

奇安信基于 Apache Doris 的 日志安全分析系统升级实践

舒鹏

奇安信 服务端技术专家

目录

- 1 公司介绍与业务背景
- 2 架构升级之旅
- 3 Apache Doris 2.0 查询提速实践经验
- 4 规划与展望

1 公司介绍与业务背景

奇安信科技集团股份有限公司



中国企业级网络安全市场的领军者

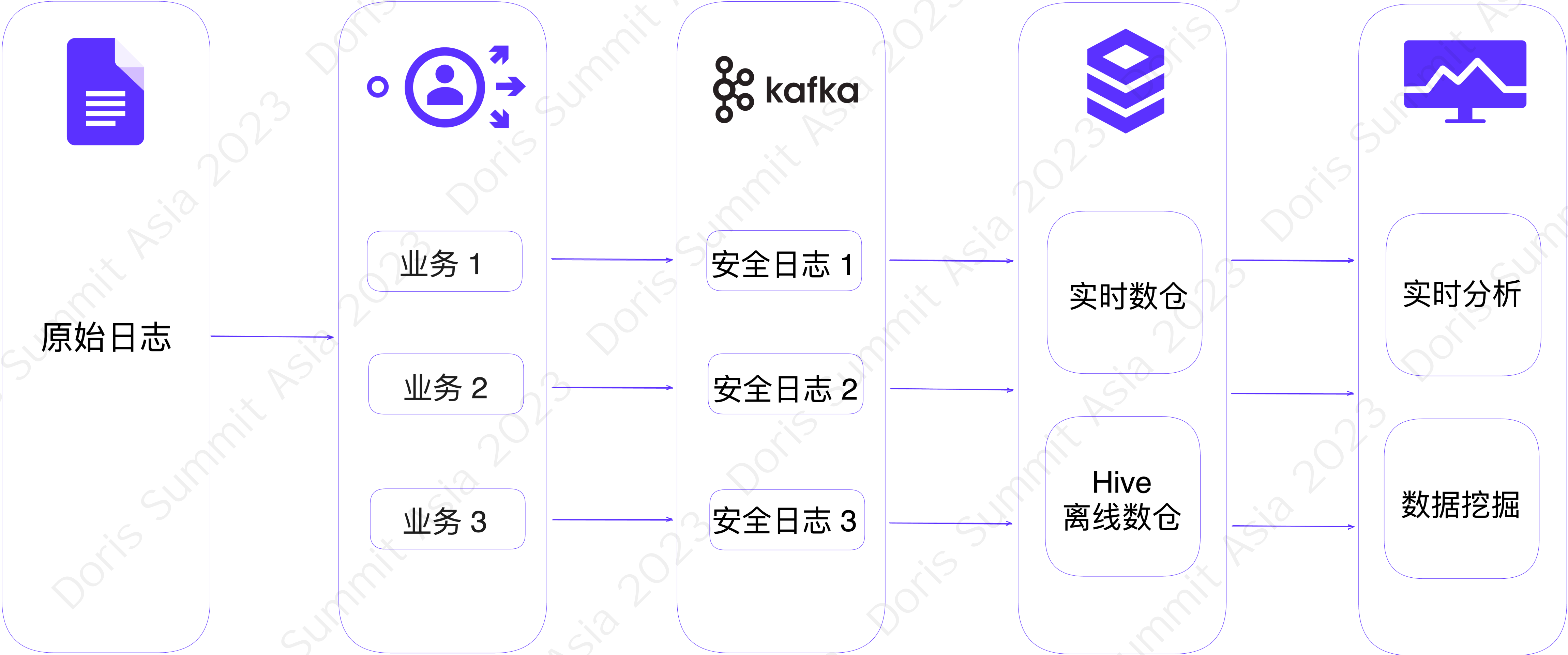
专注于为政府和企业用户提供新一代网络安全产品和服务。

2023 年，奇安信在数据安全领域全面发力。目前核心产品天擎终端安全系统在国内已有 4000 万政企用户部署、全国部署服务器超过 100 万台、服务超 40 万大型机构。

作为网络安全国家队，奇安信立志为国家构建安全的网络空间，在终端安全、云安全、威胁情报、态势感知等领域的技术研发持续领先。

“十四五”规划开局起步，数字化转型全面铺开推动网络安全需求井喷，奇安信将继续为政府与企业等用户提供全面、有效的网络安全解决方案，向成为“全球第一的网络安全公司”的愿景目标不断奋进

原日志安全分析系统



架构主要痛点

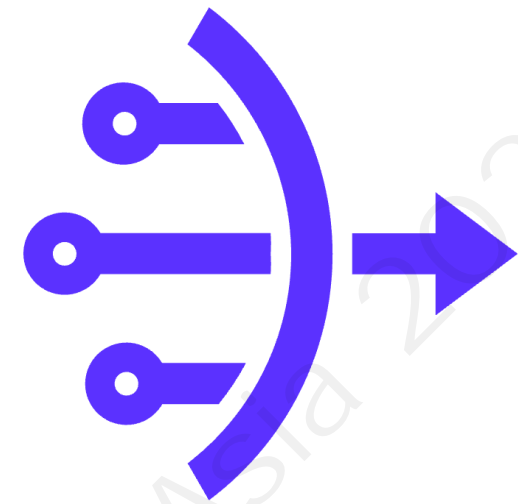
写入效率

- 每天所生产的安全日志数据达到千亿级，且每天新增日志量在不断增长中。
- 原架构系统入库速率逐渐降低，集群写入压力过大、高峰期数据挤压严重、稳定性造成影响
- 集群多次扩容（3 节点至 13 节点），机器成本已超预期的情况下，写入效率没有得到改善

查询性能

- 原系统只能通过 SQL LIKE 进行全量扫描和暴力匹配。
- 原系统不支持对文本字段，如 URL、payload 等关键字进行模糊匹配
- 千亿级数据量的查询耗时近分钟级，当遇到并发查询，性能还会进一步恶化

系统建设目标



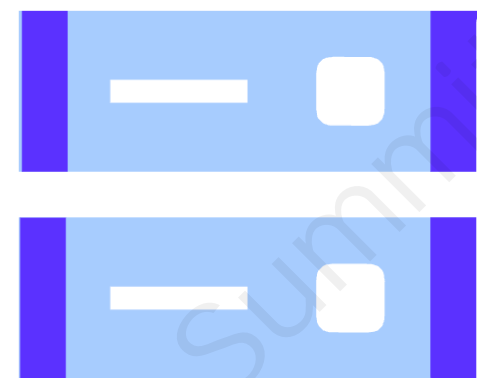
实时写入性能

- 海量病毒查杀事件实时写入与存储
- 基于日志数据 Schema Free 特性支持丰富数据类型的写入与变更



倒排索引 & 模糊查询

- 支持对字符串提供模糊查询的能力
- 能够灵活创建且类型丰富的索引，例如倒排索引以加速筛选过滤数据



高性价比存储成本

日志数据价值密度低，但存储的规模很大、存储周期相对较长



简易运维 & 高效管控工具

- 系统自身的运维简易程度
- 系统是否具备合适的管控工具

2 架构升级之旅

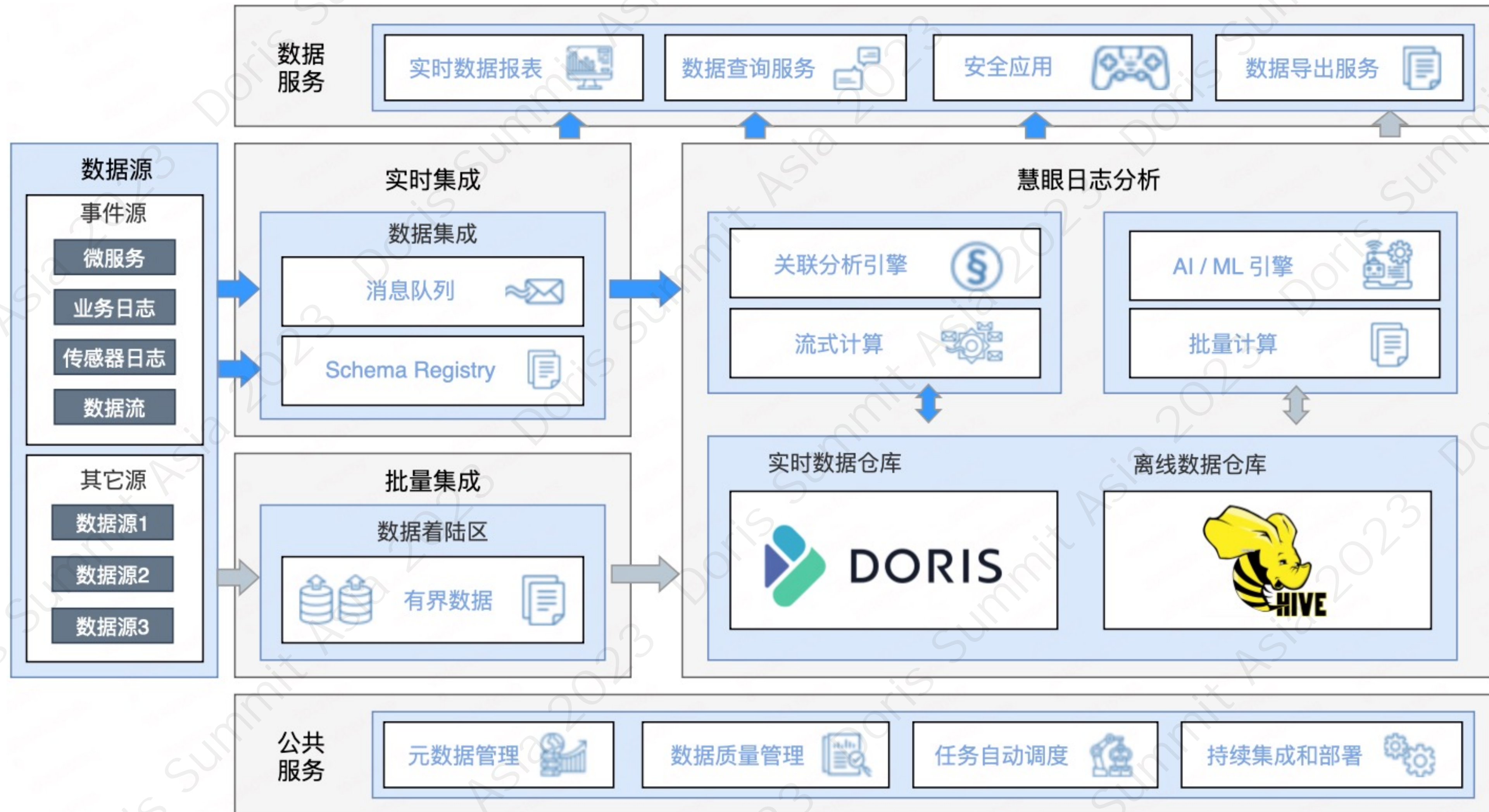
为什么选择 Apache Doris

主要研究探索 OLAP 数据库中擅长日志分析系统的组件，构建一体化日志存储分析平台，赋能网络安全。

选择 Doris 的理由：

1. 支持标准 SQL 语法，与 MySQL 高度兼容；
2. 具备倒排索引、Ngram BloomFilter 索引等检索特性；
3. 支持聚合、多表 Join、子查询、窗口函数、UDF、物化视图等功能；
4. 能够在线毫秒级删减字段、按需增减索引、按需更改类型；
5. 支持 Text、JSON、Array、Map、Variant 等多种数据类型；
6. 采用 ZSTD 压缩算法优化存储空间占用；

全新日志安全系统架构



写入与查询性能提升

写入性能提升

200%

原系统写入吞吐

- 13台服务器
- CPU 利用率 30%
- 写入吞吐 30 万/条

Apache Doris 写入吞吐

- 3台服务器
- CPU 利用率 100%
- 写入吞吐 108 万/条

查询性能提升

690%

原系统查询耗时

- 79 条 SQL 查询语句
- 整体耗时 1757.32s

Apache Doris 查询耗时

- 79 条 SQL 查询语句
- 整体耗时 253.12s

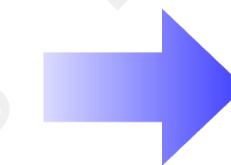
存储成本降低

存储成本降低

40%

1. 数据压缩比

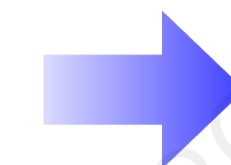
1: 4.3



1: 5.7

2. 数据膨胀率

3-5 倍



1 倍

ZSTD 是一个优秀的新型压缩算法，使用了智能优化算法，相较于常见的 GZIP 算法，ZSTD 具有**更高的压缩率更和更快的压缩解压速度**，尤其在处理日志场景时表现非常出色。

运维与管控成本降低

Cluser Manager For Apache Doris

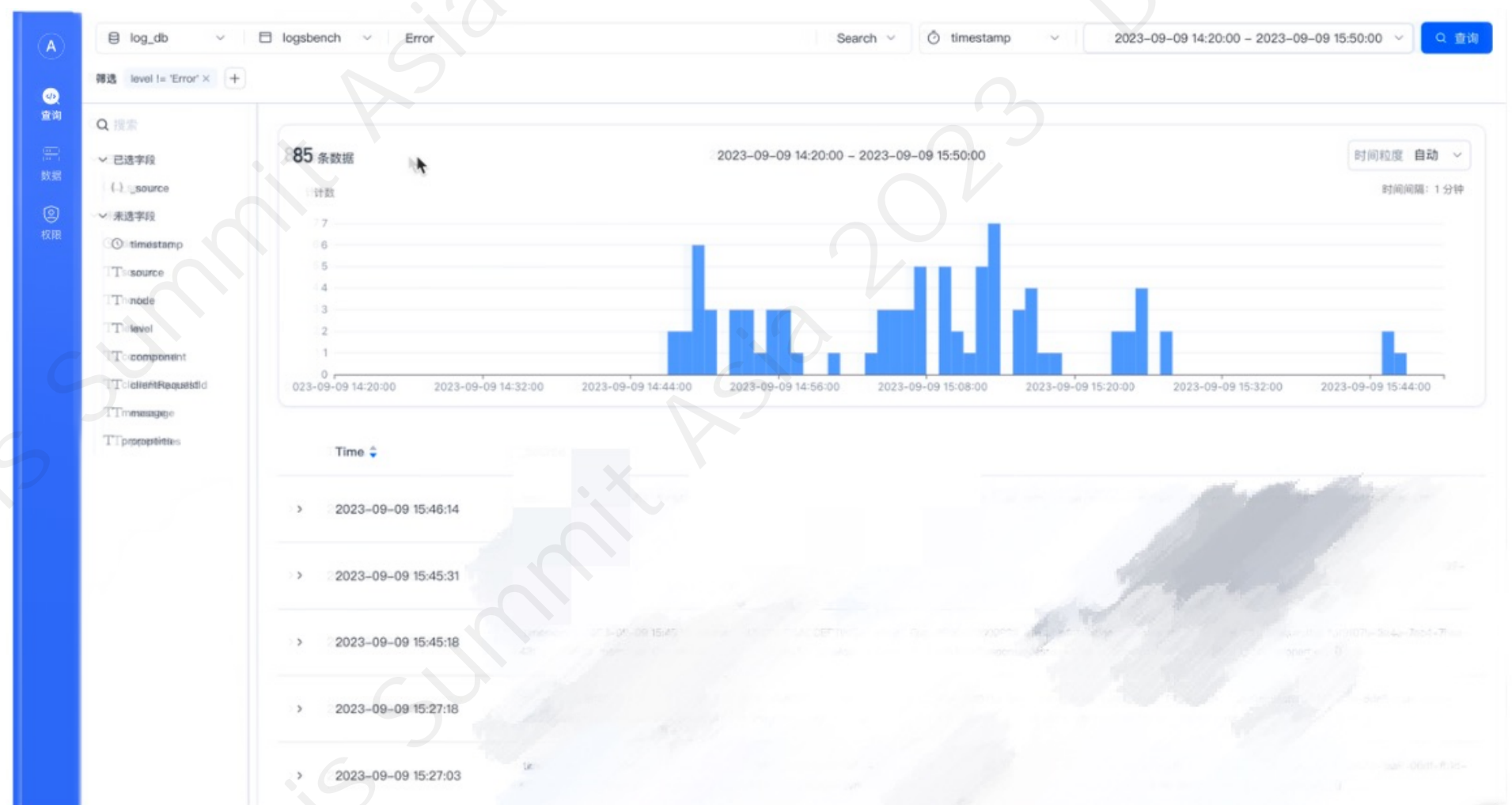
由飞轮科技免费开放

可视化集群管理工具

满足日常运维中集群监控、巡检、修改配置、扩缩容、升级等操作。
降低登陆机器手动操作的麻烦和误操作风险。

可视化日志探索分析 WebUI

支持关键词检索、趋势图展示、趋势图拖拽日期范围、明细日志平铺和折叠展示、字段值过滤等交互方便的探索式分析。
对于习惯 ELK 日志分析的用户非常友好，契合日志场景探索下钻分析需求。



4 Apache Doris 2.0 查询提速实践经验

SQL查询对比

线上去重分析 79 条 SQL，在同一天 1000 亿条总数据、同样 10 BE 节点集群规模上对比测试查询耗时。

所有查询语句均有明显提升，整体性能查询提升近 7 倍，26 条 SQL 查询语句性能提升 10 倍以上：

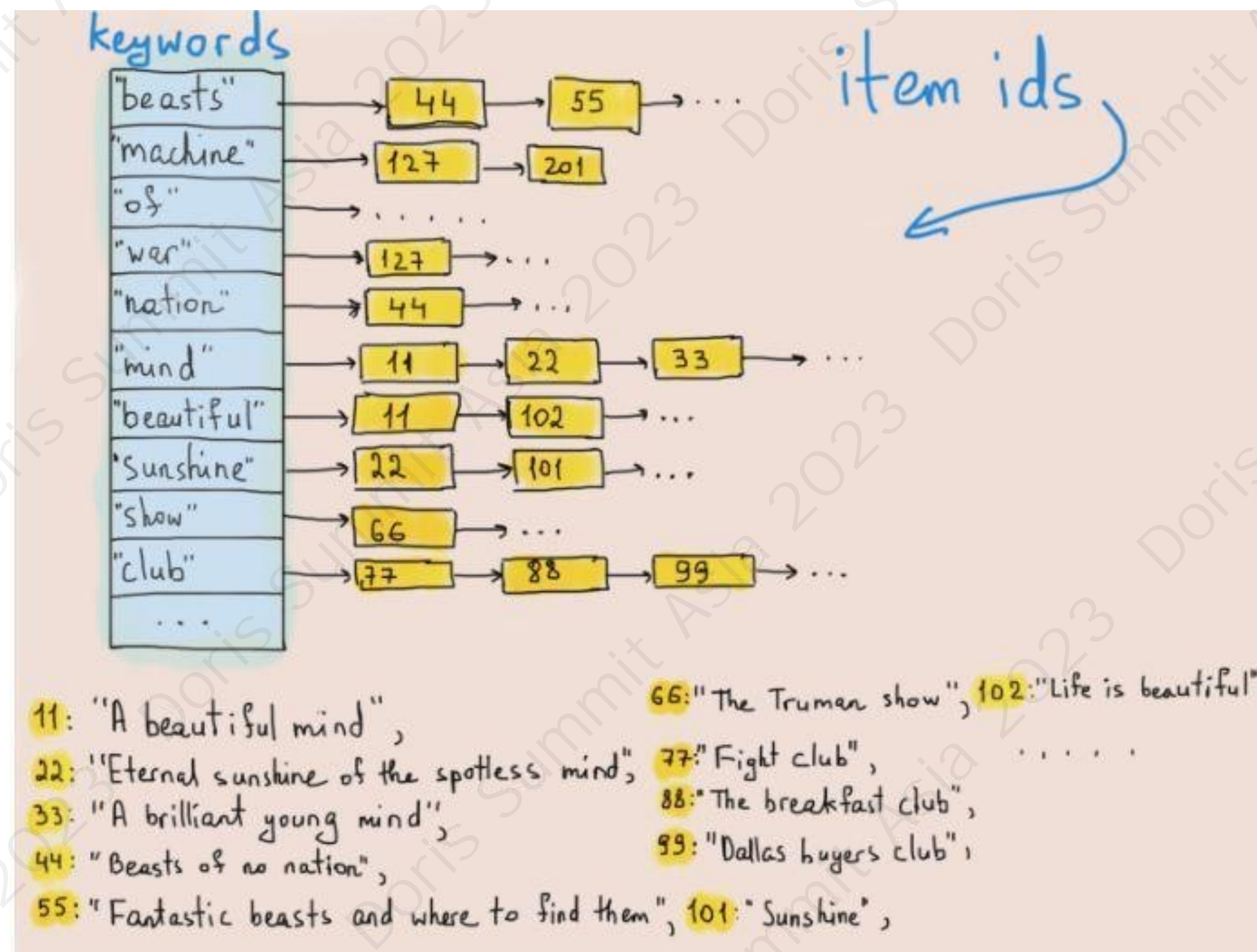
- 8 条 SQL 查询提升 10-20 倍
- 14 条 SQL 查询提升 20-50 倍
- 4 条 SQL 查询提升 50 倍以上
- 其中差异最高达 88 倍，在原系统执行接近 1 分钟，在 Apache Doris 中仅需不到 1 秒

Query	原系统耗时 (s)	Apache Doris 耗时 (s)	性能提升倍数
1	13.83	5.29	2.6
2	6.05	0.65	9.4
3	7.57	1.68	4.5
4	1.56	0.74	2.1
5	7.08	0.28	25.2
6	0.8	0.12	6.9
7	5.15	0.3	17.1
8	17.16	4.84	3.5
9	11.89	3.13	3.8
10	16.14	5.03	3.2
11	16.22	5.46	3
12	3.64	1.39	2.6
13	9.4	1.49	6.3
14	6.4	1.05	6.1
15	7.12	0.89	8
16	2.42	0.55	4.4
17	3.02	0.89	3.4
18	400.43	44.69	9
19	12.22	1.51	8.1
20	3.82	1.23	3.1
21	391.74	38.52	10.2
22	26.89	0.53	50.3
23	12.31	0.2	62
24	24.93	1.12	22.2
25	59.21	12.54	4.7
26	3.98	1.12	3.5
27	3.46	0.34	10.1
28	3.05	0.1	31.6
29	3.88	1.89	2.1
30	5.78	0.14	42
31	3.65	0.1	35.6
32	14.54	6.12	2.4
33	5.15	0.66	7.8
34	5.26	1.03	5.1
35	16.32	1.91	8.6
36	19.8	2.55	7.8
37	23.18	15.8	1.5
38	3.36	0.32	10.6
39	4.77	0.12	38.5
40	2.9	0.31	9.4

Query	原系统耗时 (s)	Apache Doris 耗时 (s)	性能提升倍数
41	8.84	2.02	4.4
42	4.87	0.23	21
43	58.66	0.67	88.2
44	34.61	0.78	44.4
45	7.63	1.63	4.7
46	5.45	1.54	3.5
47	5.79	1.53	3.8
48	2.33	0.53	4.4
49	61.77	18.38	3.4
50	25.18	0.74	33.9
51	40.06	15.64	2.6
52	27.55	0.54	51.3
53	7.69	0.61	12.6
54	5.55	0.54	10.3
55	5.12	0.33	15.4
56	33.62	12.48	2.7
57	9.5	0.6	15.7
58	5.64	0.78	7.2
59	5.91	0.97	6.1
60	7.54	3.92	1.9
61	8.7	1.08	8.1
62	6.52	1.02	6.4
63	7.03	1.12	6.3
64	6.4	3.29	1.9
65	5.27	3.26	1.6
66	6.13	0.69	8.8
67	4.74	1.05	4.5
68	4.62	0.8	5.8
69	4.9	1.03	4.7
70	8.08	0.93	8.6
71	4.45	0.15	30.6
72	4.65	0.72	6.4
73	3.88	0.76	5.1
74	3.16	0.76	4.2
75	24.96	0.56	44.4
76	23.15	0.69	33.6
77	40.29	1.94	20.8
78	30.78	0.84	36.4
79	10.27	1.33	7.7
Total	1,757.32	253.12	6.9

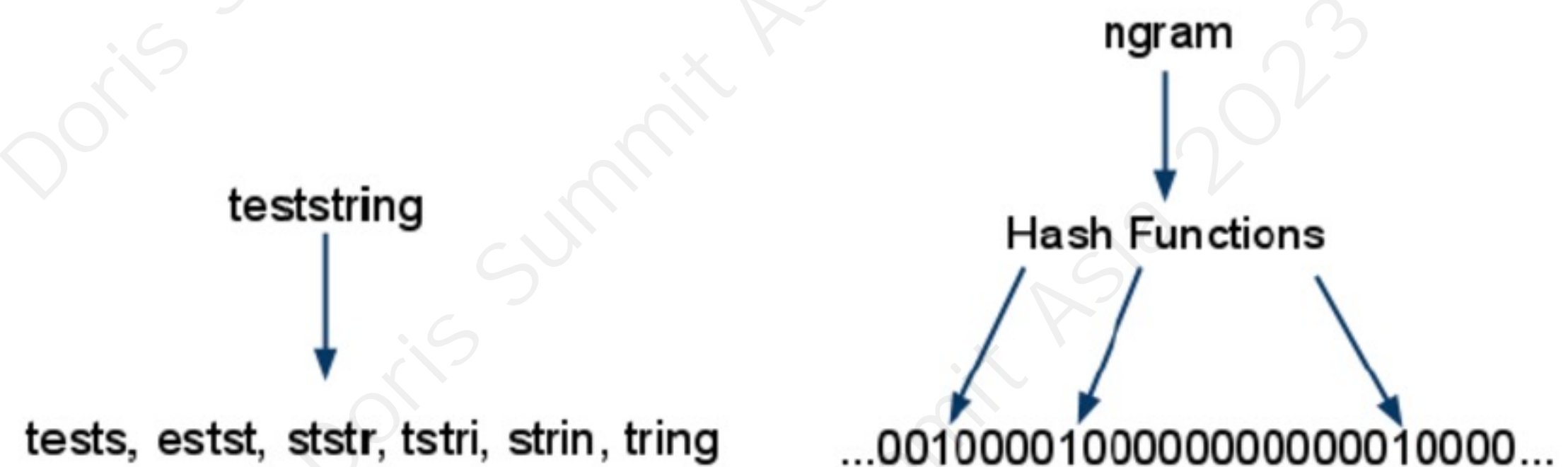
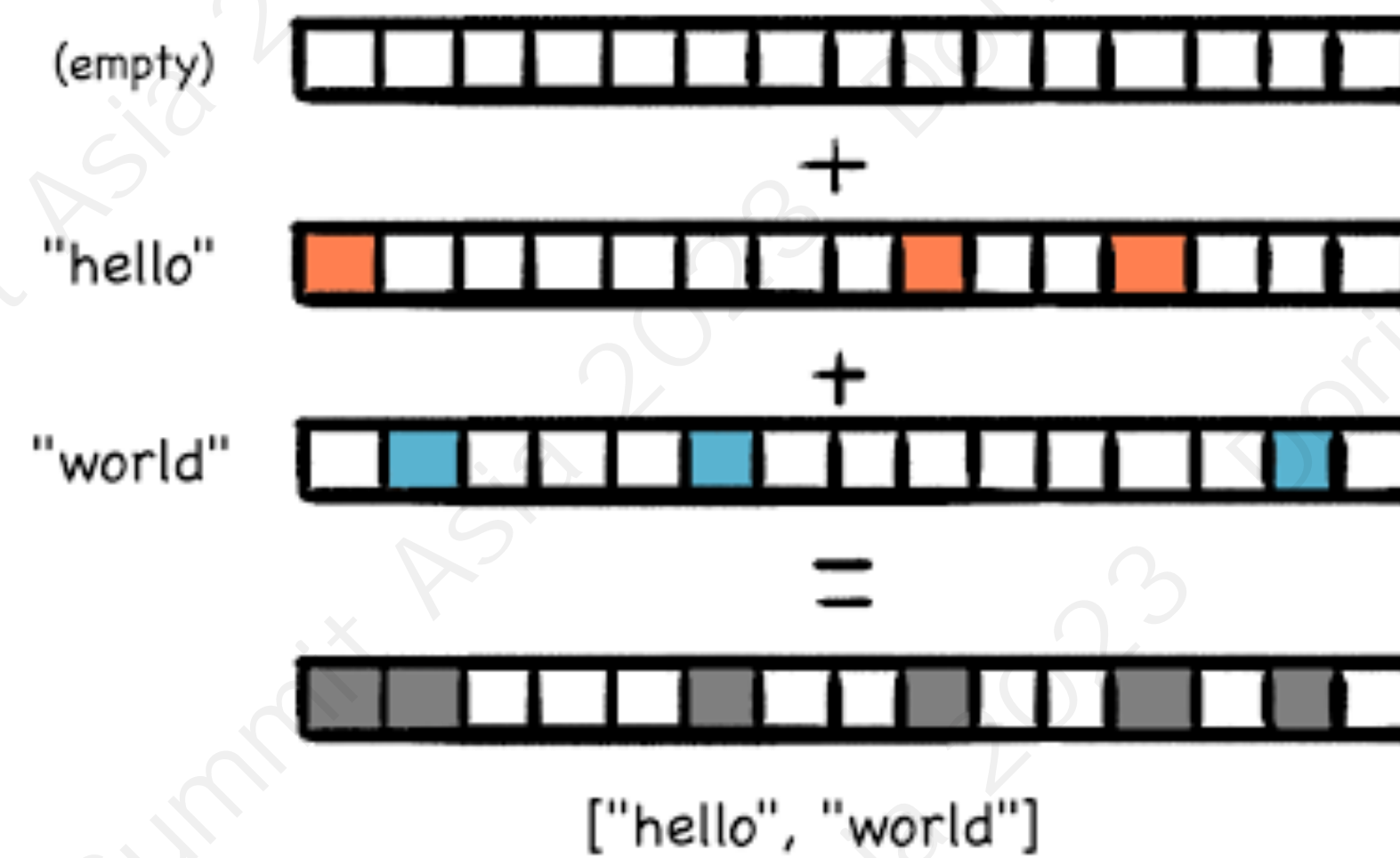
倒排索引 — 对关键词查找加速

-- 例如 Q43 提升 88.2 倍



NGram BloomFilter 索引 — 对 LIKE 加速

-- 例如 Q75 提升 44.4 倍



TopN — 日志明细查询优化

-- 例如 Q22, 提升50.3倍



5 规划与展望

扩大 JSON 数据类型的应用



Array

Object

Variant

Boolean

Number

引入 Variant 可变数据类型

- 支持存储任意结构的 JSON 数据
- 支持字段个数与类型变化
- 更灵活地定义特殊字符，更好地实现半结构数据 Schema Free 分析需求



获取更多社区动态与最佳实践

Apache Doris 官方平台:

- Apache Doris 官网: doris.apache.org
- Apache Doris GitHub: github.com/apache/doris/

获取更多峰会资料:

- Doris Summit 峰会官网: doris-summit.org.cn
- Doris Summit 峰会回放: <https://space.bilibili.com/1196172099/channel/collectiondetail?sid=1824324>