GOTC

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

OPEN SOURCE , OPEN WORLD

Milvus：探索云原生的向量数据库

郭人通 2021年07月10日

GOTC

郭人通

兴趣领域：

分布式系统、数据库、异构计算

Milvus 系统架构师

CCF 分布式计算与系统专委会委员

华中科技大学

计算机软件与理论博士

ZILLIZ

合伙人 & 架构师

Why Vector Database

Data are Increasing Horizontally : Types

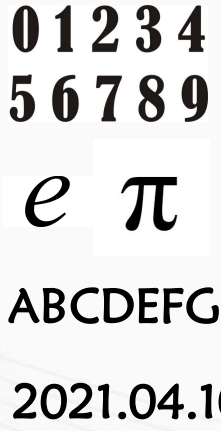
Int, float, string, …

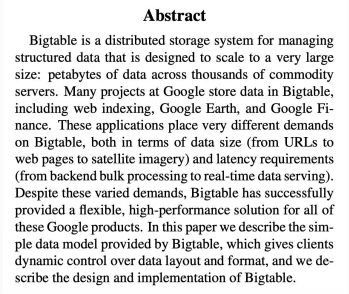
text

json

image
video
audio

domain specific

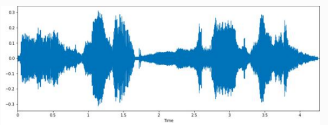


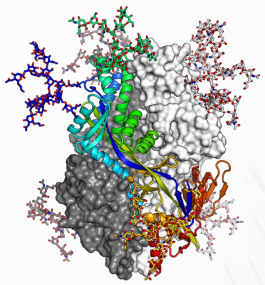


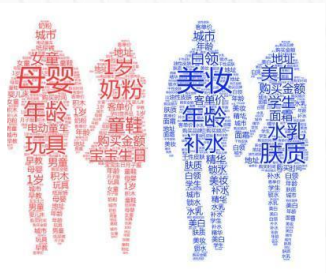
Abstract

Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Bigtable, both in terms of data size (from URLs to web pages to satellite imagery) and latency requirements (from backend bulk processing to real-time data serving). Despite these varied demands, Bigtable has successfully provided a flexible, high-performance solution for all of these Google products. In this paper we describe the simple data model provided by Bigtable, which gives clients dynamic control over data layout and format, and we describe the design and implementation of Bigtable.

```
{
  "firstName": "John",
  "lastName": "Smith",
  "isAlive": true,
  "age": 27,
  "address": {
    "streetAddress": "21 2nd Street",
    "city": "New York",
    "state": "NY",
    "postalCode": "10021-3100"
  },
  "phoneNumbers": [
    {
      "type": "home",
      "number": "212 555-1234"
    },
    {
      "type": "office",
      "number": "646 555-4567"
    }
  ],
  "children": [],
  "spouse": null
}
```





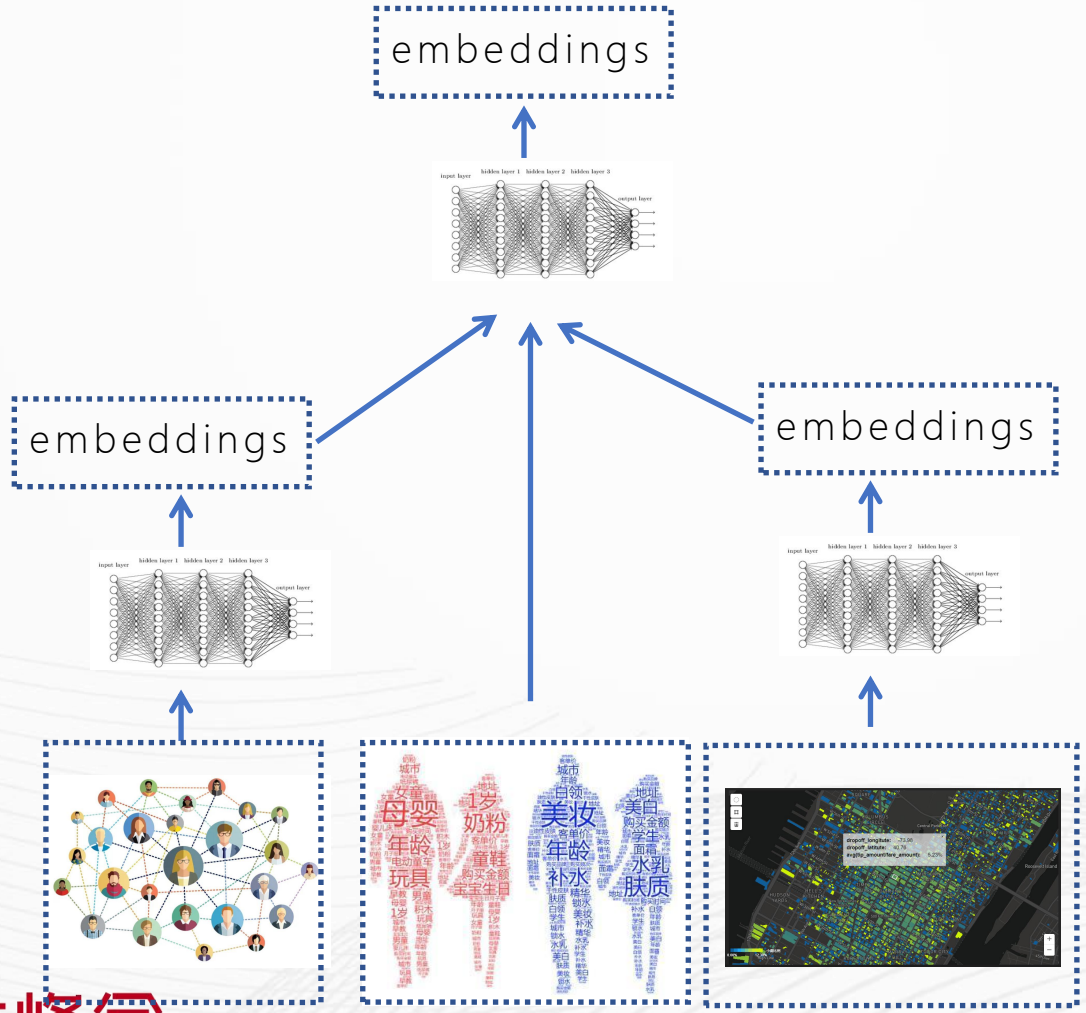


Structured data

Unstructured data

Why Vector Database

Data are Increasing Vertically : Semantics

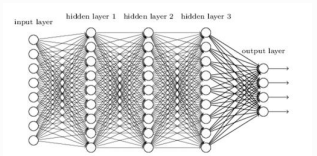


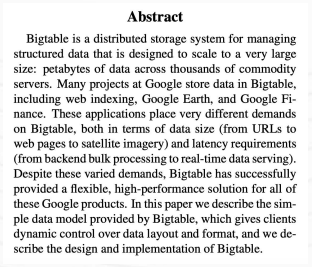
Richer semantics

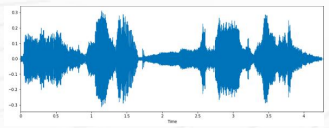
Why Vector Database

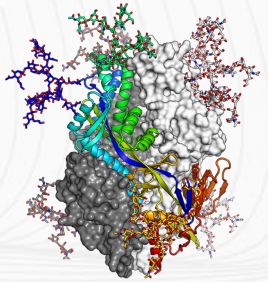
boolean Int float string ... embedding

Types with high abstraction



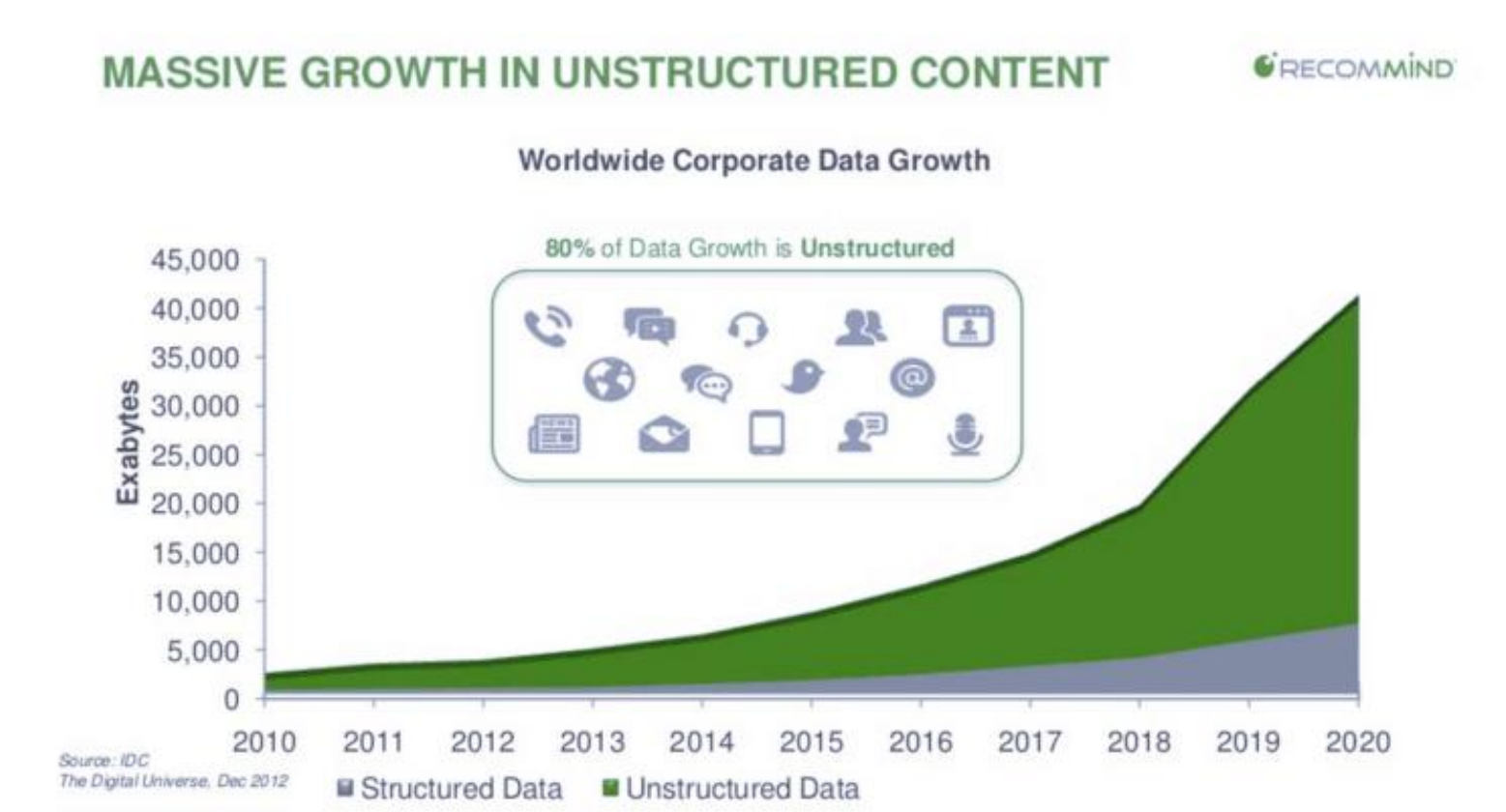




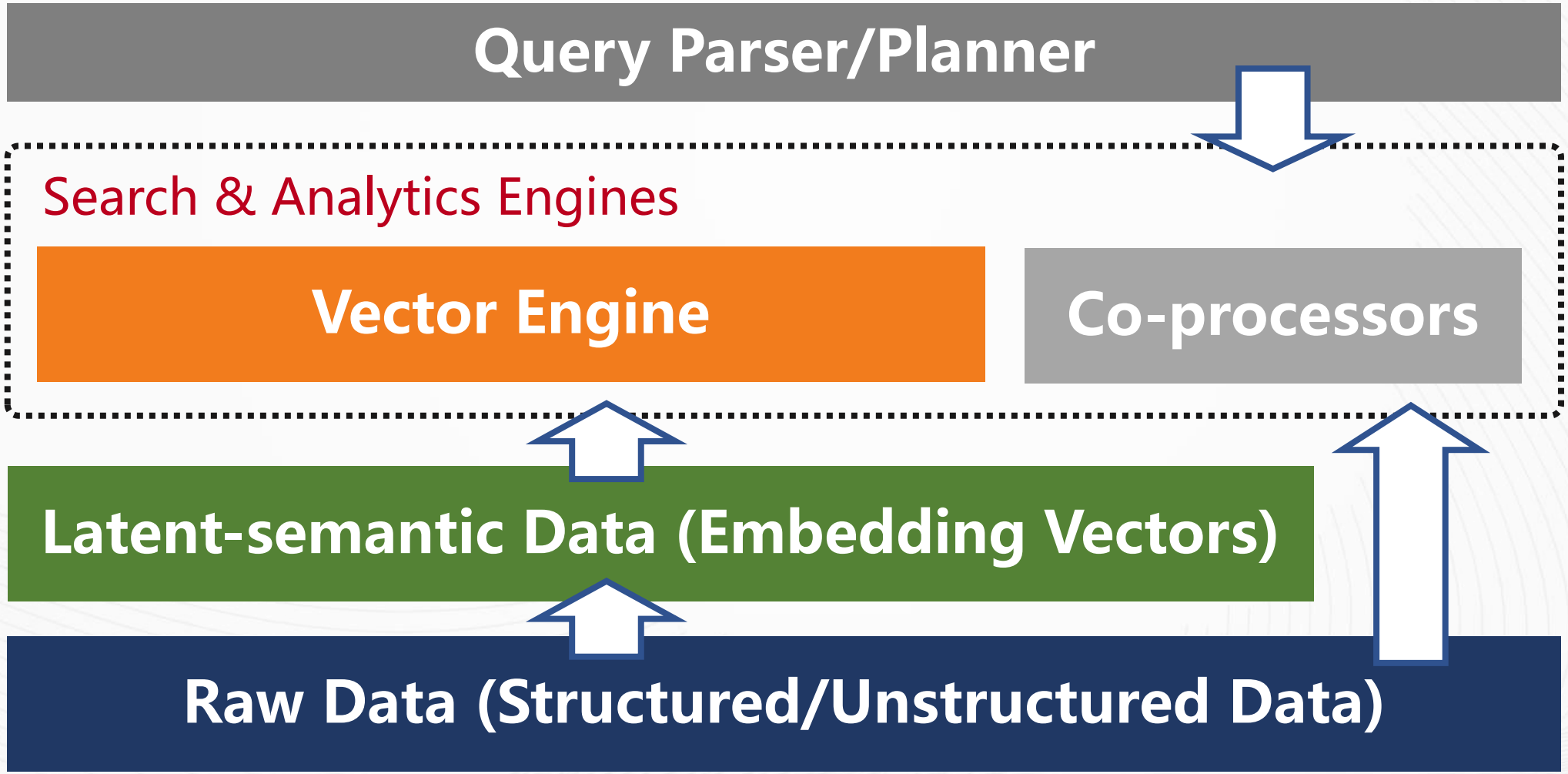


Why Vector Database

80% data growth is unstructured, over 40,000 Exabytes per year

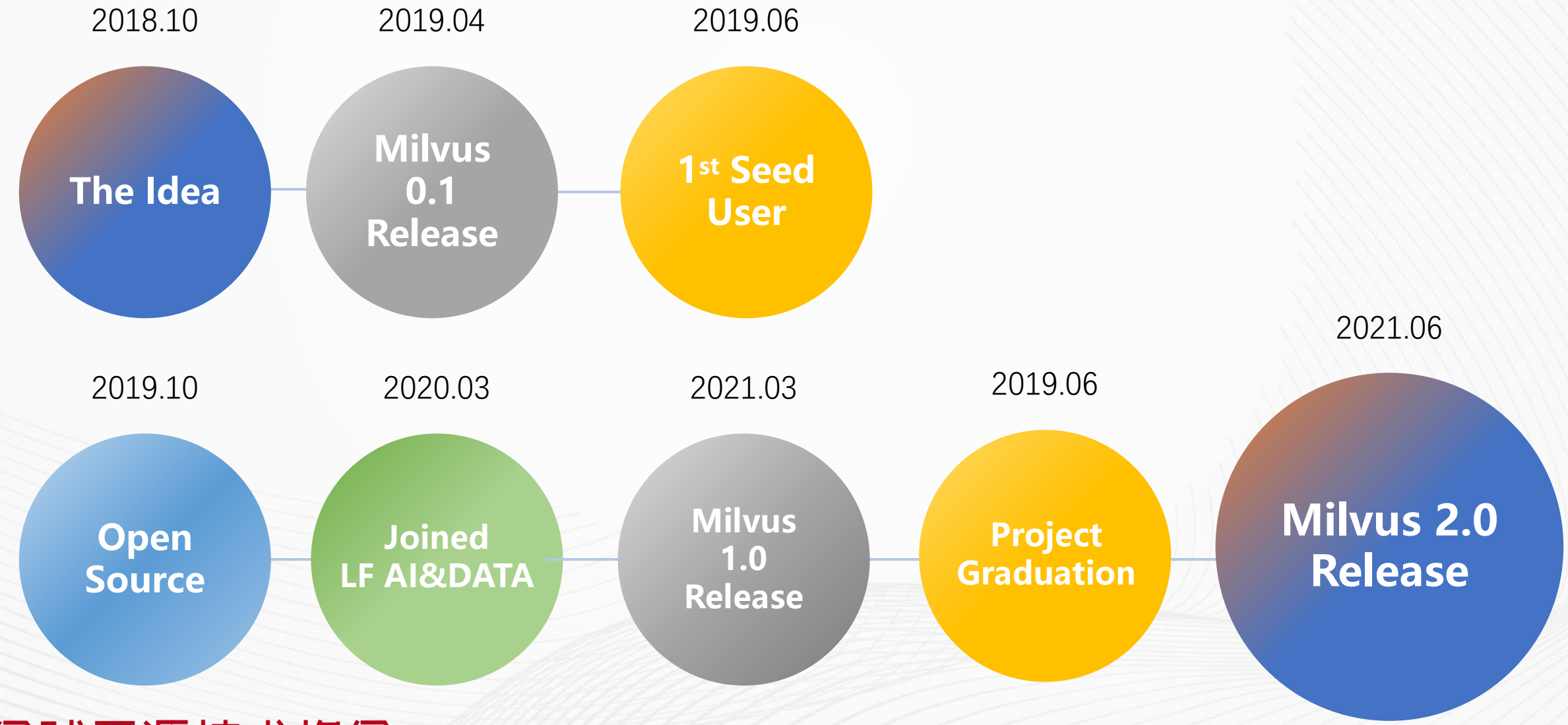


The big picture



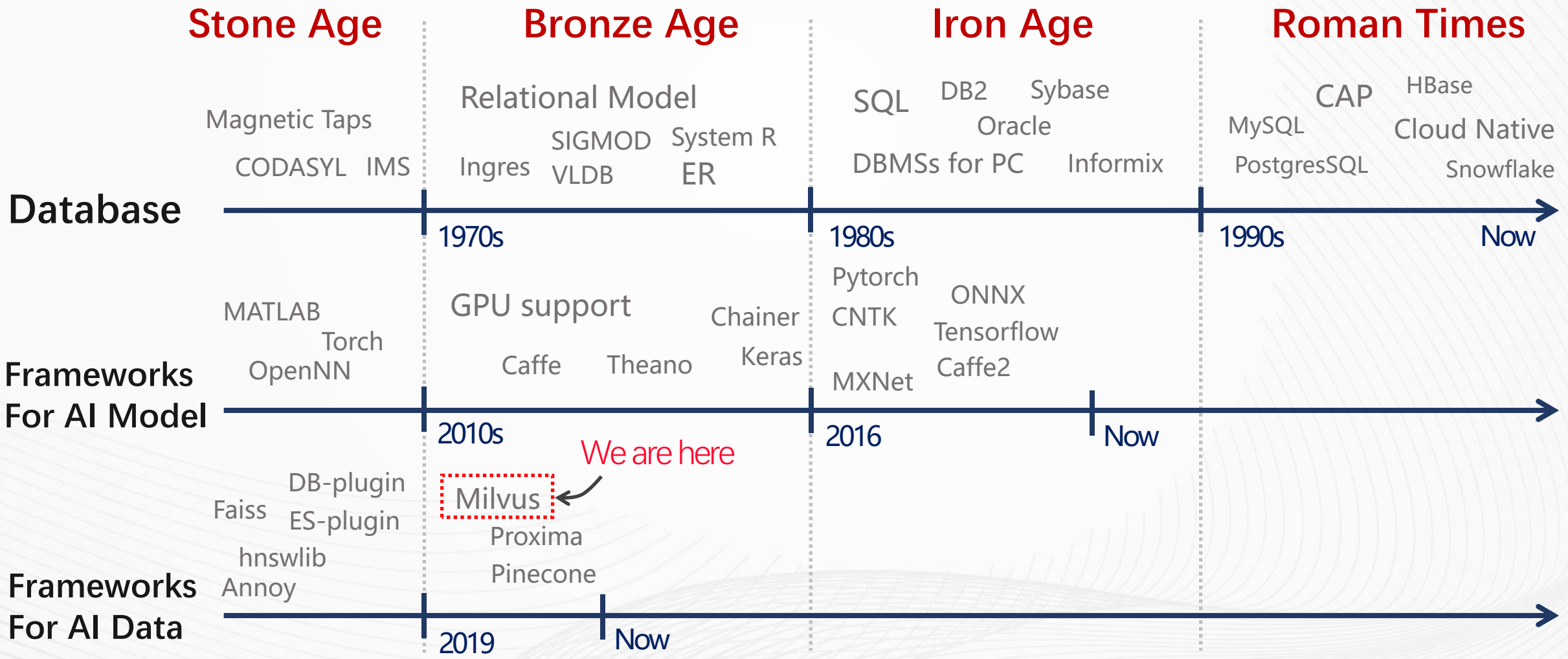
About Milvus

GOTC



A brief history

GOTC



Key challenges

01 **Fast System Evolution**

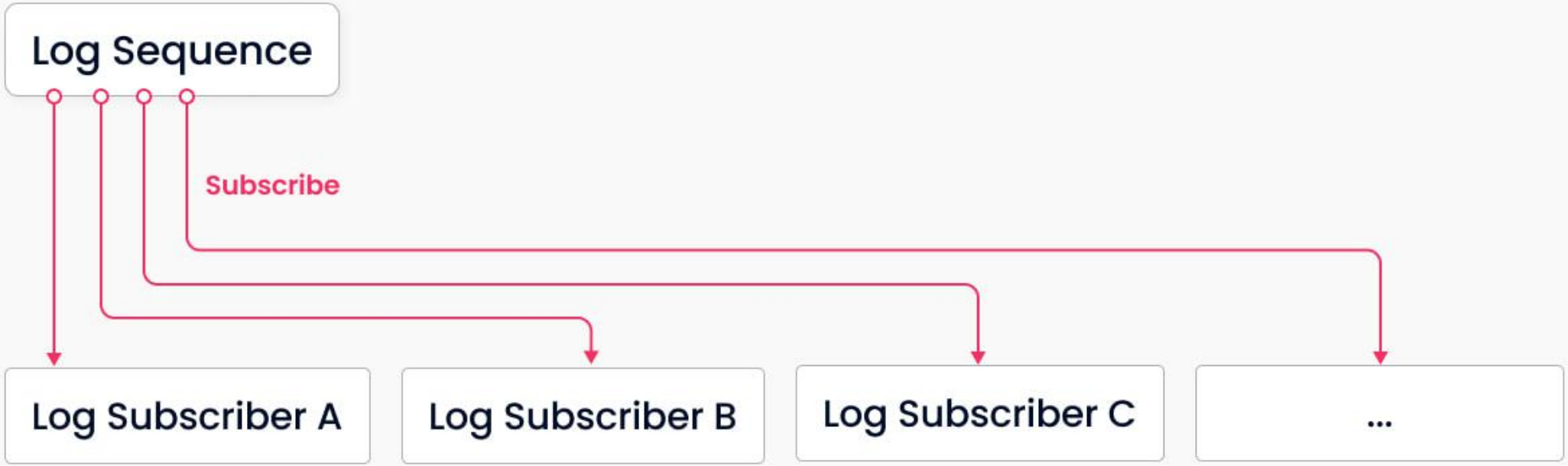
02 **Multi-environment Deployment**

03 **Hardware Cost**

04 **Diverse Workloads**

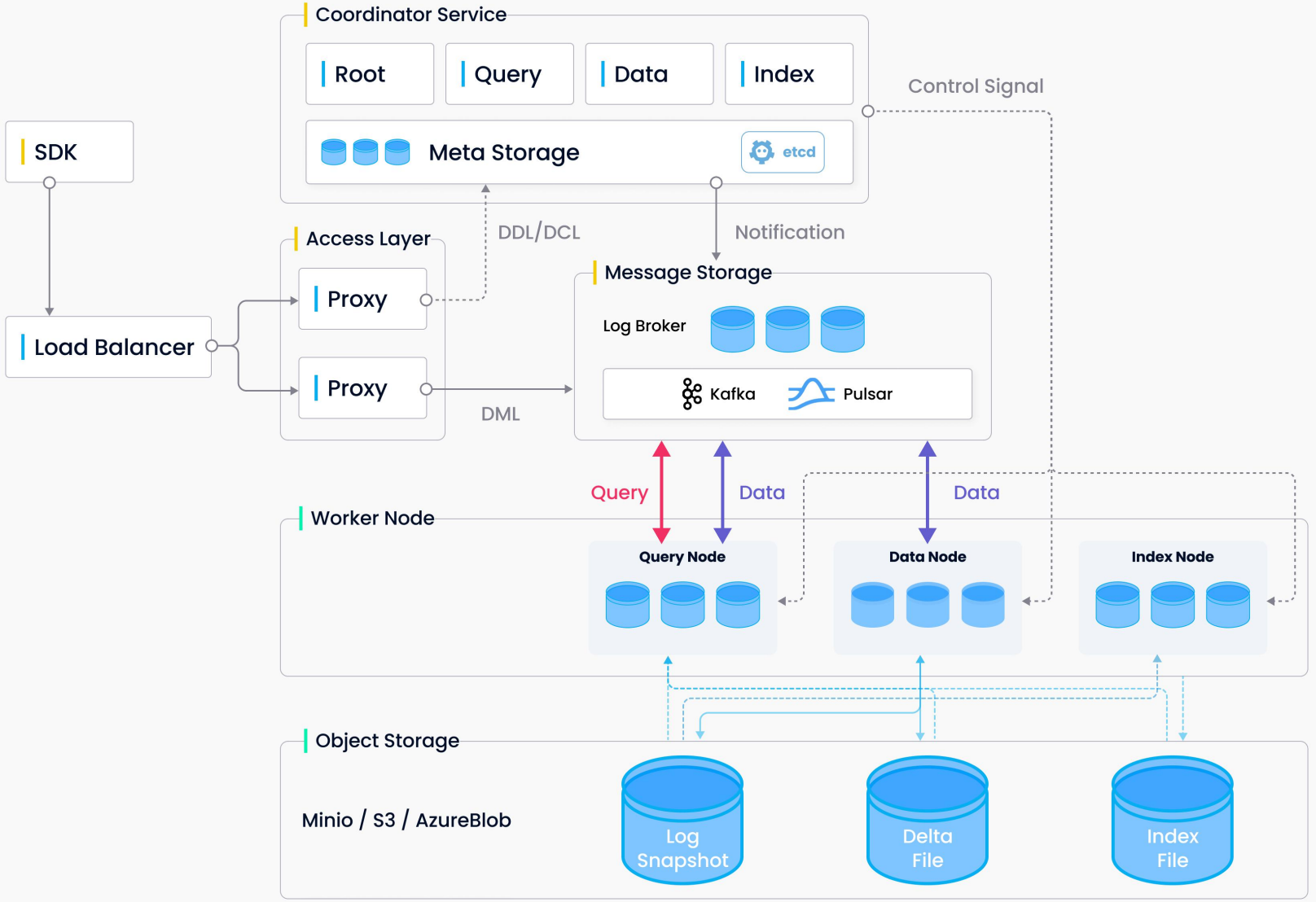
05 **Complex, Hybrid Query**

Architecture: logical log as the system backbone



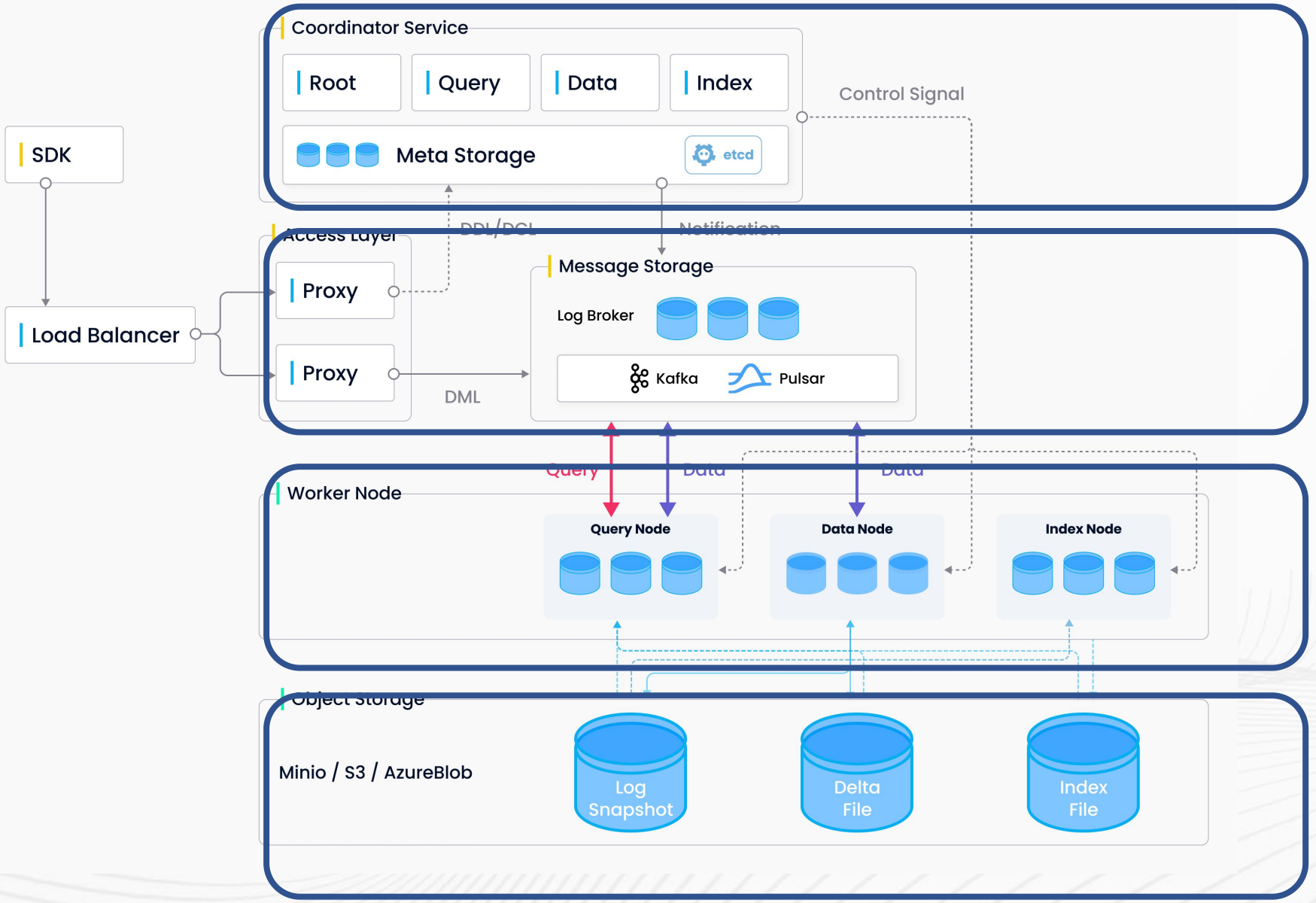
Architecture: take a closer look

GOTC



Architecture: take a closer look

GOTC



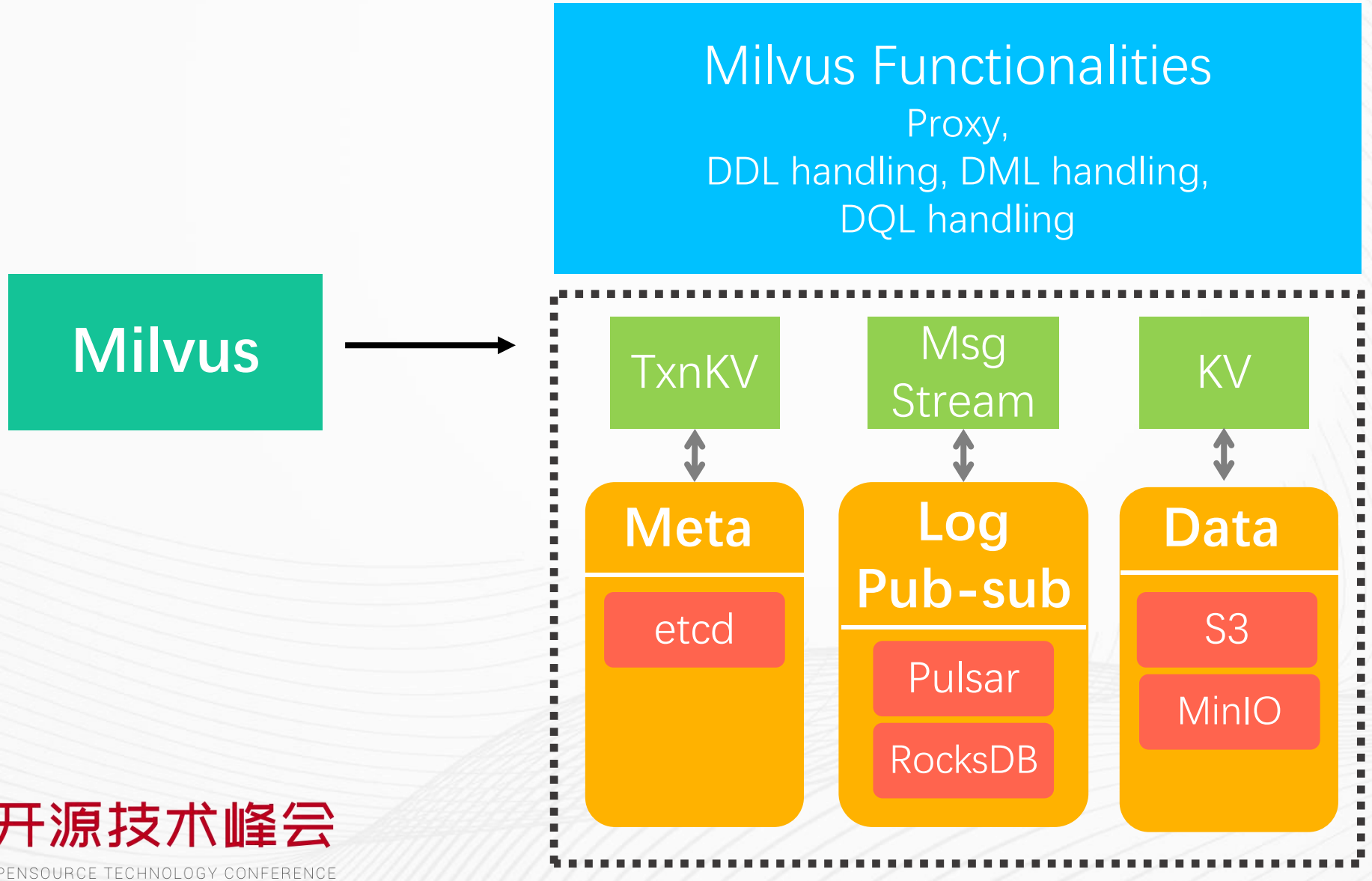
Coordinators

Log Broker

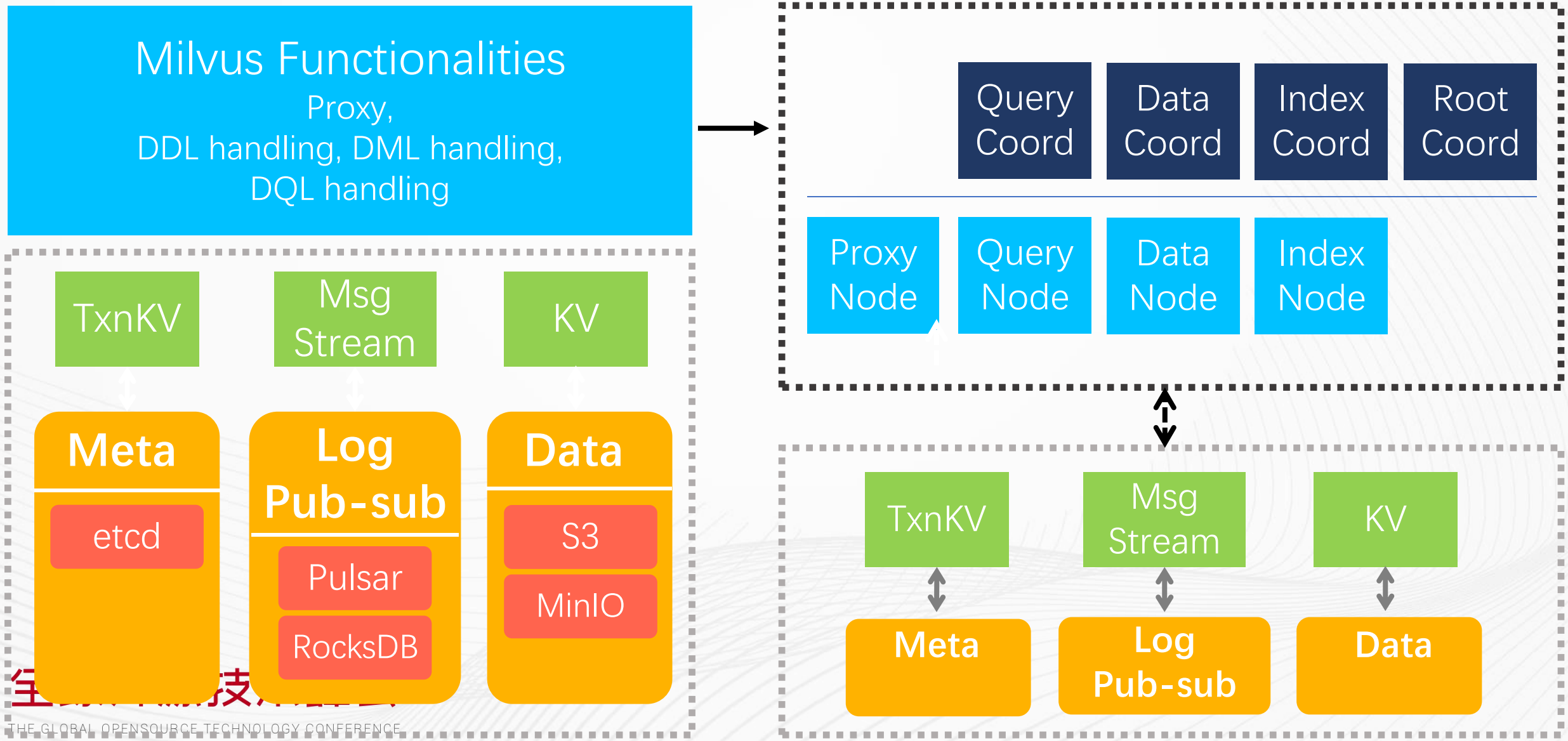
Log Subscribers

Storage

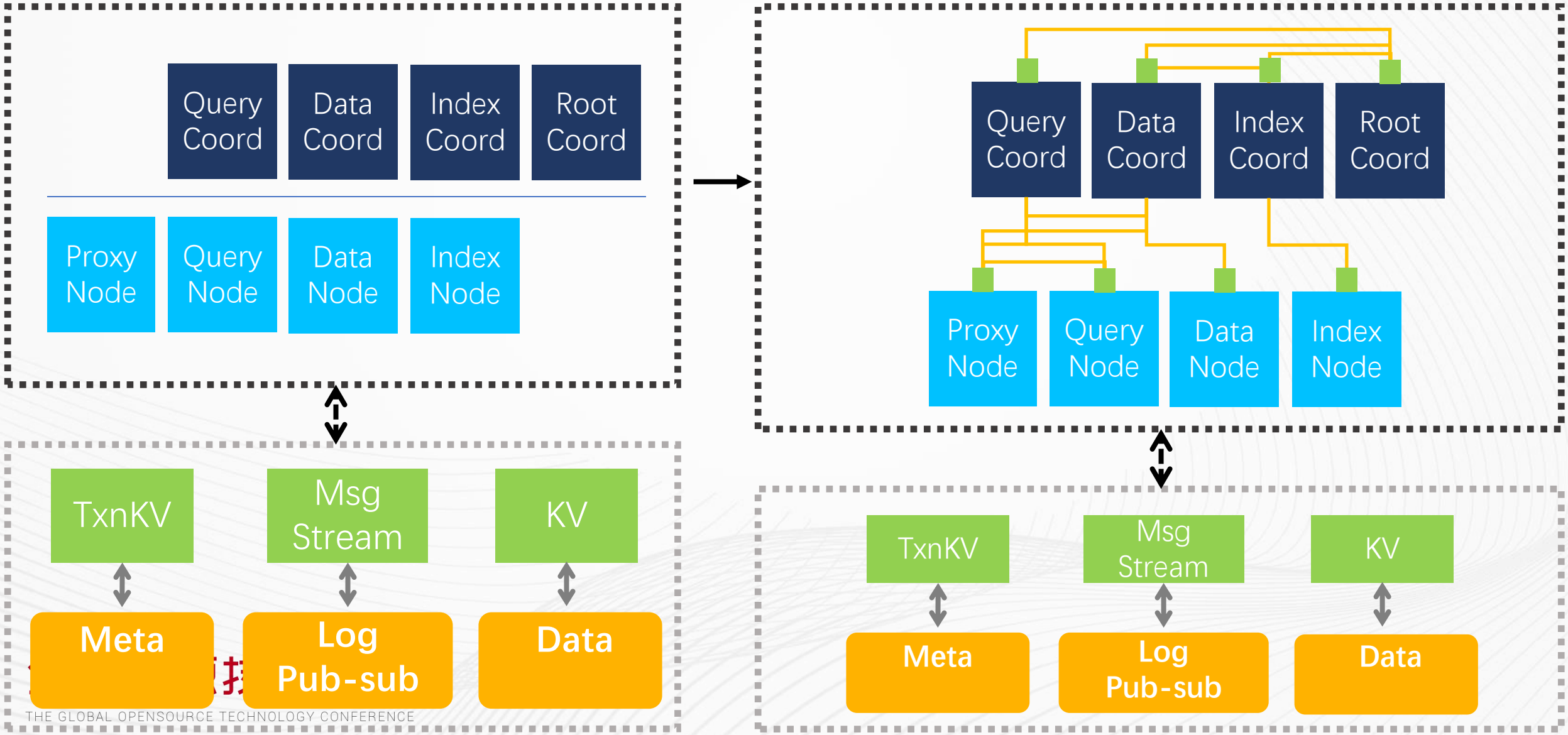
From Milvus 1.0 to Milvus 2.0



From Milvus 1.0 to Milvus 2.0

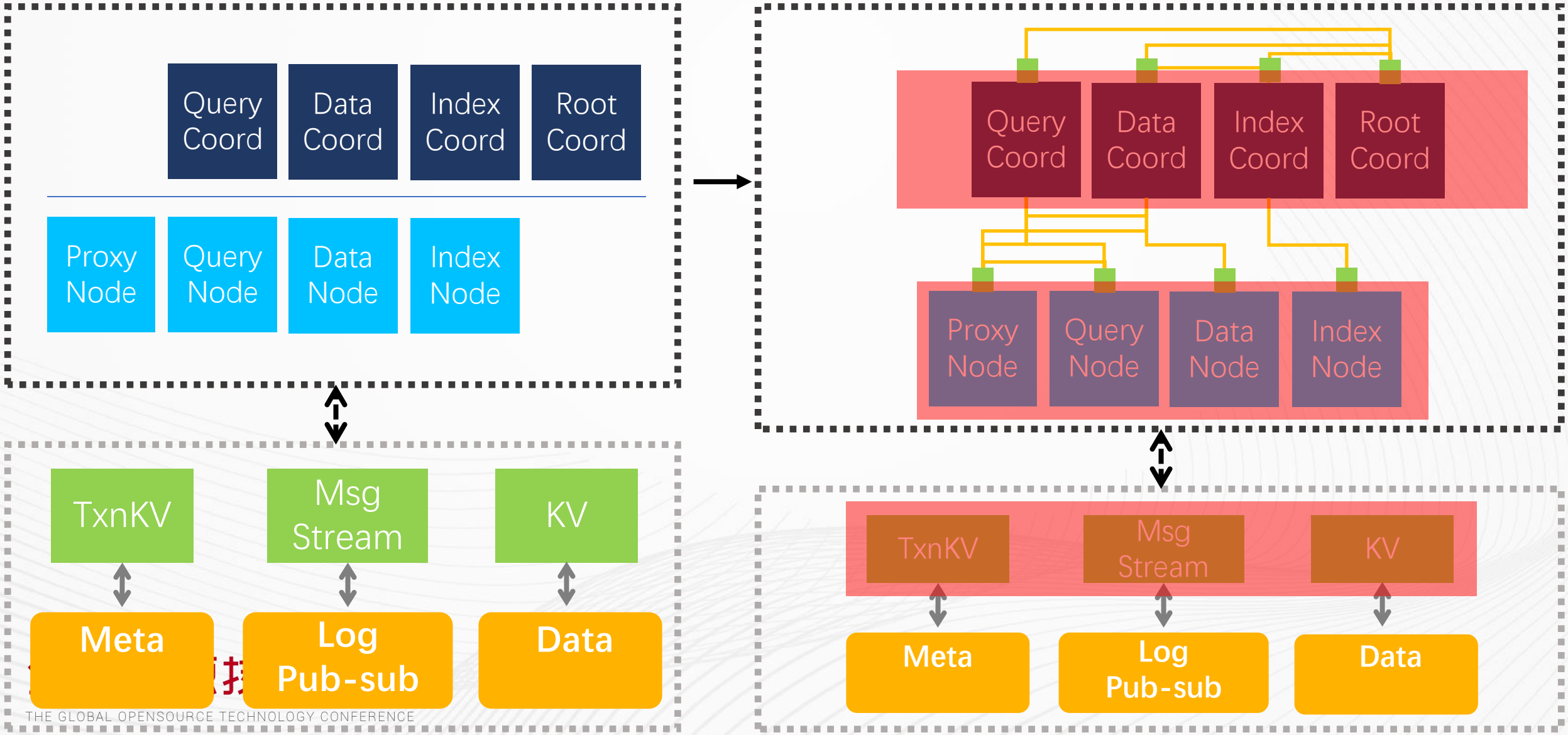


From Milvus 1.0 to Milvus 2.0

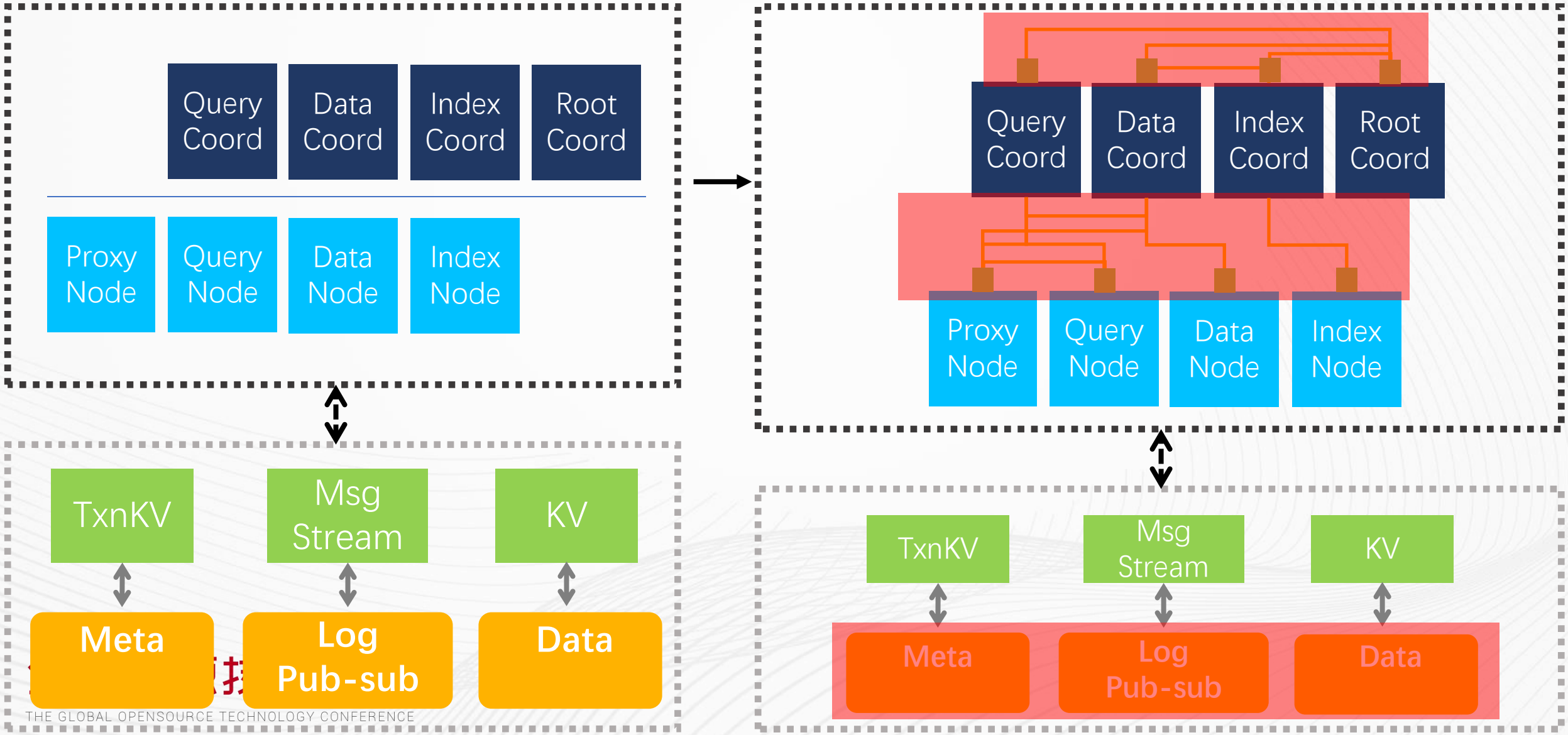


From Milvus 1.0 to Milvus 2.0

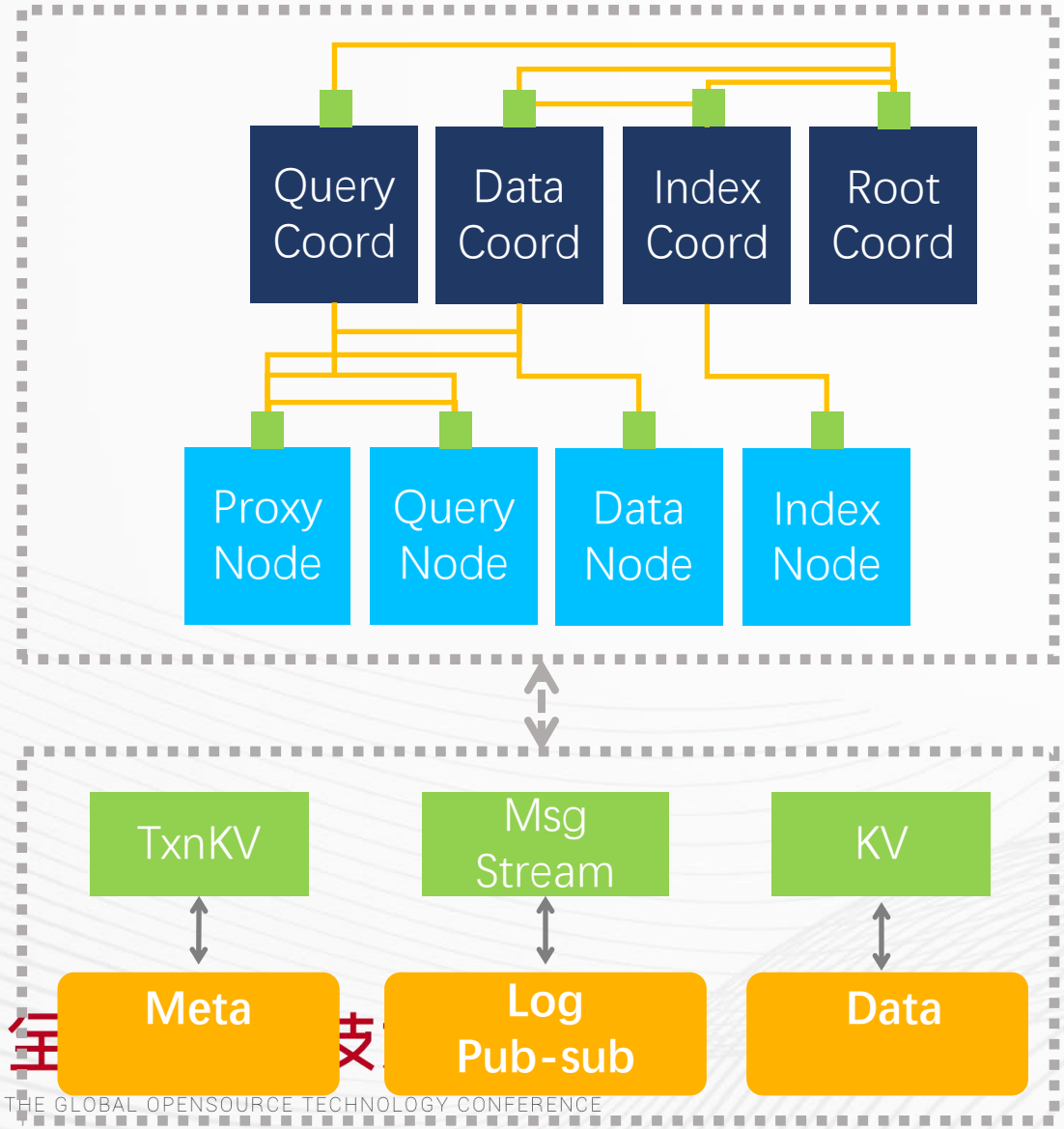
GOTC

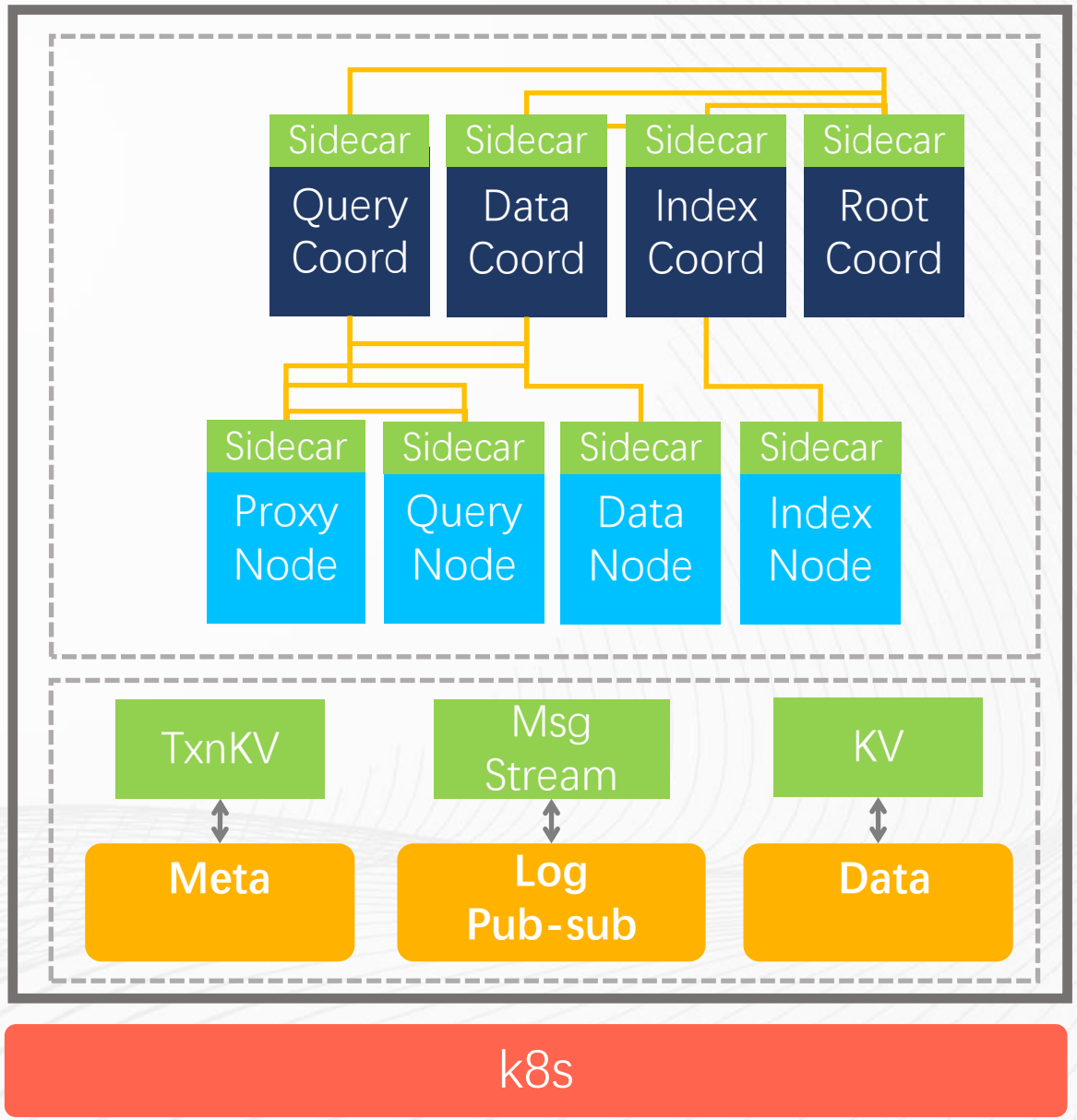


From Milvus 1.0 to Milvus 2.0



From Milvus 1.0 to Milvus 2.0





GOTC

THANKS

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE