什么是大数据

大数据产生的背景：计算广告

序号	公司名	Q2 广告营收	Q2 同比变化	Q2 广告营收占比	2021上半年 广告营收	2020上半年 广告营收	上半年 同比变化
1	阿里巴巴	810.02	13.74%	39.37%	1446	823.4	75.61%
2	腾讯	228.33	23.08%	16.51%	446.53	362.65	23.13%
3	百度	208.28	17.75%	66.44%	389.22	319.31	21.89%
4	京东	189.85	35.10%	7.48%	331.05	235.8	40.39%
5	拼多多	180.80	63.55%	78.45%	321.92	165.47	94.55%
6	快手	99.62	156.15%	52.05%	185.20	71.62	158.58%
7	美团	72.16	66.91%	16.49%	128.81	71.87	79.23%
8	小米	45	46.20%	5.13%	84	86	-2.33%
9	微博	32.65	47.48%	87.51%	58	40.17	44.39%
10	爱奇艺	18.25	15.07%	23.99%	37.42	31.23	19.82%
11	三六零	/	/	/	30.59	32.99	-7.28%
12	唯品会	13.81	53.79%	4.66%	25.58	17.27	48.12%
13	哔哩哔哩	10.49	200.57%	23.34%	17.64	5.63	213.32%
14	搜狗	8.91	-43.15%	92.91%	17.23	33.87	-49.13%
15	虎牙	3.83	189.86%	12.93%	5.96	2.69	121.56%
16	知乎	2.48	48.42%	38.87%	4.62	2.93	57.68%

大部分2C的互联网公司都是以广告的方式盈利

互联网广告的优势：可计算，准确度高。
传统广告只能进行抽样调查，统计数据存在误差。

转化漏斗

大数据的特点

- 来源丰富(Variety)。可以是传统关系型数据库里的结构化数据，也可以是日志、视频、图片等各种半结构化或非结构化数据。
- 体量庞大(Volume)。从传统的GB上升至了TB、PB、EB级的数据量；
- 商业价值(Value)。采集用户的数据可以做各种分析，挖掘出蕴含的商业价值。在流量为王的时代里，BAT等各大互联网企业很大一部分收入都来自于广告。
- 处理时效性高(Velocity)。海量的数据不仅仅局限于离线分析，天猫双十一实时大屏上的几千亿的交易额都精确到秒。

要解决的问题

- 如何存储
- 如何计算
- 如何分配资源
- 如何索引查询

谷歌与开源生态

- 存储：GFS vs. HDFS
- 计算：MapReduce vs. Spark
- 索引：BigTable vs. HBase

https://static.googleusercontent.com/media/research.google.com/zh-CN//archive/gfs-sosp2003.pdf

https://storage.googleapis.com/pub-tools-public-publication-data/pdf/16cb30b4b92fd4989b8619a61752a2387c6dd474.pdf

https://static.googleusercontent.com/media/research.google.com/zh-CN//archive/bigtable-osdi06.pdf

Google系列论文

云产品

BigQuery | Cloud DataFlow | GAE | Cloud Bigtable | Cloud Spanner

OLAP/OLTP

计算

存储/数据库

PAIali-NPU | TPU 2017

TensorFlow | PAI | TensorFlow 2016

Drill(Parquet) Impala Presto | ODPS-AliORC | Dremel 2010

ODPS-R

Mahout | ODPS-Xlib

Spark | ODPS-Spark

Hama/Giraph | ODPS-Graph Geabase | Pregel 2010

Hive | ODPS-SQL | Tenzing 2011

ODPS-MR

ODPS-DAG

UDAP | Goods 2016

PowerDrill 2012 | ADS | Druid

MillWheel 2013 | Blink | Flink/Storm

FlumeJava 2010

Pig/Dryad | Sawzall 2003

Hadoop Yarn | Fuxi | MapReduce 2004

Mesos

ODPS-Lighting | Mesa 2014 | Kudu

Shasta 2016

F1 2013

OceanBase | TiDB | CockroachDB | TiKV

Percolator 2011

Spanner 2012

MegaStore 2011 | Lindorm

Bigtable 2006 | OTS | HBase

K8S | Sigma | Borg 2015 | Omega 2013

GFS 2003 | Pangu | HDFS

Chubby 2006 | Nuwa | Zookeeper

CPI2 2013 | GWP 2010

The Datacenter as a Computer 2009

GoogleCluster 2003

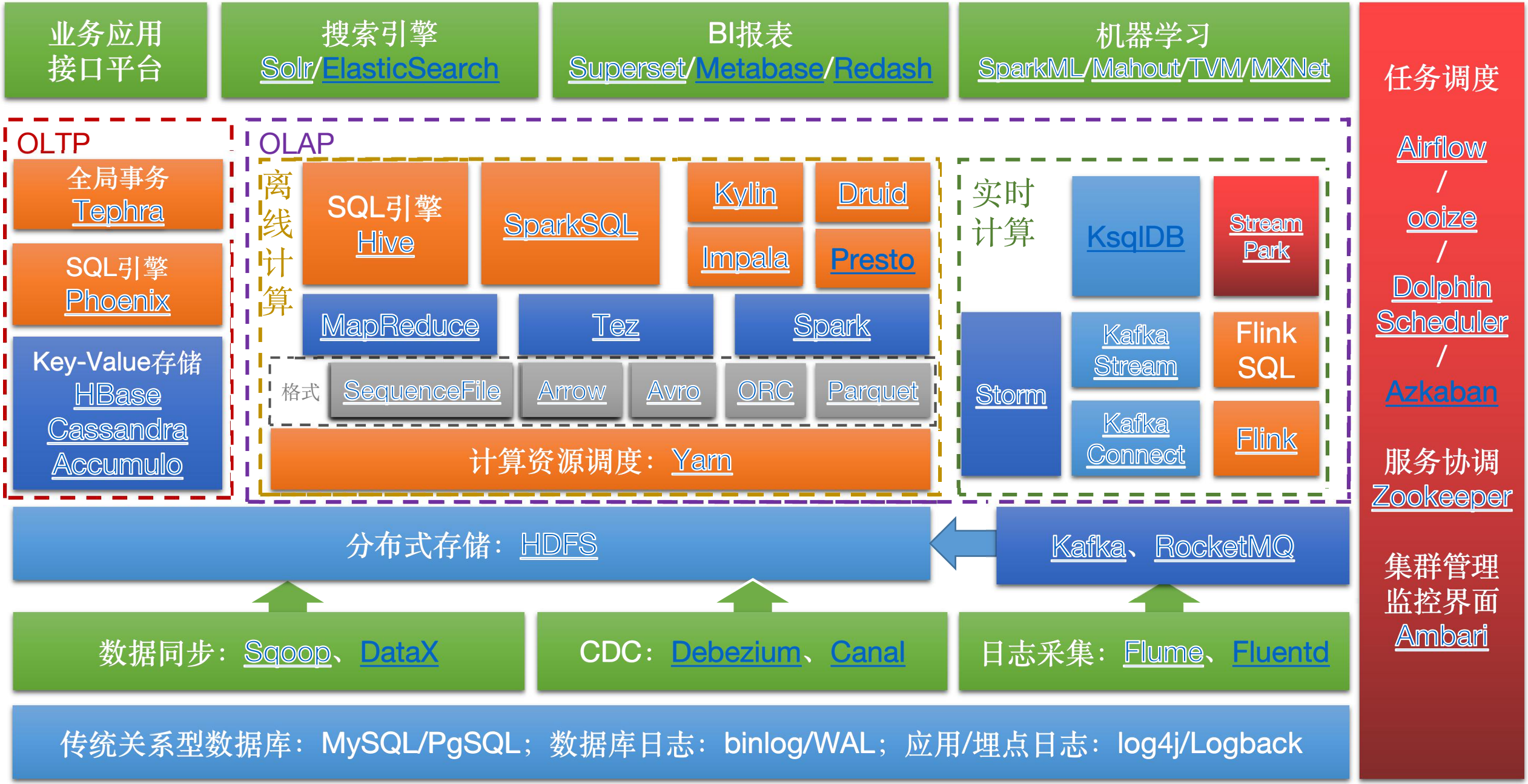
Dapper 2010 | EagleEye | Zipkin

NetO | B4 2013

Google Web Search Engine 1998

谷歌 | 阿里 | 开源

Hadoop生态

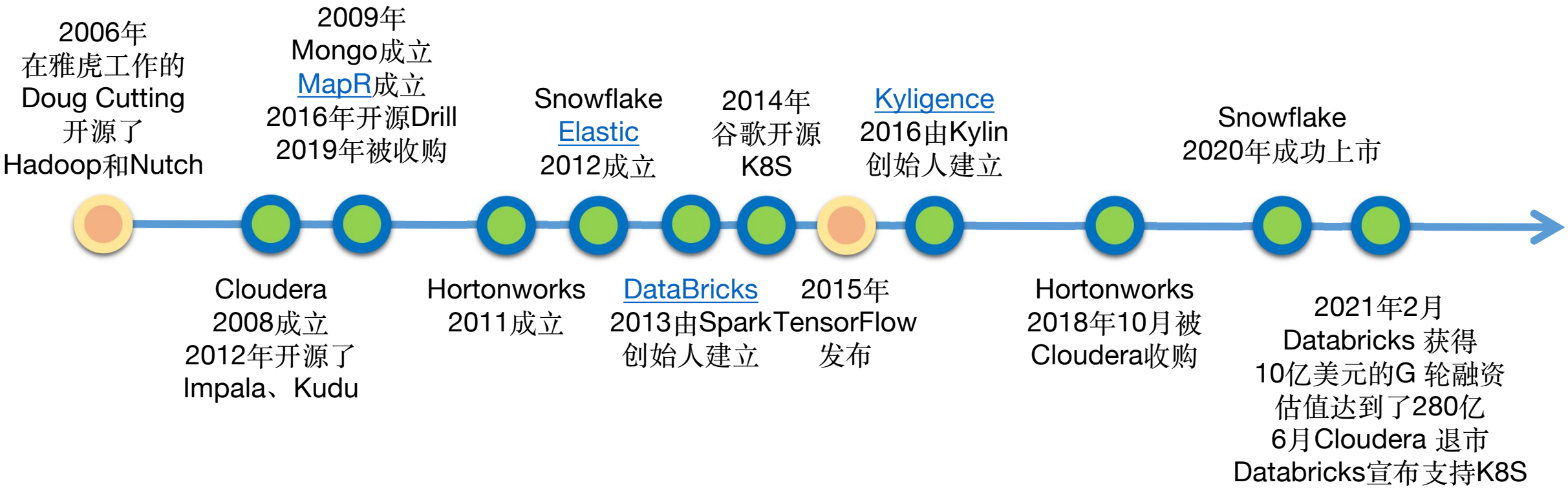

业务应用
接口平台
搜索引擎
Solr/ElasticSearch
BI报表
Superset/Metabase/Redash
机器学习
SparkML/Mahout/TVM/MXNet
任务调度
Airflow
/
ooize
/
Dolphin
Scheduler
/
Azkaban
服务协调
Zookeeper
集群管理
监控界面
Ambari
OLTP
全局事务
Tephra
SQL引擎
Phoenix
Key-Value存储
HBase
Cassandra
Accumulo
OLAP
离线计算
SQL引擎
Hive
SparkSQL
Kylin
Druid
Impala
Presto
MapReduce
Tez
Spark
格式
SequenceFile
Arrow
Avro
ORC
Parquet
计算资源调度：Yarn
实时计算
KsqlDB
Stream
Park
Storm
Kafka
Stream
Flink
SQL
Kafka
Connect
Flink
分布式存储：HDFS
Kafka、RocketMQ
数据同步：Sqoop、DataX
CDC：Debezium、Canal
日志采集：Flume、Fluentd
传统关系型数据库：MySQL/PgSQL；数据库日志：binlog/WAL；应用/埋点日志：log4j/Logback

国内公司开源的大数据产品

- 百度：Apache Doris(Palo OLAP)=>Selectdb&Starrocks、BaikalDB(HTAP)
- 阿里：Apache Flink、Apache RocketMQ、DataX、Canal、SREWorks
- 腾讯：Apache InLong(TubeMQ)
- 网易：Apache Kyuubi、Arctic
- 华为：Apache CarbonData、OpenLooKeng
- 滴滴：DDMQ、KnowStreaming
- 易观数科：Apache DolphinScheduler(DAG)
- 其他社区项目：Apache Kylin(OLAP)

Hadoop时代落幕

大数据未来的发展趋势



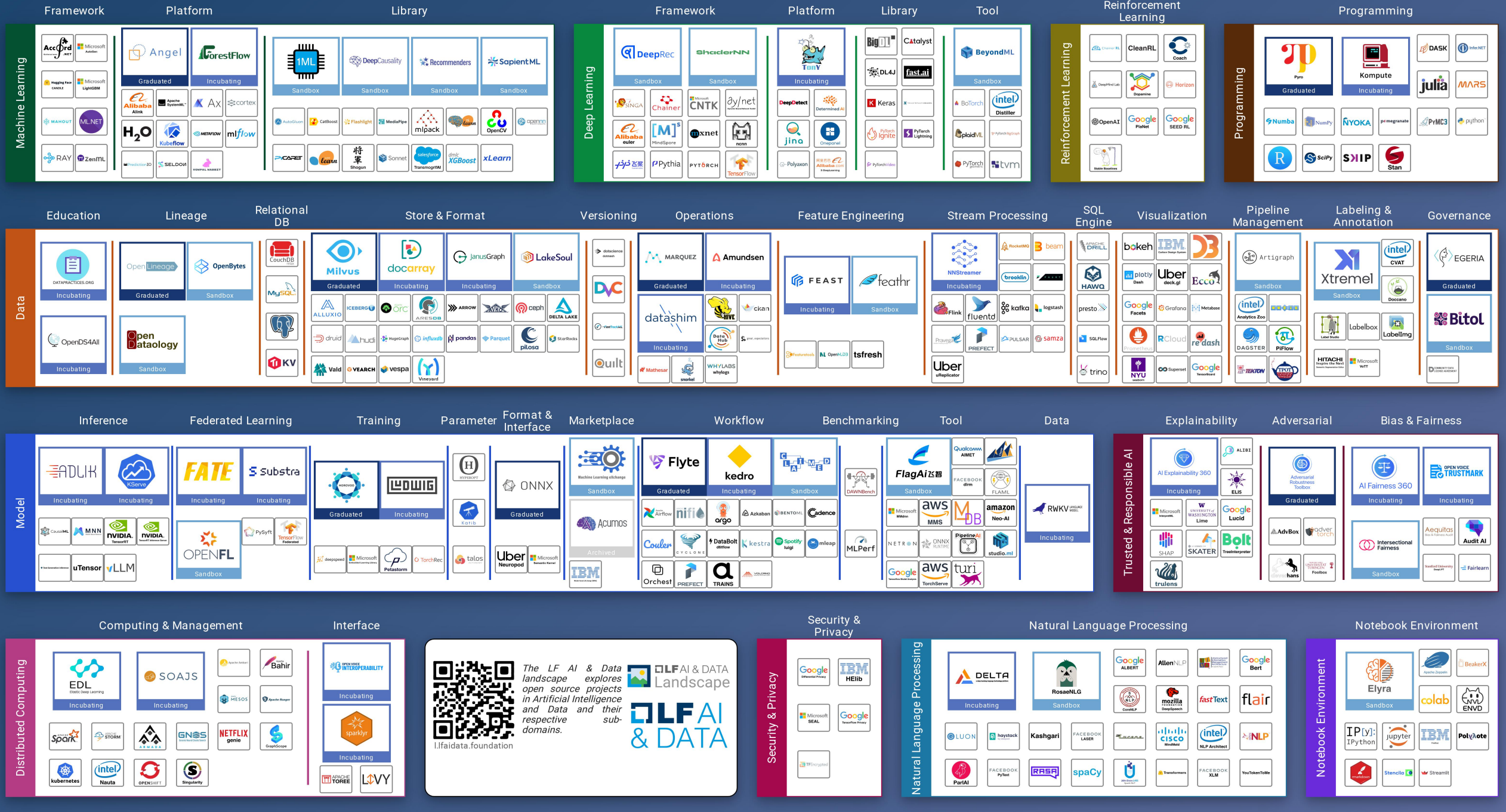
https://www.cloudera.com/products/open-source/apache-hadoop/key-cdh-components.html
https://www.zhihu.com/question/508594228

Linux Foundation AI & Data Landscape

1.0

See the interactive landscape at l.lfaidata.foundation

Greyed logos are not open source



OLTP vs. OLAP

	OLTP	OLAP
使用场景	在线业务服务	分析统计、数据挖掘、机器学习、人工智能
数据来源	应用写入RDB的业务数据	来源广泛。经过清洗的日志数据，批量同步或实时同步的业务库数据
事务要求	ACID要求高	对事务要求不高，数据可以通过同步进行重建
并发要求	高并发	低并发
索引存储方式	B+树按行存储，方便原地修改	存储方式多样，为了方便统计时扫描大量数据，通常按列存储
技术实现	MySQL、PostgreSQL	Spark、Kylin、Druid、ClickHouse

小公司的小产品如何借鉴

Hadoop生态臃肿、运维成本大

2016 年俄罗斯的搜索引擎公司Yandex开源

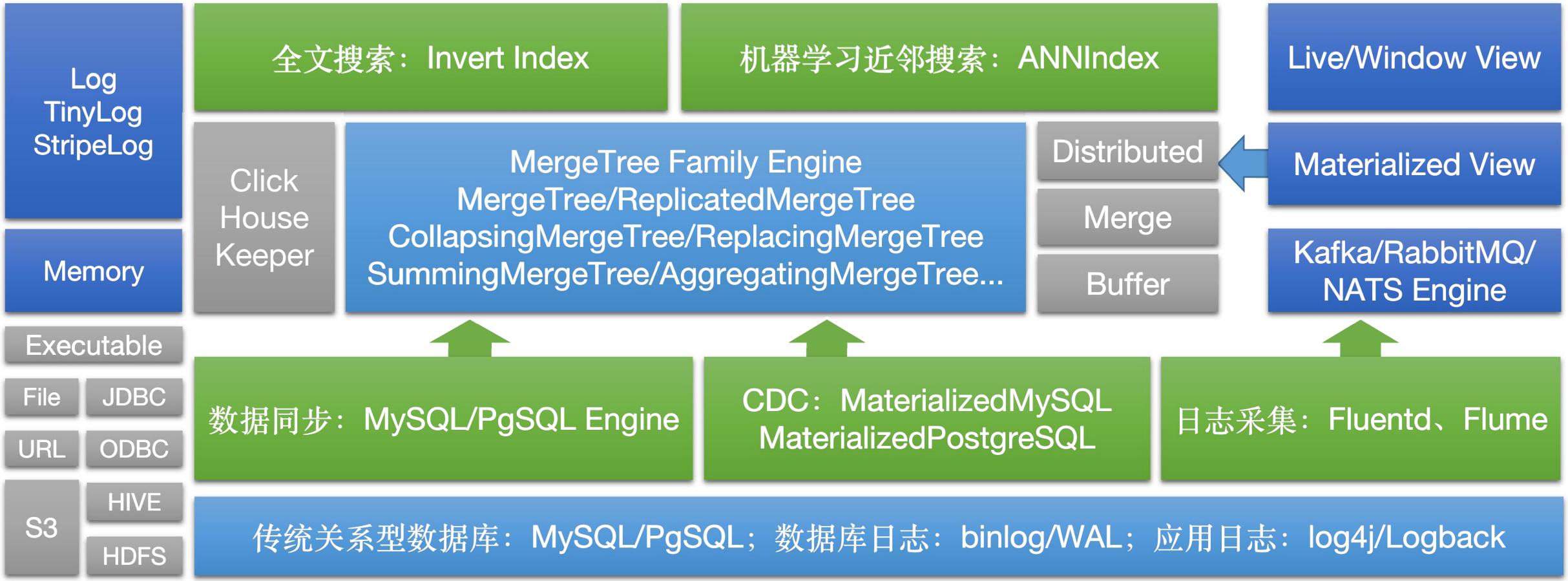
优点：

1、独立生态，不依赖Hadoop的存储计算体系
2、列式存储。方便压缩、统计
3、查询和写入速度快。
4、充分利用多核CPU并行计算
5、功能完善。

缺点：

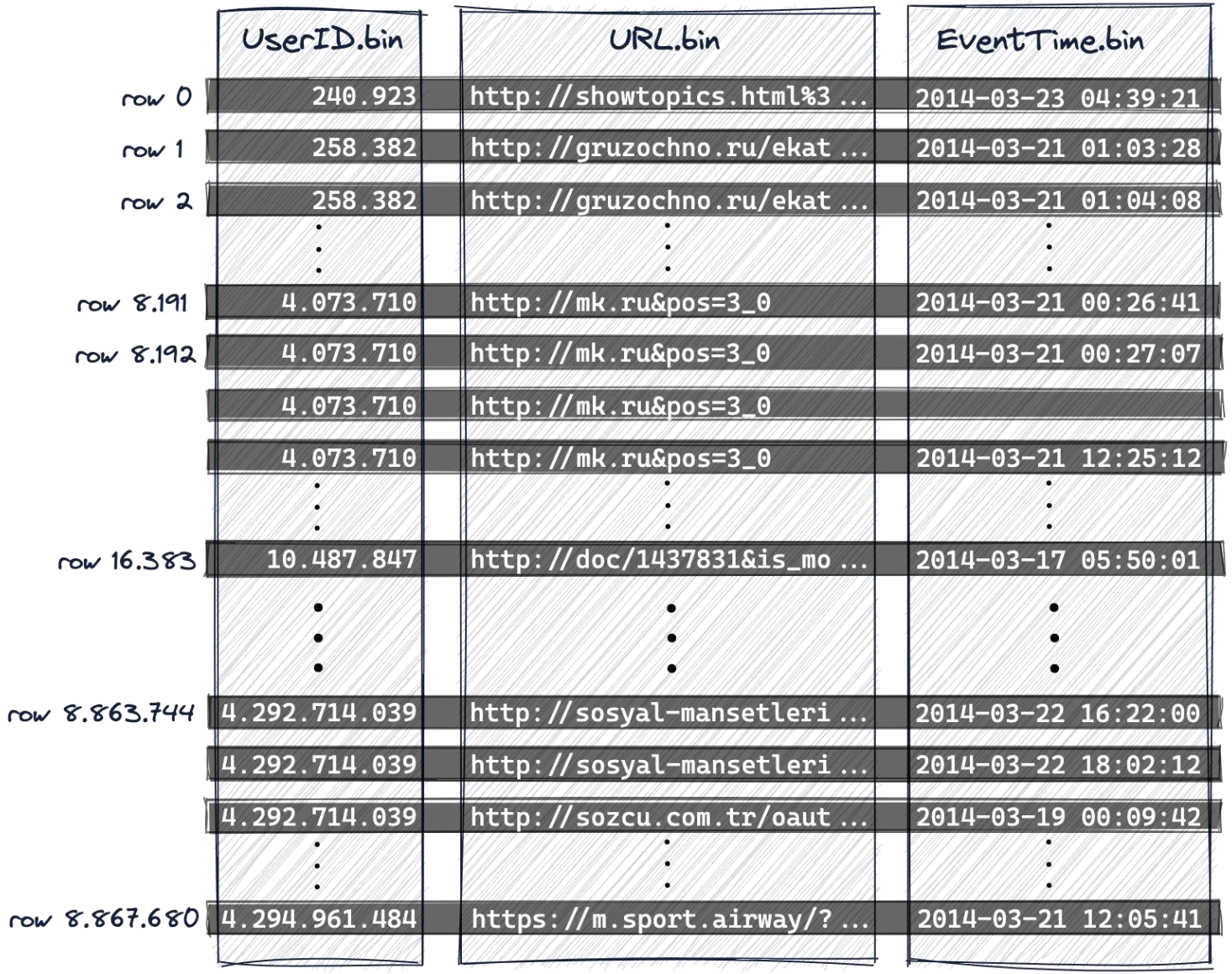
1、并发度低。单机默认最大并发100
2、事务支持度低。只支持批量插入的原子性
3、删除操作重。使用LSM-Tree结构，异步删除
4、只能用批量插入，否则容易出现too many part

ClickHouse核心功能

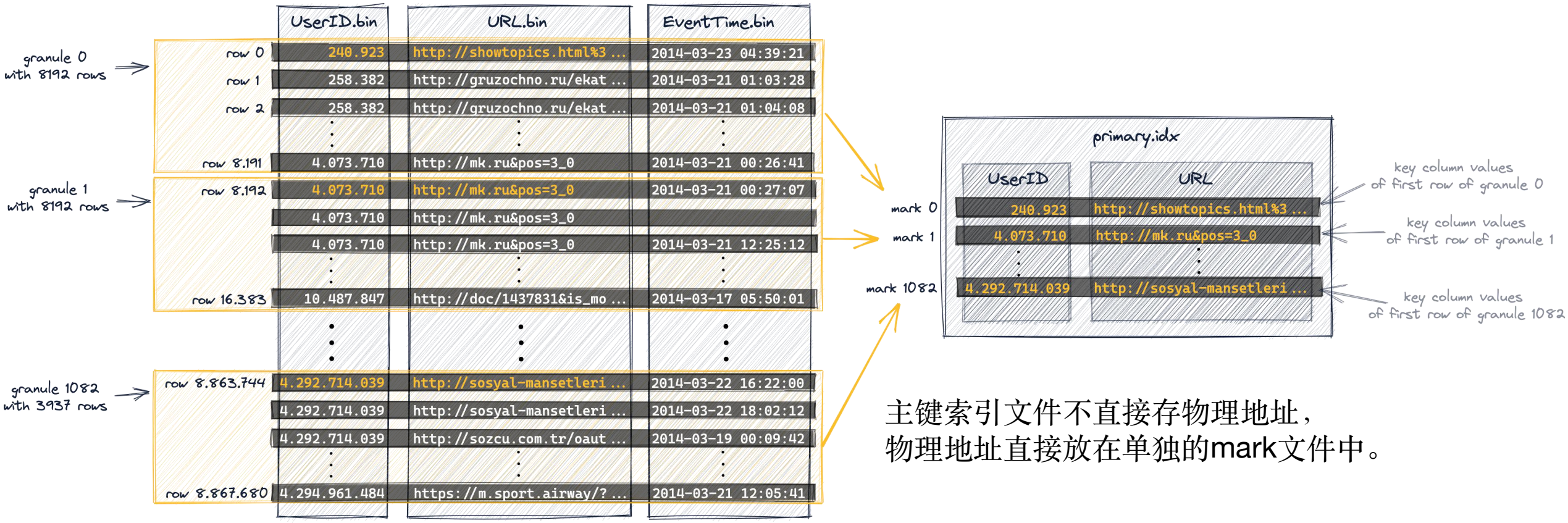


ClickHouse的列式存储

```
CREATE TABLE hits_UserID_URL
(
    `UserID` UInt32,
    `URL` String,
    `EventTime` DateTime
)
ENGINE = MergeTree
PRIMARY KEY (UserID, URL)
ORDER BY (UserID, URL, EventTime)
SETTINGS index_granularity = 8192,
                index_granularity_bytes = 0;
```

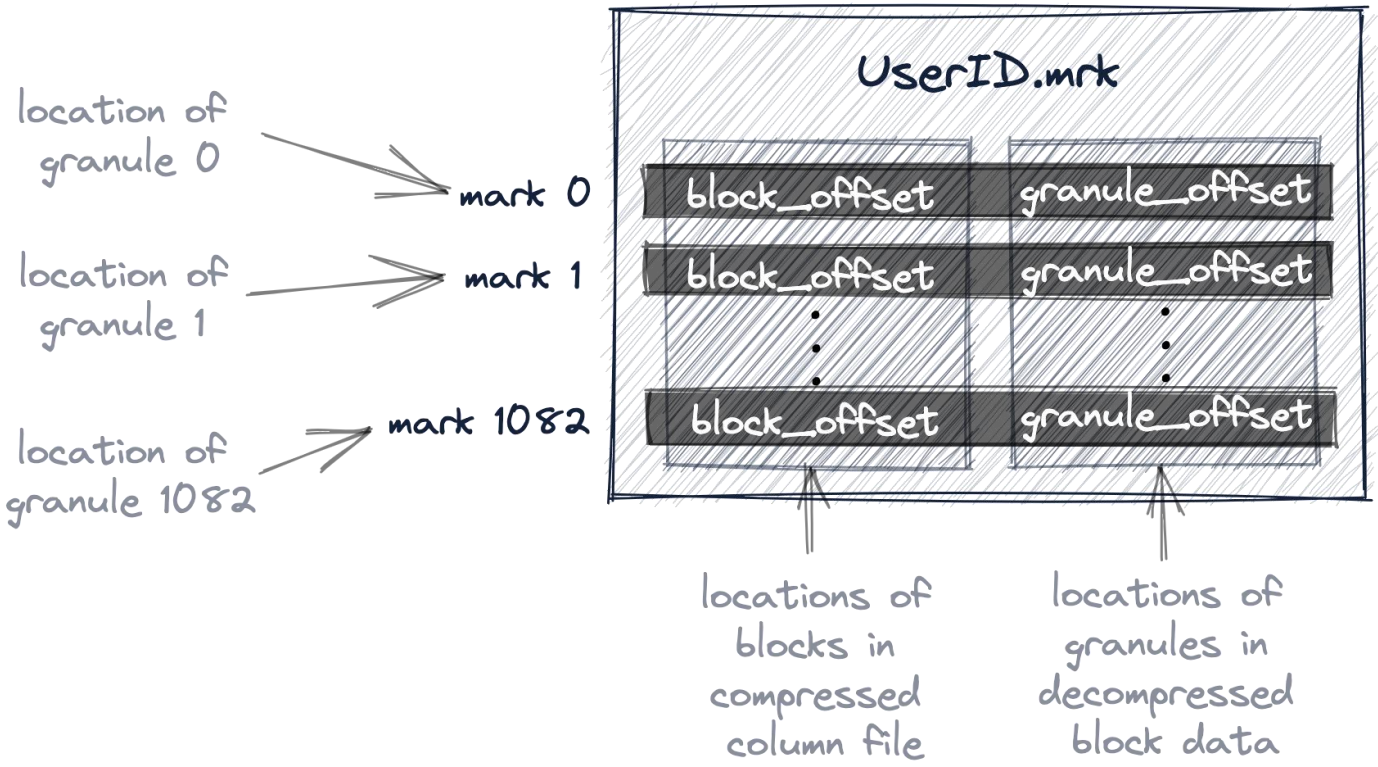


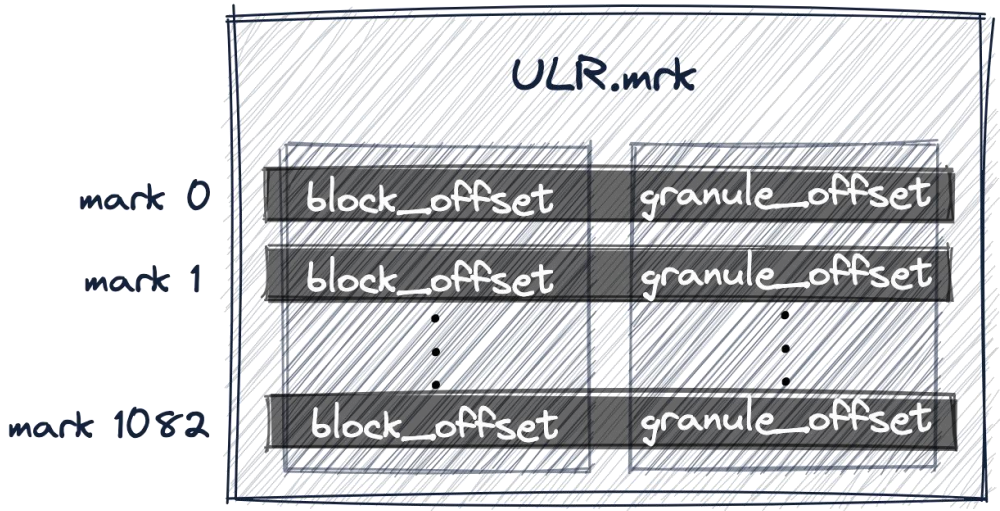
ClickHouse的索引设计



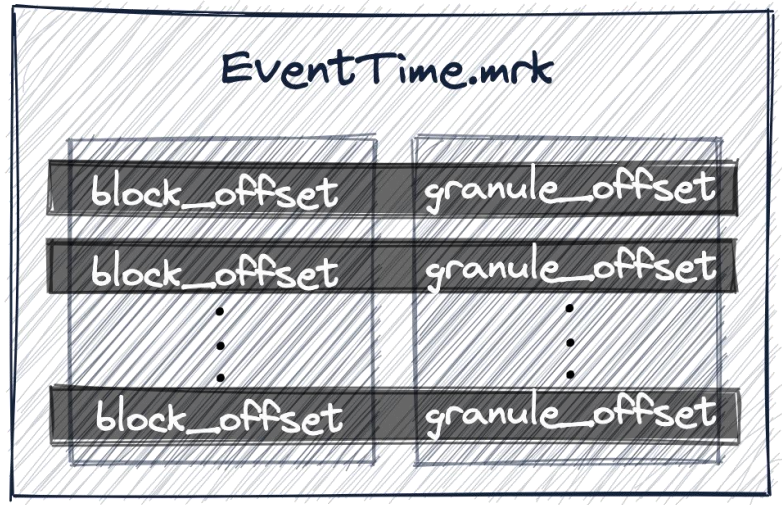
主键索引文件不直接存物理地址，
物理地址直接放在单独的mark文件中。

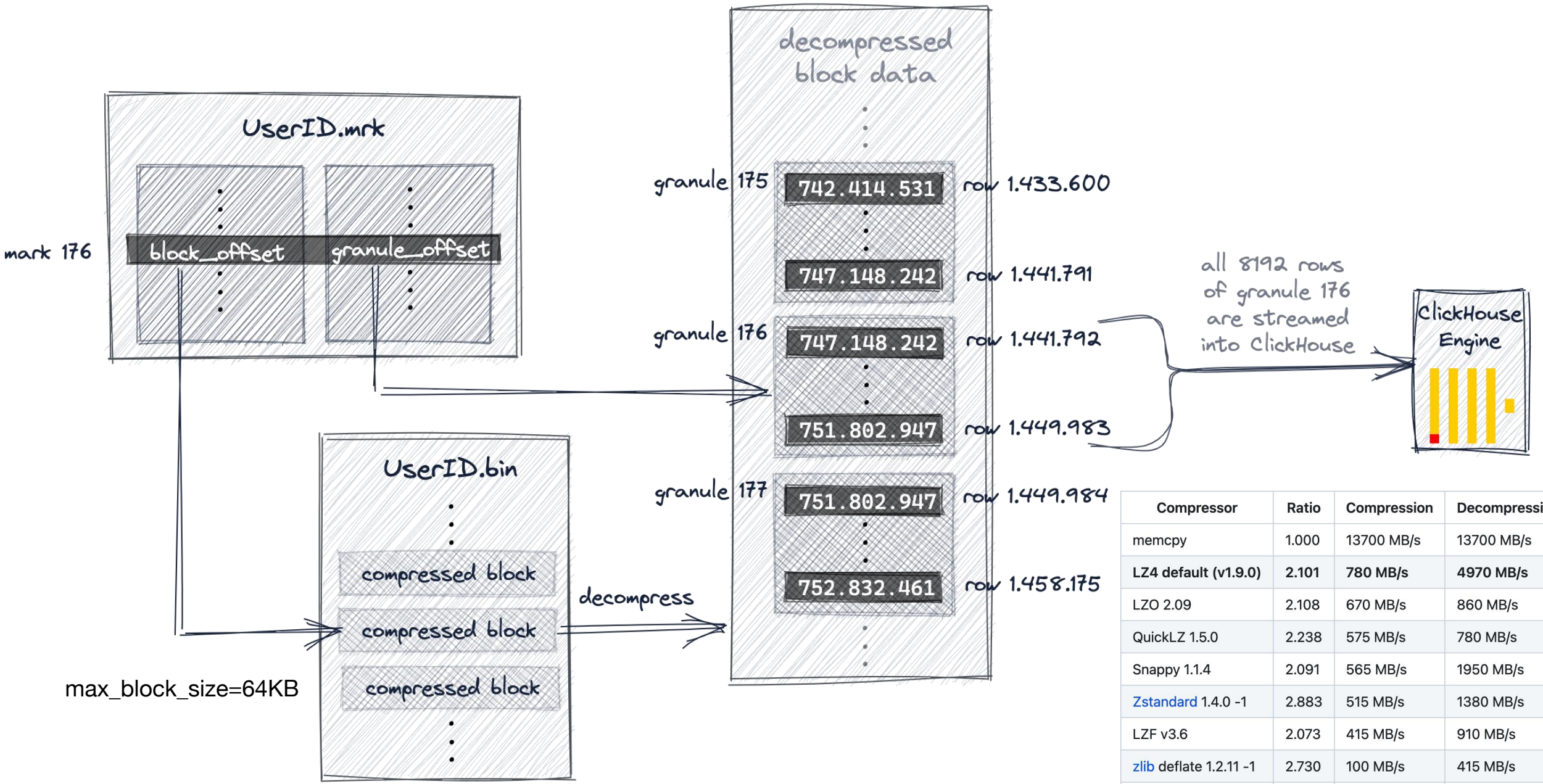
refs: primary-indexes-design， primary-indexes-multiple





1、mark文件格式固定，根据granuleID可以直接定位物理地址
2、减少primary key的大小，方便缓存在内存中，加快查询





Compressor	Ratio	Compression	Decompression
memcpy	1.000	13700 MB/s	13700 MB/s
LZ4 default (v1.9.0)	**2.101**	**780 MB/s**	**4970 MB/s**
LZO 2.09	2.108	670 MB/s	860 MB/s
QuickLZ 1.5.0	2.238	575 MB/s	780 MB/s
Snappy 1.1.4	2.091	565 MB/s	1950 MB/s
Zstandard 1.4.0 -1	2.883	515 MB/s	1380 MB/s
LZF v3.6	2.073	415 MB/s	910 MB/s
zlib deflate 1.2.11 -1	2.730	100 MB/s	415 MB/s
LZ4 HC -9 (v1.9.0)	**2.721**	41 MB/s	**4900 MB/s**
zlib deflate 1.2.11 -6	3.099	36 MB/s	445 MB/s

PostgreSQL生态

优点

1、支持DAG图调度，可以编排任务依赖关系

2、支持丰富的数据源和任务类型

3、日志、监控相对完善

目前存在的痛点

1、多环境发布不方便

2、多版本控制支持较弱，开发不方便