



GOPS2018
Shenzhen

数据安全

数据加密/数据脱敏

数据使用严格审批和监管

Hadoop自研：
支持账号+IP双重认证



GOPS2018
Shenzhen

| 数据共享

聚焦的问题



· 公司到底有**哪些**数据?



· 公司数据应该怎么**申请**?



· 如何**使用**这些数据?



GOPS2018
Shenzhen

| 数据共享

公司数据自动发现

- 公司数据自动发现
- 数据的认领和标注

搜狗数据云 退出

标记基本信息

* 标题

* 描述

* 所属项目

* 自定义标签

* 管理员

查看者

Copyright © 2018 SOGOU.COM 京ICP证050897号



GOPS2018
Shenzhen

数据共享

- 数据索引与检索
- 任务关联分析

搜狗数据云 退出

搜狗数据云 退出

AUTO zeus:/storage/sogou/cdn/nginx/ 最近更新时间: 2018年03月09日 10:18:31

数据类型 文件数量 数据大小 最近访问 2018年03月09日 10:18:31

路径 /storage/sogou/cdn/nginx/dl.bizhi.sogou.com

数据分区 预览数据 使用概况

2018-03-09

使用Host	槽位数	96	用时	2018-03-09 01:05:07-2018-03-09 01:06:36	✓ 已完成
使用Host	槽位数	96	用时	2018-03-09 01:30:09-2018-03-09 01:36:20	✓ 已完成
使用Host	槽位数	96	用时	2018-03-09 01:00:07-2018-03-09 01:02:56	✓ 已完成
使用Host	槽位数	96	用时	2018-03-09 02:00:08-2018-03-09 02:04:46	✓ 已完成



GOPS2018
Shenzhen

数据共享

- 数据标准化
统一的数据仓库
和demo数据

搜狗+ 首页 文档中心 产品中心 数据中心 数据字典 管理中心

数据中心 总查询量: 13498 今日查询量: 112

表详情 dh_daohang_pb

字段名	类型	长度	描述
user_id	STRING		用户id
user_ip	STRING		用户ip
user_sduv	STRING		sduv
user_yyid	STRING		yyid
user_action	SMALLINT		1访问2点击3搜索5show
version	STRING		版本
reg	STRING		渠道号
refer	STRING		访问来源
sogou_hp	STRING		nginx服务器收到的http请求地址
ptype	STRING		ptype
pcode	STRING		pcode
pageurl	STRING		pageurl
module	STRING		pbflag
url	STRING		url
uriname	STRING		uriname
ctime	STRING		日志时间
ckid	STRING		ckid
ua	STRING		浏览器信息
loc	STRING		省份城市位置
stuv	STRING		ssuv
sk_name	STRING		皮肤名称



GOPS2018
Shenzhen

[3.0时代]

AI驱动/产品商业化
大数据时代

新问题 [2] ?



从数据产生到数据使用,
流程很长



数据使用门槛很高
清洗成本、协议处理成本很高



数据平台对外的易用性不够,
用户不会用, 使用出错

如何让大数据平台使用更简单

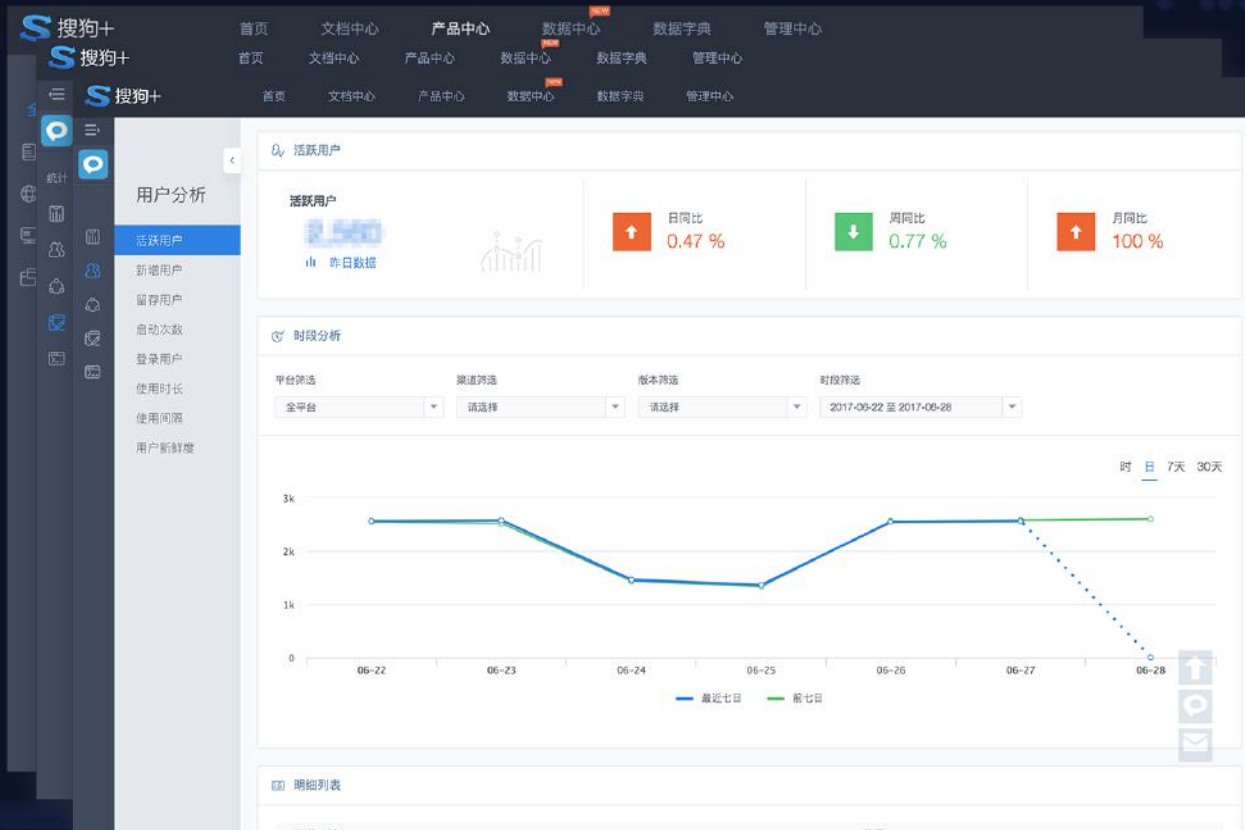


GOPS2018
Shenzhen

| 用户体验更简单

提供SAAS级的产品级数据解决方案

- 配置SDK数据可查
- 自定义埋点





| 用户体验更简单

数据处理简单化

一键完成数据清洗

搜狗通行证接口访问日志 数据云

最近更新时间: 2018年03月09日 00:20:18

管理员 申请权限

文件数量

数据大小

最近访问 2018年03月09日 11:00:19

清洗到数据仓库

接口 用户 通行证

搜狗通行证用户登录态日志 数据云

最近更新时间: 2018年03月09日 00:19:09

管理员 申请权限

文件数量

数据大小

最近访问 2018年03月09日 04:50:15

清洗到数据仓库

用户 登录态 通行证



| 用户体验更简单

类Google/BigQuery
的交互式查询界面

The screenshot shows the '搜狗+' data center interface. The top navigation bar includes '首页', '文档中心', '产品中心', '数据中心', and '数据字典'. The '数据中心' tab is active. On the left sidebar, there are options for '新建查询', '新建报表', '历史查询', '收藏', '申请权限', and '创建新应用'. Below these are sections for '私有数据集' (listing '测试demo', '支付DEMO', 'kevin test', '天天健康', '资金系统测试应用', 'test_123', '平台产品') and '公共数据集' (listing '数据仓库DEMO应用'). The main content area shows a query editor with the SQL: `1 SELECT "yyid", "version" FROM`. Below the editor are buttons for '执行查询', '格式化', and '收藏'. A '查询' section below the editor has tabs for '历史' and '收藏'. The '历史' tab is active, displaying a table of query history:

SELECT "yyid", "version" FROM "10138:demo_db".demo_table LIMIT 1000	2018-01-15 20:07:03
SELECT "yyid", "version" FROM "10138:demo_db".demo_table LIMIT 1000	2018-01-12 16:01:42
SELECT "yyid", "version" FROM "10138:demo_db".demo_table LIMIT 1000	2018-01-12 16:01:41
SELECT "yyid", "version" FROM "10138:demo_db".demo_table LIMIT 1000	2018-01-12 16:01:41
SELECT "yyid", "version" FROM "10138:demo_db".demo_table LIMIT 1000	2018-01-12 16:01:40
SELECT yyid, version FROM "10138:demo_db".demo_table LIMIT 1000	2018-01-12 15:44:18
SELECT "yyid", "version" FROM "10138:demo_db".demo_table LIMIT 1000	2017-10-24 16:38:48
SELECT yyid, version FROM "10138:demo_db".demo_table LIMIT 1000	2017-10-18 11:59:15



GOPS2018
Shenzhen

| 用户体验更简单

简单设置完成报表

搜狗+ 首页 文档中心 产品中心 数据中心 数据字典 管理中心 退出

创建报表 1 新建SQL 2 变量替换 3 报表设置 4 生成报表

appid	app_name	en_name	icon	category_id	category_name	class_id	class_name	cc
2441	微信	WeChat	https://cdn.questmobile.cn/img/icons/2441.png	1	移动社交	190	即时通讯	腾
2441	微信	WeChat	https://cdn.questmobile.cn/img/icons/2441.png	1	移动社交	190	即时通讯	腾
2441	微信	WeChat	https://cdn.questmobile.cn/img/icons/2441.png	1	移动社交	190	即时通讯	腾
2441	微信	WeChat	https://cdn.questmobile.cn/img/icons/2441.png	1	移动社交	190	即时通讯	腾
2441	微信	WeChat	https://cdn.questmobile.cn/img/icons/2441.png	1	移动社交	190	即时通讯	腾
2441	微信	WeChat	https://cdn.questmobile.cn/img/icons/2441.png	1	移动社交	190	即时通讯	腾

报表设置

应用
搜狗通行证

报表名称
最多128个字符

为了支持报表上钻、下钻数据，系统已自动识别所有字段的属性，其中时间维度已默认加上，如下命名为day，无需再添加时间维度，其他字段若有判断错误您可进行调整

time_stamp
 别名 值

local_ip
 别名 值

passport_id
 别名 值

id_type
 别名 值

interface
 别名 值



| 用户体验更简单

数据可视化

- 类tableau的
数据可视化系统

安卓新日志-主动-分版本分渠道

报表数据

聚合方式 无聚合 日期 < 2018-03-09 ~ 2018-03-09 > 继承筛选条件 下载报告数据

版本	渠道	主动用户数
4430	求和	1698
4430	最大值	1299
4430	最小值	1272
4430	日均值	0
4430		1924
4210		0
4430		1271
4430		1300
4430		1656
4300		1272
4430		1774
4430		1867
4300		1299
4300		1698
4300		1271

< 上一页 1 2 3 4 5 ... 17 下一页 >



GOPS2018
Shenzhen

用户体验更简单

易用性持续改进



新手上路引导



数据透明，进度透明



GOPS2018
Shenzhen

数据使用更简单

移动端支持



关键数据提醒



查看数据报表



推送汇总信息



GOPS2018
Shenzhen

[3.0时代]

AI驱动/产品商业化
大数据时代

新问题 [3] ?



如何让大数据平台发挥出更大的价值?