# 项目简介-技术组

自动化运维和数据管理

唐瀚林

厦门大学 经济学院

2016年12月10日

- 1 自动化运维系统
  - 自动化部署
  - 自动化管理
- 2 数据管理策略
  - 日常管理
  - 后续管理
- ③ 小结



- 1 自动化运维系统
  - 自动化部署
  - 自动化管理
- 2 数据管理策略
  - 日常管理
  - 后续管理
- 3 小结

# 自动化部署

● 使用方法:根据需求输入一行指令,即可完成多台服务器上 的同步部署

### set host:arg

- arg=all 针对所有的服务器
- arg=tmall 针对分配天磁任务的服务器
- arg=imail 针对分配京东任务的服务器
- arg-slaver 针对除了master外所有的服务器
- arg=master针对作为master任务的服务器
- arg=master fTXHFAImaster社shthlikiAni
   arg=人名 对应人名的服务器

## fab set\_host deploy

- 更新相应服务器上的代码
- e.g fab set\_host:all deploy 更新所有服务器上的代码

### fab set host remove

- 删除相应服务器上的代码
- e.g fab set\_host:all deploy 删除所有服务器上的代码(ERROR!!!)

#### fab set host restart

- 若由于特殊原因有节点中断,使用该指令重启,数据累加
- e.g fab set\_host:zhangyongjie restart 将zhangyongjie服务器上的爬虫重启

### fab set\_host repeat

- 若由于特殊原因(如特殊节日)一天需要能取多次、使用資指令重复启动、需等待第一次能虫完毕之后运行。考数据重命名当 数据不冲空
- e.g fab set\_host:tmall repeat 将分配tmall任务的服务需重复启动

### fab set\_host:master add\_master\_cron

绘master服务器添加manager定时任务必须指定参数为master

### fab set\_host add\_slaver\_cron

给指定服务器添加slaver, procmon, setlink定时任务

### fab set host init

• 重新分配任务或者list有更新后,运行此脚本

### fab set\_host add\_key

生成master的id\_rsa.pub后将文件移动根目录下,将master公钥批量添加

## setting 文件设置

### 添加或服务器

. HOST, env, shop, shop\_HOST, choices

### 新开服务器设置须知

- 安装pyenv及python2.7.9
- 安装python第三方库
- 安徽rar sudo apt-get install rar
- 更改ssh权限 sudo chown -R ubuntu:ubuntu .ssh
- 生成公钥 ssh-keygen 然后一直按enter
- cd .ssh 打开id\_rsa.pub, 将内容复制到github项目的deploy keys内
- cd .ssn f1)ffid\_rsa.pub, 特內普級期間github相目的deploy keysiyi
- 将master的d\_rsa,pub添加到.ssh/authorized\_keys内, 也可以在master通过fab添加
   ssh T git (stgithub.com)

## redis内数据结构情况(shop指所爬取的目标网站或app)

### shophash

• 用于存放告子节点能取uri的情况

### shop

- 用于存放危权所需的uni
- shop\_success

  用于存放膨胀成功后的url

## shop\_failed

用于存效范围生物并日生收次数在一定范围内的url

#### shop threw

田干森林都原生除土日生物学教研は一中学教の。

#### shop failed bucket

• 用于存放url胞取失败的次数

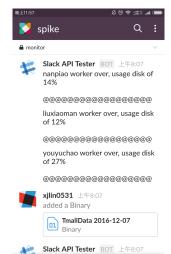
#### task

• 用于存放每个子节点和纯要胞取的商店

- 1 自动化运维系统
  - 自动化部署
  - 自动化管理
- 2 数据管理策略
  - 日常管理
  - 后续管理
- 3 小结

# 自动化管理

● 管理工具: slack——便捷 的企业协作工作软件







- 1 自动化运维系统
  - 自动化部署
  - 自动化管理
- 2 数据管理策略
  - 日常管理
  - 后续管理
- 3 小结

# 日常管理

• 管理任务: 检查数据是否正常,将数据收理归类

• Windows端管理: Winscp

- 1 自动化运维系统
  - 自动化部署
  - 自动化管理
- 2 数据管理策略
  - 日常管理
  - 后续管理
- 3 小结

# 后续管理

- 任务:将原始数据整理归类,并输入数据库
- 主要难点: 原始数据量级过大时, 硬件容易出现瓶颈

# 小结

- 主要优点:
  - 操作难度极低,几乎不需要什么计算机基础
- 待解决问题:
  - 部署任务依然需要使用命令行
  - 数据管理步骤仍然繁琐
- 解决方案:
  - 部署任务将会转移到slack上
  - 直接把数据输入数据库,进一步减少操作复杂度



◆ロ ト ◆ 個 ト ◆ 差 ト → 差 ・ 夕 Q ペー