



IT大咖说  
知识分享平台

OSF 2017 开源数据库论坛(北京)  
OPEN-SOURCE DATABASE FORUMS (OSDF)

# 开源数据库正在改变世界

2017年8月24日-25日 北京-京仪大酒店



# 运维数据可视化探索

## ——DBA的游戏

古雷 央视网MySQL专家

# 目录

## CONTENTS

第1关 DBA出发

第2关 Spark Streaming

第3关 Gephi

第4关 MariaDB ColumnStore

第5关 PHP, Echarts, Neo4j

01

# DBA出发

大数据、机器学习、人工智能风起云涌，  
管理数据的库的我们要出发



# DBA出发

1. 新技术新趋势
  - 大数据、人工智能、机器学习、数据可视化
2. 数据库运维工程师所处环境
  - OLTP、OLAP
  - 由运维产生的数据（比如监控数据）
3. 针对运维数据的分析
4. 收集、运算、展示——快速、可重复



# 数据分析思路

1. 结论从总体到局部
  - 由大量细节数据得到总体结论
  - 比如故障发生时，首先确定故障范围
2. 机器学习中的监督学习
  - 建立现象与人类概念的联系
  - 建立数据与人类概念的联系
3. 本次探索决定从tcpdump抓包开始
  - 表达了机器或者应用间的联系
  - 将来可以考虑从交换机上抓包



# Spark Streaming

很像在用SQL



# Spark Streaming

1. 为什么要用Spark Streaming
  - 为大规模使用做准备
  - 探索阶段，不希望过多占用存储
  - 与nc配合收集数据的例子，不需要另外的agent
  - 比较方便收集机器上的任何数据
2. Scala函数式编程
  - 貌似能找到的Scala的例子比Python多
  - 有趣的函数式编程思想（在Javascript里也有了）
  - Spark的并行计算思想





# Spark Streaming数据处理

1. tcpdump加nc把数据传给Spark Streaming
2. Spark Streaming
  - 服务端端口保留，客户端端口用0替换
  - 算出IP端口之间包数的TopN
    - 1.1.1.1:80->2.2.2.2:0 500
    - 3.3.3.3:8080->4.4.4.4:0 300
3. 用Scala代码做了什么？
  - 类似count group by、sum、join
4. 输出结果不直观——要有图



# Gephi

图 (graph) 可视化分析神器



# Gephi

## 1. 演示一

- 一个node表示一个IP
- 从MySQL导入数据
- 社区发现、着色、degree决定顶点大小
- Lay out: Yifan Hu proportional

## 2. 演示二

- 一个node表示一个IP加一个端口

## 3. 学习资源——YouTube上有很多讲解视频

04

# MariaDB ColumnStore

MySQL 列式存储引擎



# MariaDB ColumnStore

## 1. MySQL分支MariaDB下的存储引擎

- 前身是InfiniDB
- 列式存储
- 擅长统计型SQL（各种group by，大join）

## 2. 依据分析需求跑各种大SQL

- 比如源IP、源端口、目的IP、目的端口不同维度统计
- 导入到gephi中看效果
- 或者作为最终结果在ECharts中展示



05

# PHP, ECharts, Neo4j

直观展示给最终用户



# PHP, ECharts, Neo4j

1. 展示给用户
  - 前端ECharts(javascript库)
  - 后端PHP加MySQL
2. ECharts演示 (IP改得面目全非了)
  - 8月14日一个例子
  - 7月10日一个例子
  - 7月10日另一个例子
3. 用不同颜色表示不同机房或者不同状态
4. 用Chrome跟踪js代码, 绕过ECharts小问题



# Neo4j (neo是电影《黑客帝国》男主)

1. 使用图数据库找找感觉
2. 演示：这台机器出问题，可能会影响哪些机器？
  - `match (c)-[*1..5]->(s:IP {name:'0.0.2.117'}) return c,s`
  - `match (c)-[*1..5]->(s:IP {name:'0.0.0.90'}) return c,s`
3. 演示：这台机器依赖了哪些机器（服务）？
  - `match (c:IP {name:'0.0.1.91'})-[*1..5]->(s) return c,s`
  - `match (c:IP {name:'0.0.2.167'})-[*1..5]->(s) return c,s`
4. 和SQL挺像？
  - 图查询语言，叫Cypher（《黑客帝国》里那个叛徒）





# DBA发展新方向?

1. 曾经有一个DBA的群
  - 突然有一天大家都说自己不是DBA
  - 而是DA
2. Big DA
  - DBA身边有大量运维数据可供分析
  - Spark某种程度上是个更大的数据库
  - 可视化提供某种直观的角度
  - 机器学习, 智能运维
3. 快速搭建原型展现给潜在用户

# Thanks

关注开源数据库论坛