

Elasticsearch 在企业协作服务中的应用实践





我是谁

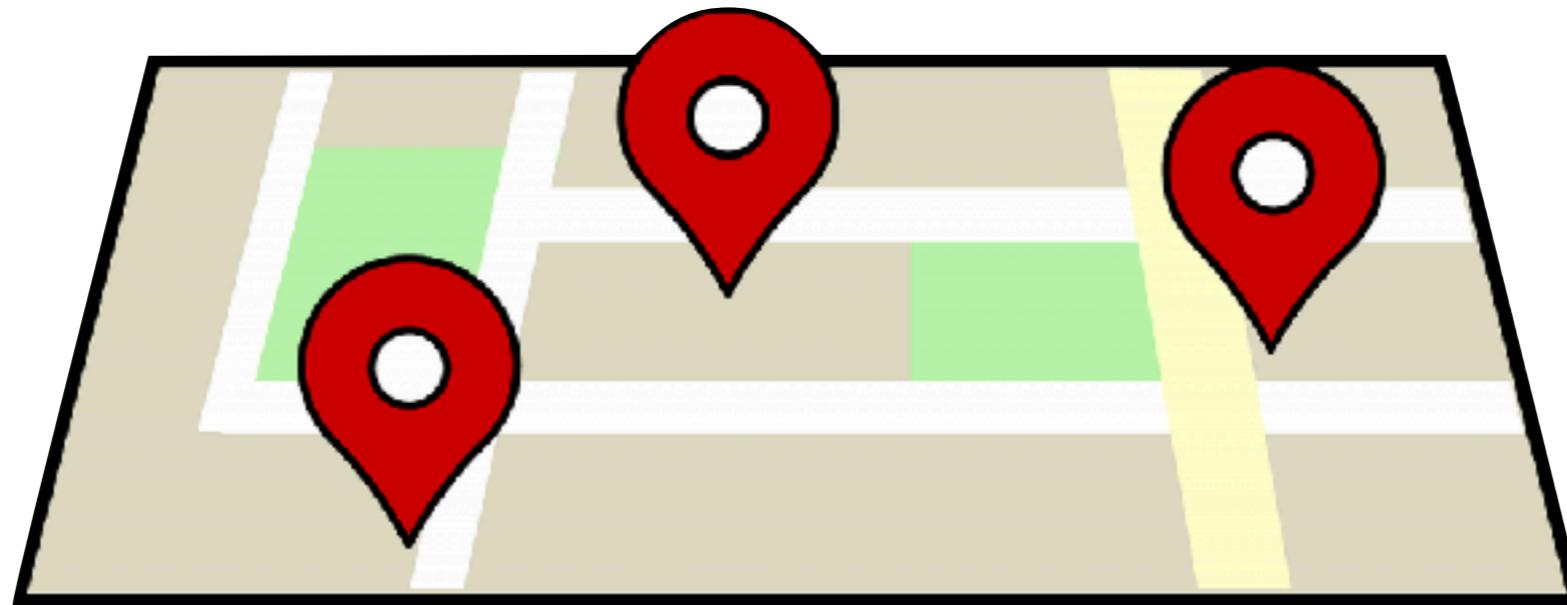
Who am I ?

- 宓文捷
- github.com/orangemi
- weixin: orangemi
- 一个喜欢研究各种技术的谷粉



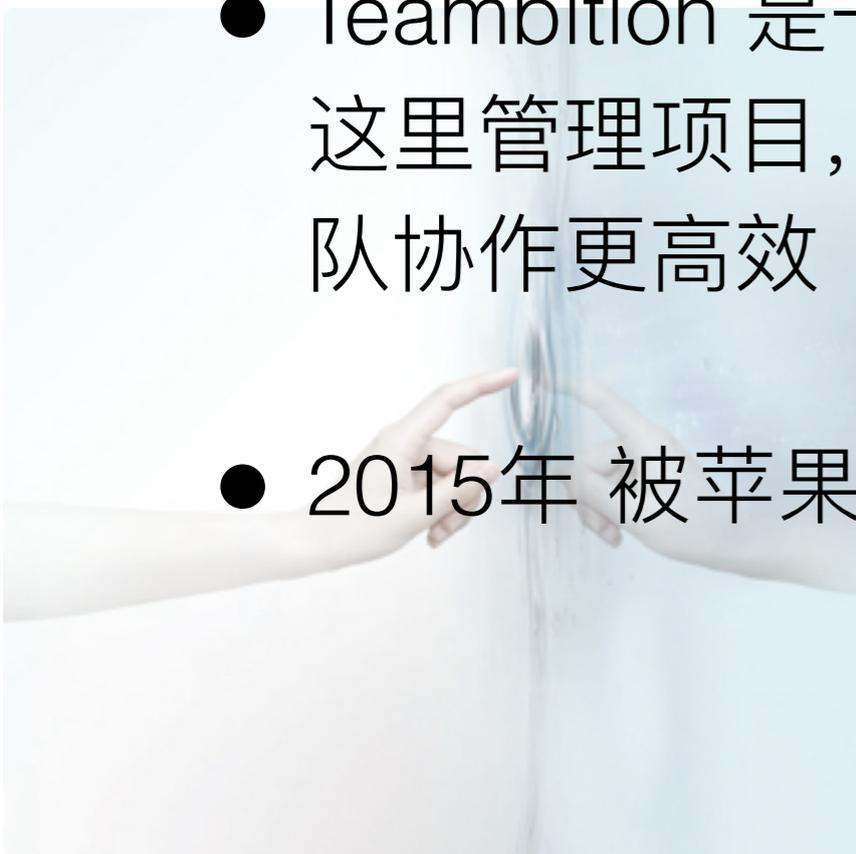
我来自哪里

Where do I come from ?



teambition

- <https://www.teambition.com>
- 团队协作工具创导者
- Teambition 是一个简单，高效的项目协作工具，你可以在这里管理项目，跟踪任务进度，存储项目文件，让你的团队协作更高效
- 2015年 被苹果公司评为年度最佳应用



用户情景

搜索机票

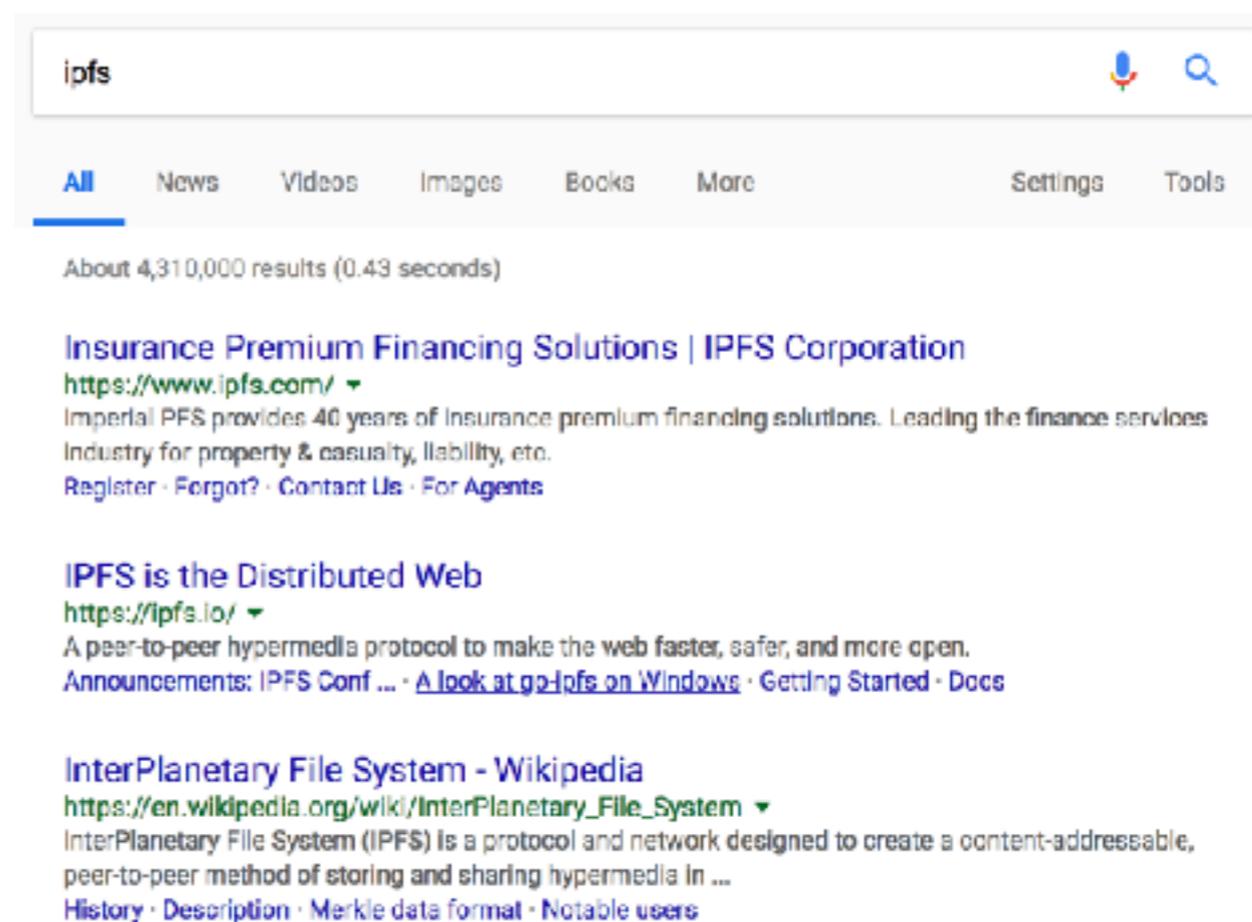
往返 出发城市 法兰克福(FRA) 到达城市 上海(SHA) 出发日期 2018-06-20 明天 返回日期 2018-06-27 星期三

成人 1 儿童 0 婴儿 0 舱位等级 经济/超级经济舱 航空公司 (选填)中文/代码

- 数据：机票数据
- 条件：航班，起飞降落，时间范围，价格范围，国内/国际
- 排序：时间，价格，航空公司
- 聚合：无

搜索文献

- 数据：文献数据
- 条件：包含文字，作者
- 排序：相关度，偶尔有时间，作者
- 聚合：作者，标签



业务价值&产品需求

- 大海捞针，获得洞察
- 易用的任务筛选需求、降低任务管理维护成本
- 精细的任务导出需求，低成本关注成员进展
- 跨项目任务汇总需求，商业报告应用支撑

搜索任务

- 数据：任务
- 条件：包含文字，截止时间，参与人员，评论内容，自定义字段（时间字段，文本字段，数字字段）
- 排序：相关度，时间，数字字段，自定义字段
- 聚合：按时间维度，按数字范围，按人员，按自定义字段

什么是任务

The screenshot shows a task management interface with several red annotations:

- Fixed Fields:** A red box highlights the task title "已付费的企业通过批量导入新员工不成功, 控制台有报错", the assignee "Not Assigned", and the date range "Start Date - Due Date".
- Custom Fields:** A red box highlights the "Note" field containing text like "企业名称: Facevisa" and "企业 ID: 5b19e34b42fcd70001a45689".
- Comment Dynamic:** A red box highlights the bottom comment input area with a "Reply" button.

固定字段:
人员, 时间, 文本

自定义字段:
文本, 选项, 时间, 数字

隐藏字段: 权限

评论动态

A snippet of a "Followers" list showing 4 users with profile pictures and a plus sign for more.

筛选条件

Task

Matching range

- Title
- Note
- Comments

Project

All

Followers

All

Start Date

All

Due Date

All

Creation Time

All

Clear Filter

固定字段搜索条件

Due Date

Last 7 days

Creation Time

From: 6/13

Last 7 days

Last month

Last 3 months

Custom

2018.06.13 - Not set

2018-06-20

June 2018

Su	Mo	Tu	We	Th	Fr	Sa
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

清除 确认

时间字段过滤器

Search

是否需要数据同步 / 别组测试
统一的客户语言

实际工时
链车金服实施项目-模板项目

计划完成时间
人民日报

预计成交金额 (元)
SMB · Shenzhen

Customer Segment
Farming Projects

时长
运动会

+ Add Filter

按照自定义字段搜索
可能有文本，数字，日期

问题

Problems

梳理产品需求，把握需求难点

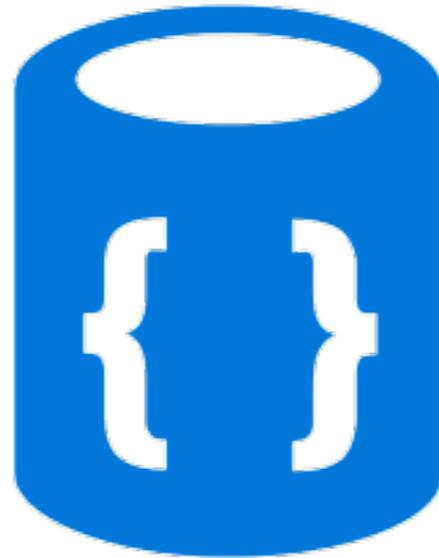
权限

Privileges

- 不可以访问的任务
- 可以访问的任务
- 管理员
- (允许用户建立自己的权限模型)
- 权限模型变化的数据结构不太容易处理



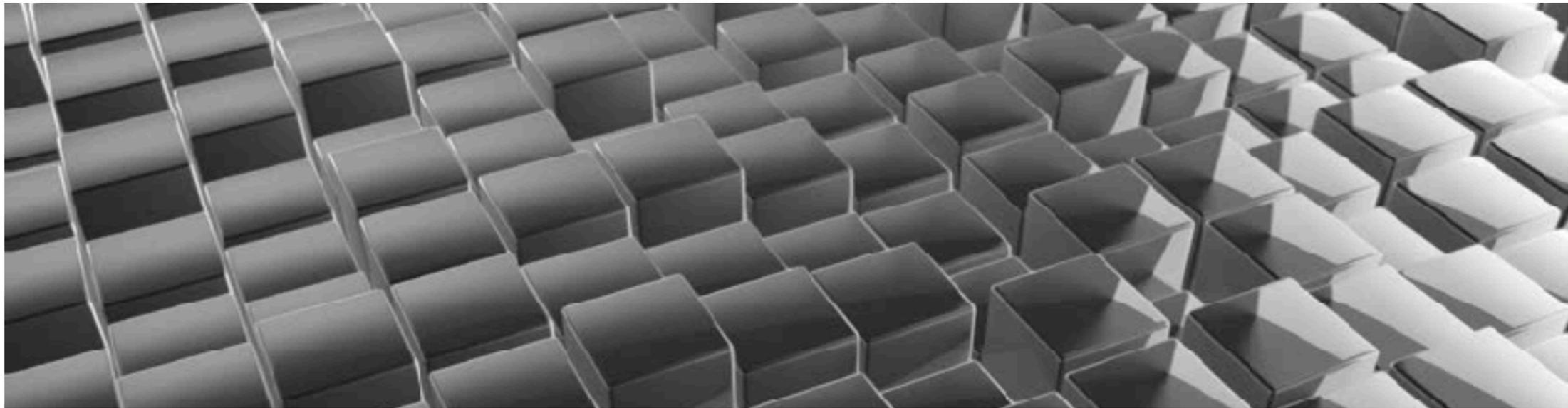
Schemaless Database



- Teambition的搜索 = 固定字段 + 自定义字段。
- 用户在任务管理系统里创建的内容几乎是无规则的

数据量

- > 80M Documents
- > 10k Fields



解决方案

Solutions

梳理产品需求，把握需求难点，难点转化操作

(存哪里)

存什么

怎么存

怎么查



elasticsearch

- Elasticsearch是一个高度可扩展开源全文搜索和分析引擎，它允许用户快速，实时地存储，搜索和分析大量数据。
- ES通常用作具有复杂的搜索功能和要求的应用程序提供的底层引擎/技术。

数据模型

Model

固定字段

```
1  ---
2  # task mapping
3  properties:
4  |   title:
5  |     type: text
6  |   content:
7  |     type: text
8
```

自定义字段

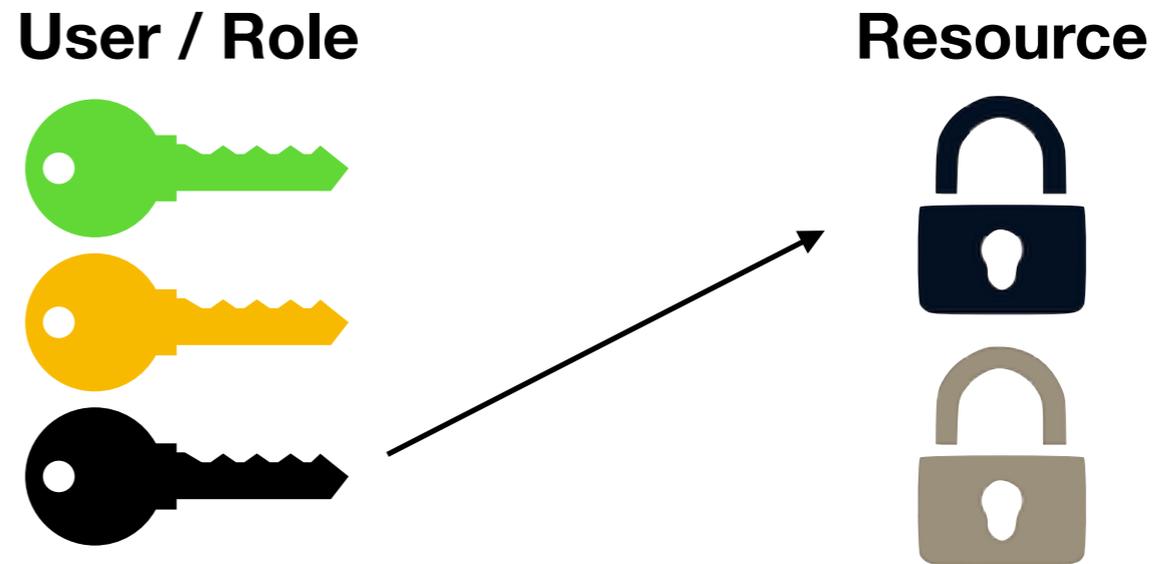
fid	string	date	keyword	number
1	abc			
2			ae29c	
3		2017-05-04		
4				10.4

```
17 # any extendable field should
18 cf:
19   type: nested
20   properties:
21     fid:
22       type: keyword
23     string:
24       type: text
25     number:
26       type: double
27     date:
28       type: date
29     keyword:
30       type: keyword
31     boolean:
32       type: boolean
```

- Nested Object
- 不同类型分别存
- 性能问题?

权限模型

- 用户持钥匙
- 资源带有锁
- 任意把对上
- 就是有权限



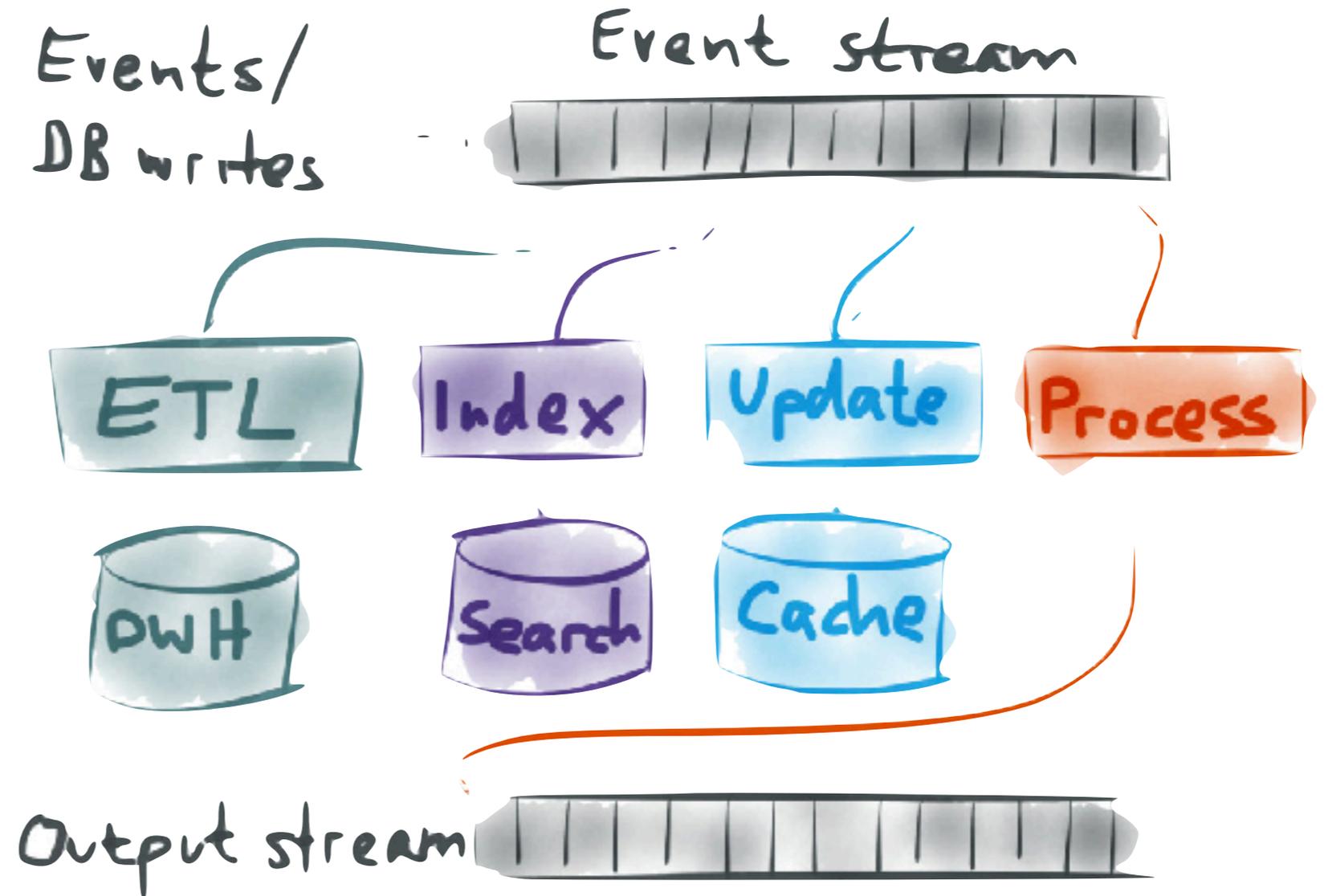
```
9      # for privileges
10     involveMembers:
11     |   type: keyword
12     isOnlyInvolveVisible:
13     |   type: boolean
14     projectId:
15     |   type: keyword
```

数据同步

Data Streaming

Event Sourcing

- Command and Query Responsibility Segregation
- 根据事件，获取数据当前状态(扩展字段)
- 保证最终一致性
- Kafka / RabbitMQ
- binlog / oplog

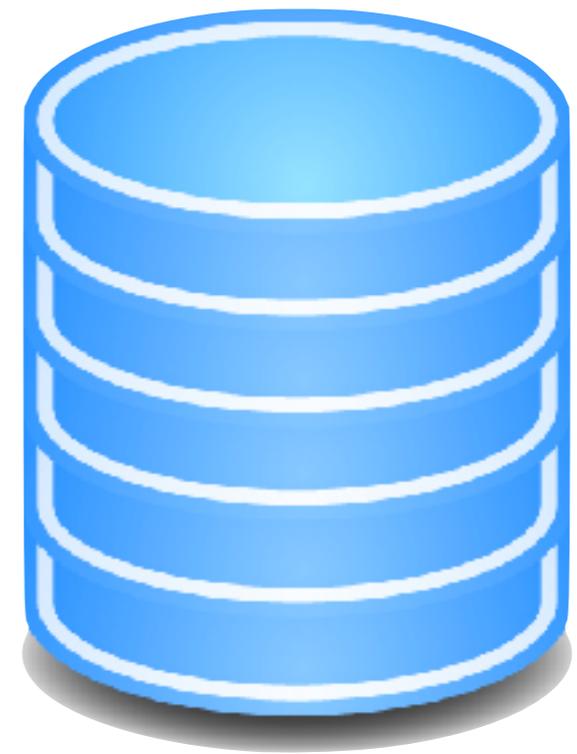


查询

Query

TQL的建立

- Teambition Query Language
- SQL / JQL / Elasticsearch SQL
- 简化和优化
- 开发喜欢，甚至直接给用户

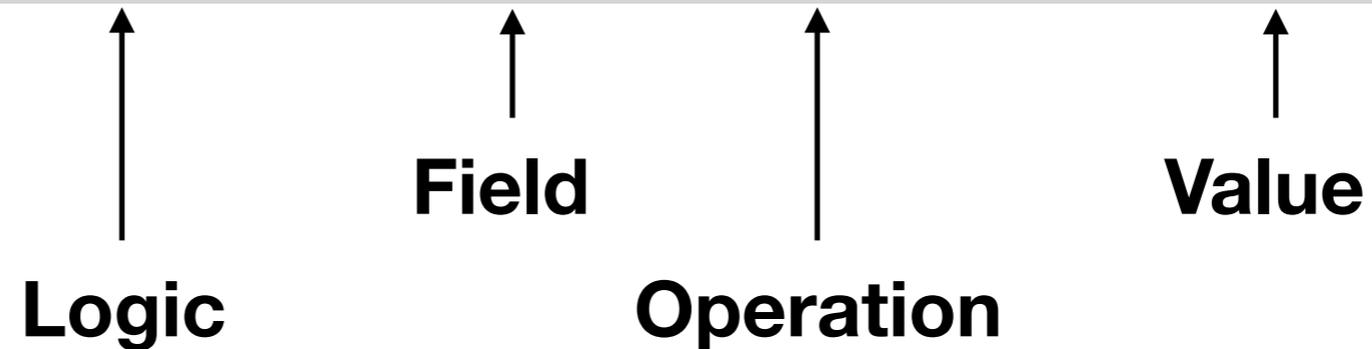


看一些例子

- `dueDate < 7d`
- `dueDate < 7d AND executorId = currentUser()`
- `cf:阶段 IN (TODO, DOING)`
- `text ~ 关键词`
- `GROUP SUM(storyPoint) BY executorId`

TQL的组成

text ~ 搜索 AND executorId = currentUser()



- 条件表达式 [conditionExpr]
- 聚合表达式 [groupExpr], 排序表达式 [orderExpr]
- <https://orangemi.github.io/teambition-tql/tql.html>

TQL与SQL的差异

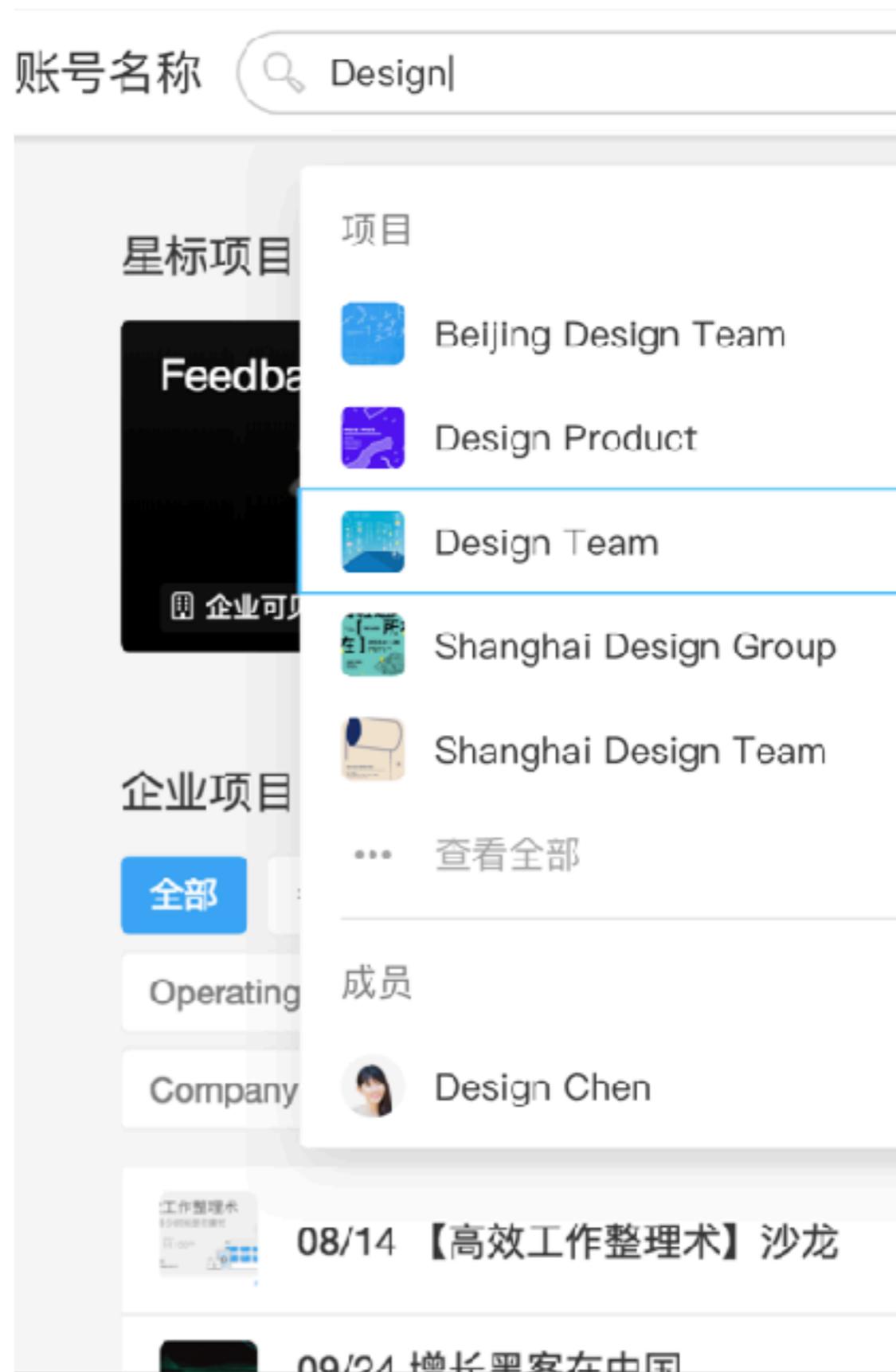
- 没有SELECT，结果只返回数据ID，明确服务边界
- Logic语法中没有NOT前缀，但会出现在操作符中间，更接近自然语言语意
- 输入符号是不友好的，没必要时不需要引号
- 优化日期表达式，接近自然语义

业务应用

Usage

快速搜索

- 优先显示项目，成员
- 优先搜索任务ID
- 分步加载任务列表
- 拼音搜索模糊匹配



36256 results | Default | Export **Pro** | Task

- 任务-任务列表** 测试用例库
前置条件 项目中有多个任务列表
 操作步骤 1.查看任务列表 2.完成任务 3.点击任务 4.添加任务 5.移动任务 (拖...
- 任务-复制任务** 测试用例库
前置条件 进入任务详情, 打开菜单, 点击复制
 操作步骤 1.选择项目, 分组, 列表 2.选择「原任务接受...
- 任务-移动任务** 测试用例库
前置条件 进入任务详情
 操作步骤 1.点击任务标题上方分组和列表移动任务 (只支持项目内部移动) 2....
- 任务-任务详情 任务动态** 测试用例库
前置条件 进入任务详情 (以下操作都是有权限的操作, 相关权限看权限说明)
 操作步骤 1.查看任务动态
...
- 任务-任务详情 关联任务** 测试用例库
前置条件 进入任务详情 (以下操作都是有权限的操作, 相关权限看权限说明)
 操作步骤 1.关联
 预期结...
- 子任务任务列表独立** 测试用例库 
子任务支持不同任务列表 ## 现有规则 - 跨分组移动任务, 子任务跟随父任务移动并处于同一任务列表 - 跨项...

Matching range

- Title
- Note
- Comments

Project

All

Followers

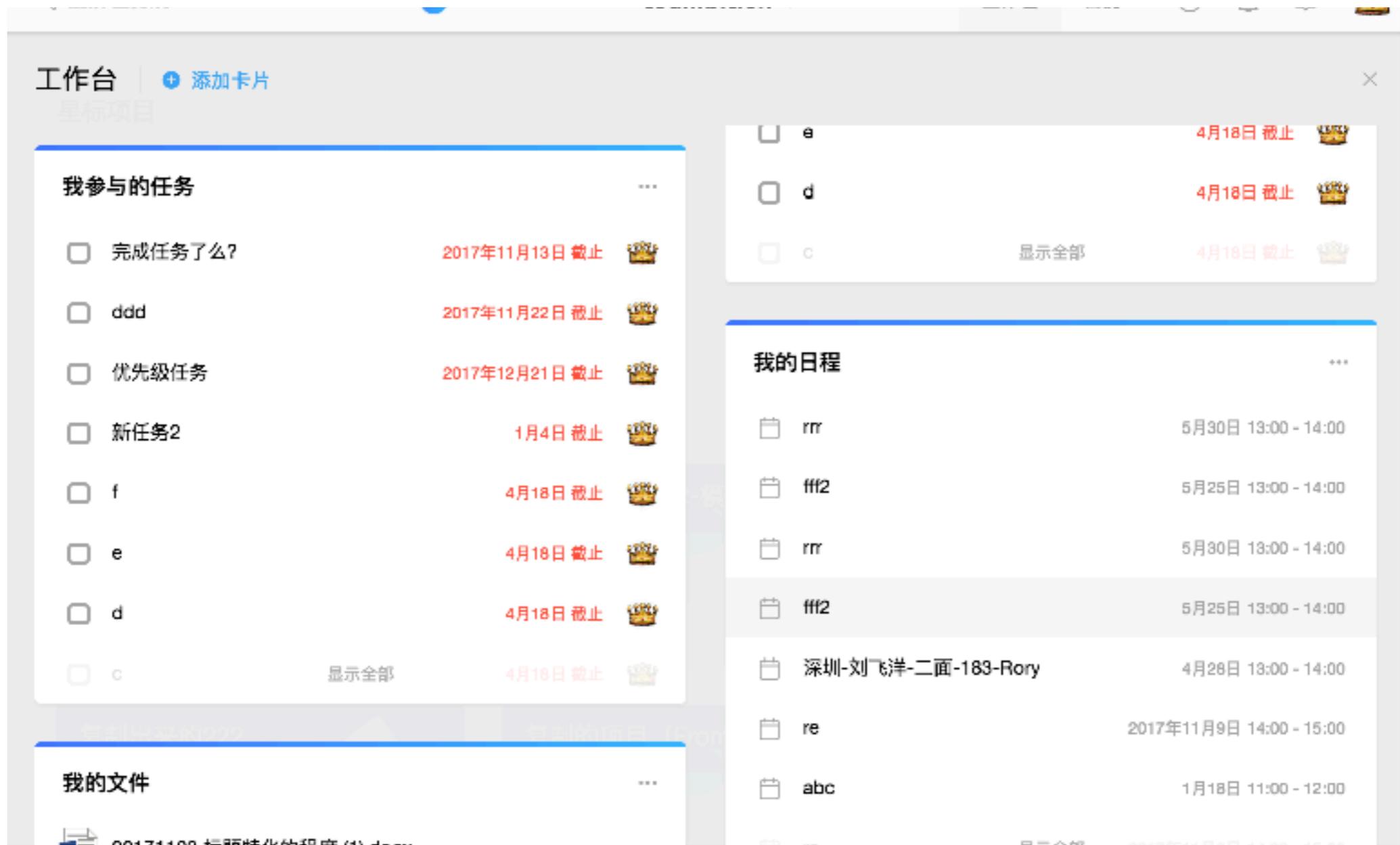
All

Start Date

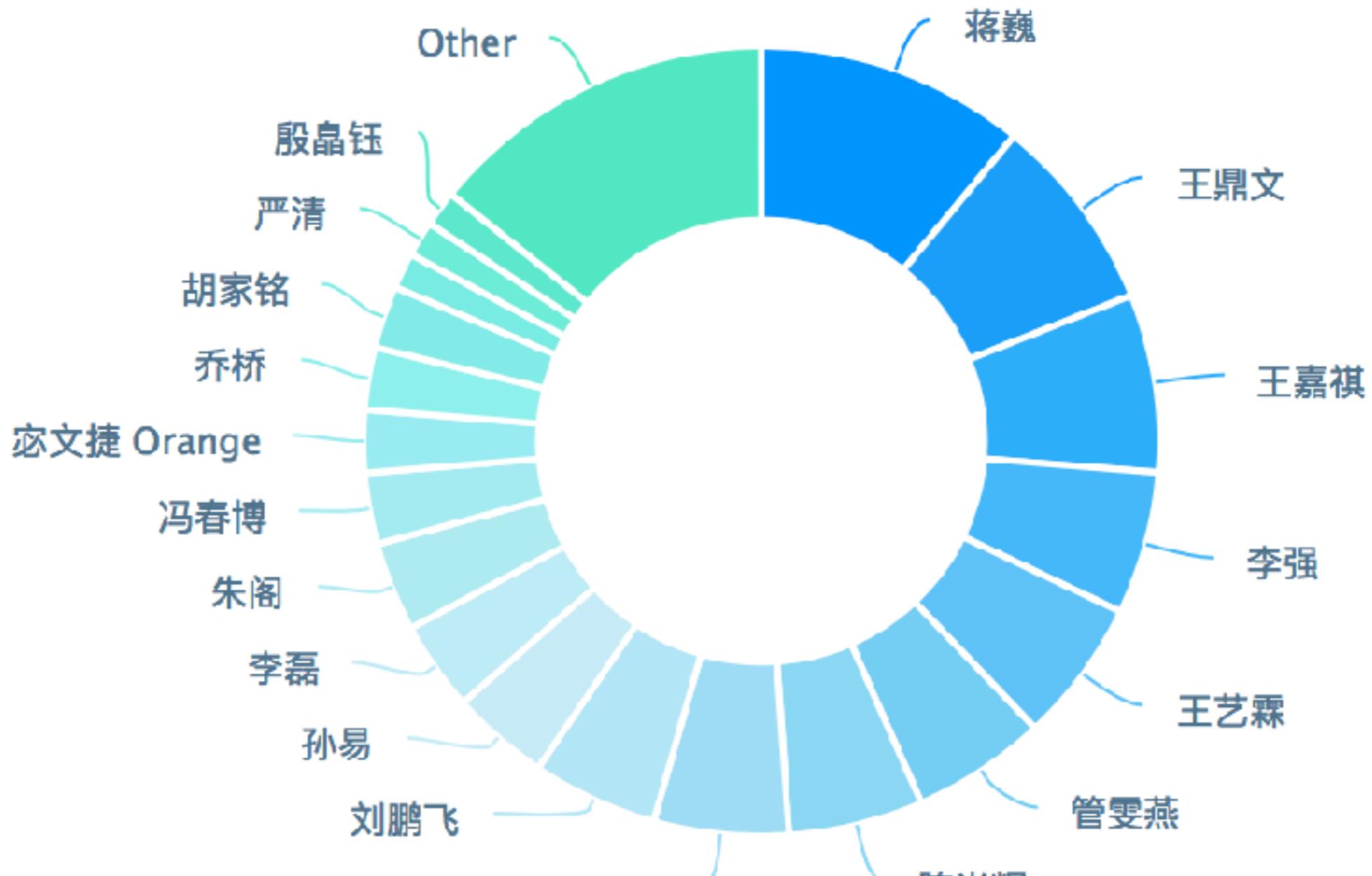
All

高级搜索

Search



用户自定义Dashboard



统计分析

Statistics & Analytics



二次开发

ES 线上情况

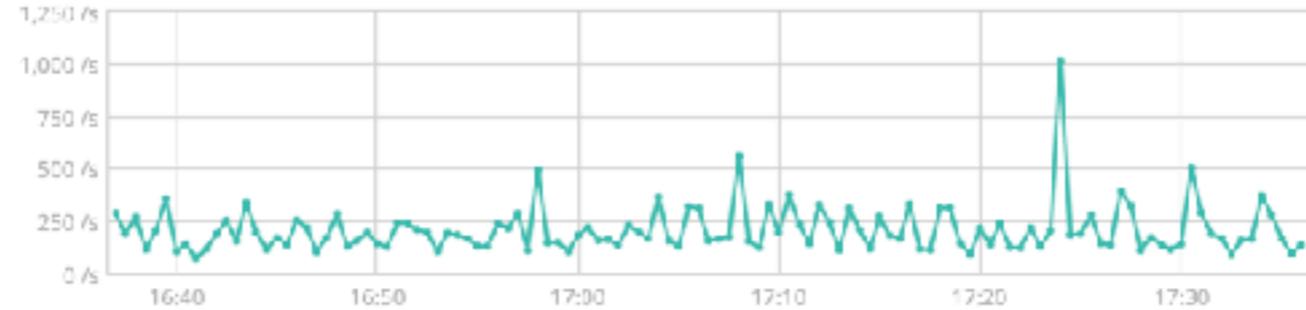
业务数据库
Nodes: 3
Indices: 70+
Docs: 150M+
Largest Index: 70M+

Clusters / teambition-kbs / Elasticsearch

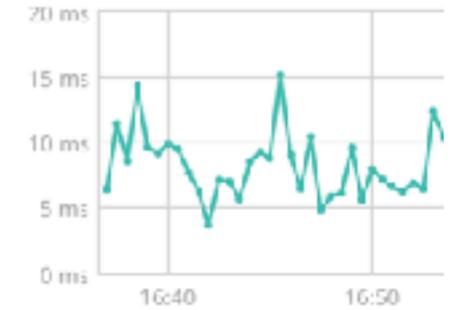
Overview Nodes Indices

Nodes: 3 Indices: 78 Memory: 36GB / 48GB Total Shards: 340 Unassigned Shards: 0 Documents: 157,405,325 Data: 264GB

Search Rate (/s)



Search Latency (ms)



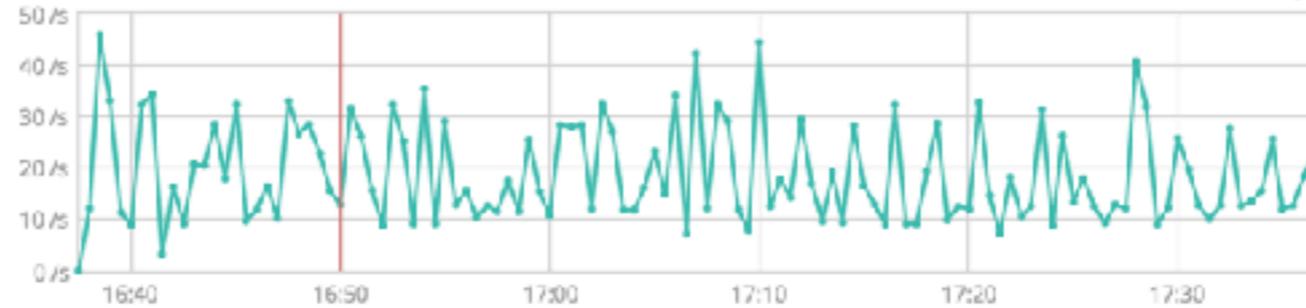
日志数据库
Nodes: 2
Indices: 1000+
Docs: 1B+

Clusters / elasticsearch / Elasticsearch

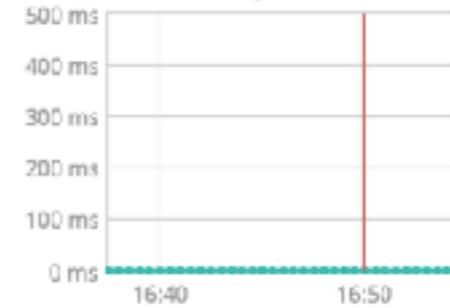
Overview Indices Nodes

Nodes: 2 Indices: 1012 Memory: 8GB / 12GB Total Shards: 4963 Unassigned Shards: 0 Documents: 1,006,701,624 Data: 466GB

Search Rate (/s)



Search Latency (ms)





Im小Mi 

Shanghai Pudong New Area



Q&A

Thank you



elastic
中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区
<https://elasticsearch.cn/>