基于 阿里云 SelectDB 构建
高性能、低成本、开放的可观测性平台

飞轮科技 2025.04

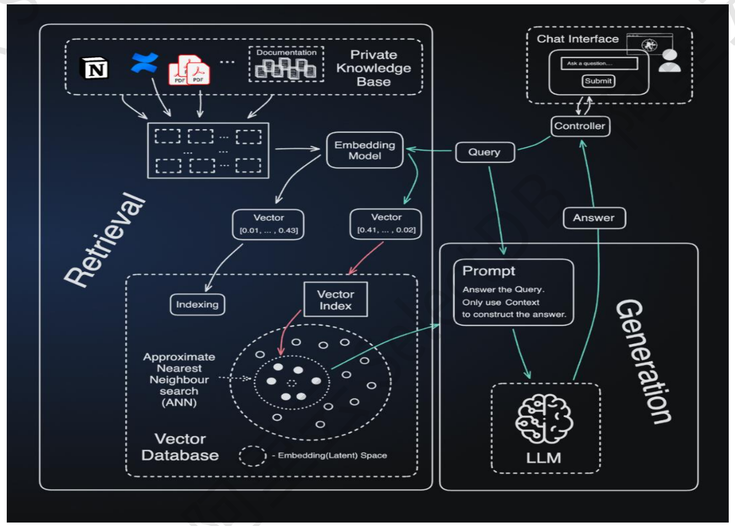
目录

1 可观测性场景需求与挑战

可观测性的典型应用场景

故障排查

保障服务稳定 提升用户体验



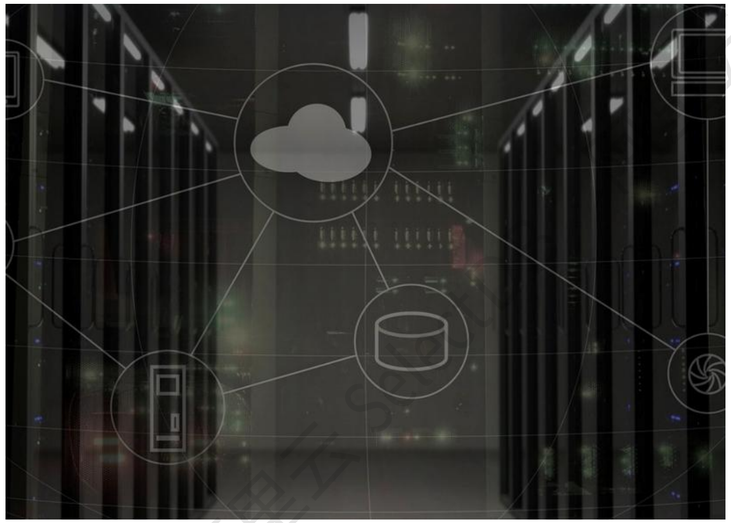
GenAI & LLM

大模型应用全链路可观测

运维开发

高效协同 提升开发和运维效率

可观测性数据的特点



Volume – 数据量大

PB级海量存储、存储周期长
对存储成本敏感

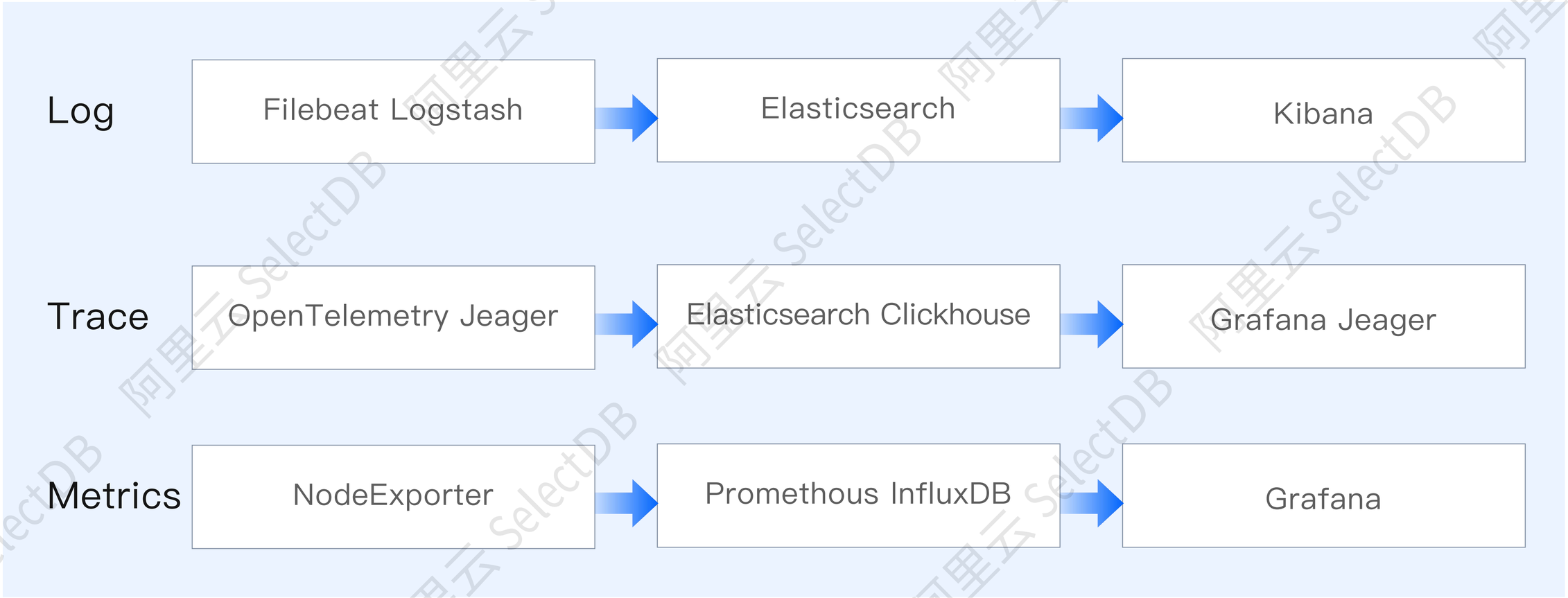
Velocity – 实时写入与检索

百万条/s GB/s 高吞吐、秒级实时写入
秒级实时交互式检索分析

Variety – 数据多样化

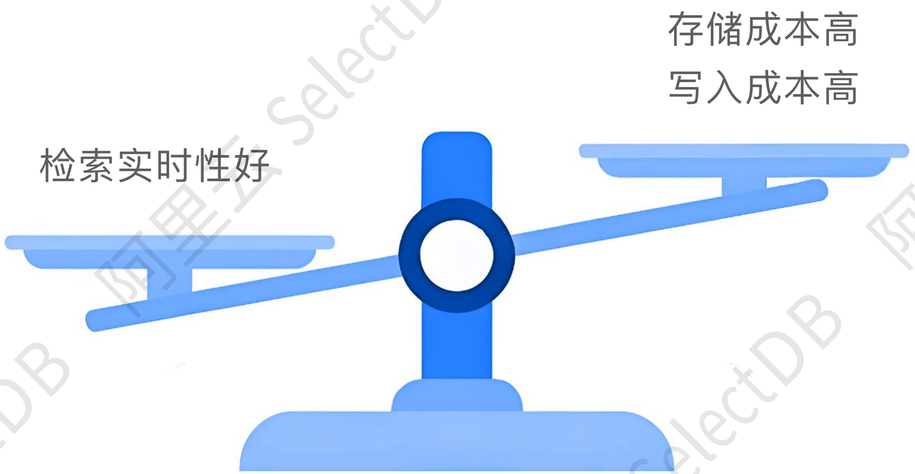
需要对接类型多样的数据
Text 和 JSON Schema 复杂多变

典型可观测性平台架构

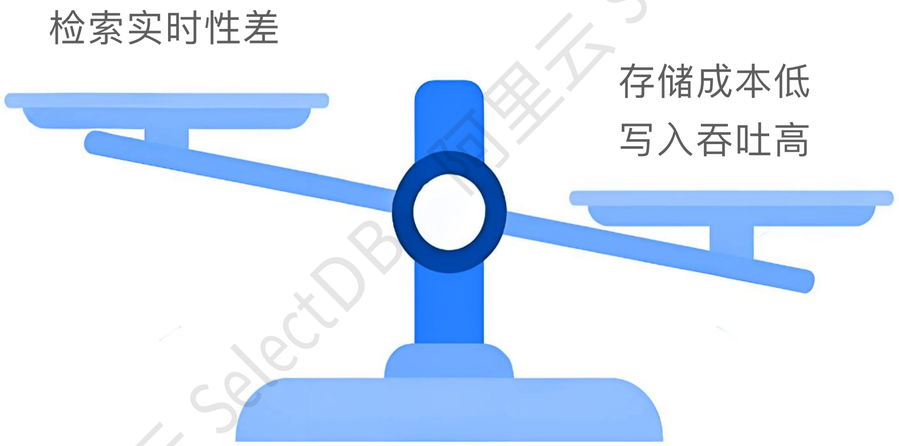


挑战1 — 实时性与成本的矛盾

Elasticsearch 为代表的稠密倒排索引架构



Loki Clickhouse 为代表的稀疏跳数索引架构



挑战2 — 半结构化数据处理困难

Log 扩展字段


```
time: “2024-04-09 16:06:03.684886”
source: "fluentd_tcp"
host: ”host1"
status: info
filepath: ”path1"
container_id: ”xxx"
container_name: ”xxx"
message: "....."
ext: {
"nameSpace": "kube-public",
"level": "info",
"service": "console-media",
"category": "debug",
"tag": "[]",
"serviceID": "console-media-6b57fc7758",
"date_ns": 886517,
"index": "default",
"cluster": "UcloudKubernetes"
}
```


Trace 属性字段


```
time: “2024-04-09 16:13:51.381141”
source: "opentelemetry"
service: "iov-fleet-mqtt"
parent_id: "0"
resource: "fleet/ListFleets"
span_id: ”xxxx"
trace_id: ”xxxx"
duration: 3810
start: 1712650431181241
attrs: {
"rpc_system": "grpc",
"rpc_method": "ListFleets",
"service_sub": "iov-fleet-mqtt",
"k8s_cluster": "sh-test-uk8s",
"date_ns": 141210,
"rpc_service": "fleet.FleetService",
"service_name": "iov-fleet-mqtt",
"rpc_grpc_status_code": 0
}
```


- 字段个数不固定
- 字段类型不固定
- 需要高性能过滤和分析

挑战3 — 可观测性生态开放多样

数据采集生态

可视化分析生态

3 基于 SelectDB 的可观测性平台

基于 SelectDB 构建 高性能、低成本、开放的可观测性平台

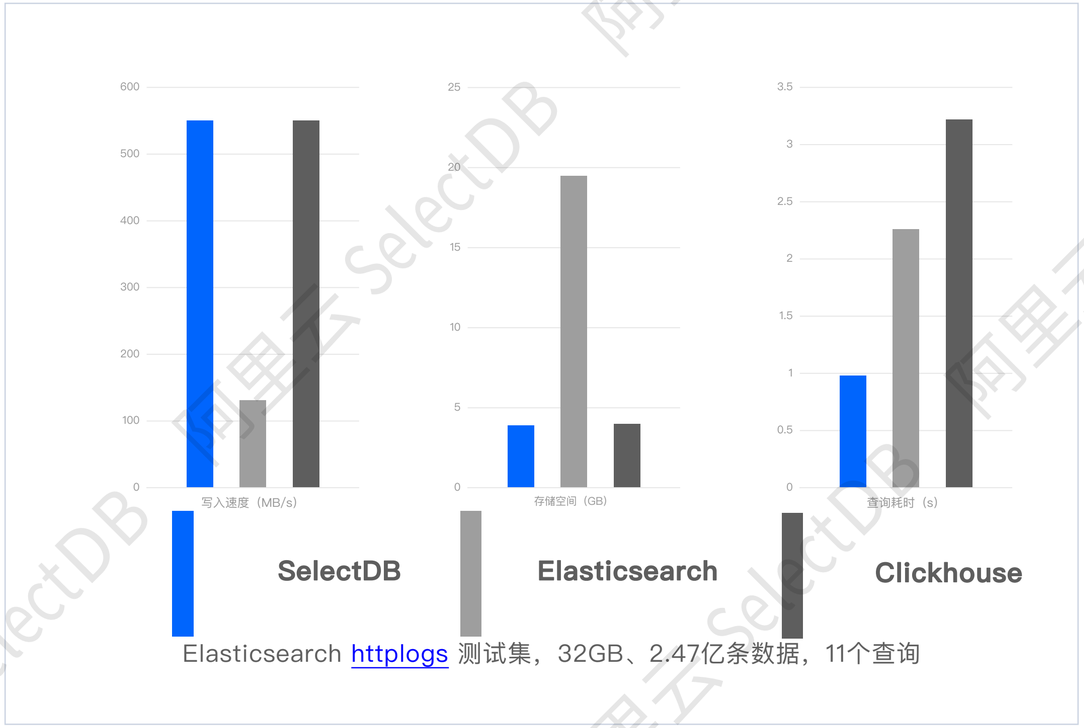
开放的数据接入方式

高性能、低成本、开放的
统一存储底座

兼容 MySQL 的开放生态

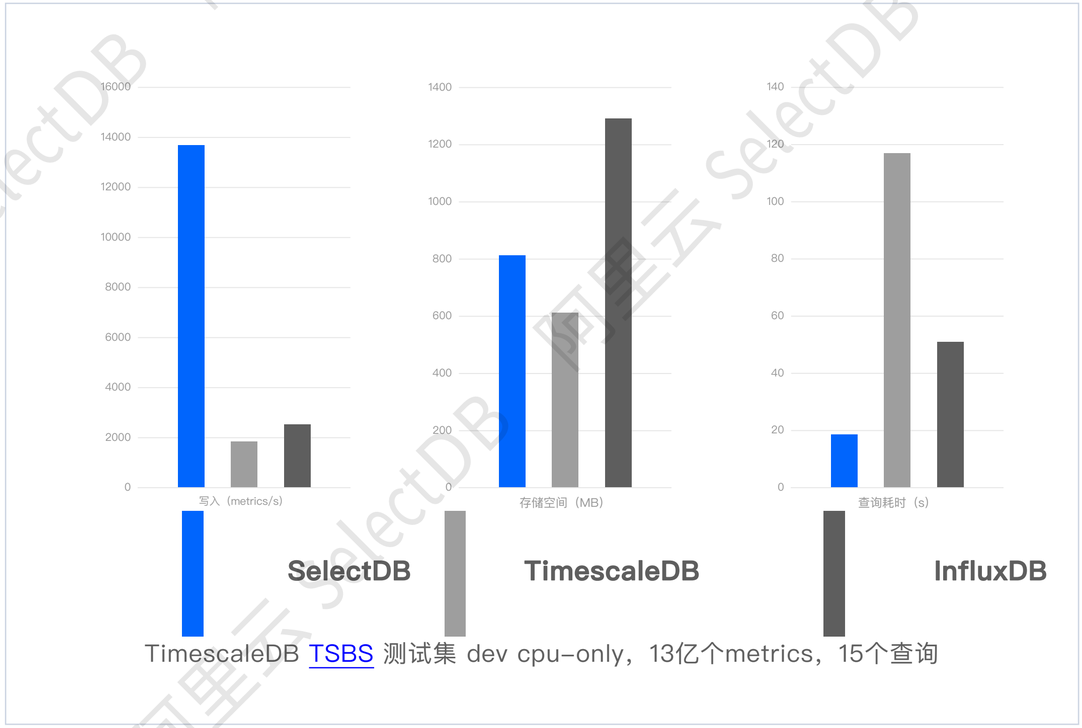
优势1 — 高性能、低成本

Log Trace 相对于ES 写入性能 3~5倍，80% 存储空间降低，查询性能 2~3倍，相对 CK 查询性能 3倍



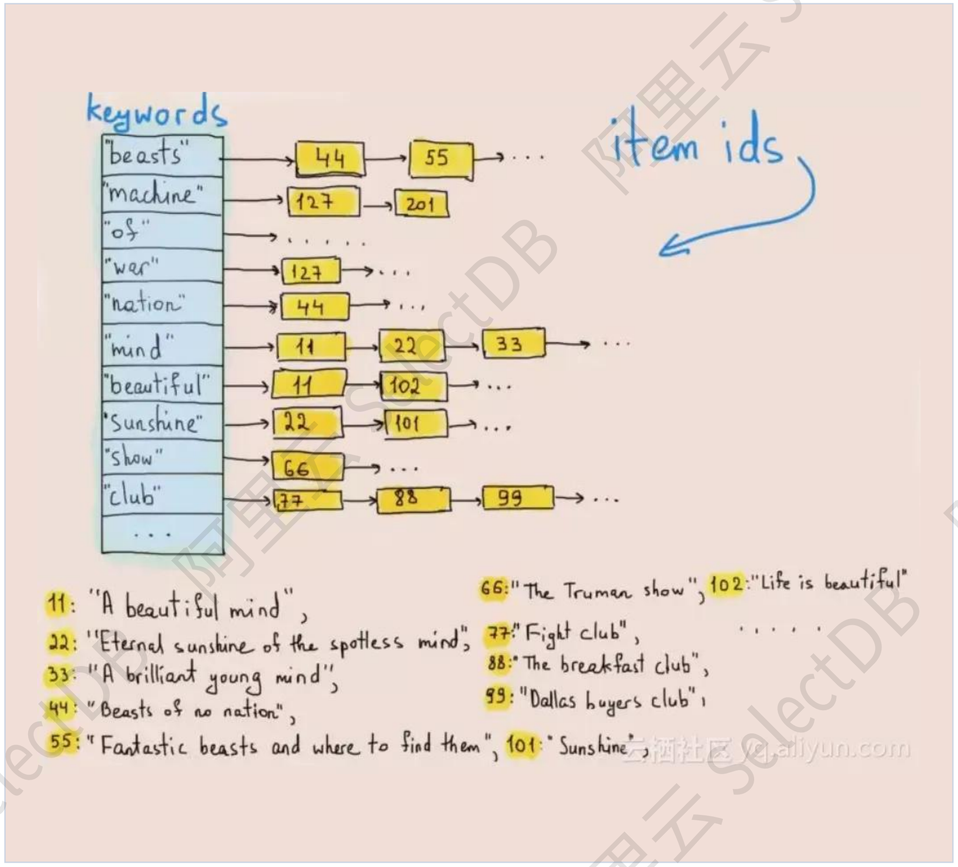
Elasticsearch [httplogs](#) 测试集，32GB、2.47亿条数据，11个查询

Metrics 相对于 TimescaleDB InfluxDB 写入性能 5~7倍，查询性能 2~6倍



TimescaleDB [TSBS](#) 测试集 dev cpu-only，13亿个metrics，15个查询

关键技术 — 高性能倒排索引




```
CREATE TABLE httplog
(
  `ts` DATETIME,
  `clientip` VARCHAR(20),
  `request` TEXT,
  INDEX idx_clientip (`clientip`) USING INVERTED,
  INDEX idx_request (`request`) USING INVERTED PROPERTIES("parser" = "unicode")
)
DUPLICATE KEY(`ts`)
...

-- 查看最新的10条数据
SELECT * FROM httplog ORDER BY ts DESC LIMIT 10;
-- 查询clientip为'8.8.8.8'的最新10条数据
SELECT * FROM httplog WHERE clientip = '8.8.8.8' ORDER BY ts DESC LIMIT 10;
-- 检索request字段中有error或者404的最新10条数据
SELECT * FROM httplog WHERE request MATCH_ANY 'error 404' ORDER BY ts DESC LIMIT 10;
-- 检索request字段中有image和faq的最新10条数据
SELECT * FROM httplog WHERE request MATCH_ALL 'image faq' ORDER BY ts DESC LIMIT 10;
```

关键技术 — 泛时序数据查询优化

```
SELECT * FROM log
WHERE ts >= t1 AND ts <= t2 AND message MATCH 'error'
ORDER BY ts DESC LIMIT 100
```

挑战 从海量日志中全文检索关键词

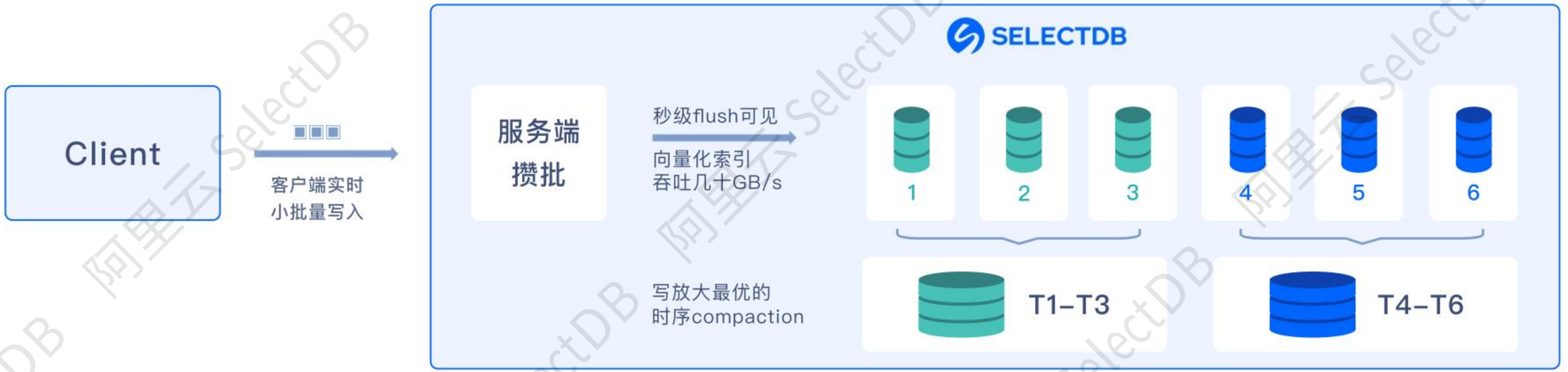
基于分区、主键的时间范围快速跳过
基于倒排索引的全文检索精准定位

挑战 按时间排序取满足条件的最新N条日志

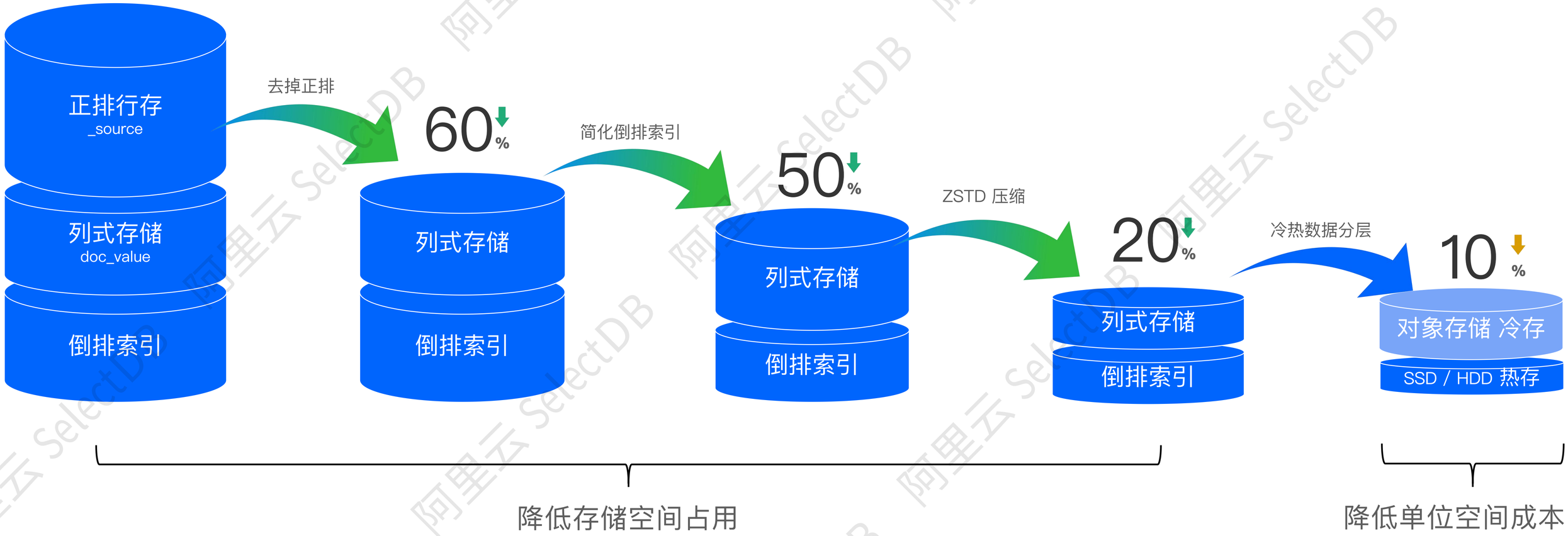
按时间排序的时序存储模型
基于动态剪枝的TopN查询算法

效果：百亿日志检索秒级响应

关键技术 — 高吞吐低延迟导入



关键技术 — 极致存储成本


正排行存
_source
列式存储
doc_value
倒排索引
去掉正排
60%
列式存储
倒排索引
简化倒排索引
50%
列式存储
倒排索引
ZSTD 压缩
20%
列式存储
倒排索引
冷热数据分层
10%
对象存储 冷存
SSD / HDD 热存
降低存储空间占用
降低单位空间成本

优势2 — 灵活的半结构化数据类型 VARIANT

JSON数据自适应

- 自动识别JSON数据中的字段名和类型
- 自动将频繁出现的字段采用列式存储
- 自动将不频繁的字段合并存储，避免类似ES的mapping爆炸

支持一个字段多个类型

- 允许一个字段有多种类型
- 更好满足业务发展中字段类型变化的需求

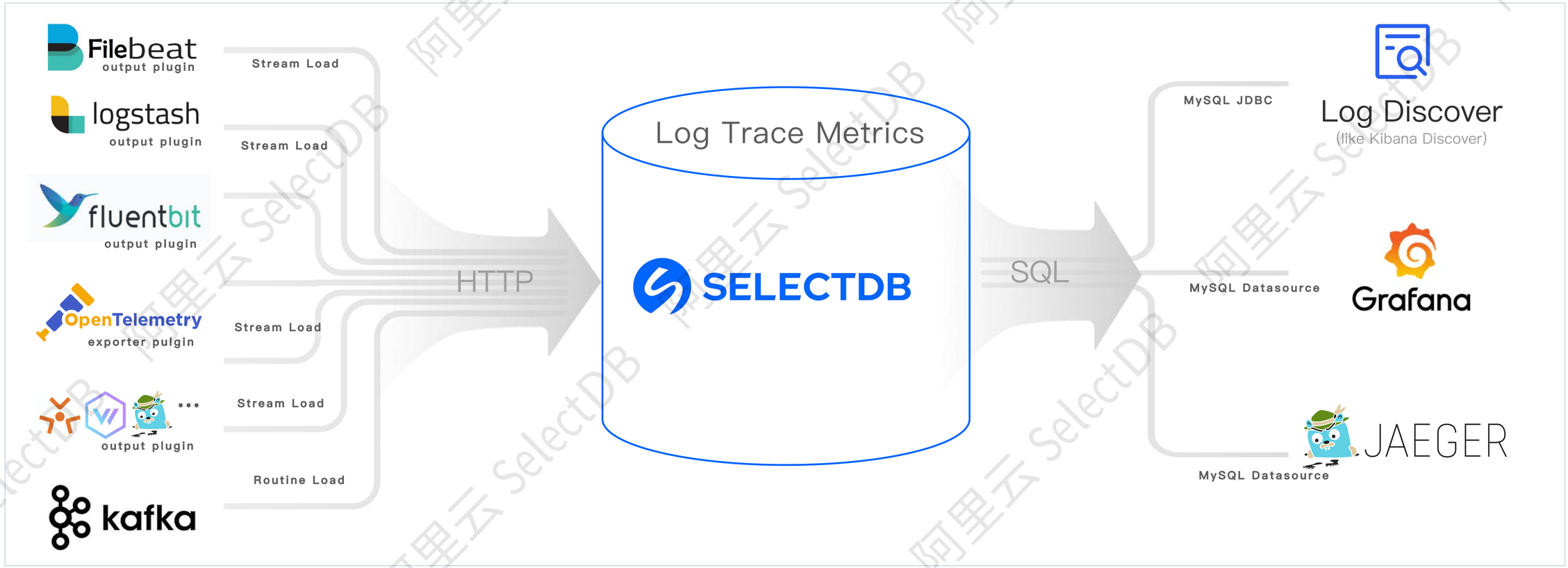
支持倒排索引

- 为variant字段创建索引，子字段自动创建
- 可指定文本字段是否分词、分词类型等参数
- 后续版本将支持各个子字段索引灵活定义

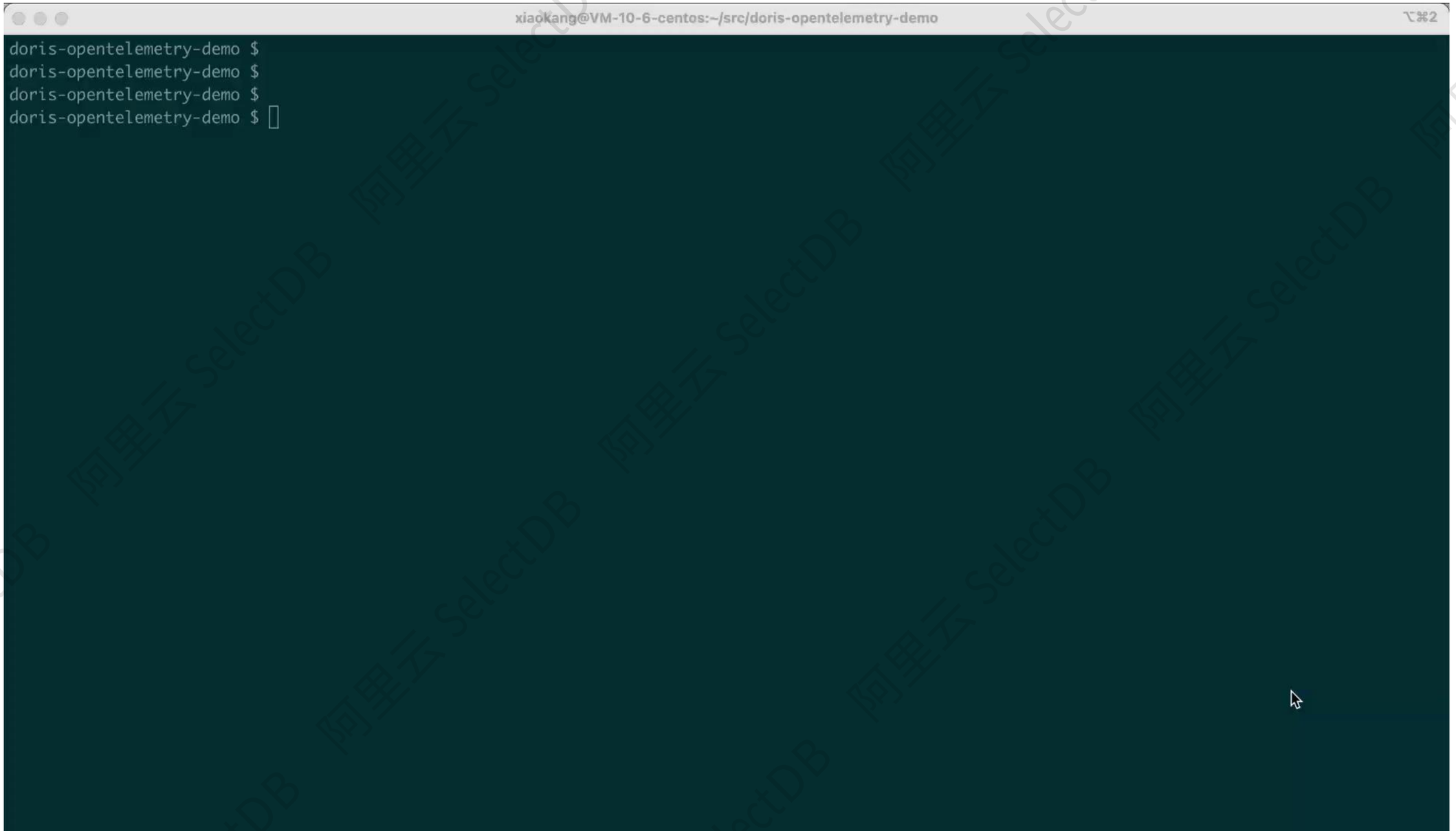
```
-- 创建了三个VARIANT类型的列， actor, repo和payload
-- 创建表的同时创建了payload列的倒排索引idx_payload
-- USING INVERTED 指定索引类型是倒排索引，用于加速子列的条件过滤
-- PROPERTIES("parser" = "english") 指定对子列采用english分词
CREATE TABLE IF NOT EXISTS github_events (
    id BIGINT NOT NULL,
    type VARCHAR(30) NULL,
    actor VARIANT NULL,
    repo VARIANT NULL,
    payload VARIANT NULL,
    public BOOLEAN NULL,
    created_at DATETIME NULL,
    INDEX idx_payload (`payload`) USING INVERTED PROPERTIES("parser" = "english")
)
```

```
mysql> desc github_events;
+-----------------------------------------------------+------------+------+-------+---------+-------+
| Field                                               | Type       | Null | Key   | Default | Extra |
+-----------------------------------------------------+------------+------+-------+---------+-------+
| id                                                  | BIGINT     | No   | true  | NULL    |       |
| type                                                | VARCHAR(*) | Yes  | false | NULL    | NONE  |
| actor                                               | VARIANT    | Yes  | false | NULL    | NONE  |
| actor.avatar_url                                    | TEXT       | Yes  | false | NULL    | NONE  |
| actor.display_login                                 | TEXT       | Yes  | false | NULL    | NONE  |
| actor.id                                            | INT        | Yes  | false | NULL    | NONE  |
| actor.login                                         | TEXT       | Yes  | false | NULL    | NONE  |
| actor.url                                           | TEXT       | Yes  | false | NULL    | NONE  |
| created_at                                          | DATETIME   | Yes  | false | NULL    | NONE  |
| payload                                             | VARIANT    | Yes  | false | NULL    | NONE  |
| payload.action                                      | TEXT       | Yes  | false | NULL    | NONE  |
| payload.before                                      | TEXT       | Yes  | false | NULL    | NONE  |
| payload.comment.author_association                  | TEXT       | Yes  | false | NULL    | NONE  |
| payload.comment.body                                | TEXT       | Yes  | false | NULL    | NONE  |
....
+-----------------------------------------------------+------------+------+-------+---------+-------+
```

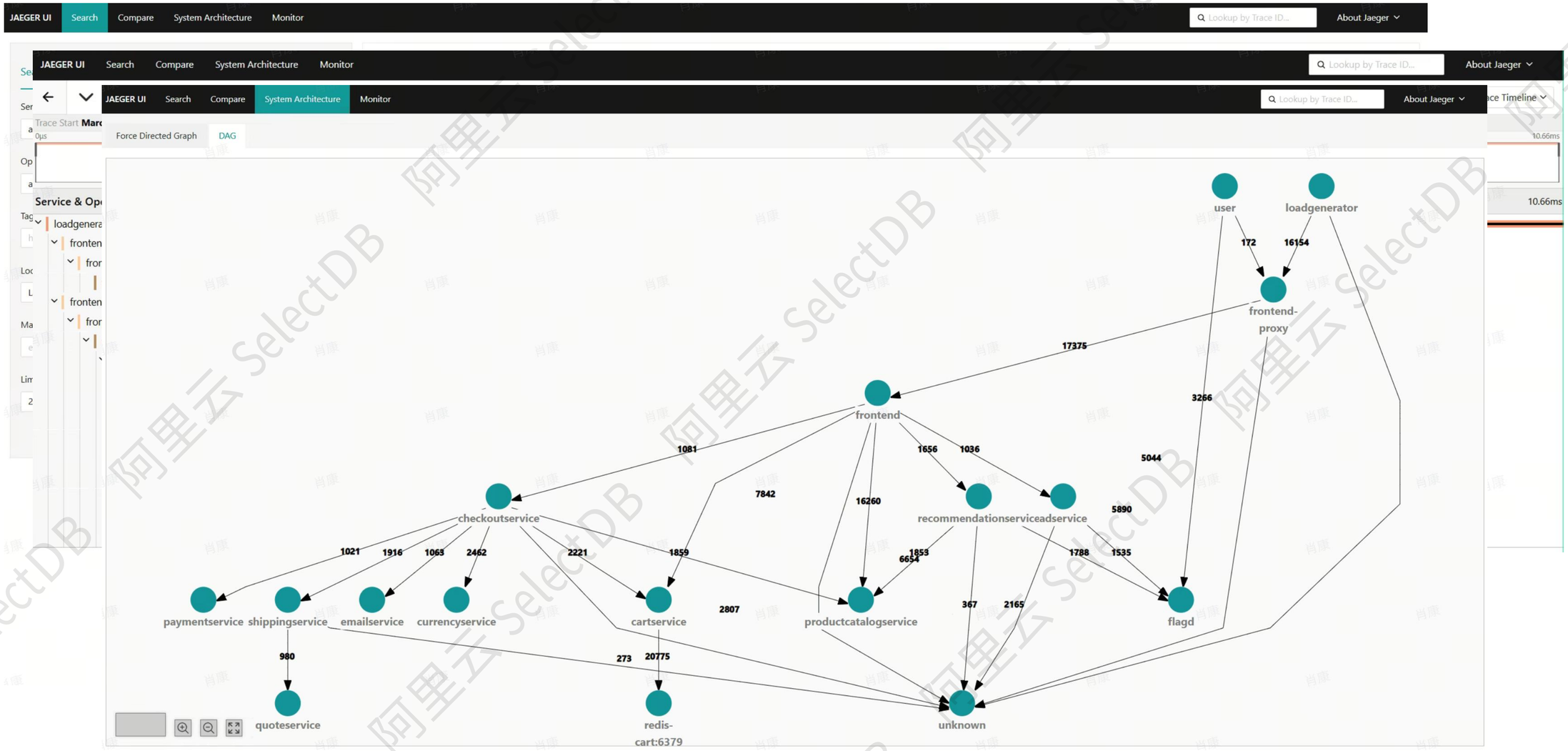
优势3 — 开放丰富的生态对接



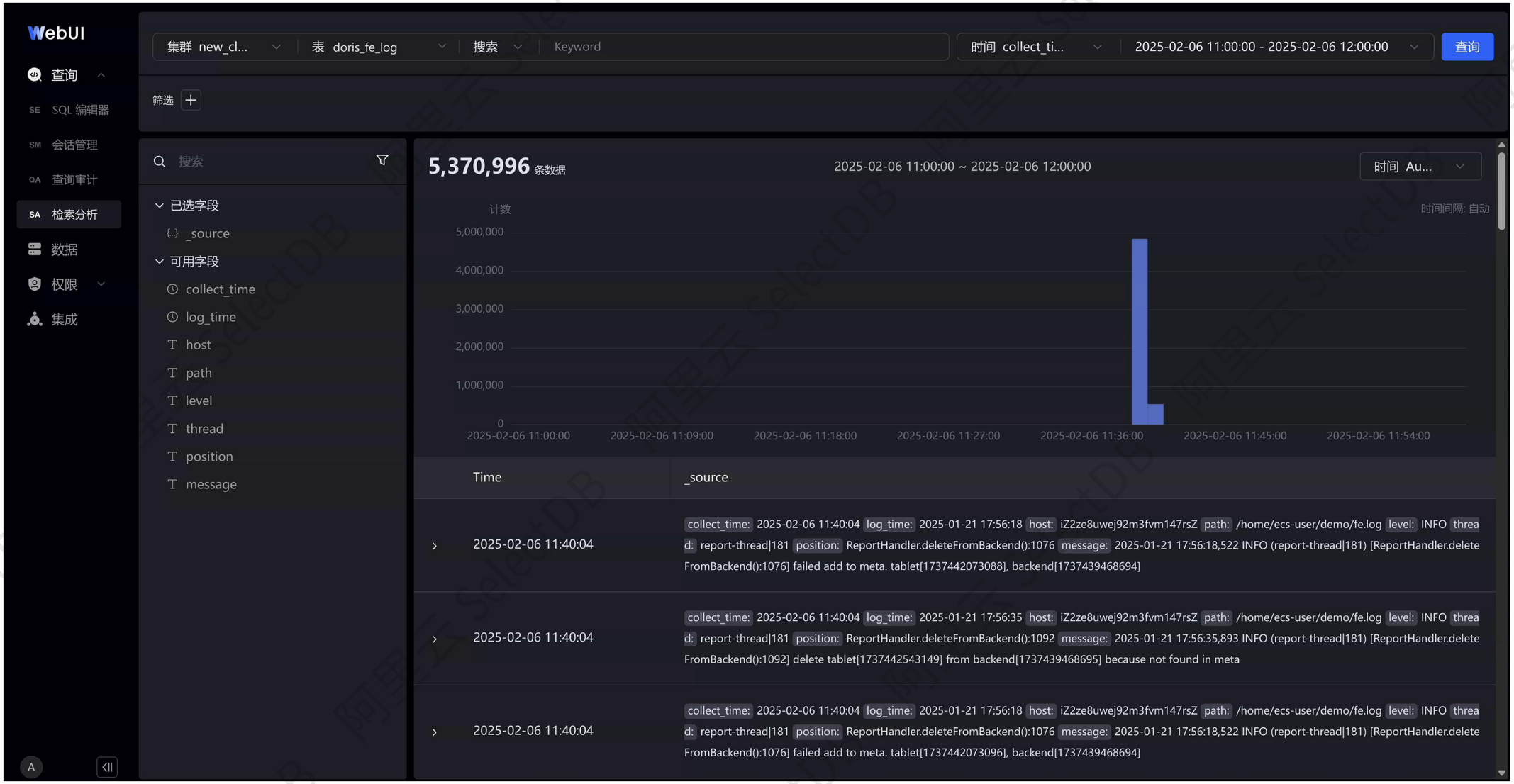
生态对接 OpenTelemetry & Grafana



生态对接 Jeager


JAEGER UI
Search
Compare
System Architecture
Monitor
Lookup by Trace ID...
About Jaeger
Force Directed Graph
DAG
user
loadgenerator
frontend-proxy
frontend
checkoutservice
recommendationservice
adservice
paymentservice
shippingservice
emailservice
currencyservice
cartservice
productcatalogservice
flagd
quoteservice
redis-cart:6379
unknown
172
16154
17375
3266
1081
1656
1036
5044
7842
16260
5890
1021
1916
1063
2462
2221
1859
6654
1853
1788
1535
2807
367
2165
980
273
20775

生态对接 SelectDB WebUI



优势4 — 开放易用 数据统一

3大可观测数据统一

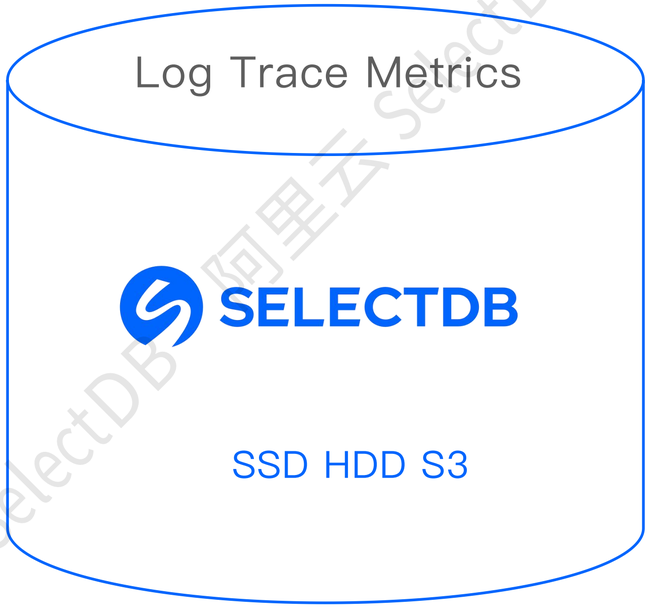
- Log
- Trace
- Metrics

历史和实时数据统一

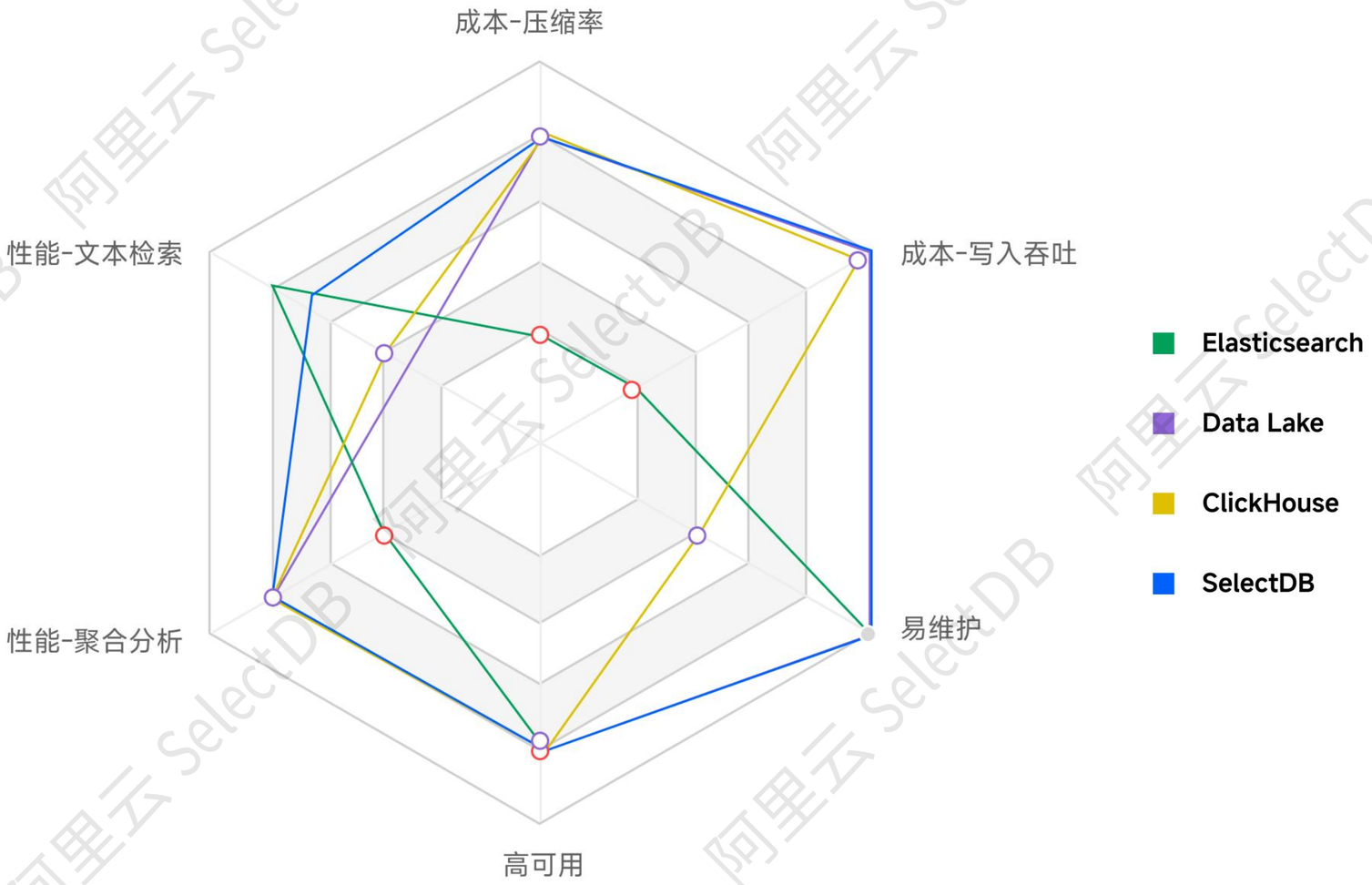
- 支持SSD HDD S3 HDFS 多种存储介质
- 定义好数据生命周期，自动热转冷、过期删除
- 没有写入两个系统开销、数据一致性问题

维护和使用简单

- 一套系统部署和数据处理流程
  - 维护一套存储系统和数据处理流程
  - 支持SSD、HDD、S3 三种存储介质
  - 定义好数据生命周期，自动热转冷、过期删除
- 一套查询接口和查询界面
  - 一个SQL查询接口，冷热数据查询方式一致
  - 可以跨冷热分界混合查询



多种方案对比



4 实践案例

实践案例1 — 移动互联网

“Log Trace 场景能够完全支持，标志着 Doris 几乎能扛住 **抖音集团** 绝大部分场景的导入性能需求”

集群规模：3000+ core

数据增量：每天新增8000亿条日志、500TB

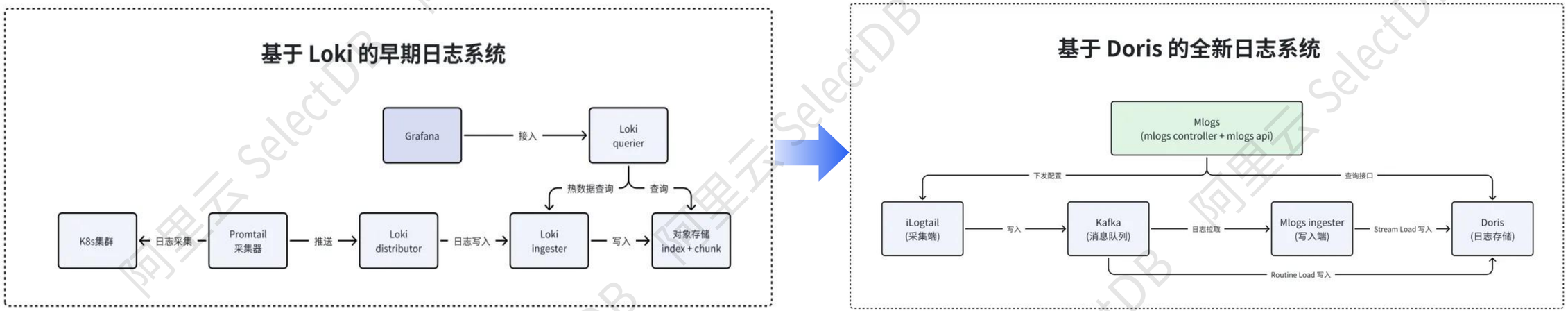
数据总量：总共7PB、24万亿条

写入性能：线上平均1000w/s、60GB/s

峰值1500w/s，90GB/s

实践案例2 — AI 大模型

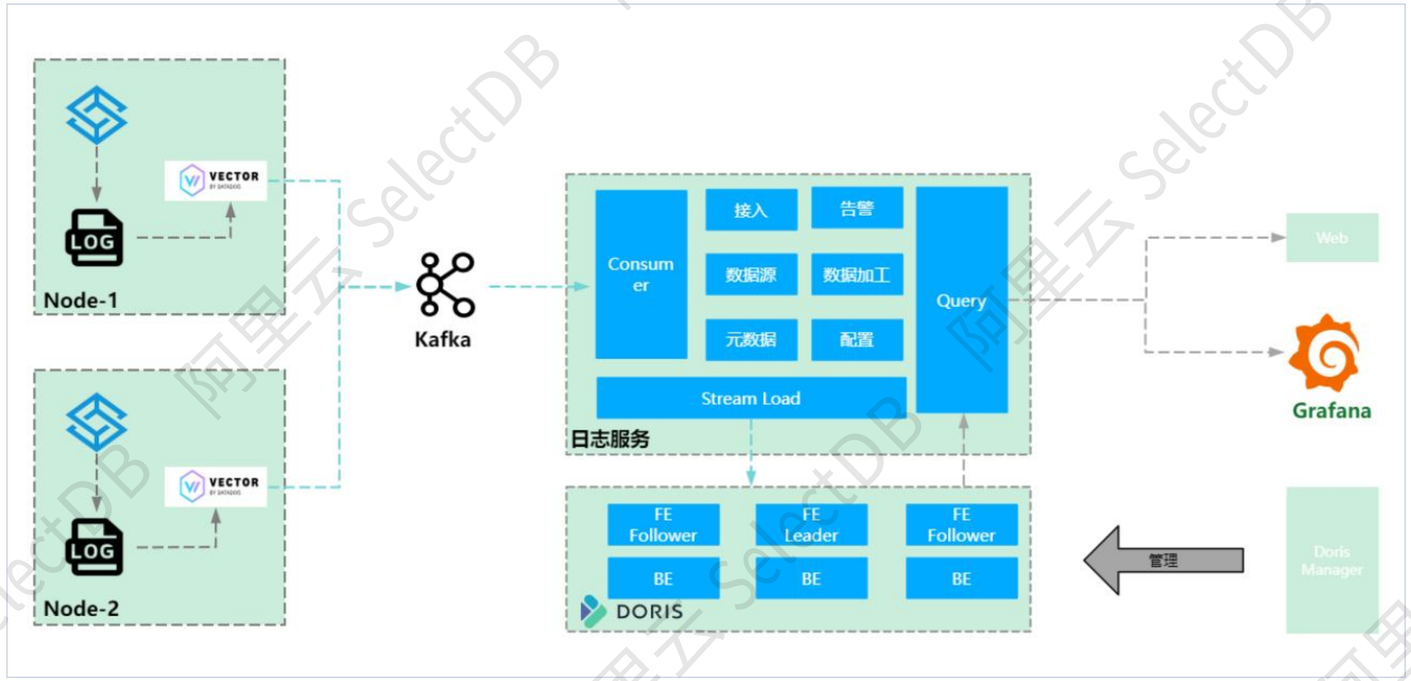
“基于 SelectDB 的新系统已接入 MiniMax 内部所有业务线日志数据，满足实时写入和查询的需求
通过存算分离比自建 Doris 计算资源降低 40%，热数据存储资源降低 50%”



数据规模：PB级　写入吞吐：10GB/s　查询：秒级响应　冷热分层：7天热30天冷

实践案例3 — AI 大模型

“科大讯飞 将可观测性存储底座从 ES Loki 升级到 Doris，成本降低 60%，性能提升 10倍
使用 VARIANT 类型存储 Log Trace 扩展字段”



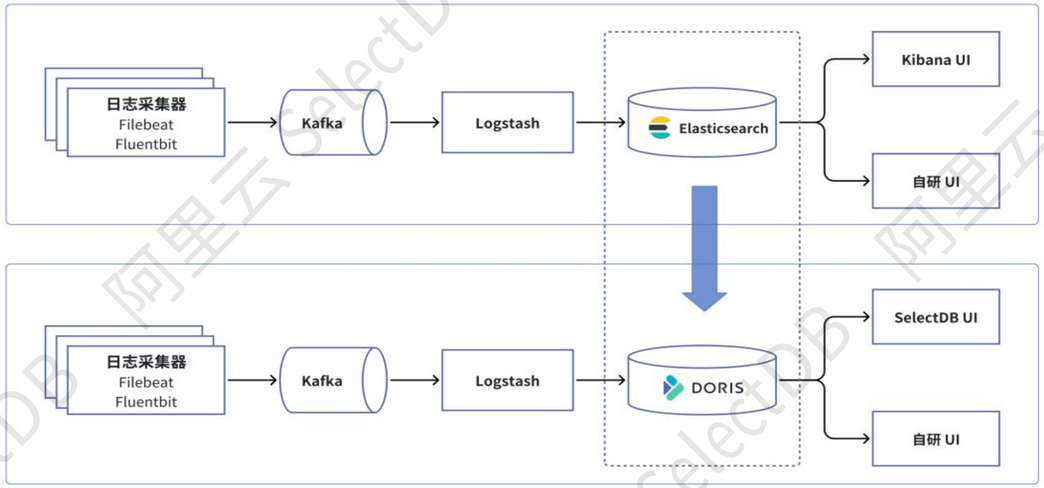
Log: 明细模型

Trace: 主键模型

Metrics: 聚合模型

实践案例4 — 金融

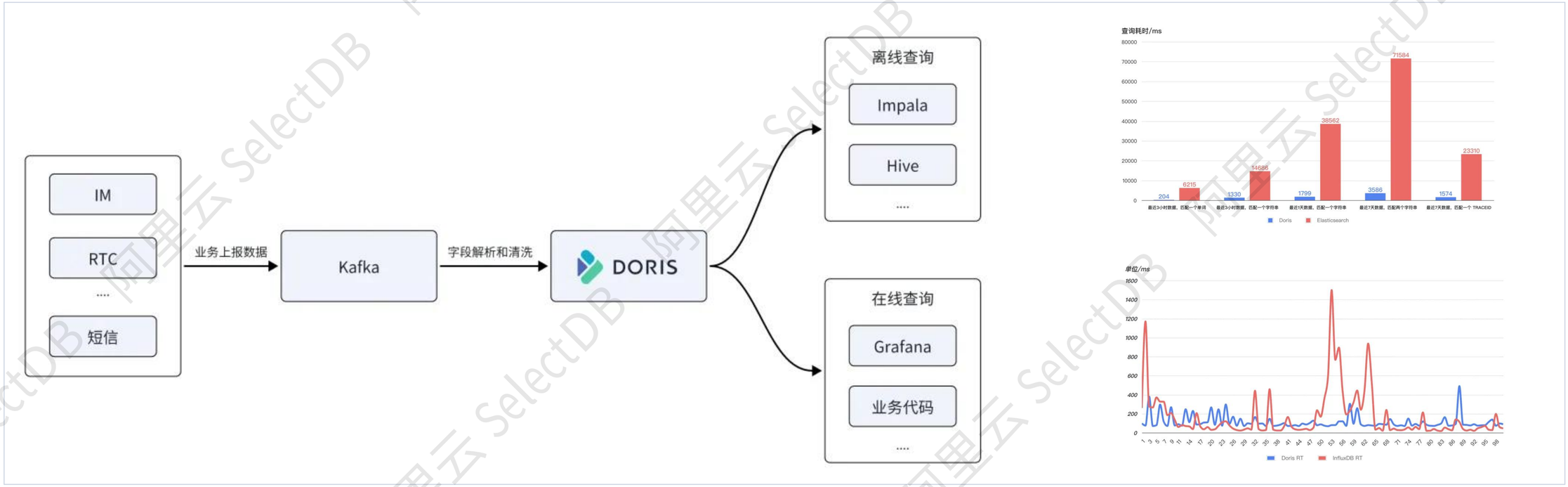
“中信信用卡中心 基于 Doris 构建 PB 级日志平台，比 ES 资源节省 50%，查询性能提升 2～4倍
实现 Log Trace 联动”



性能对比	组件	节点数量及规格	同等日志量下 占用磁盘空间	同等日志量下 TPS 写入峰值	同等日志量下 查询耗时
替换前	Elasticsearch	9数据节点 9×16C32G(HDD X86)	1.22TB/天	9.1k/s	499ms
替换后	Doris	4数据节点 4×8C32G	0.51TB/天	13.5k/s	308ms

实践案例5 — 互联网

“网易日志数据存储空间降低到 ES 的1/3，查询效率获得10倍提升，查询性能更加平稳
时序场景替代 InfluxDB, 服务器节省50%，存储空间降低67%”



实践案例6 — 可观测性 & 新能源汽车

“SelectDB 提供了灵活半结构化类型 VARIANT，成本相比云上ES节省70%，全文检索提升2-3倍”

数据增量：每天新增400亿条数据、80TB，SelectDB压缩后16TB（包括倒排索引，压缩率1:5）

数据总量：1副本保存30天，总共480TB、1.2万亿条

写入性能：线上平均50w/s、1GB/s，峰值100w/s，2GB/s，秒级实时写入，P90 查询延迟 200ms

更多案例

- MiniMax GenAI 可观测性分析：基于阿里云 SelectDB 构建 PB 级别日志系统
- 金融场景 PB 级大规模日志平台：中信银行信用卡中心从 Elasticsearch 到 Apache Doris
- 从 Elasticsearch 到 Apache Doris 统一搜索与分析引擎，腾讯音乐内容库升级实践
- 查询平均提速 700%，奇安信基于 Apache Doris 升级日志安全分析系统
- 从 Elasticsearch 到 SelectDB，观测云实现日志存储与分析的 10 倍性价比提升
- 从 Elasticsearch 到 Apache Doris，统一日志检索与报表分析，360 企业安全浏览器升级实践
- 查询提速11倍，资源节省 70%，Apache Doris 在网易日志和时序场景的落地实践
- 从 ClickHouse 到 Apache Doris：在网易云音乐日增万亿日志数据场景下的落地
- 从 Elasticsearch 到 Apache Doris，10 倍性价比的新一代日志存储分析平台
- 查询性能提升10 倍、存储空间节省 65%，Apache Doris 半结构化数据分析方案及典型场景
- 为什么 Apache Doris 是比 Elasticsearch 更好的实时分析替代方案？

阿里云

扫码加入阿里云 SelectDB 产品交流群