Open Source AceCon 2021 智能云边开源峰会
AI x Cloud Native x Edge Computing
人工智能 × 云原生 × 边缘计算

Milvus: 探索云原生的向量数据库

郭人通
Partner at Zilliz

About Me

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

CONTENTS

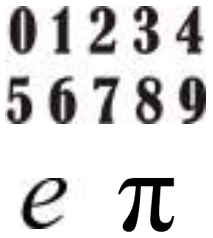
01

What is Milvus?

Milvus is an open source vector database for Unstructured Data Search and Analytics

Unstructured Data Trend

Int, float, string, …



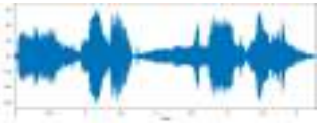
ABCDEFG

2021.04.10

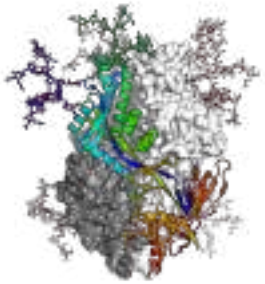
text

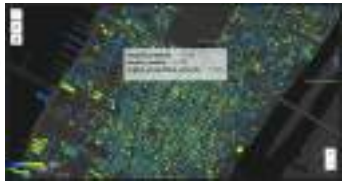
json

image
video
audio



domain specific

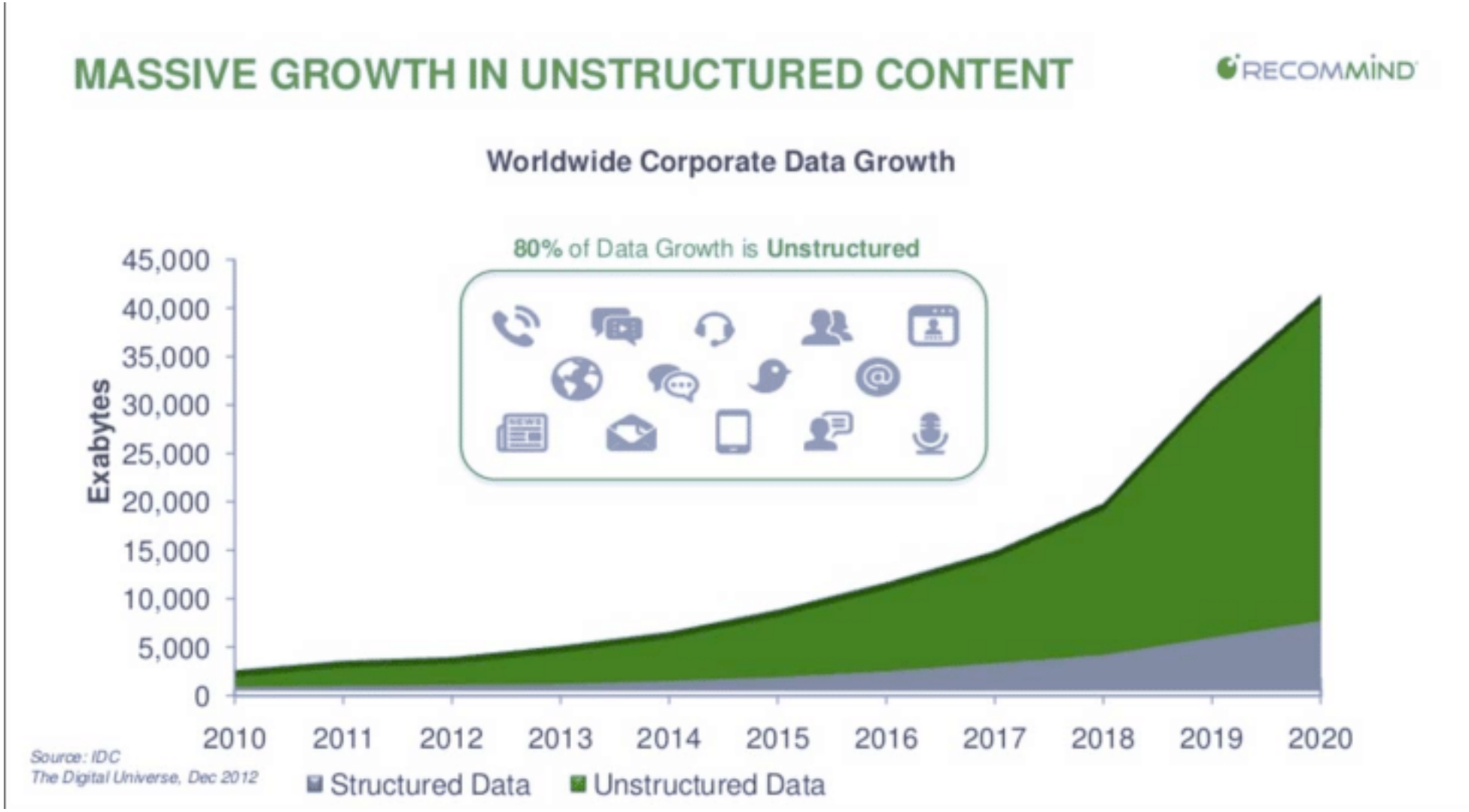




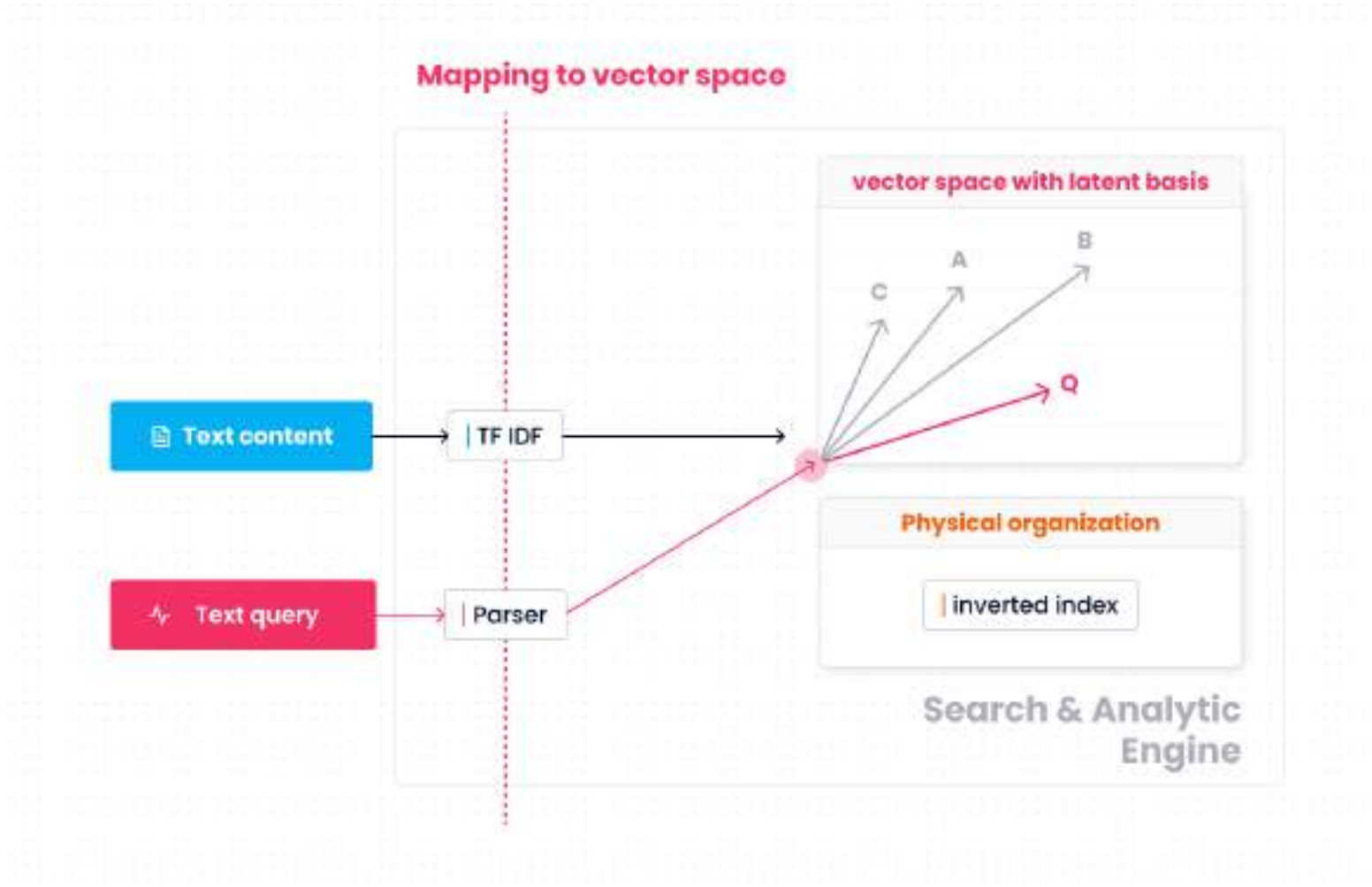
Structured data

Unstructured data

Unstructured Data Growth

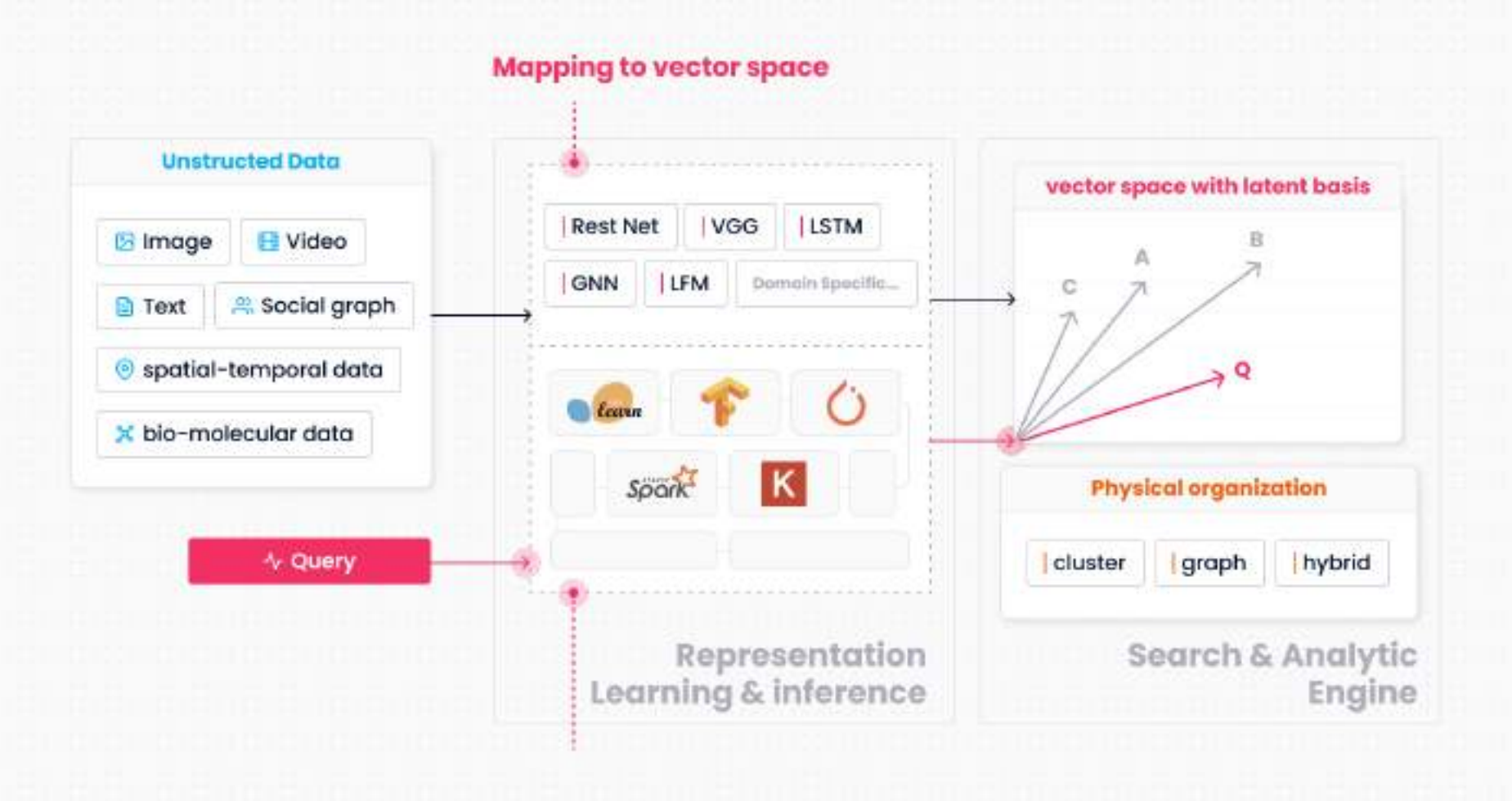


AI-driven Data Search and Analytics



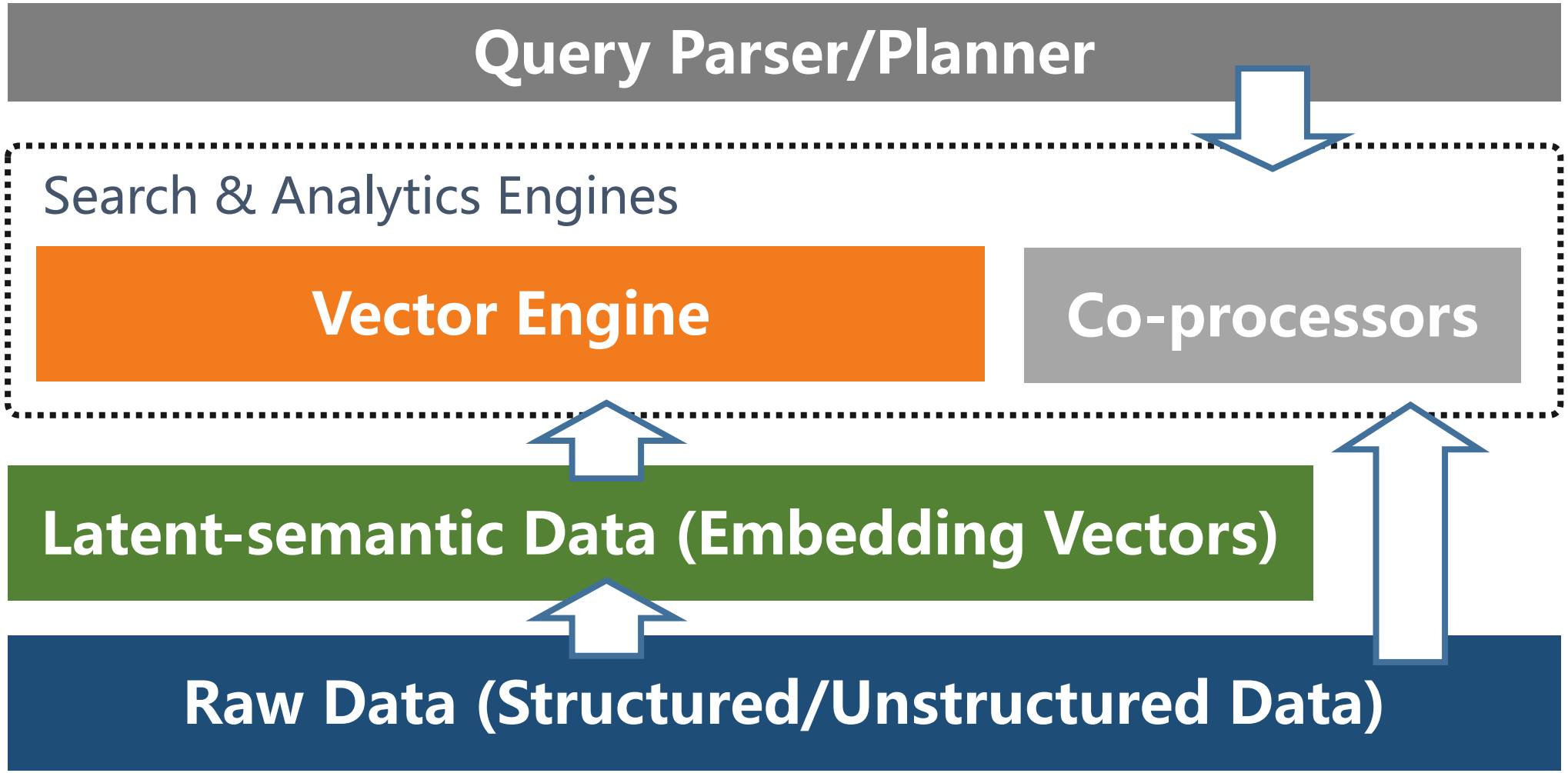
ElasticSearch Perspective

AI-driven Data Search and Analytics



Milvus Perspective

The Big Picture



Milvus is an LF AI & Data Graduated Project

LF AI & Data Foundation Interactive Landscape

LF AI & DATA

The LF AI & Data Foundation landscape (png, pdf) is dynamically generated below. It is modeled after the CNCF landscape and based on the same open source code. Please open a pull request to correct any issues. Greyed logos are not open source. Last Updated: 2021-09-07 23:04:31Z

You are viewing 7 cards with a total of 42,422 stars.

Landscape | Card Mode | LF AI & Data Members | Companies Hosting Projects

Tweet 151

Graduated LF AI & Data Projects (7)

Acumos ★17 LF AI & Data Foundation	Angel-ML ★6,279 LF AI & Data Foundation	Egeria ★565 LF AI & Data Foundation	Horovod ★11,333 LF AI & Data Foundation	Milvus ★6,445 LF AI & Data Foundation
ONNX ★10,808 LF AI & Data Foundation	Pyro ★6,975 LF AI & Data Foundation			

Crunchbase data is used under license from Crunchbase to CNCF. For more information, please see the license info.

02

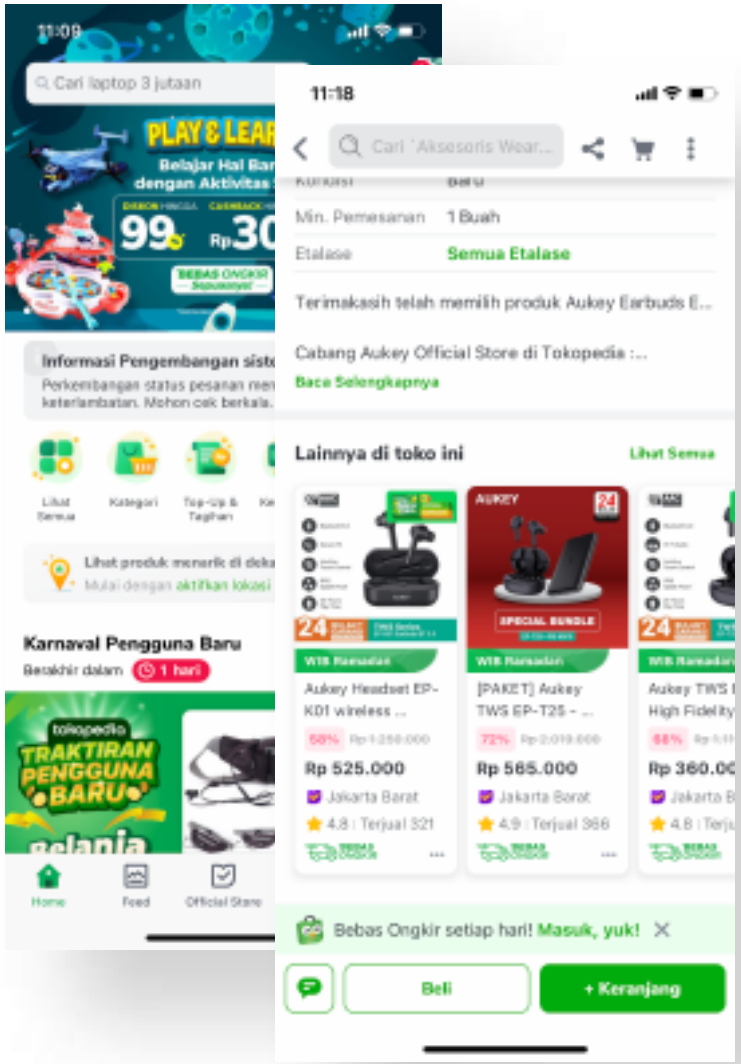
Real-World Cases

Users

1000+ Enterprise users around the global

dailyhunt
AXA
Lucidworks
TREND MICRO
moj
OPERA software
THE CLEVELAND MUSEUM OF ART
MessageBird
LINE
PrivatBank
identy
COMPASS
SmartNews
tokopedia
FINNOS
WISE BITS
Neoway
acesso digital
快手
mi
鹏华基金 PENGHUA FUND
唯品会 品牌特卖
丁香园 WWW.DXY.CN
Haier
Hisense
新浪微博 weibo.com
Alibaba Group
中国建设银行 China Construction Bank
招商银行 信用卡
爱建信托 AJ TRUST
iQIYI 爱奇艺
中国太平 CHINA TAIPING
泰康人寿 Taikang Life
永辉超市
爱云动
阳光 阳光出行
平安银行 PING AN BANK
喜马拉雅
蜻蜓FM
科大讯飞 IFLYTEK
平安科技 PING AN TECHNOLOGY
中国电信 CHINA TELECOM
用友 yonyou
什么值得买
企查查 qichacha.com
交通银行 BANK OF COMMUNICATIONS
TRANSSION
WPS

Related product search and recommendations



ASCII representation

	b	r	e	a	d
bread	098	114	101	097	100
	t	o	a	s	t
toast	116	111	097	115	116

Elasticsearch for keyword search with ASCII codes.

The code we know about these two arrays of numbers is that bread not equal to toast.

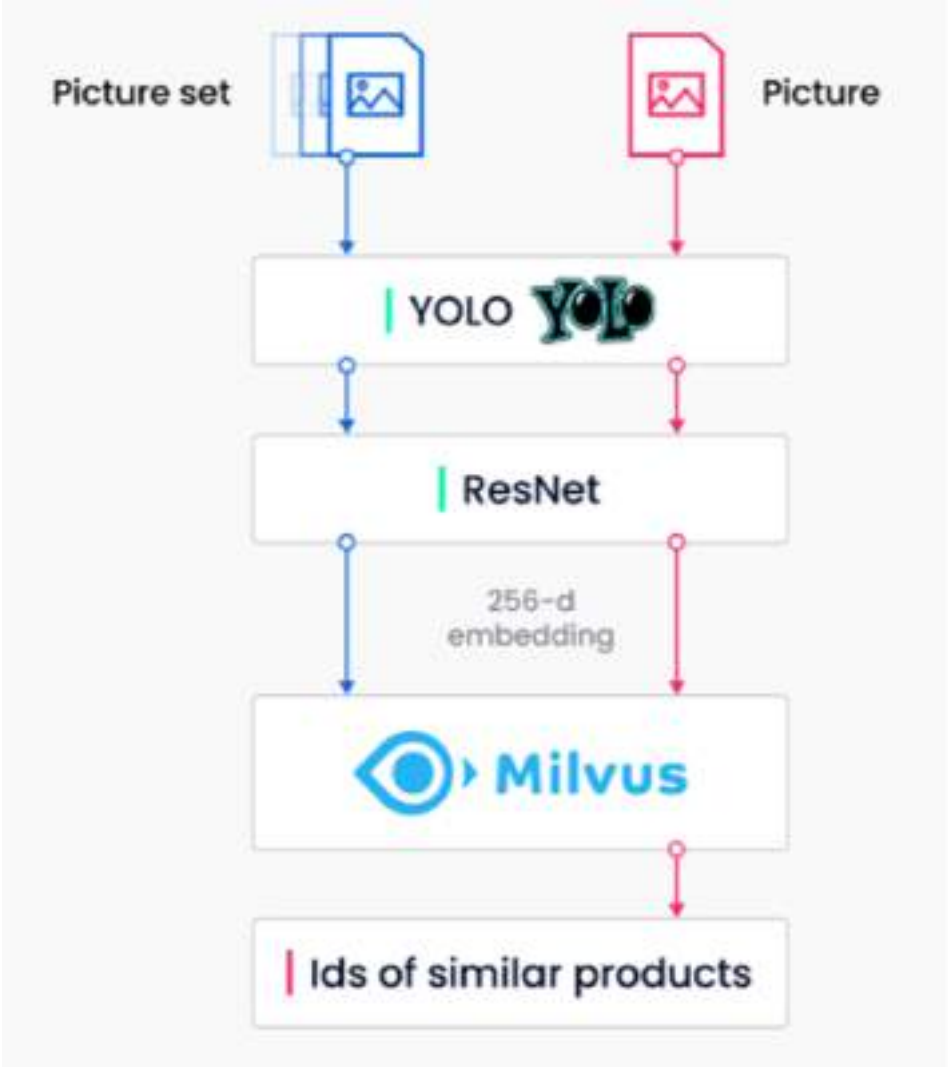
We assume that similar contexts represent similar things, and try to compare them using mathematical methods. We could even find a way to encode whole sentences by their meaning.

Vector representation

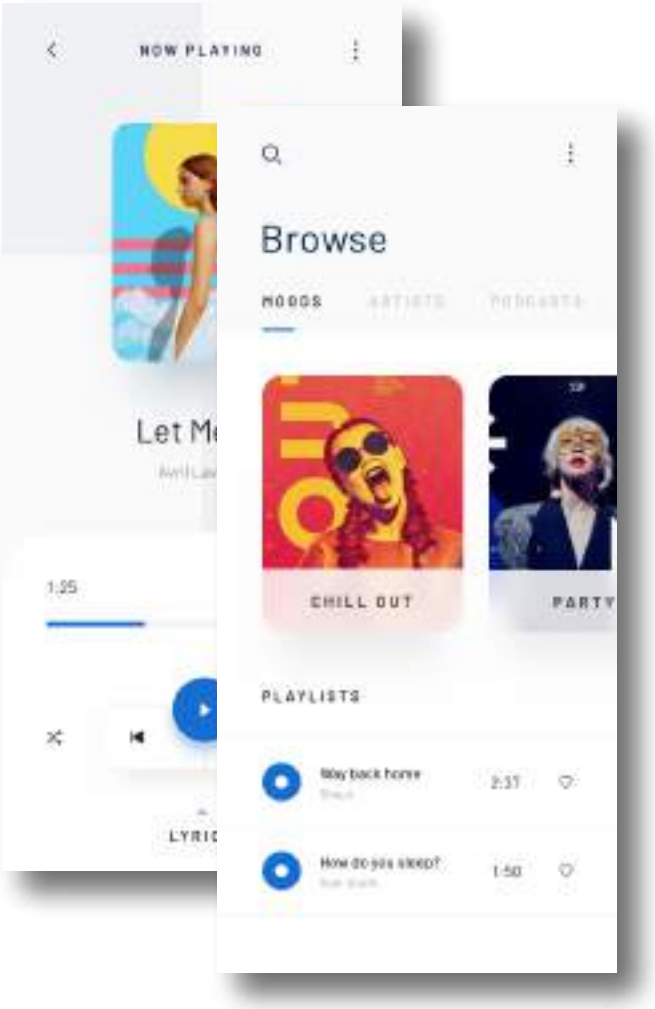
	butter	car	sun	brekfast	cat
bread	0.80	0.11	0.05	0.93	0.20 ...
toast	0.76	0.22	0.15	0.95	0.12 ...

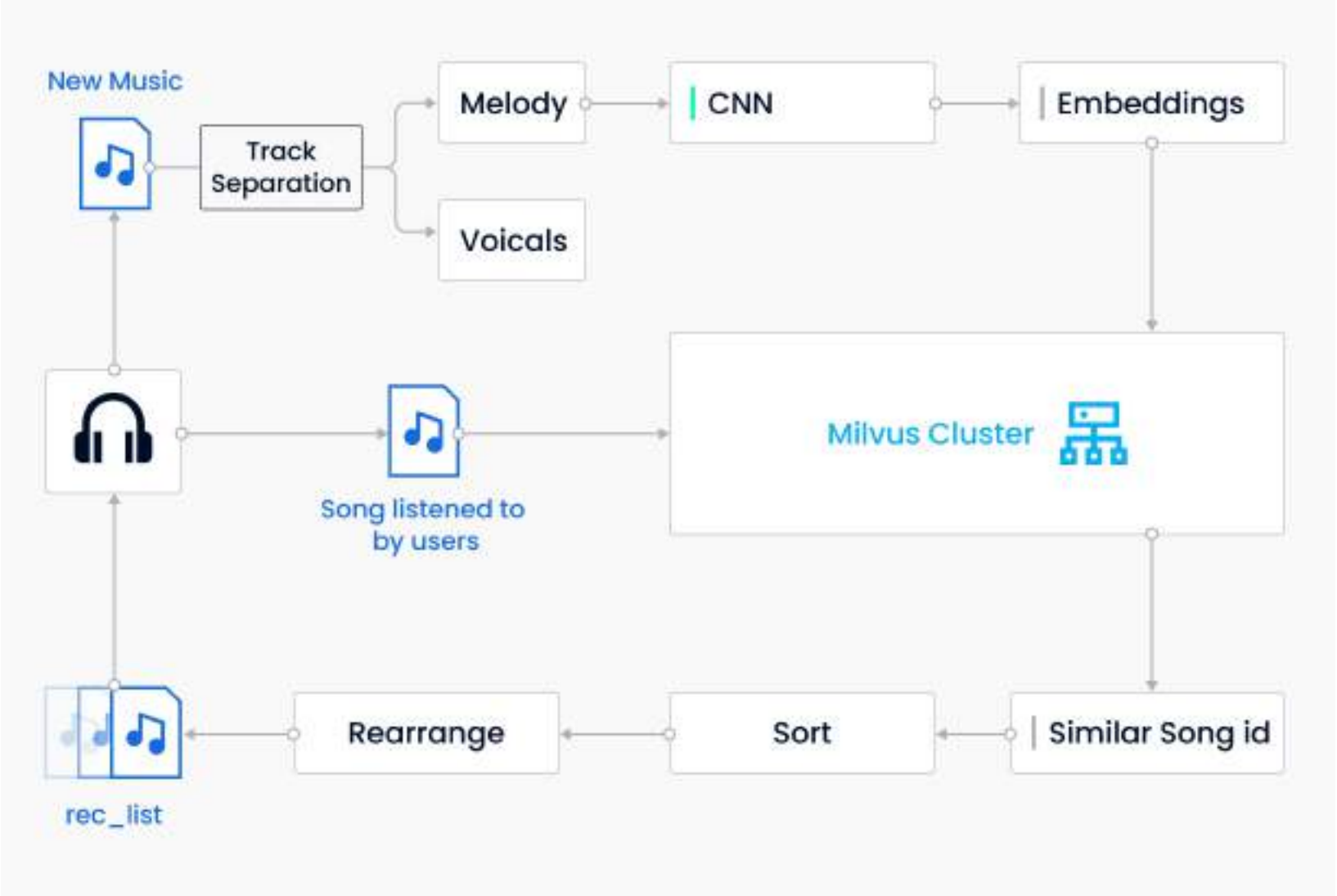
The word bread is often used together with words butter, breakfast, and rarely with words car, sun and cat. So, is the toast.

Reverse Image Search

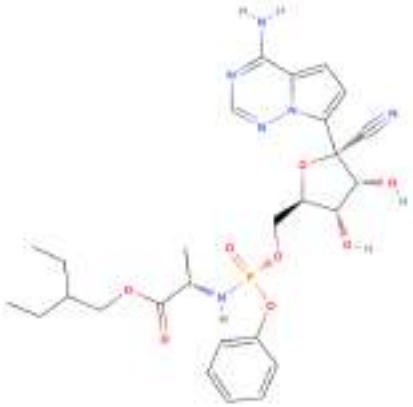


Music Recommendations



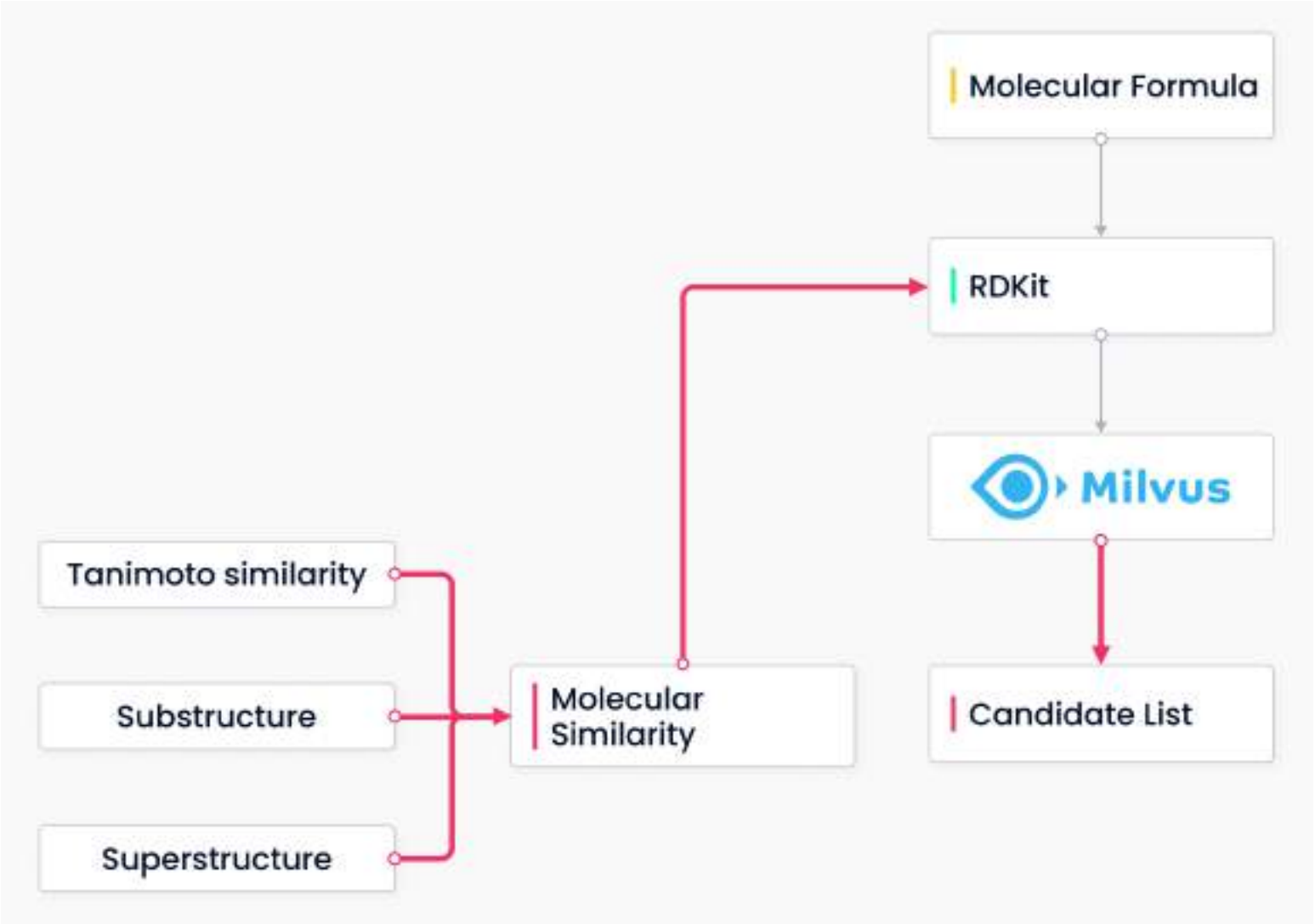


Pharmaceutical Molecular Analysis



CC(=O)Nc1ccc(S(=O)(=O)NCC(=O)N2 CCS(=O)CC2)cc1

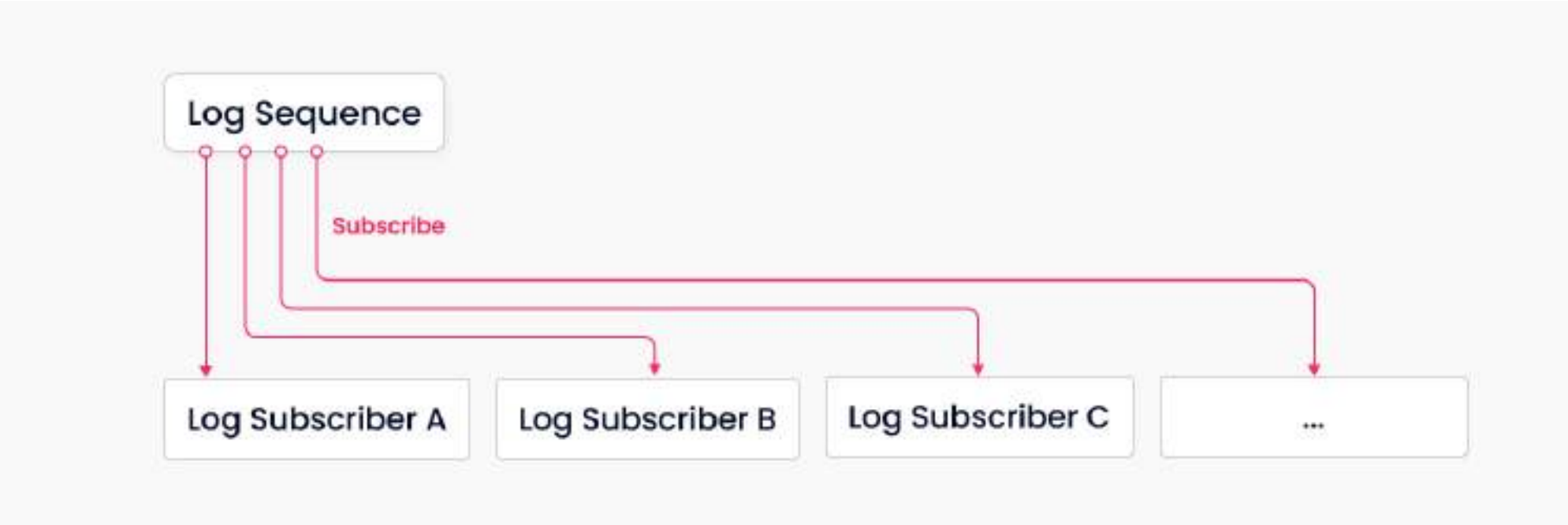
Molecular fingerprint: 1024 bits
00001100...10000000



03

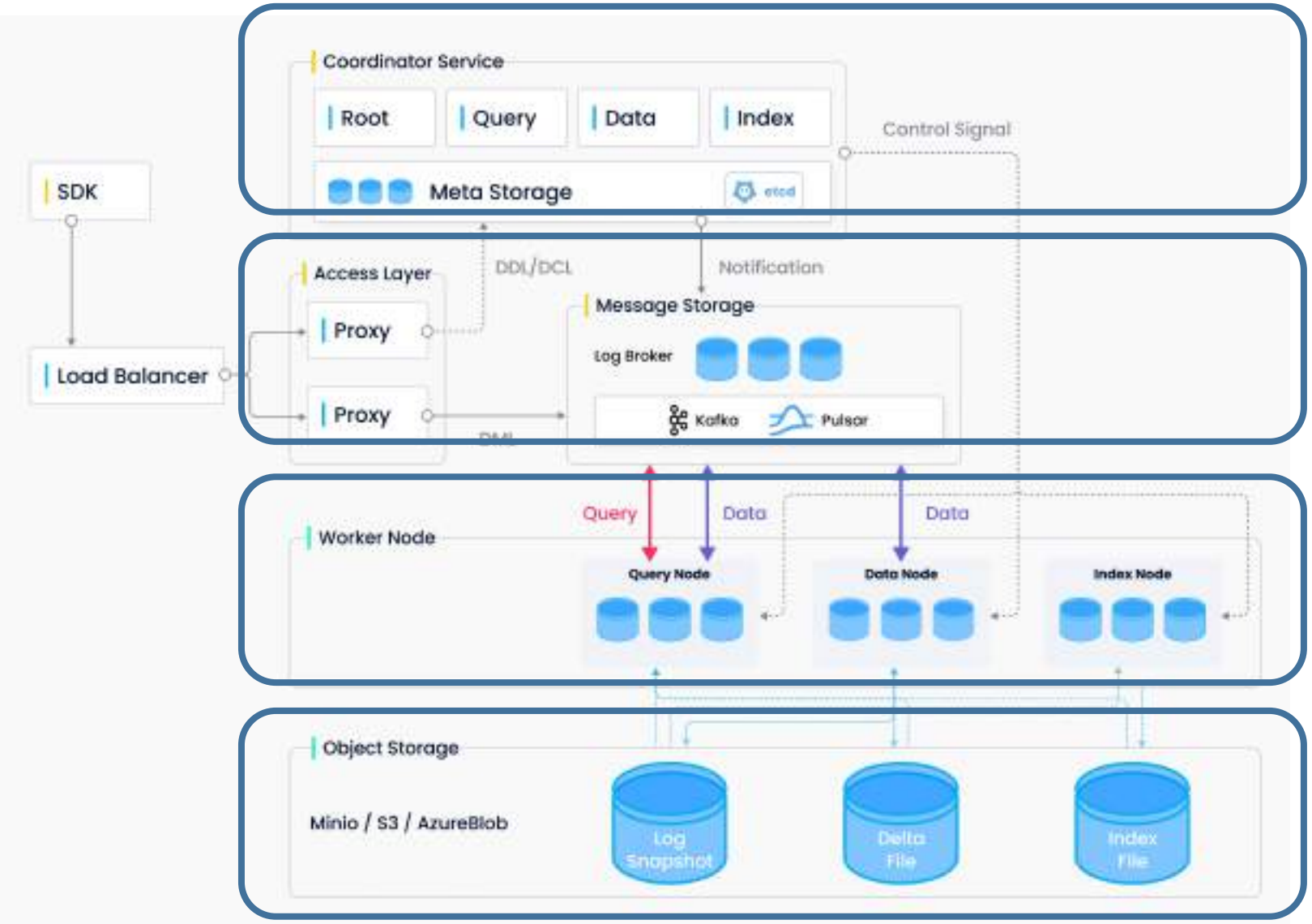
Milvus2.0 Architecture Overview

Log Sequence Pub-sub as System Backbone



- Disaggregate Log and database, make failure recovery easy and fast
- Guarantee data durability
- Make System extendable
- Reduce system complexity

Architecture



Coordinators

Log Broker

Log Subscribers

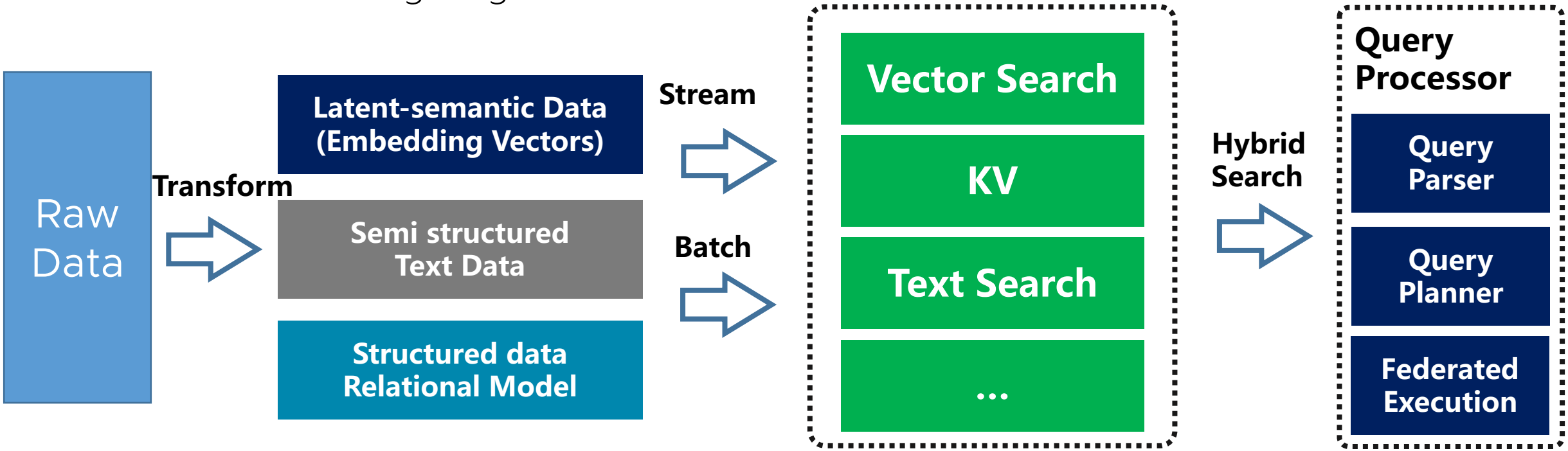
Storage

The Bazaar Architecture

How to further extend the system?

Solution

The 'bazaar' architecture. Loosely coupling multiple execution engines with different functionalities through logs.

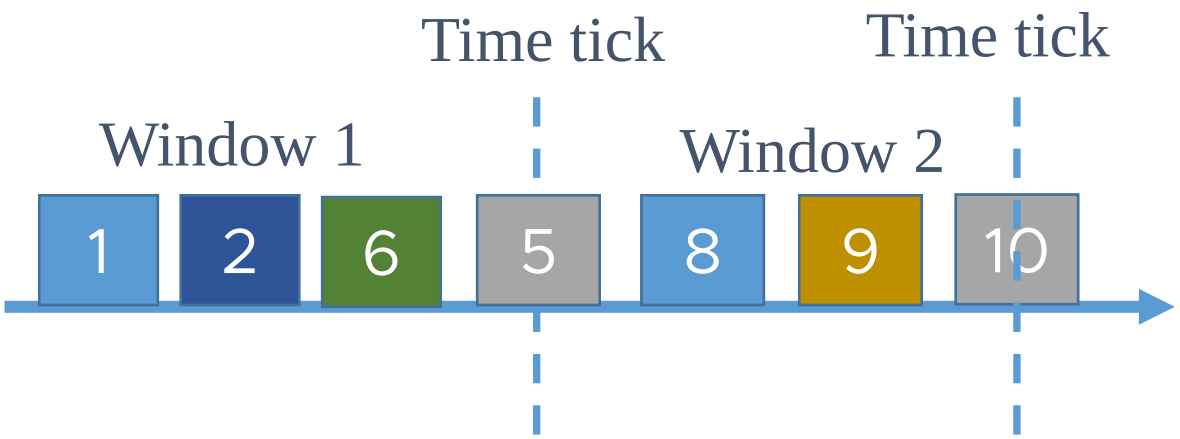


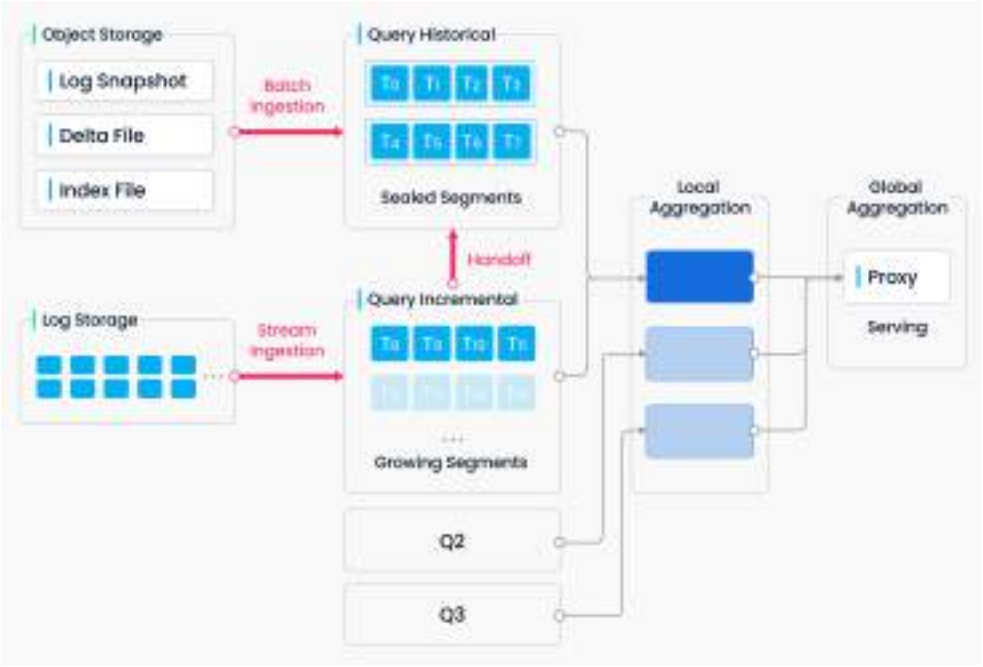
Combine Streaming and Batching

Relying only on log stream for reads is not practical (too slow)

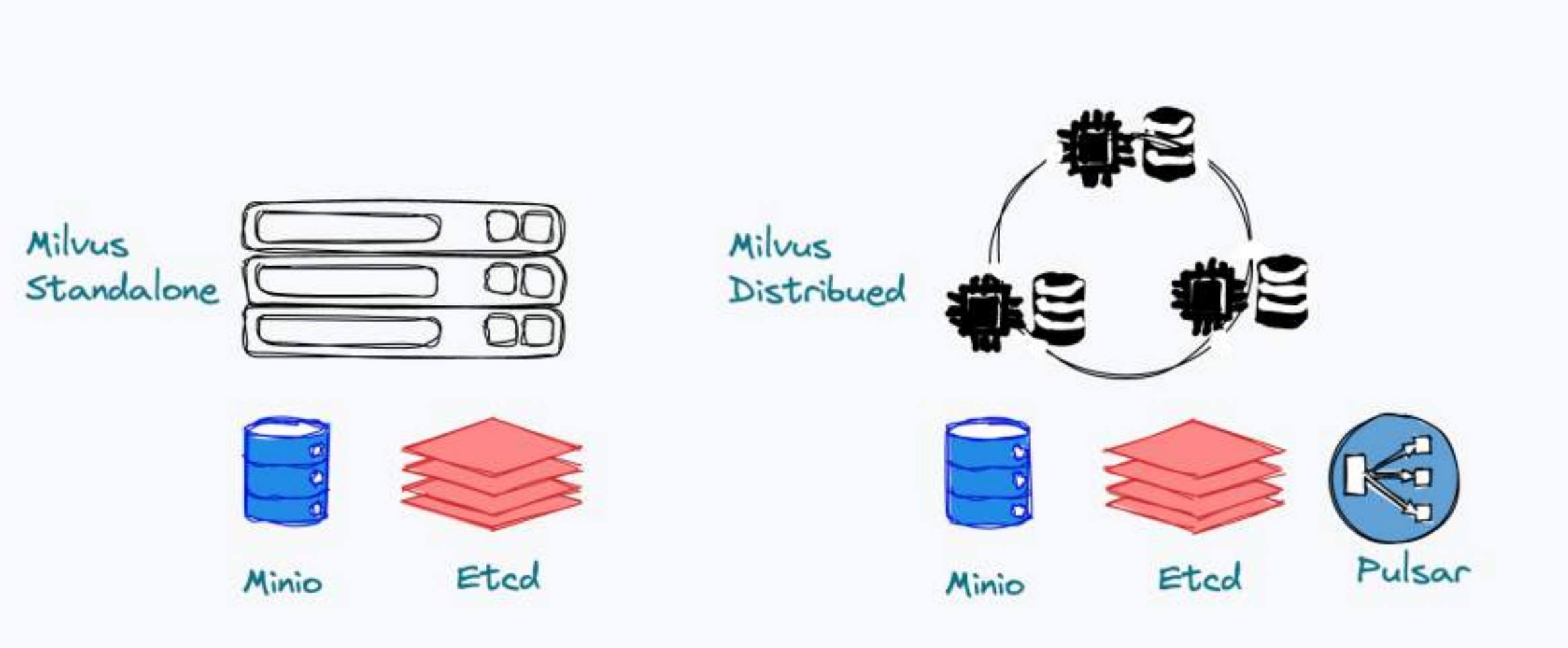
Solution

Periodically backfill history data to segments, just like flush in LSM tree, and handoff growing segments to historical. Merge incremental and historical on Read to maintain data completeness





Milvus Deployment



04

RoadMap

History of Milvus

2018.10
The Idea

2019.04
Milvus 0.1 Release

2019.06
1st Seed User

2019.10
Open Source

2020.03
Joined LF&AI

2021.03
Milvus 1.0 Release

2021.06
Milvus 2.0.0rc1 Release

Milvus 2.0 GA will coming in Oct. 2021

Support String Data Type

Urgent needs from users

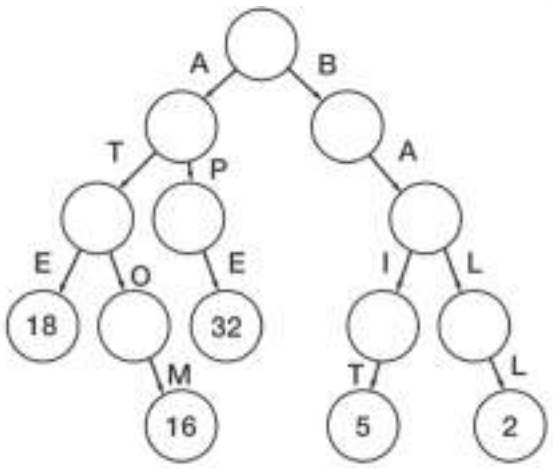
Scalar filtering on string field

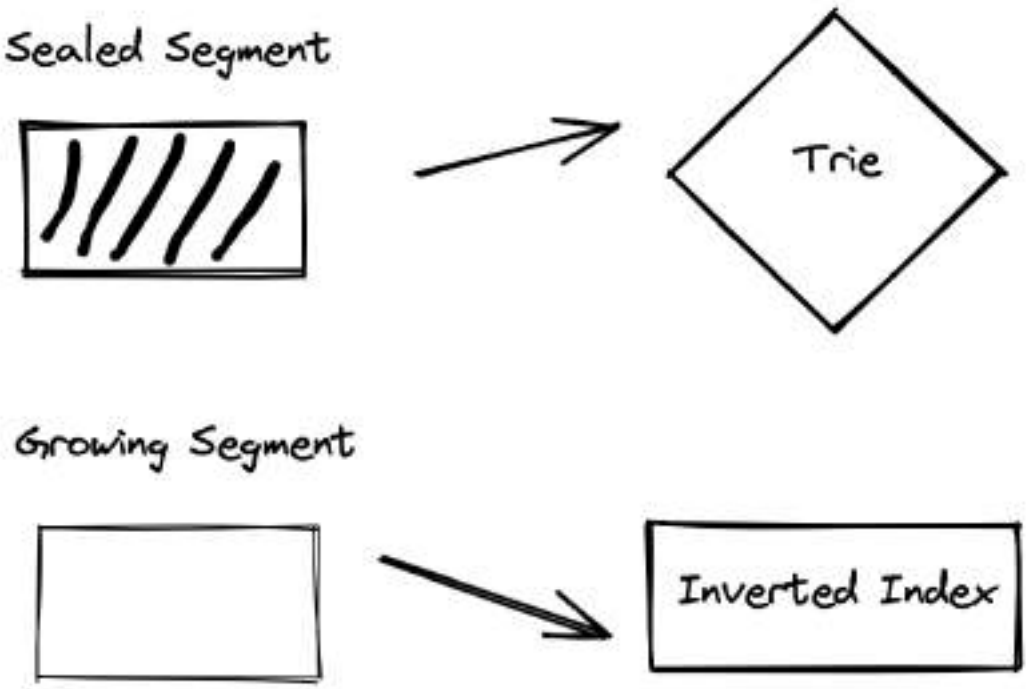
Retrieve origin string

Memory consumption matters

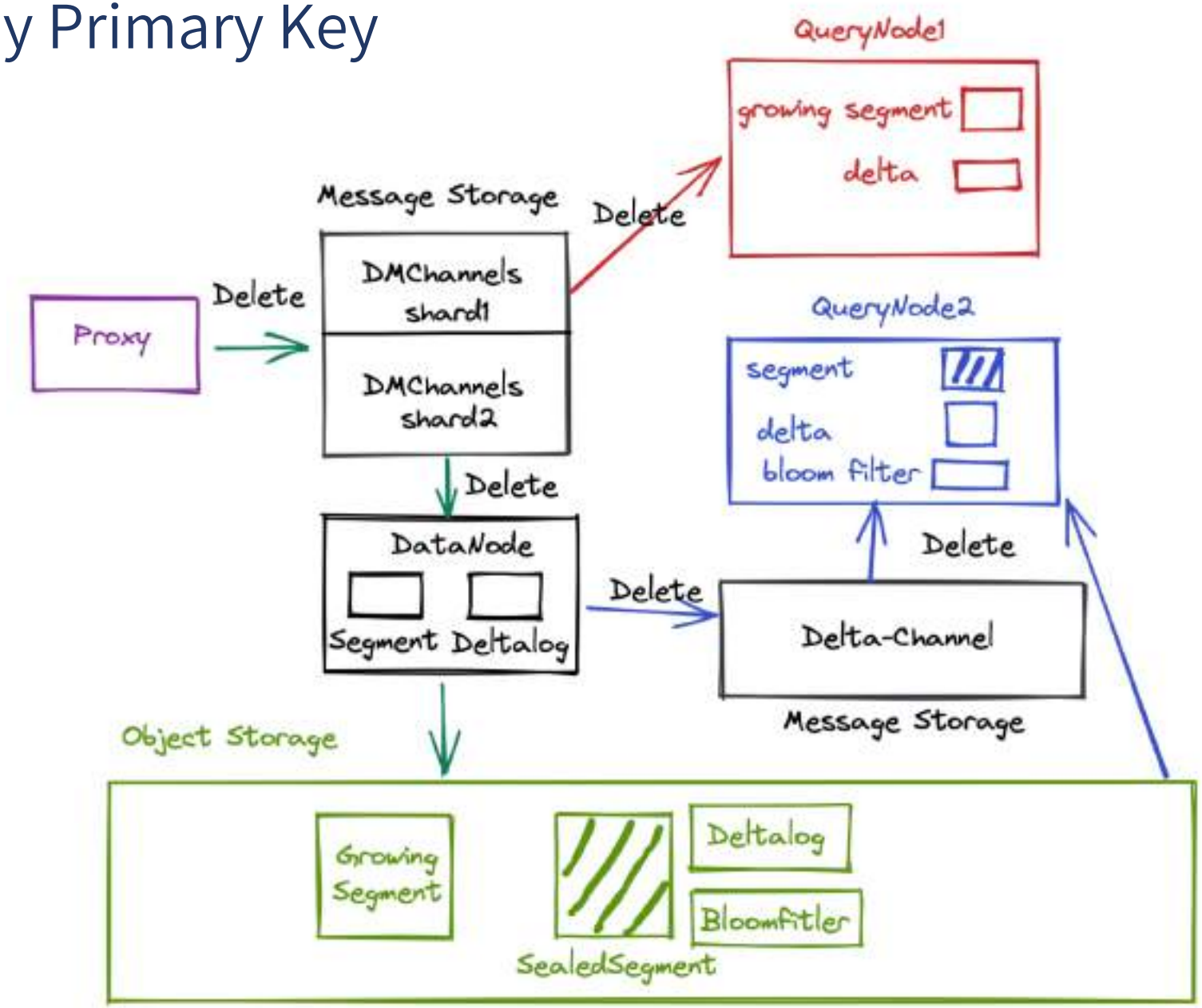
3 million unicode russian words

DataStructure	MemoryUsage
Python-Dict	600MB
Python-List	300MB
PAT-Trie	242 MB
HAT-Trie	125 MB
DA-Trie	101 MB
Marisa-Trie	11 MB





Support Delete by Primary Key

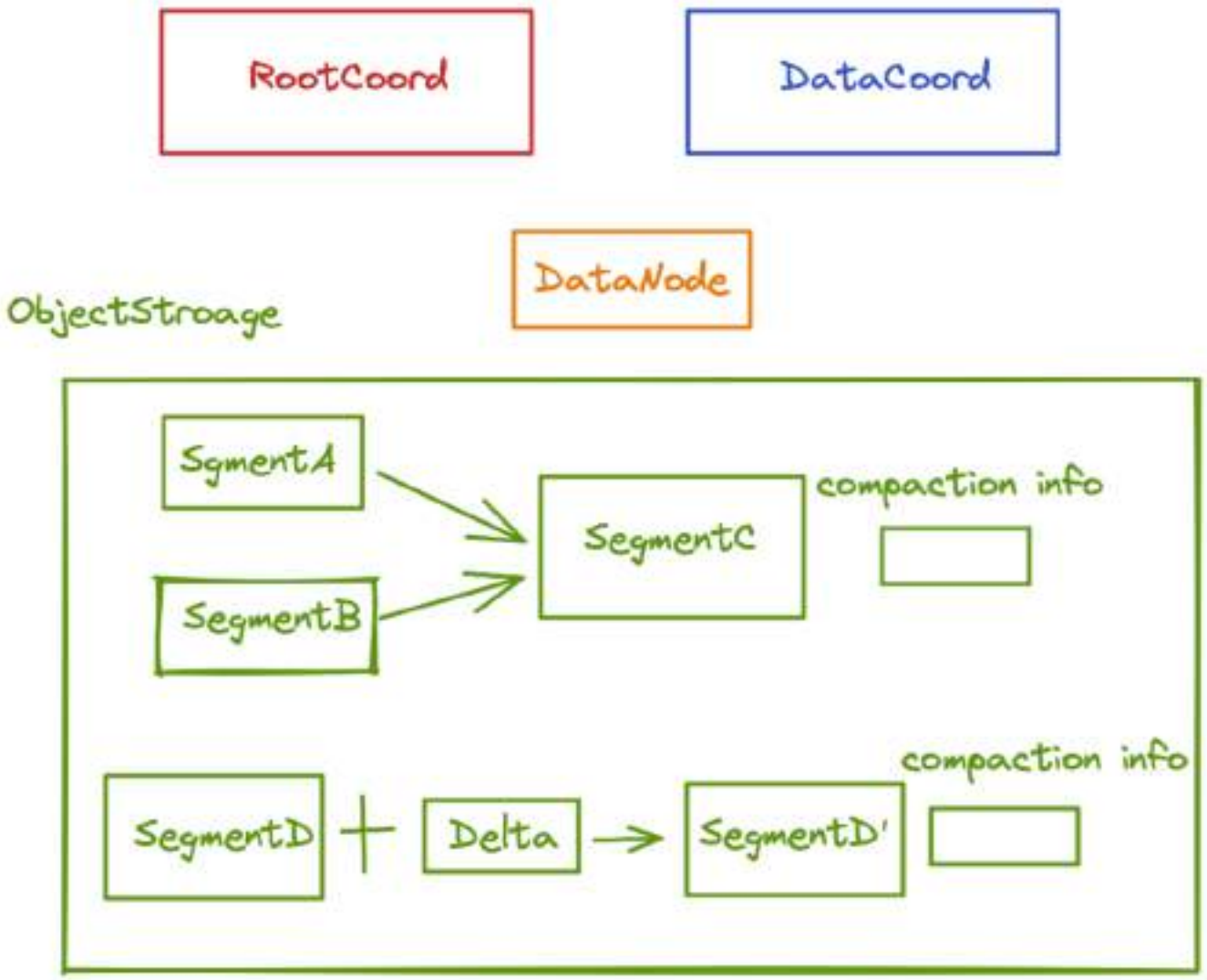


Segment Compaction

Segment size varies

Delete operation make segment and index sparse

Indexes of large segments are more efficient



Search/Query with Expression

Search

A set of criteria that results in a relevancy-ordered list that match the query.

Query

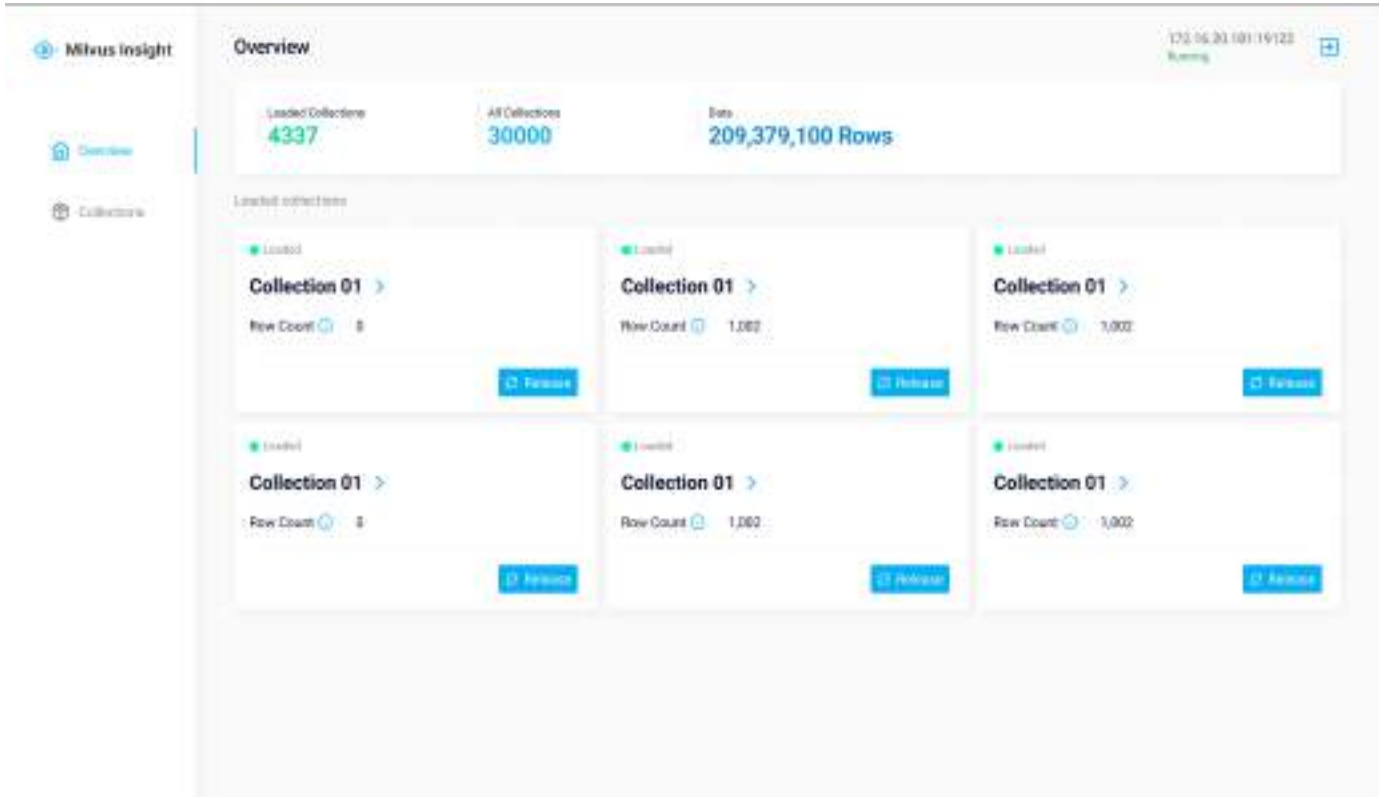
A set of criteria that results in a list of records that match the query exactly, returned in order of particular field values

Operator	Description	Examples
Relational operators	Relational operators use symbols to check for equality, inequality, or relative order between two expressions. Relational operators include `>`, `>=`, `<`, `<=`, `==`, and `!=`.	• A > 1 • B >= 2 • C < 3 • D <= 4 • E == 5 • F != 6
Logical operators	Logical operators perform a comparison between two expressions. The supported logical operators are: AND, && OR,	, and NOT.
IN operator	The IN condition is satisfied when the expression to the left of the keyword IN is included in the list of items.	• FloatCol in [1.0, 2, 3.0] • Int64Col in [1, 2, 3]

Milvus combines scalar and vector search, such as "Find top 10 drama films similar to Forrest Gump"

- Support scalar datatypes columnar storage
- Filter scalar data by arithmetic and bool expressions
- Retrieve field data on query/search

Milvus Insight



- Cluster state visualization
- Meta Management
- Data Query
- Health Diagnosis
- Open source, Please join us

https://github.com/milvus-io/milvus-insight

Milvus CLI

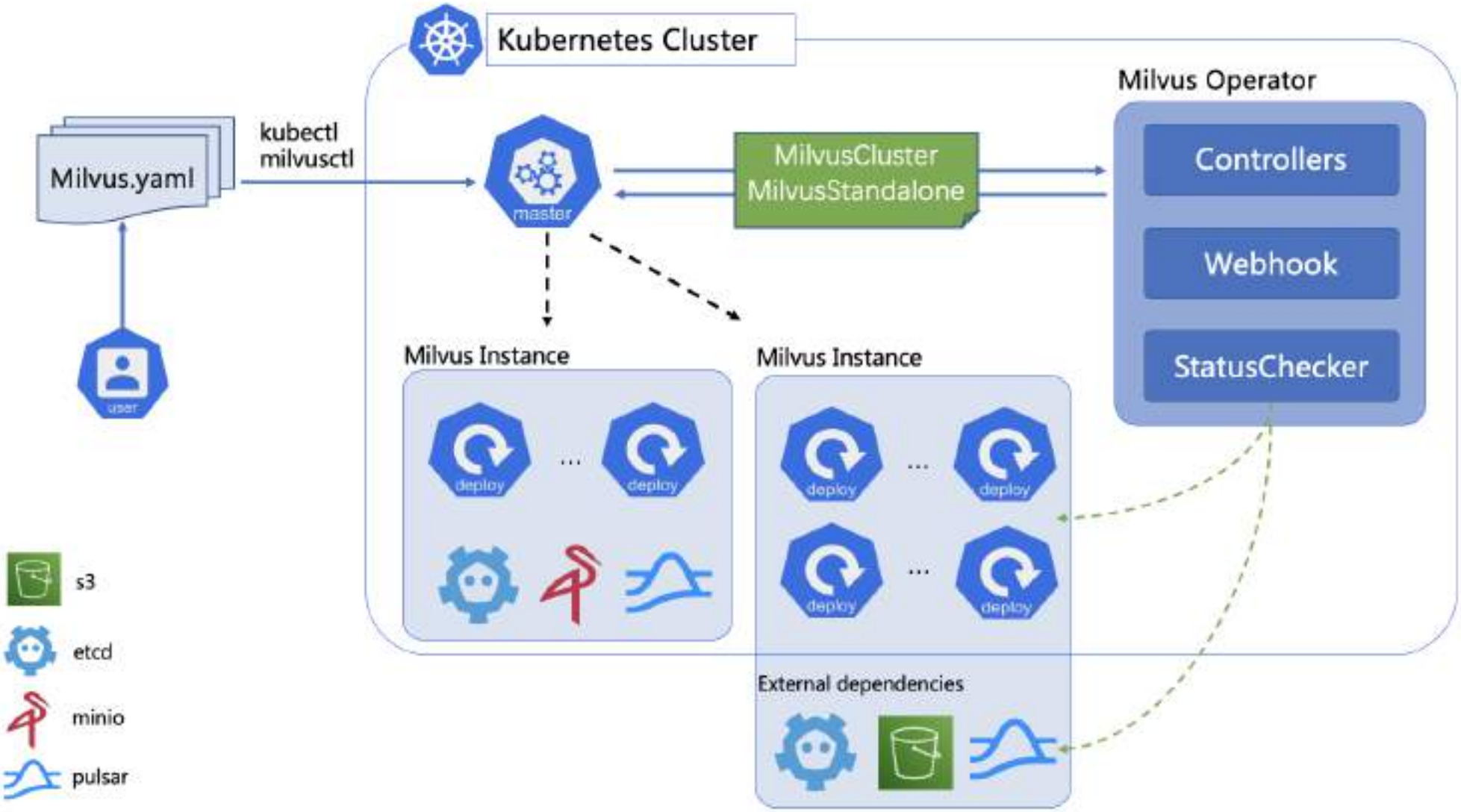

```
milvus_cli > help
Usage:  [OPTIONS] COMMAND [ARGS]...

  Milvus CLI

Commands:
  clear     Clear screen.
  connect   Connect to Milvus.
  create    Create collection, partition and index.
  delete    Delete specified collection, partition and index.
  describe  Describe collection or partition.
  exit      Exit the CLI.
  help      Show help messages.
  import    Import data.
  list      List collections, partitions and indexes.
  load      Load specified collection.
  query     Query with a set of criteria, and results in a list of...
  release   Release specified collection.
  search    Conducts a vector similarity search with an optional boolean...
  show      Show connection, loading_progress and index_progress.
  version   Get Milvus CLI version.
```

https://github.com/milvus-io/milvus_cli

K8s Operator



Open Source AceCon

2021 智能云边开源峰会

AI x Cloud Native x Edge Computing

人工智能 × 云原生 × 边缘计算

Thank You