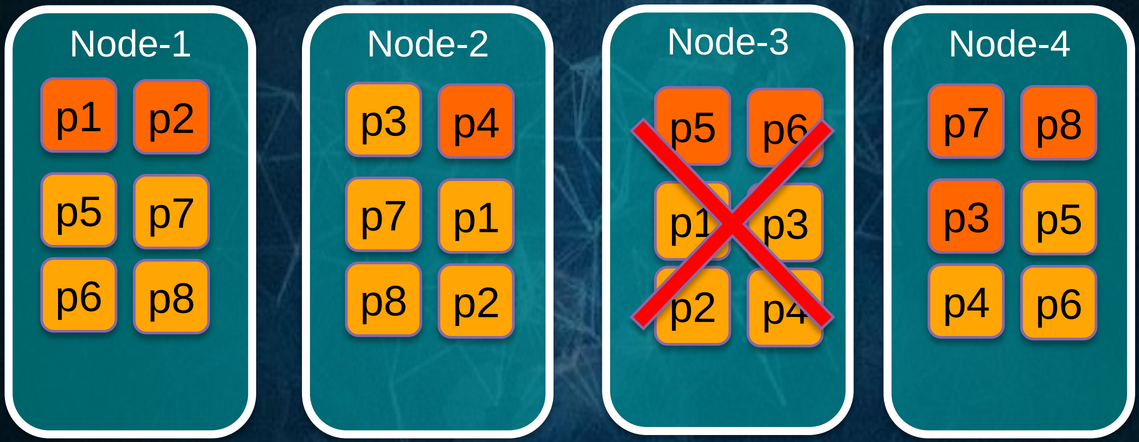
p3 p4
p7 p1
p8 p2

Node-3

p5 p6
p1 p3
p2 p4

Node-4

p7 p8
p3 p5
p4 p6

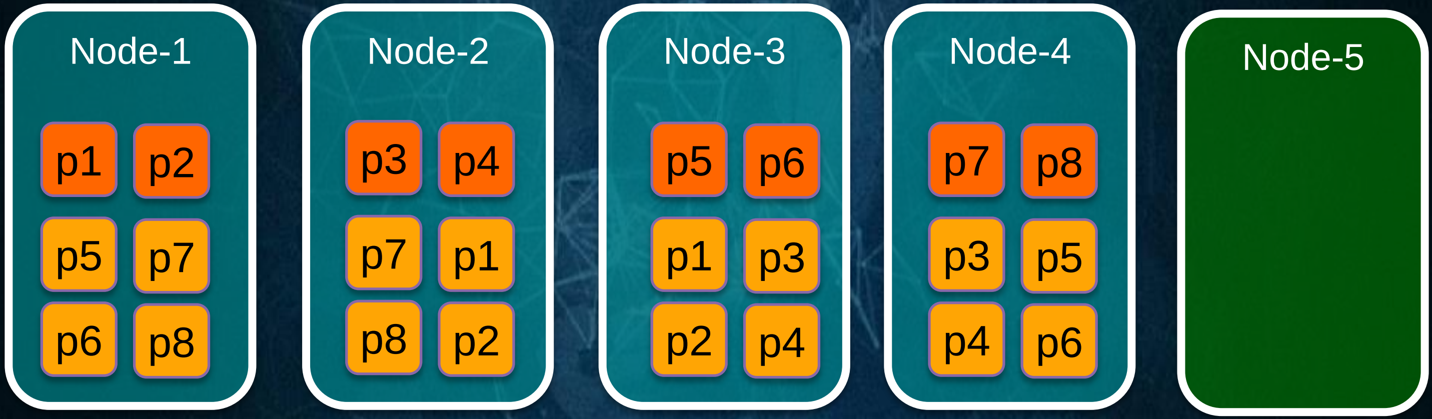


Node-1
p1 p2
p5 p7
p6 p8

Node-2
p3 p4
p7 p1
p8 p2

Node-3
p5 p6
p1 p3
p2 p4

Node-4
p7 p8
p3 p5
p4 p6



Node-1

p1 p2
p5 p7
p6 p8

Node-2

p3 p4
p7 p1
p8 p2

Node-3

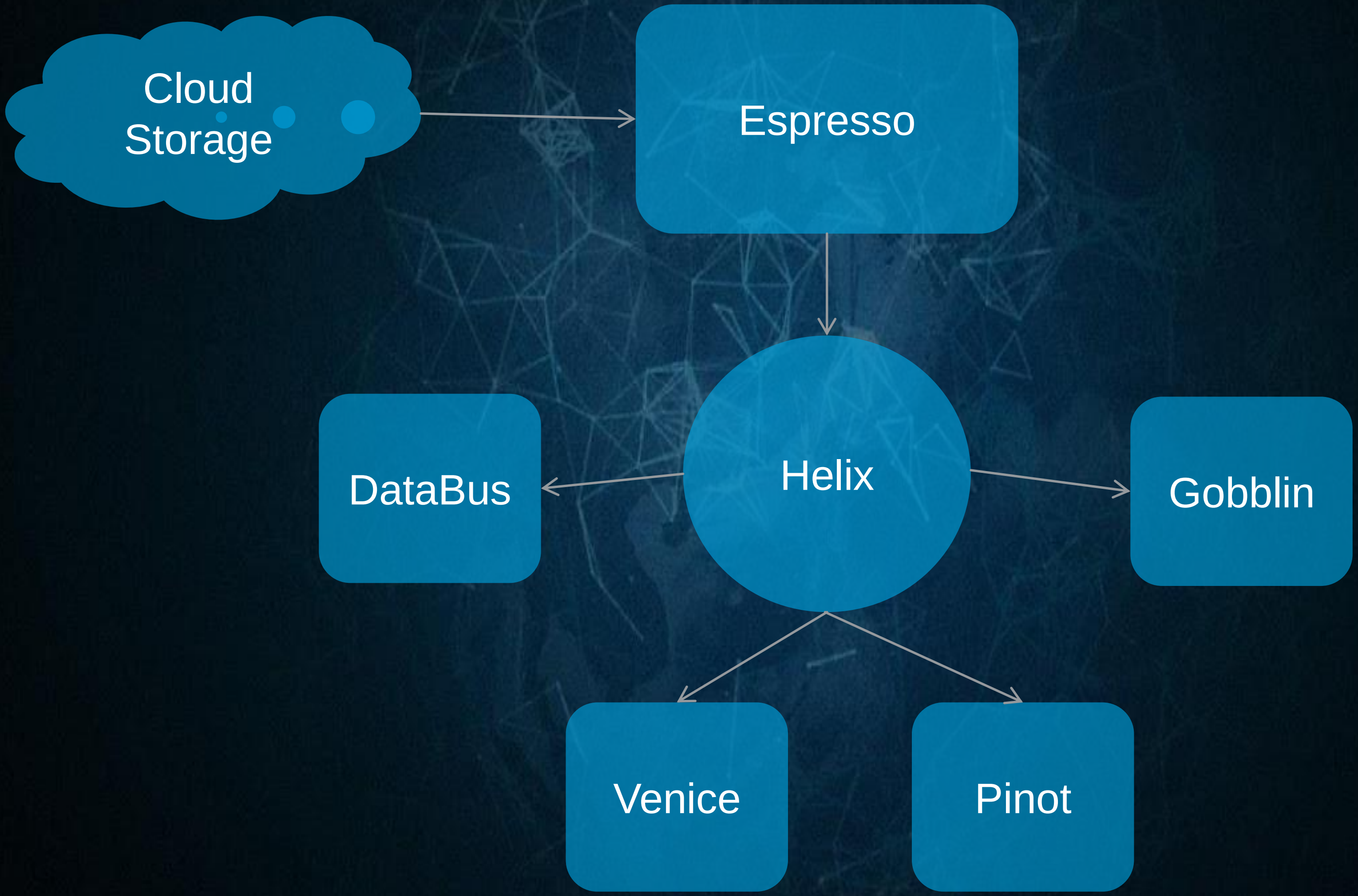
p5 p6
p1 p3
p2 p4

Node-4

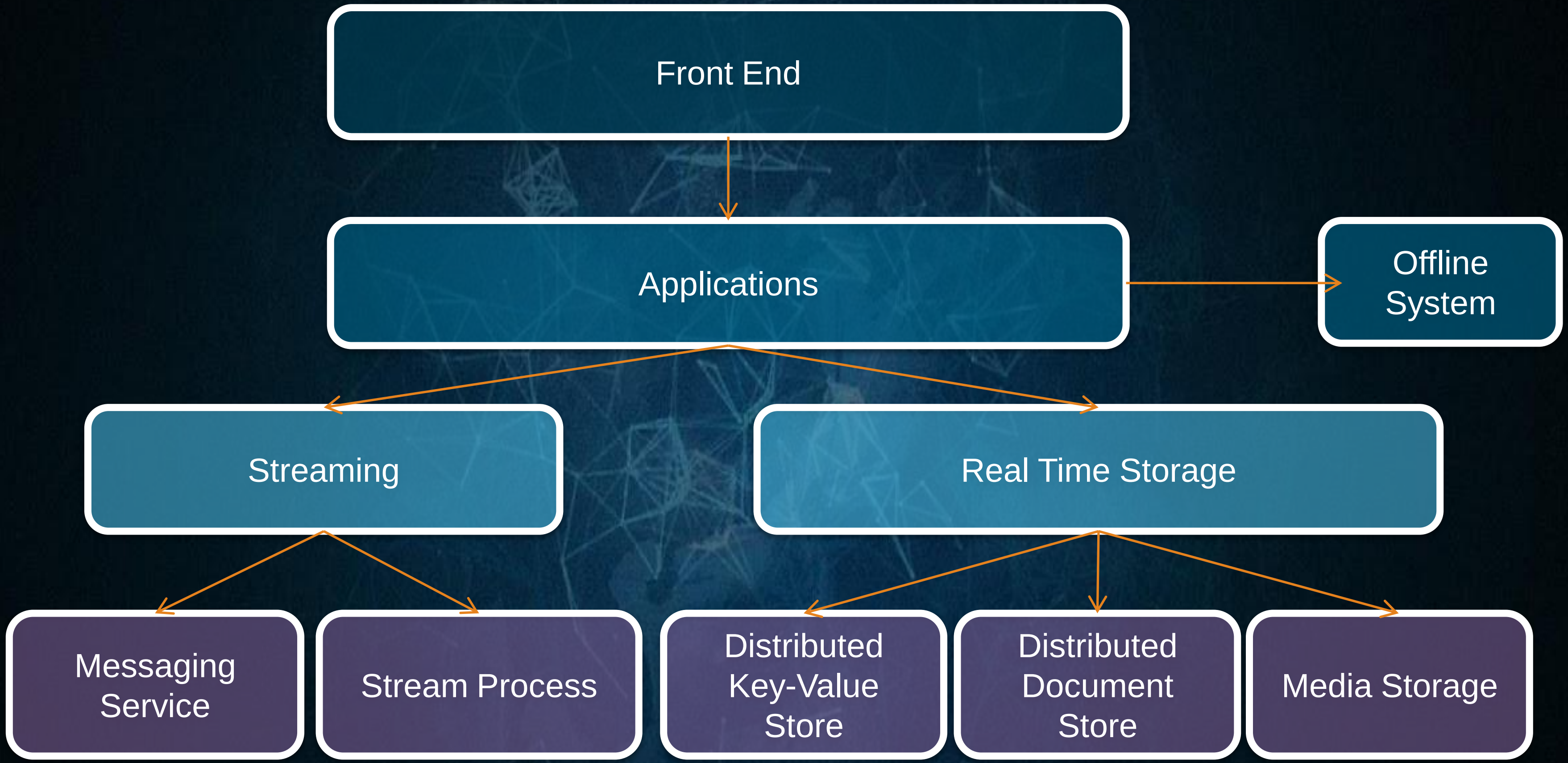
p7 p8
p3 p5
p4 p6

Node-5

Why Helix


Cloud Storage
Espresso
Helix
DataBus
Gobblin
Venice
Pinot

Online Ecosystem



Front End

Applications

Offline System

Streaming

Real Time Storage

Messaging Service

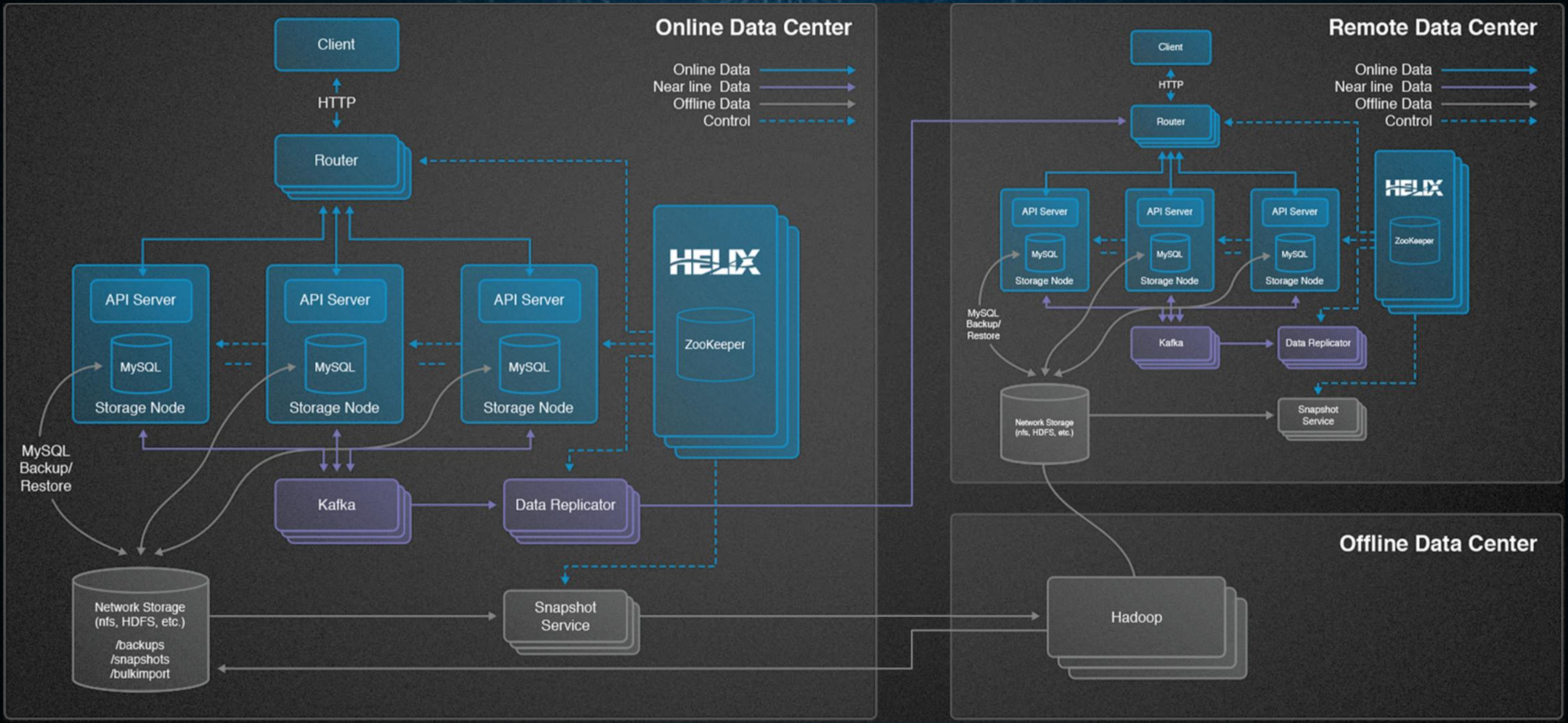
Stream Process

Distributed Key-Value Store

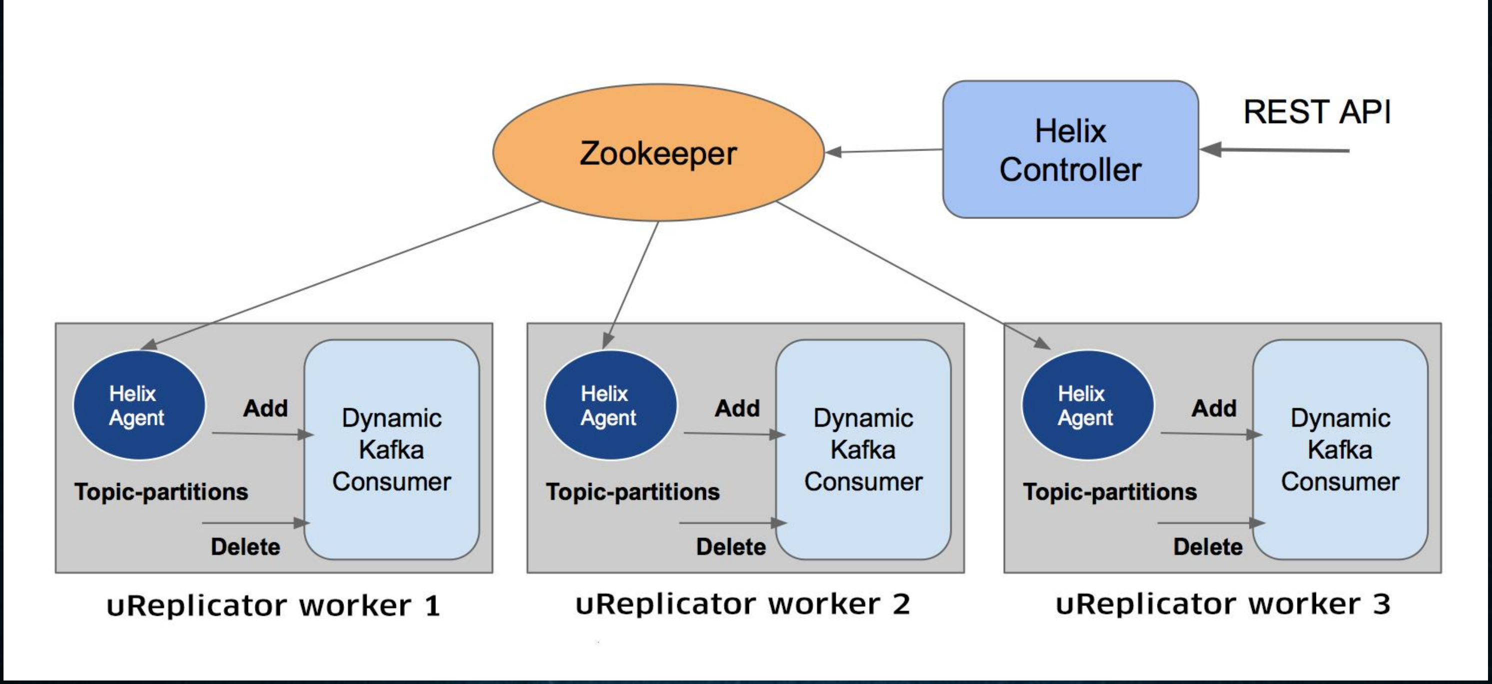
Distributed Document Store

Media Storage

Espresso : LinkedIn Distributed, Fault-Tolerant Database

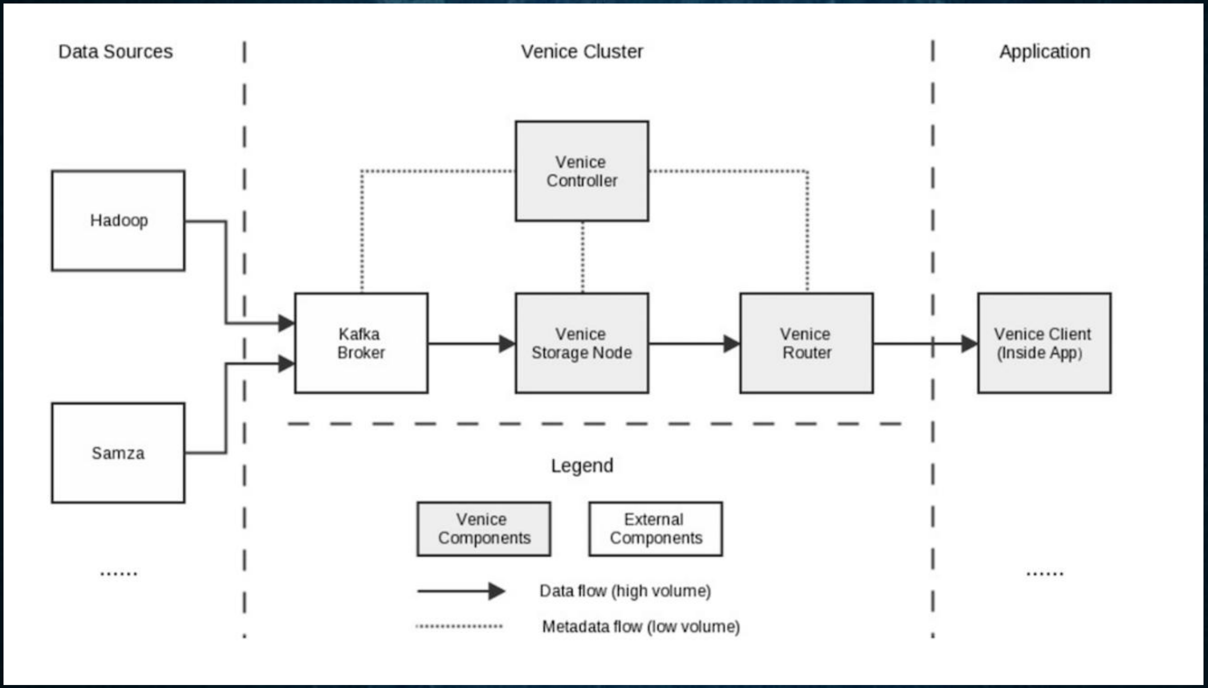


Helix based Kafka replication



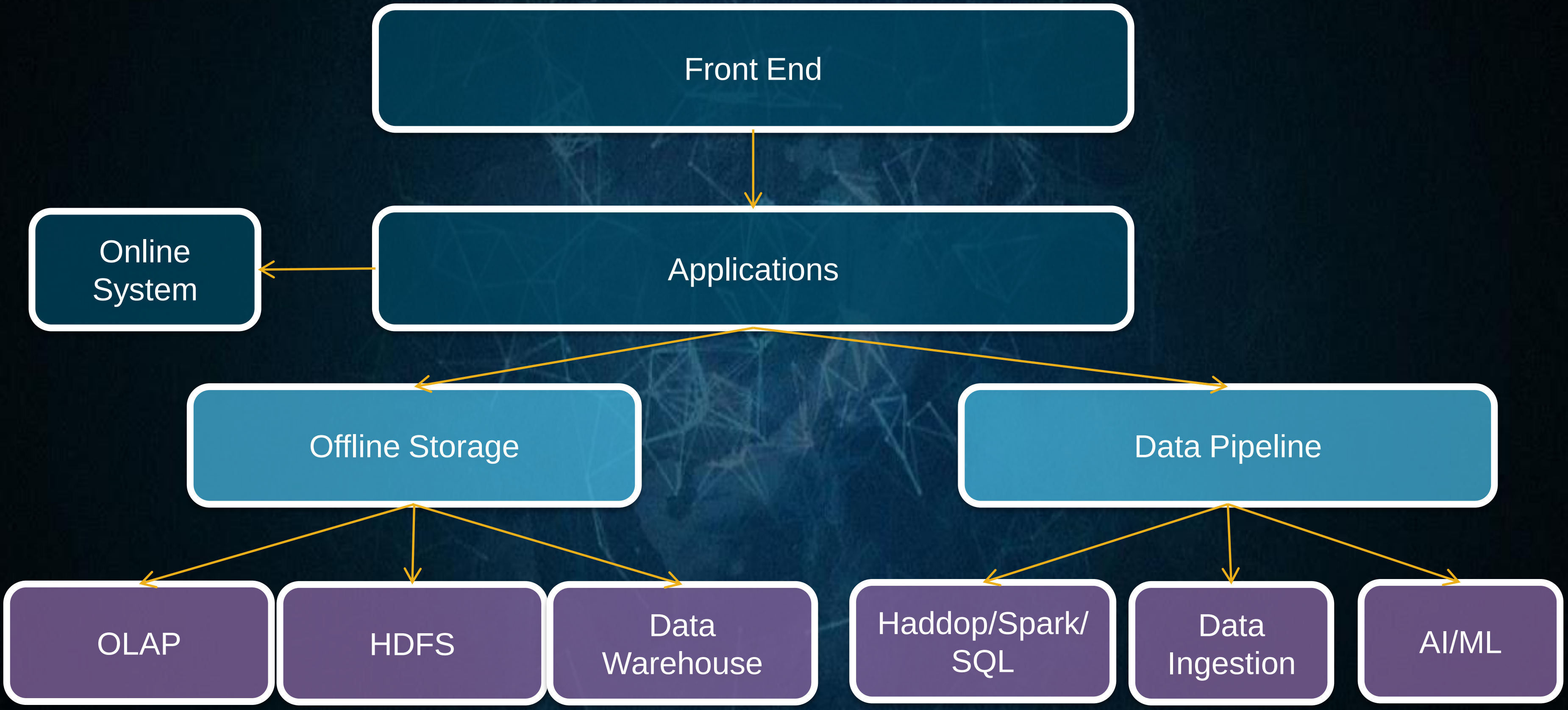
CHINMAY SOMAN, YUANCHI NING, XIANG FU & HONGLIANG XU. *UREPLICATOR: UBER ENGINEERING'S ROBUST KAFKA REPLICATOR*, Uber Engineering Blog

Venice : next- generation derived data serving platform



YAN YAN. *Building Venice with Apache Helix,* LinkedIn Engineering Blog

Offline Ecosystem



Front End

Online System

Applications

Offline Storage

Data Pipeline

OLAP

HDFS

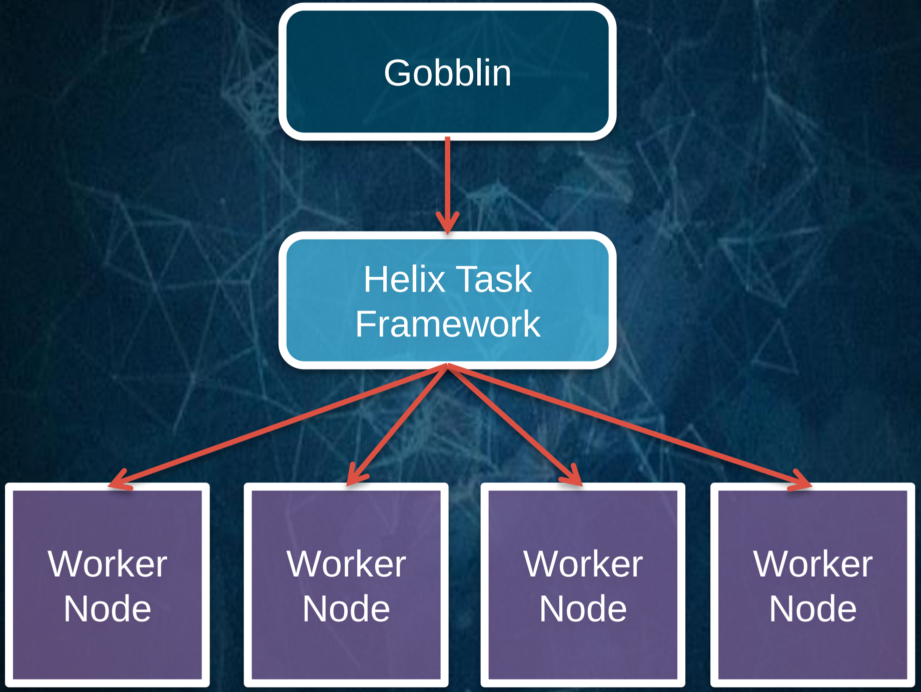
Data Warehouse

Haddop/Spark/ SQL

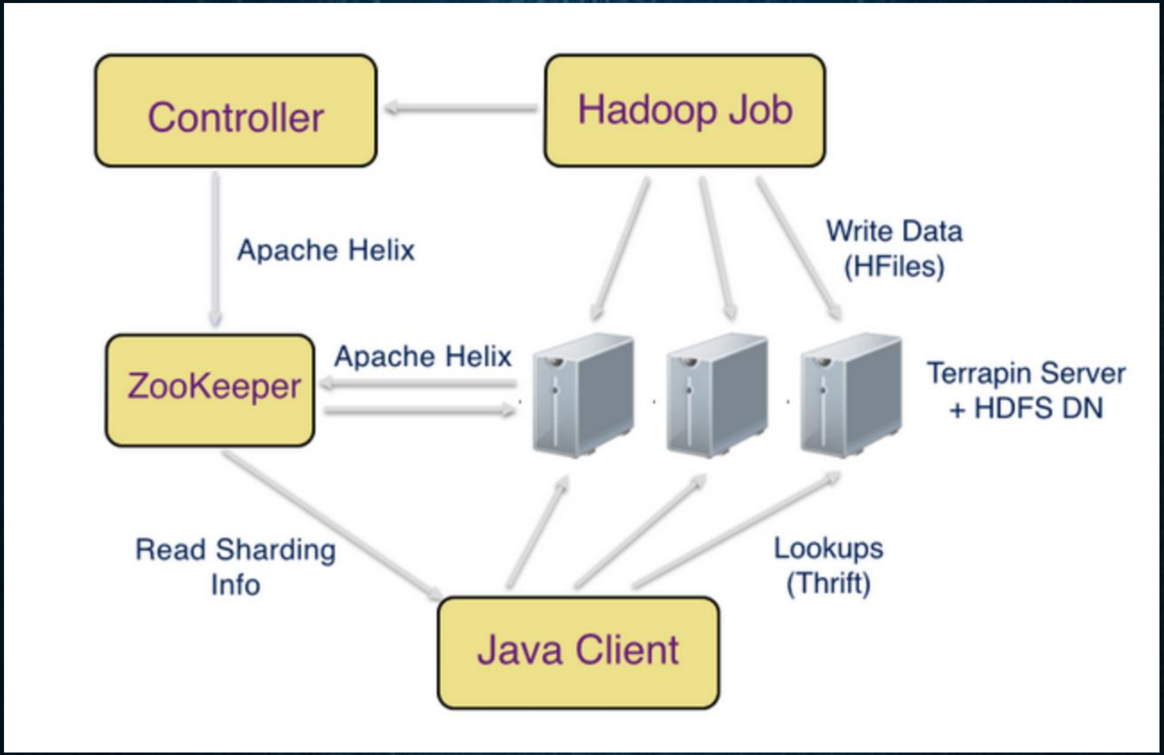
Data Ingestion

AI/ML

Gobblin: a distributed data ingestion framework



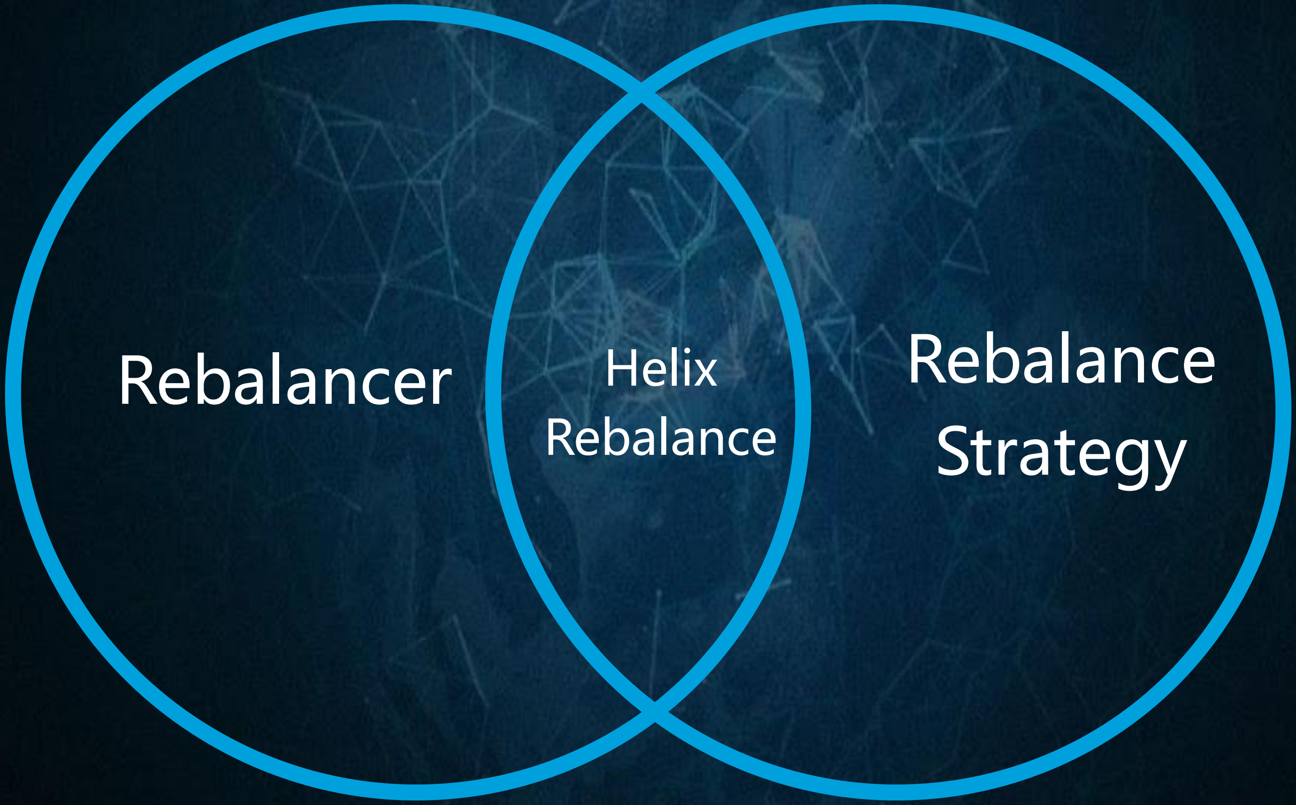
Helix based data batching system



VARUN SHARMA, *Open-sourcing Terrapin: A serving system for batch generated data*, Pinterest Engineering Blog

Rebalance Strategy Algorithms

What is rebalance in Helix?



Rebalancer

Helix Rebalance

Rebalance Strategy

Rebalancer Types

Customized Rebalancer

Semi Auto Rebalancer

Full Auto Rebalancer

Delayed Auto Rebalancer

Supported rebalance strategy algorithms

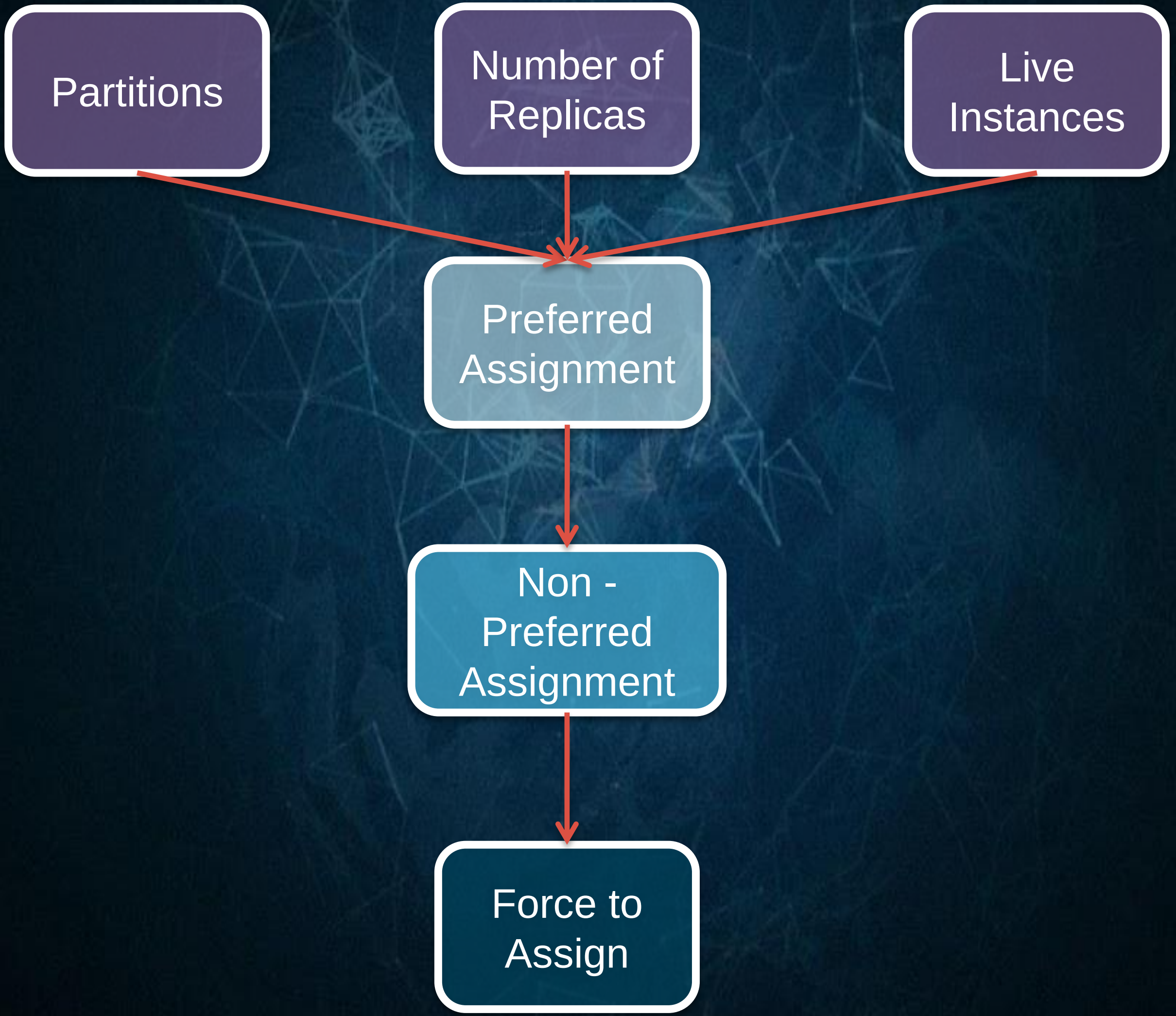
Greedy algorithm

Consistent hashing algorithm

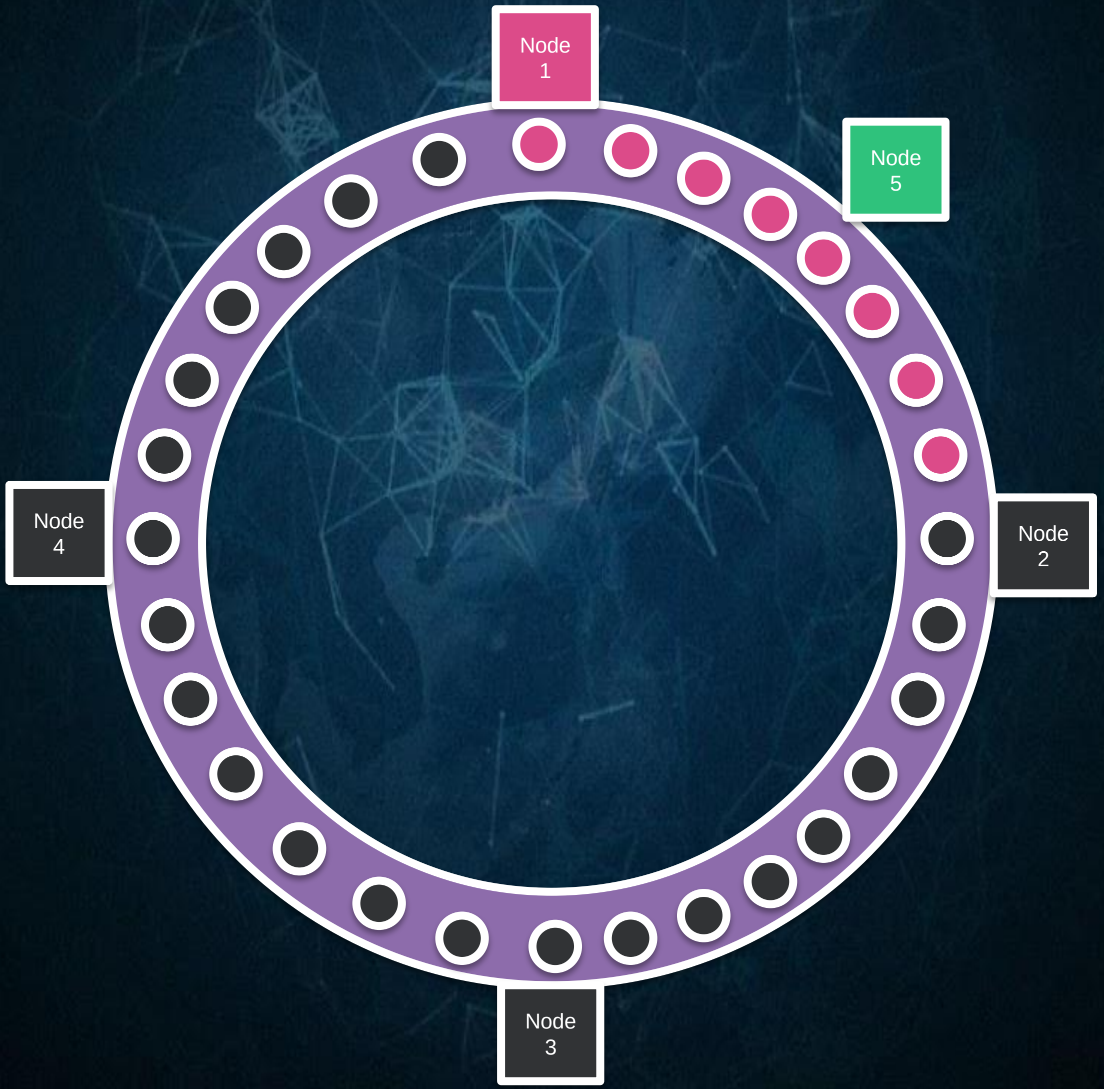
CRUSH algorithm

Multi-round CURSH algorithm

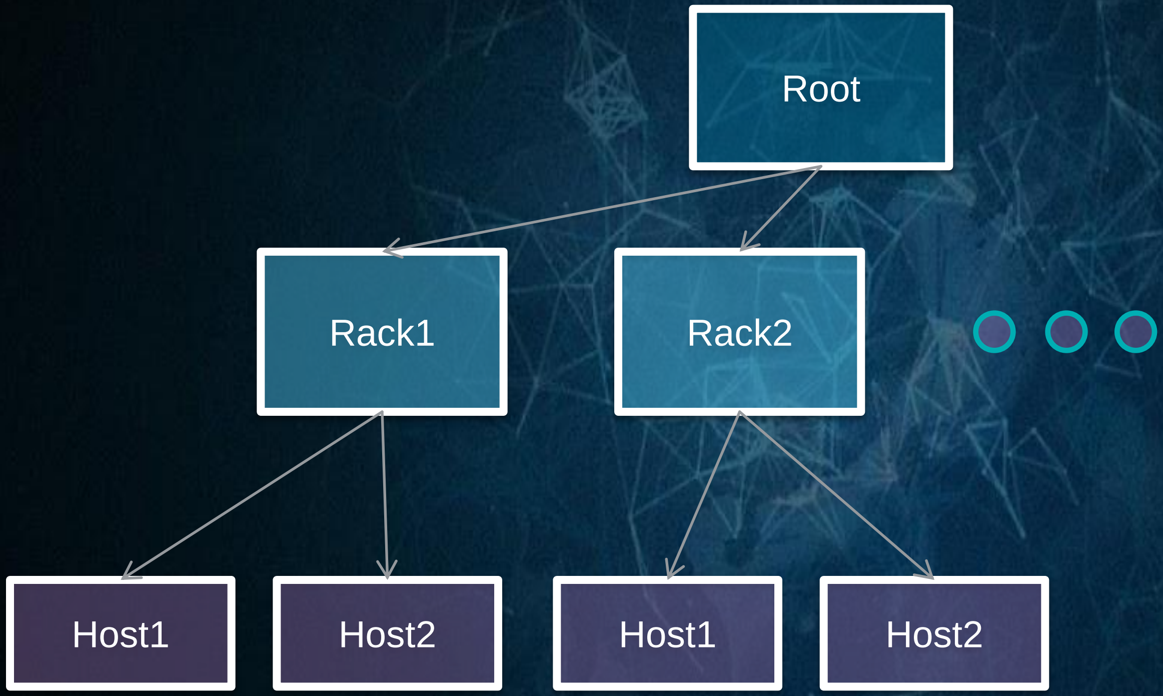
Greedy Algorithm



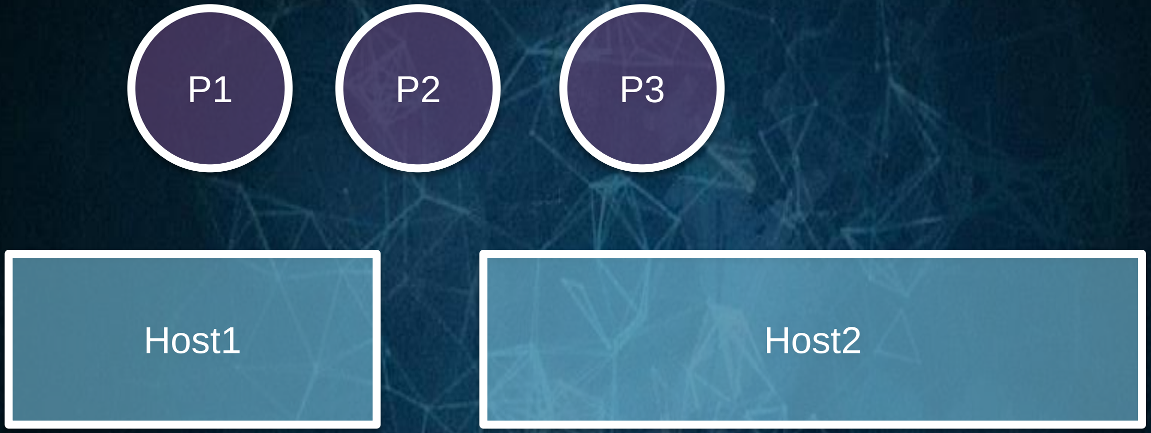
Consistent Hashing Algorithm



CRUSH Algorithm Map Structure

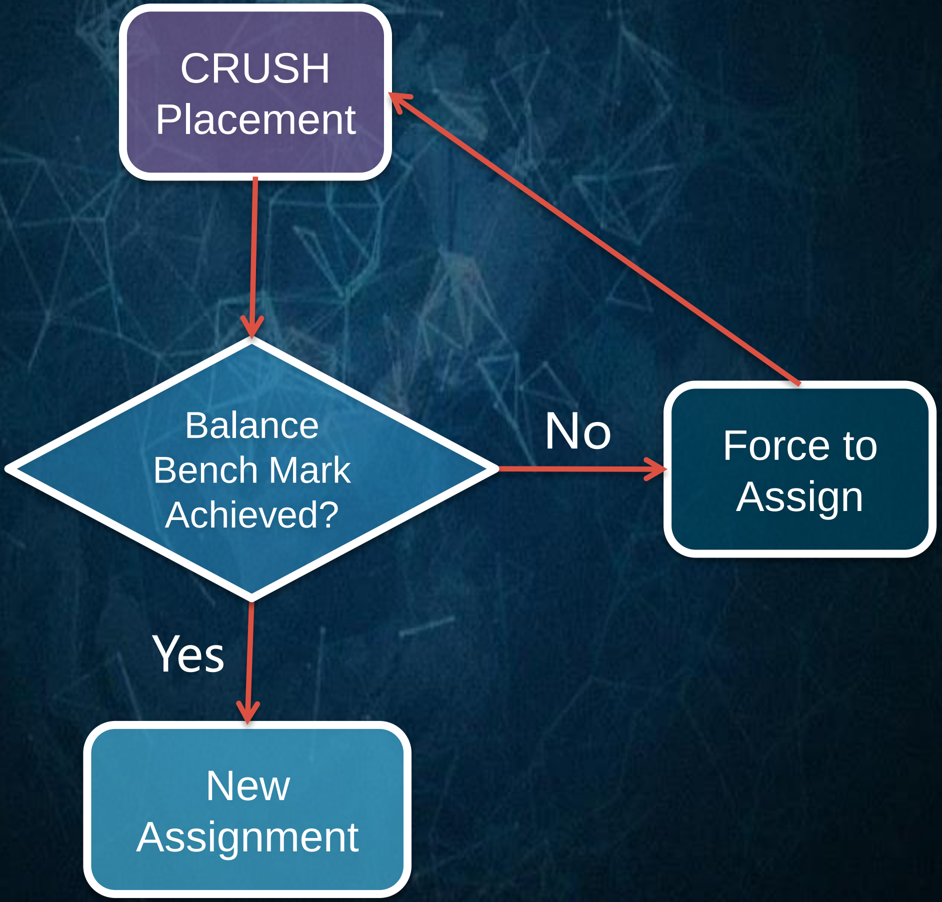


CRUSH Algorithm Placement



$$c(r,x) = max(f(wi)hash(x,r,i))$$

Multi-Round CRUSH Algorithm Placement

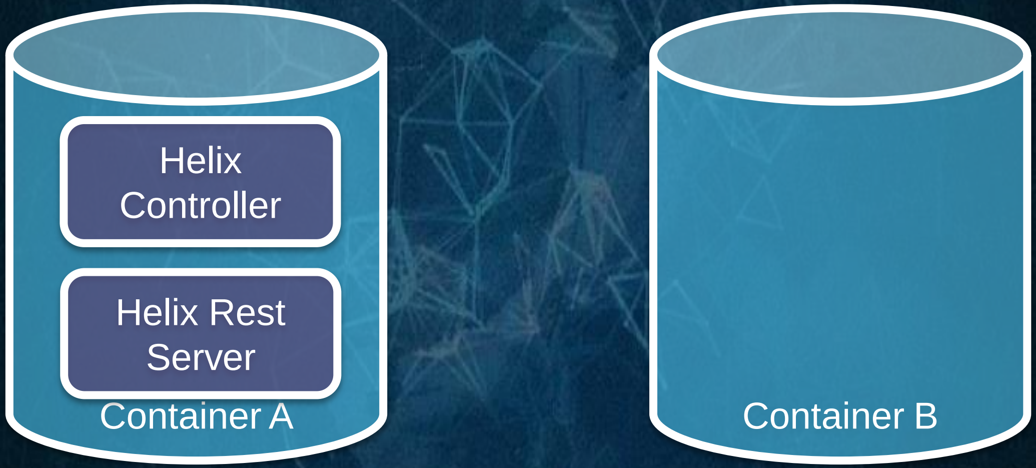


Helix & Container

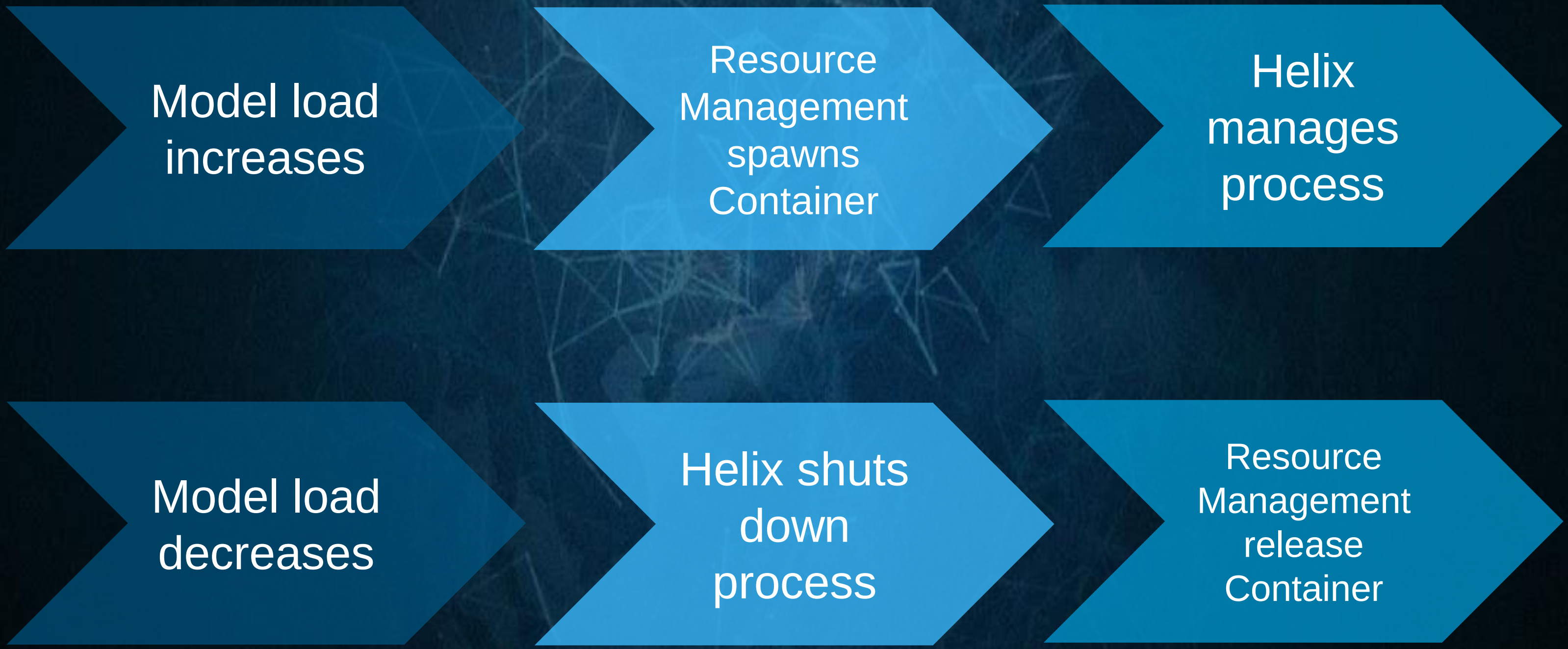
Kubernetes VS Docker

Features	Kubernetes	Docker
Setup	Combine with YAML, API and clients with manual configuration	Docker API but limited functions, but faster setup
Load Balancing	Need manually enable pods are defined as services	Internal load balancing any nodes in the cluster
Data Volumes	Share data across same pod	Shared local volumes
Networking	More flexible in configuration	Automatically authenticated TLS

Helix Deployment



Helix Participant Auto Scaling



Model load increases

Resource Management spawns Container

Helix manages process

Model load decreases

Helix shuts down process

Resource Management release Container

Conclusion

Generic cluster management and distributed task framework for building distributed systems.

Helix widely impacts architecture Online/Offline Ecosystems.

Different RebalanceStrategy can solve varies problems.

Q&A

Thank you!