



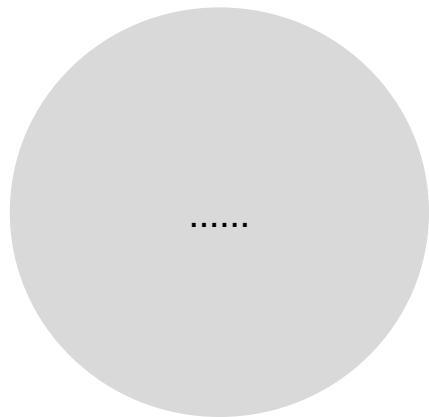
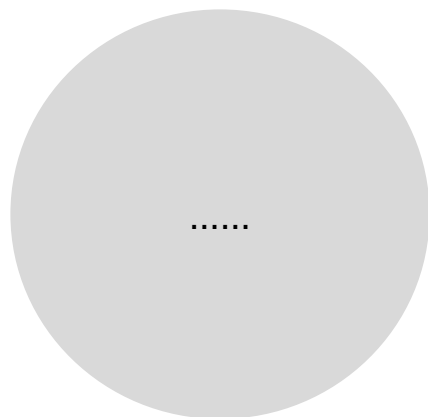
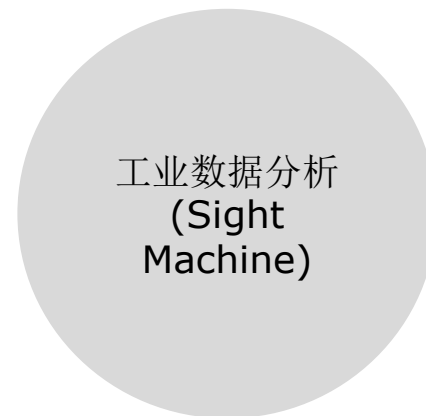
大数据分析系统及其应用实践

云朋

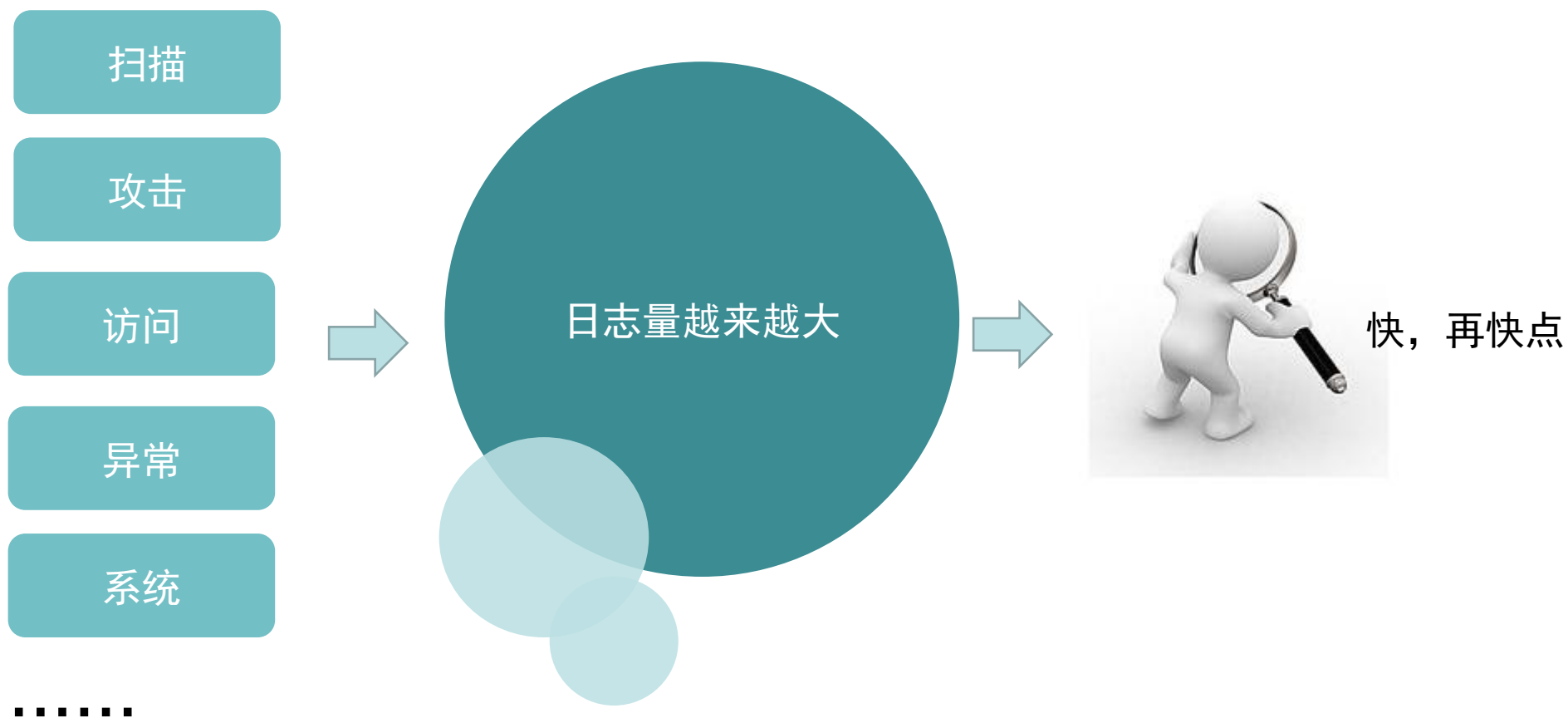
数据分析很大，我们做的事很小



只是大数据分析中的一种



为什么选择做日志分析



核心作用

- 数据鸟瞰
 - 系统正在发生着什么
- 分析问题
 - 用数据引导解决问题，促进故障改进
- 预警
 - 早期预警系统，简单实用，保证系统健康运行

基础设施需求能达到的能力

- 全文检索：不仅能返回关键字段的内容，还要能搜索在文本中的内容
- 水平扩展的能力
- 高效的读写性能：海量内容的入库、快速搜索到关键信息
- 容错：节点错误一定会发生，要有机制能Bypass
- Agent-Free
- GUI、API

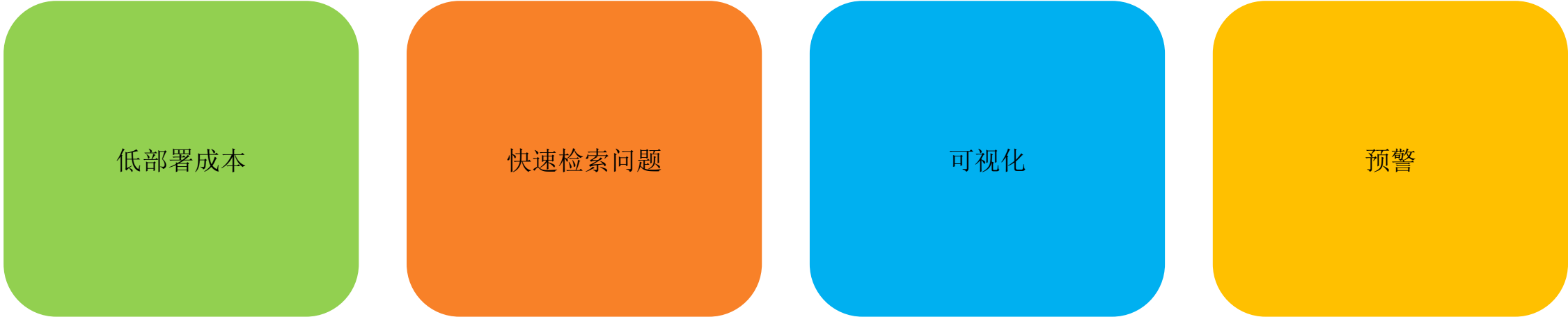
我们计划做一个什么样的系统

快速支持用户自定义日志

日处理150亿条（10TB
）

可单机 & 群集 & 公有云
& 可私有云自动化部署

满足用户场景



低部署成本

快速检索问题

可视化

预警

快速验证技术可行性

Hadoop

ElasticSearch

Storm

BAIDU

Docker

More OLTP & OLAP Model

云分析架构层

安全检测能力

规则

异常算
法

聚类算
法

人机识
别

关联分
析

统计

大数据能力

分布式消息件

分布式流式处理
引擎

分布式搜索引
擎

分布式计算

日志收集能力

syslo
g

AD日
志

http日
志

数据库
日志

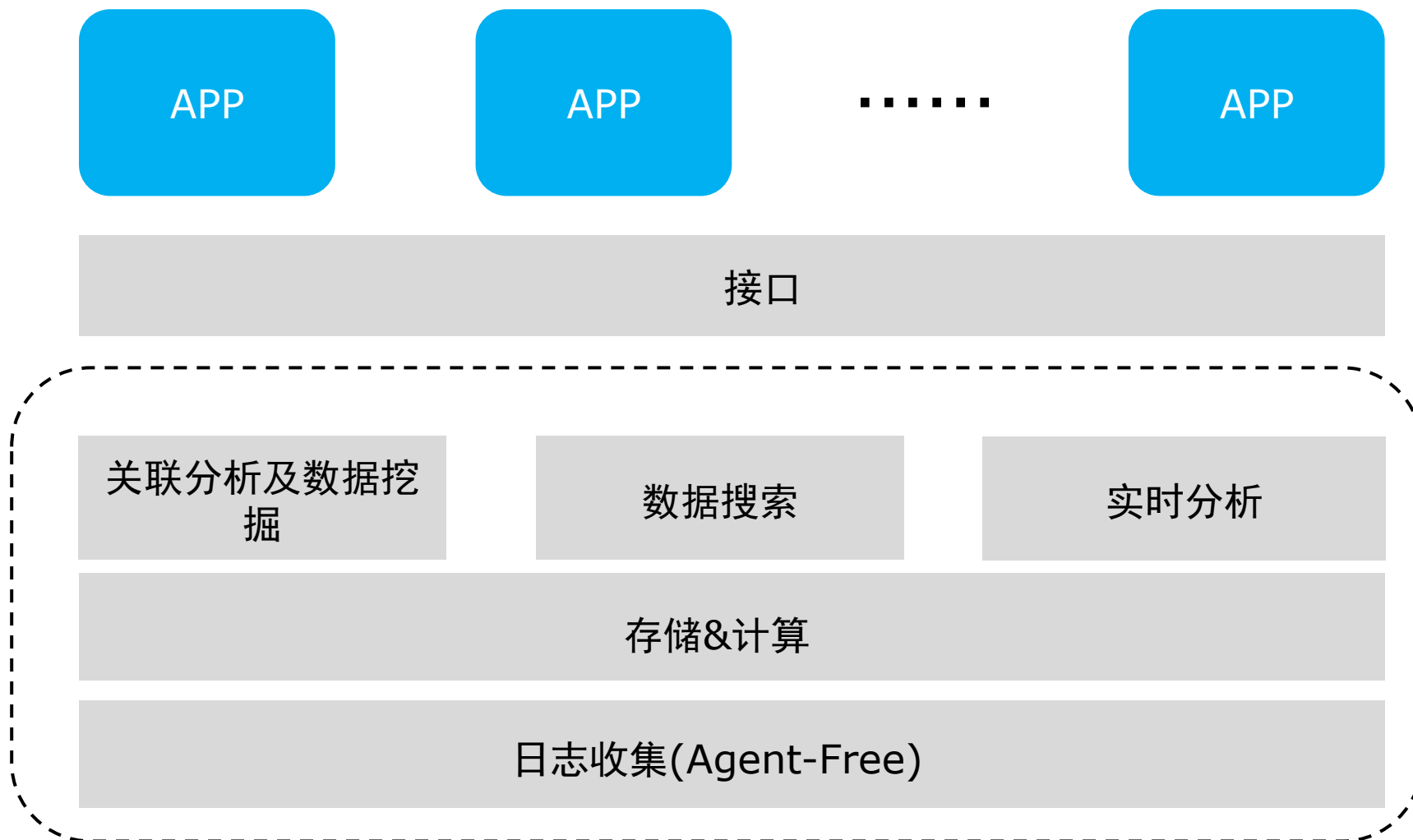
bash

DNS

Wind
ows主
机

自定义
日志

动静分离，做合适的系统



CASE1-服务器安全分析（自动化APP）

日志类型
Apache访问日志
Nginx访问日志
IIS访问日志
bash_log (bdsh)
SSH日志
AUDIT日志
Proftpd日志
davinci_log (dba_mysql)
MySQL查询日志
Windows安全日志
邮件服务器安全日志
exchange OWA日志

应用

网站安全分析App
web安全分析
数据库安全分析
主机安全分析
邮件安全分析

报警管理
报警处理
攻击还原
报警检索
报警批注

CASE1-服务器安全入侵



CASE-2邮件安全（钻取）

Figure 1: A screenshot of a data analysis interface showing a table of login events and a modal dialog for field configuration.

The table displays login events with columns: instance_id, source_id, source_type, time, server, status, eventid, clientip, logtype, tags, username, _id, and _type. The status column is highlighted in red.

The modal dialog titled "字段: status" (Field: status) is open, showing options to display the field (是) or not (否). Below the dialog, a table shows the distribution of status values:

值	计数
success	4624
failure	4625

2 查看连接email服务器最多的IP

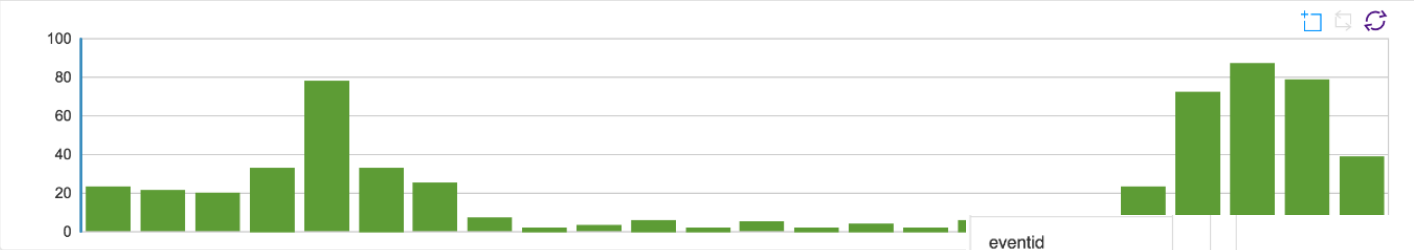
1 以失败登录为条件进行检索

[illegible]

CASE-2邮件安全

clientip:" " AND status:"success" 邮件服务器安全日志 24小时内 搜索

581个事件; 2015-06-23 13:54:44至2015-06-24 13:54:44之间的数据



选取字段

选取

- eventid
- clientip
- logtype
- tags
- username
- _id
- _type

4 找到被成功登录的email名

3 以可疑IP的成功登录为条件

g[success] 4624 已成功登录帐户。#177 #177主题:#177安全 ID:S-1 -0-0 #177帐户名 :#177帐户域 :#177登录 I D :0x0 #177 #177登录类型 :3 #177 #177新登录 :

2015-06-24 13:52:00

字段 : username

是否显示该字段 : 是 否

报表 :

时段最大值

平均值 :

值	计数
177	1
177	1
177	6
177	5
177	4

3 microsoft-windows-security-auditin
3 microsoft-windows-security-auditin
3 microsoft-windows-security-auditin
3 microsoft-windows-security-auditin
3 microsoft-windows-security-auditin

两个有价值的参考



Sumo Logic

Sumologic

DEVOPS

Do You Need To:

Streamline Your Continuous Deployment?
Monitor App KPIs & Alert On Problems?
Identify Root Causes Of Downtime?

[LEARN MORE](#)

IT INFRASTRUCTURE AND OPERATIONS

Do You Need To:

Monitor & Troubleshoot Cloud Workloads?
Manage Full Stack Log Analysis?
Improve Your App Performance & Uptime?

[LEARN MORE](#)

COMPLIANCE AND SECURITY

Do You Need To:

Quickly Complete Your PCI & Other Audits?
Get Intelligent About Threat Management?
Spend Less Time Scaling Your SIEM &
More Time Using It?

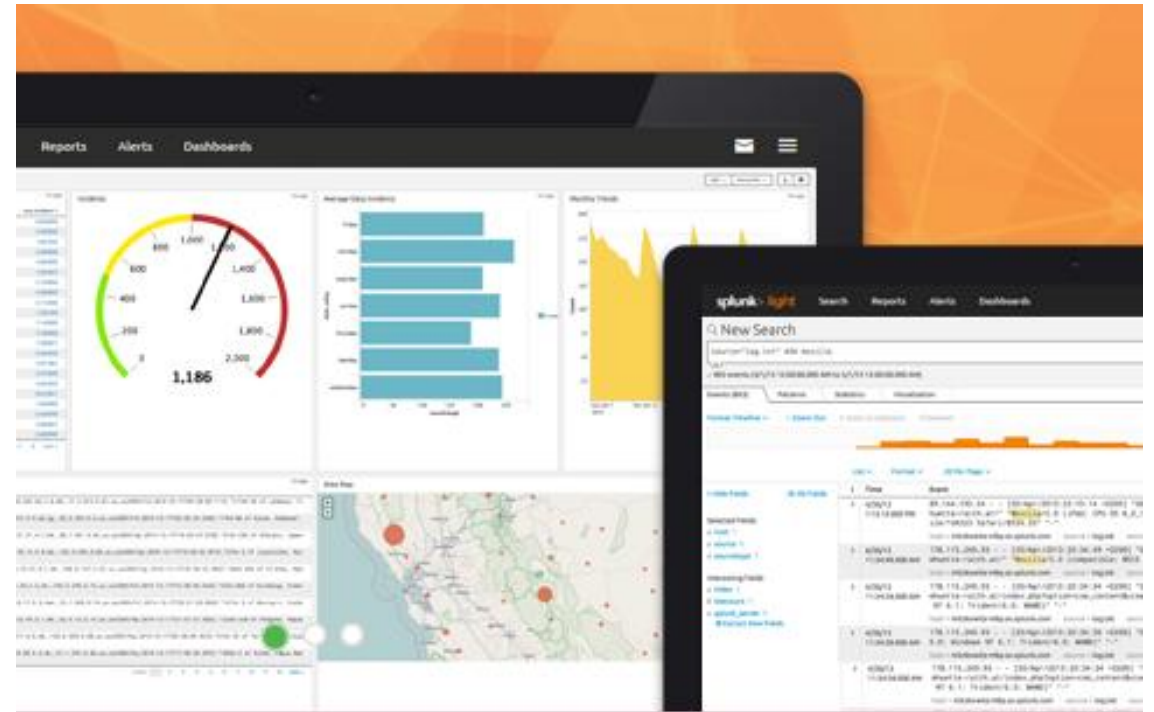
[LEARN MORE](#)

Splunk

Enterprise-Ready SaaS

What can you do with Splunk Cloud?

[Learn More](#)



Q&A

谢谢！

云朋

yunpeng01@baidu.com