

大规模集群下的 自动化部署

- 阿里巴巴云计算 朱佳文/2010-08
jiawen_zhu@hotmail.com

大规模——服务器规模急剧增长场景下的运维

挑战和变化：

- 运维理念
- 运维方式和手段
- 对运维人员的综合素质要求提高

自动化——从重复性的工作中解放出来

挑战和变化：

- 工作单调，重复，没有成就感
- 人肉的方式也搞不定

问题和挑战



- 操作简单：
 - a) 一键式安装 (CLI)
 - b) 在页面点一下 (Web UI)
 - c) 定制方便
- 良好的引导控制：
 - a) 重装系统
 - b) 无系统默认PXE引导 (新购)
 - c) 防止误装 (*)
- 友好的过程监控及控制：
 - a) 日志
 - b) 信息采集，并发，状态控制

- PXE + Kickstart
简单，灵活，易于实现
- PXE + Image
高效，标准化
- PXE + NFS Root
集群化

- 1) 网卡PXE引导 →
- 2) Memory OS启动 →
- 3) Anaconda请求kickstart →
- 4) 执行安装 →
 - I) %pre
 - II) installing
 - III)%post

install
lang en_US.UTF-8
langsupport --default en_US.UTF-8 zh_CN.GB18030 en_US.UTF-8
.....
rootpw --iscrypted \$1\$5uXyAm1Ssdfu\$20neCkGOEpg1marMta0
firewall --disabled
selinux --disabled
timezone Asia/Shanghai
url --url=http://10.0.0.10/rhel/5.4/os/x86_64
repo --name=Update --baseurl=http://10.0.0.10/yum/5Server/update/ gpgcheck=0

ks.cfg-header
ks.cfg-rootpw
ks.cfg-firewall
ks.cfg-selinux
ks.cfg-timezone

←这里的每一项都可以定制

%packages --resolvedeps --ignoremissing
@ base
@ core
@ system-tools
@ editors
@ system-tools

ks.cfg-packages

←包列表可以定制

%pre
#!/bin/sh
You can do something here before the real installation
wget -nv -O /tmp/pre-install http://10.0.0.10/pre-install
/bin/sh /tmp/pre-install

←真正执行安装前的一个shell环境

%post
#!/bin/sh
You can do something at the end of the installation
wget -nv -O /tmp/post-install http://10.0.0.10/post-install
/bin/bash /tmp/post-install

←Basic OS完成安装后的一个shell环境

模板定义:

```
- default
  - ks.cfg-header
  - ks.cfg-rootpw
  - ...
  - ks.cfg-timezone
  - ks.cfg-firewall
  - ks.cfg-partition
  - ks.cfg-packages
  - pre-install.sh
  - post-install.sh
  - firstboot.sh

- web_server
  - ks.cfg.partition
  - post-install.sh

- mysql_server
  - ks.cfg-packages
  - fistboot.sh

- ... ..
```

可以玩自动化的几个思路:

- 预定义安装模板 (profiles)
- 将ks文件拆成可定制的几个文件
- 通过PHP“拼凑”成完整的ks文件:
<http://xxx/ks.php?sn=xxx>
- 充分利用%pre阶段, 进行设备信息采集, 校验, 及安装前的准备工作 (如, 设置IPMI, 做raid)
- 通过%post阶段, 进行系统环境的初始化, 及定制软件的安装

Kickstart足够吗？

KS是否已足够满足我们的要求？ 答案是：No!

自动化的核心究竟是什么？

统一的数据源是运维自动化的核心

2010系统架构师大会

配置信息

主机名: xxx.aliyun.com

IP地址: 10.0.0.10

Mac地址: 00:26:b9:35:e9:aa

SN: 712dm2y

OOB类型: IPMI2

串口类型: serial2

OOB IP: 192.168.0.10

机型: Dell: PowerEdge R710

操作系统: rhel-54-64

安装模板: mysql_server

安装状态: 不需要安装

负责人: 朱佳文

自动化重点解决的几个问题

- 统一的数据源管理 (data source)
 - 设备信息采集，及准确性保障
 - 定制，要有数据依据
 - 行为可控，要有统一的信息支持
- OOB (out of band) 远程管理
 - 控制电源，引导顺序，串口信号
- 安装模板，及定制的实现
 - 管理要方便，操作要方便，安装过程要能自己识别
- 系统引导流程的控制
 - 新设备到货，没有OS自动进入pxe时的控制
 - 重装系统时的控制
 - 防止机器被错误重装
- 规模化部署时的并发控制
- 规模化部署时的状态跟踪

告别 “重启按**F12**”；查看安装过程，不再仅仅限于**KVM**

电源控制：

```
ipmitool -I lanplus -U root -H 192.168.0.10 power reset
```

引导顺序控制：

```
ipmitool -I lanplus -U root -H 192.168.0.10 chassis bootdev pxe
```

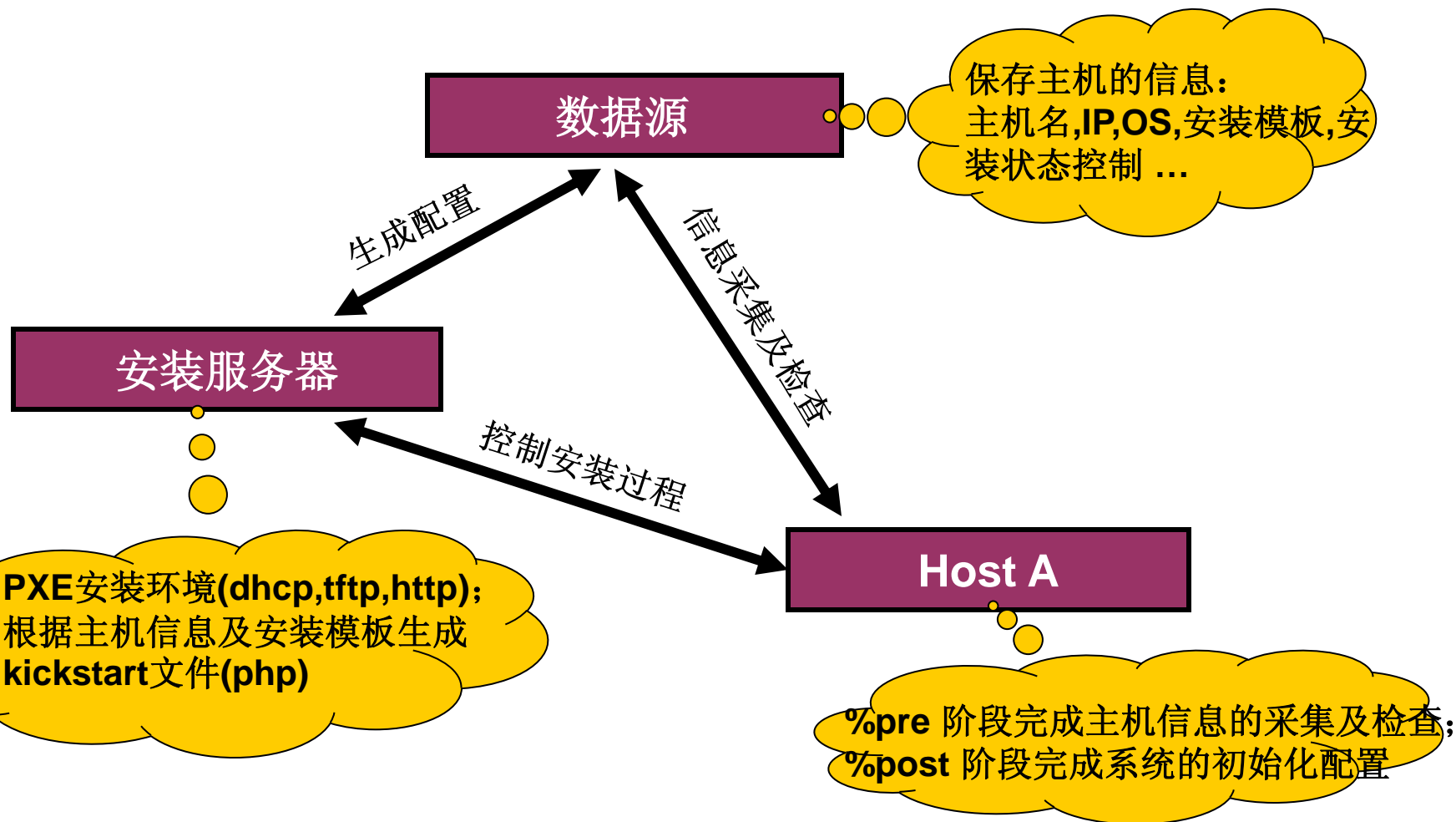
Console串口信号（需做串口重定向）：

```
ipmitool -I lanplus -U root -H 192.168.0.10 -e ^ sol activate
```

CLI封装，通过数据源获取主机信息，可以实现一键式安装。

OOB网络信息的设置，可以在**memory os**中提前设置（**%pre**）。

自动化部署的信息交互



自动化部署实践之日志跟踪



```
2010-04-24 08:29> ----- STARTING CLONE -----
2010-04-24 08:29> Clone process will take you some minutes,please wait...
2010-04-24 08:29> Auto-detect hostinfo via mac=00:26:b9:35:e9:aa
2010-04-24 08:29> Wrong SN, auto detect: 712dm2y, NOT 712dm2x    主机信息检查
2010-04-24 08:29> Setting oob: 192.168.0.10/255.255.255.0    安装前自动设置OOB
2010-04-24 08:29> Reach concurrent-clone's limit:55. Sleep 38s and then try to continue...
2010-04-24 08:29> Pre-install done.Go ahead,basic os will be installed in about 10 min.
2010-04-24 08:43> Default post-install completed    更新信息源安装状态
2010-04-24 08:43> Change host clone status as don't clone: mac=00:26:b9:35:e9:aa
2010-04-24 08:43> Reset host pxe config as harddisk booting    防止反复重装
2010-04-24 08:43> hostname:xxx.aliyun.com; os:rhel54_64; ip:10.0.0.10; profile:mysql_se
2010-04-24 08:43> The OS is going to reboot now...
2010-04-24 08:43> Auto-partition processe at firstboot, it takes about 10min per 1T disc...
2010-04-24 08:51> ----- CLONE DONE -----
```

- 安装前自动采集主机信息，并跟数据源进行校验，确保数据准确性；
- 在**OS**安装前自动设置**OOB**远程管理，确保异常时有远程控制能力；
- 控制并发安装队列，限制内部流量
- 自动更新相关安装状态

自动化部署我们还能做什么？



规模化运维场景下，我们需要确保：
安装时的系统基础环境，及运行时的系统基础环境保持一致。

系统基础环境的保障由 独立的配置管理平台 来做是很合适的。

因此：

系统安装只关注Basic OS的安装；配置管理平台完成对它的初始化及运行时的维护更新。

可以使用Image的方式来进行Basic OS的快速部署。
集群方式，可以采用NFS Root方式，集中管理Basic OS。

Thank You

欢迎交流：

MSN：

jiawen_zhu@hotmail.com