

极道分布式计算、存储系统 自动化部署工具使用手册

V0.1



极道科技（北京）有限公司

2021 年 4 月

目录

约定	3
1、概述	5
2、部署模式	5
3、硬件要求	6
4、硬件设置	8
5、网络规划	11
6、安装部署服务器	13
7、配置集群	14
7.1 使用 VI 工具修改配置文件	14
8、配置文件说明	17
8.1 集群配置部分 (cluster_config)	17
8.2 网络配置部分 (nic_config)	18
8.3 服务配置部分 (services_config)	19
8.4 补充说明	19
9、配置过程简介	20
10、配置步骤	22
11、申请 license	24
11.1 申请临时 license 方法	24
11.2 申请正式 license 方法	25
12、附录	25

约定

- 1、 极道分布式计算和存储自动化部署系统：本文主要阐述主体，极道科技自主研发的自动化部署工具，通过该工具，实现极道计算系统，存储系统的自动化部署，也是极道推荐客户采用的系统部署方式。以下简称“自动化部署系统”；
- 2、 部署服务器：1 台物理服务器或虚拟机，用来安装自动化部署系统，承担自动化安装实际计算集群系统和存储集群系统的任务；
- 3、 VirtualBox：Oracle 公司旗下的开源虚拟机软件；
- 4、 enp0sX: enp0s3，enp0s8，enp0s9，enp0s10。VirtualBox 虚拟机中的网卡名称；
- 5、 节点：计算系统集群或存储系统集群中的各个实际服务器，下文中也会称之为“目标节点”；
- 6、 内部网络：计算/存储集群内部数据交换网络，与业务网隔离；
- 7、 管理网络（MGMT）：承担计算/存储集群的管理，监控和配置变更；
- 8、 外部网络：与客户业务网络连接，承载实际业务，与内部网络隔离；
- 9、 ALAMO：极道科技自主研发的分布式文件系统；
- 10、 ANNA：极道科技自主研发的分布式统一存储系统；
- 11、 Achelous：极道科技自主研发的分布式融合计算系统；

- 12、 IPMI：能够横跨不同的操作系统、固件和硬件平台，可以智能的监视、控制和自动回报大量服务器的运作状况。
- 13、 PXE：预启动执行环境（ Preboot eXecution Environment，PXE）也被称为预执行环境，提供了一种使用网络接口（ Network Interface）启动计算机的机制。这种机制让计算机的启动可以不依赖本地数据存储设备（如硬盘）或本地已安装的操作系统。
- 14、 License：软件系统的许可证，安装软件过程中会要求提供。

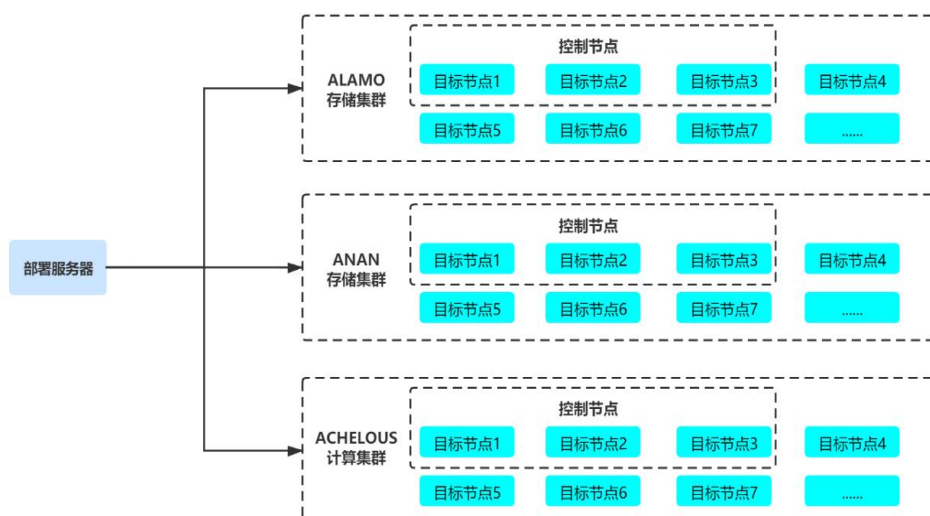
1、概述

伴随着大数据时代的到来，数据中心规模也在不断扩大，随之而来的是系统的部署复杂度提升，部署效率低下。

极道分布式计算和存储自动化部署系统（以下简称“自动化部署系统”）旨在帮助系统管理员高效、简便的一次性部署大规模计算集群和存储集群。为了实现这个目标，该系统提供了简洁的部署工具、管理工具，从而提升系统部署，监控和维护的自动化程度，提升管理效率。

2、部署模式

仅需要将自动化部署系统安装到 1 台部署服务器，经过简单配置后，部署服务器将并行对计算集群或存储集群中各个节点进行自动化安装和配置。



3、硬件要求

部署服务器	物理机	虚拟机	备注
内存	≥2GB	≥2GB	
千兆网口	≥2 个	≥2 个	用于给目标节点分配 IPMI, PXE 和 MGMT 网络
硬盘空间	≥30GB	动态分配 30GB	
USB/CDROM	需要	不需要	
目标节点	物理机	虚拟机	备注
CPU	≥16 Core, ≥2.6GHz	≥2 Core	
内存	≥16GB	≥2GB	
系统盘数量	≥1	1	推荐使用 2 块系统盘互为镜像
系统盘容量	≥300GB	动态分配 50GB	
数据盘	SATA/SAS/NL-SAS 和 SSD	添加虚拟硬盘, 8GB 即可, 数量依据 raid 组所需添加	系统盘和数据盘分离, 最好不共享 HBA 总线。例如系统盘连接在 AHCI 上, 数据盘连接在 SAS 控制器上
HBA 卡	1	N/A	不推荐使用硬件 RAID HBA。如只有硬件 RAID 控制器, 请使用 JBOD 模式。推荐使用 LSI IOC SAS 控制器
万兆网口	≥2 个 (推荐 4 个)	≥2 个虚拟网卡	硬件平台应至少包含 2 个万兆以太网口, 分别用作集群节点间内部互联和外部网络接入。推荐准备 2 个以上个的万兆网络做内部互联, 和 2 个以上万兆网络做外部网络接入

千兆网口	≥ 3 个	2 个千兆用于接受部署服务器分配 PXE 和 MGMT 网络，这 2 个系统网卡需要支持 PXE 启动。第 3 个千兆连接 BMC 网卡，用于接受部署服务器分配 IPMI 网络，并通过 IPMI 控制目标节点。
------	------------	---

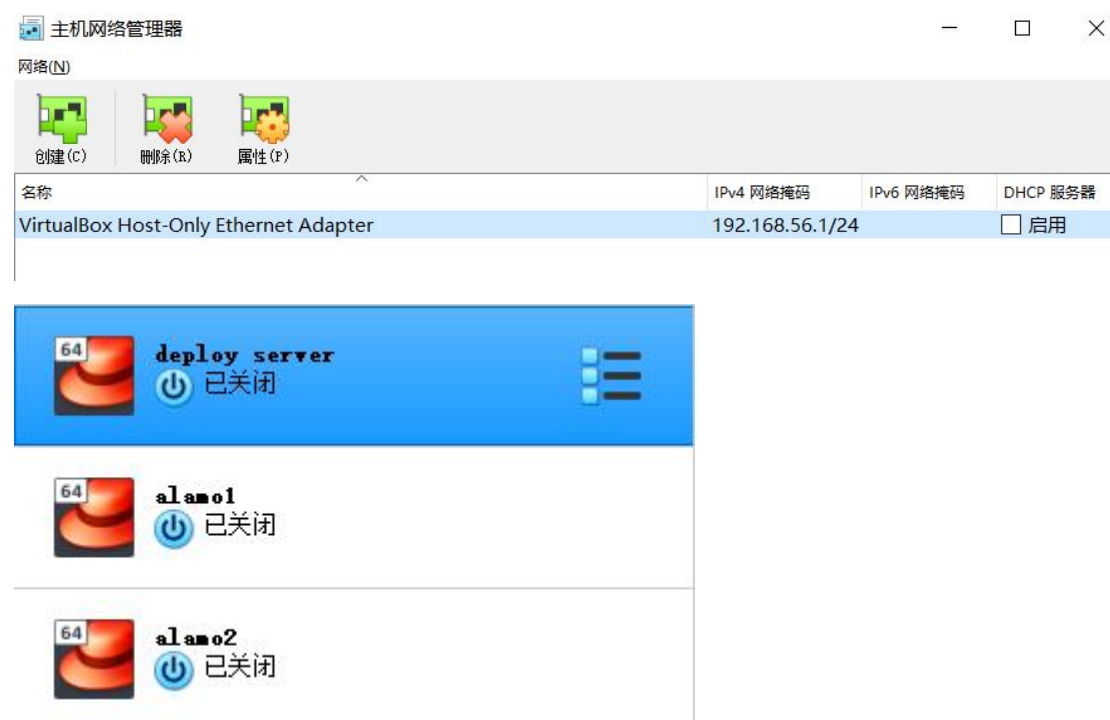
4、硬件设置

在 VirtualBox 中创建 3 台虚拟机，命名为 deploy server，alamo1，alamo2。

deploy server 为部署服务器，alamo1/2 为目标节点，配置如下：

deploy server：2Core，2GB 内存，30GB 硬盘空间（动态分配）

alamo1/2：2Core，2GB 内存，50G 磁盘空间（动态分配）



! Tips： 确保 VirtualBox 自带的 DHCP 没有启用。

deploy server (部署服务器) 网卡及启动顺序设置

deploy server	网卡名称	连接方式	混杂模式	作用
网卡 1	enp0s3	内部网络	全部允许	用于 PXE 网络给目标节点安装系统
网卡 2	enp0s8	内部网络	全部允许	用于 MGMT 网络给目标节点推送服务
网卡 3	enp0s9	桥接网络	全部允许	用于 ssh 远程登录

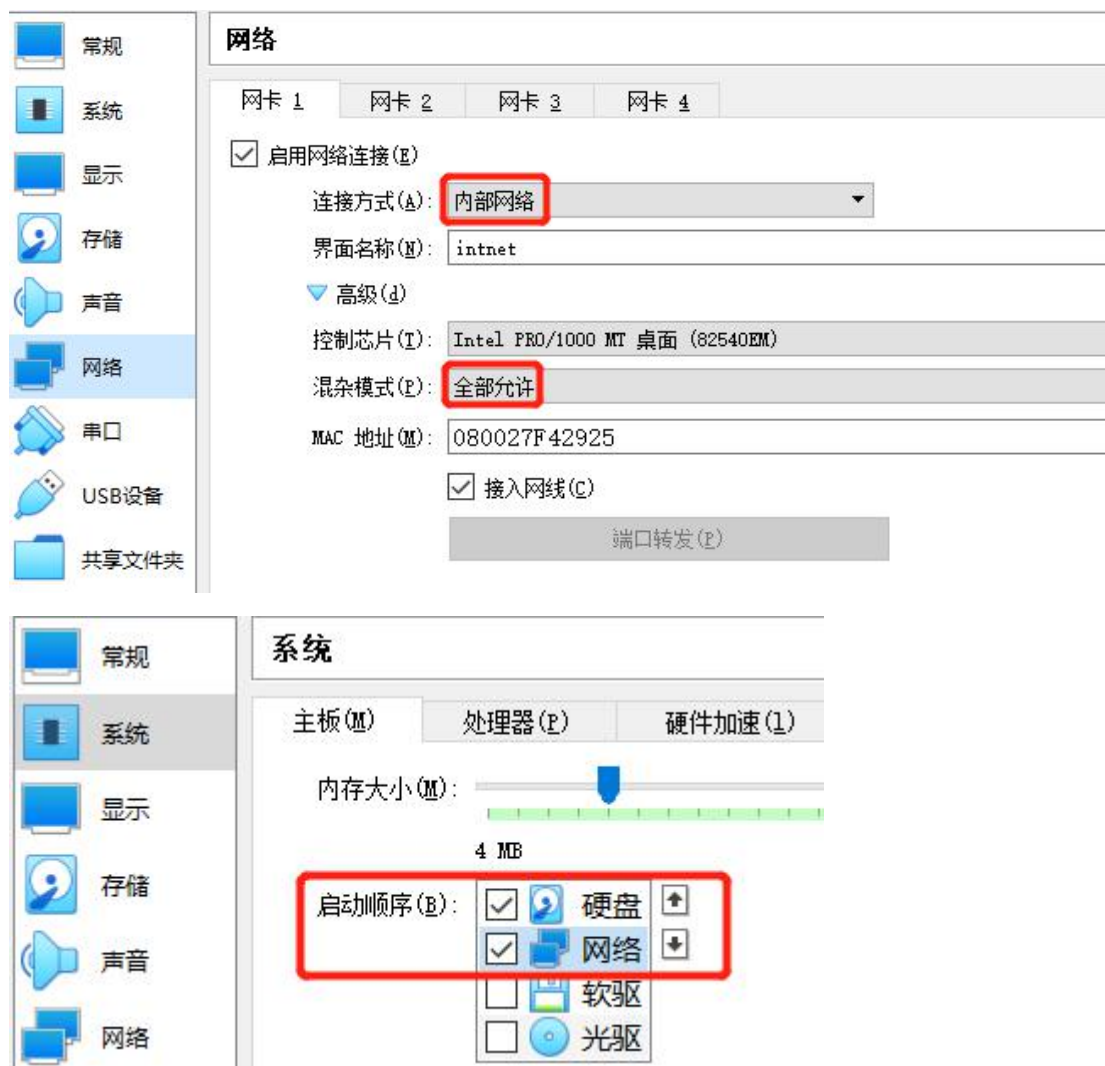
系统启动顺序：1、硬盘——2、CDROM

alamo1/2 (目标节点) 网卡设置

alamo1	网卡名称	连接方式	混杂模式	作用
网卡 1	enp0s3	内部网络	全部允许	PXE 安装系统，并且后期会用于配置存储内网和存储外网
网卡 2	enp0s8	内部网络	全部允许	MGMT 网络
alamo2	网卡名称	连接方式	混杂模式	作用
网卡 1	enp0s3	内部网络	全部允许	PXE 安装系统，并且后期会用于配置存储内网和存储外网
网卡 2	enp0s8	内部网络	全部允许	MGMT 网络

系统启动顺序：1、硬盘——2、网络

! Tips：可以看出存储内网和存储外网复用了网卡 `enp0s3`，实际部署生产系统不建议内外网复用网卡。



5、网络规划

极道系统有 4 种网络：

- 存储内网；
- 存储外网；
- 计算内网；
- 计算外网。

其中存储系统只需配存储内网，存储外网；计算系统需要配 4 种网络。实际部署生产系统的时候，也要先规划网络再部署。

PXE 网络 (enp0s3)	管理网络 (enp0s8)	登录网络 (enp0s9)
192.168.20.0/24	192.168.200.0/24	与宿主机同网段网络

目标节点	存储内网 (enp0s3)	存储外网 (enp0s3)	管理网络 (enp0s8)
目标节点 1	192.168.77.1	192.168.1.1	192.168.200.1
目标节点 2	192.168.77.2	192.168.1.2	192.168.200.2

! Tips：存储外网根据实际情况规划。

本次实际网络配置

depoly server	网卡名称	IP 地址	作用
网卡 1	enp0s3	192.168.20.254	目标节点将通过 DHCP 获取 192.168.20.0/24 网段地址，之后通过 pxe 网络从部署服务器获取系统并安装。
网卡 2	enp0s8	192.168.200.254	目标节点安装完系统后将自动获取 192.168.200.0/24 网段地址。部署服务器后期将通过管理网络推送系统服务。
网卡 3	enp0s9	与宿主机同网段网络	登陆网络，配置笔记本能访问到的网段 IP 地址，如公司 wifi 网段，192.168.15.xxx

6、安装部署服务器

将产品 iso 镜像载入 deploy server 光盘驱动器，启动 deploy server 虚拟机，部署服务器将自动安装完成。



默认用户名：root

密码：nasadmin

7、配置集群

登录部署服务器，极道已经将集群配置文件模板 template.yml 放置在

/usr/local/deploy/template.yml

复制该配置文件模板命名为 alamo.yml，并填写相应配置信息

7.1 使用 VI 工具修改配置文件

```
vi usr/local/deploy/alamo.yml
```

```
cluster_config:

  add_node: false

  cluster_type: alamo

  arch_type: x86_64

  cluster_name: xtao

  cluster_node_number: 2

  existing_cluster_node_number: 0

  existing_large_cluster_node_number: 0

  ipmi_user: ADMIN

  ipmi_passwd: ADMIN

  mgmt_nic: enp0s8

  mgmt_subnet: 192.168.200.0/24
```

```
pxe_nic: enp0s3

pxe_subnet: 192.168.20.0/24

system_disk:

    disk_name: sda

    disk_num:

    disk_size:

    disk_type:

nic_config:

    storage_internal:

        subnet: 192.168.77.0/24

        begin: 1

        interface:

            - nodes: 1

              nic: enp0s3

              bond_name:

            - nodes: 2

              nic: enp0s3

              bond_name:

    storage_external:

        subnet: 192.168.1.0/24

        begin: 1

        gateway: ""

        interface:
```

```
- nodes: 1-2

  nic: enp0s3

  bond_name:

services_config:

  alamo:

    cold_base: 1-2

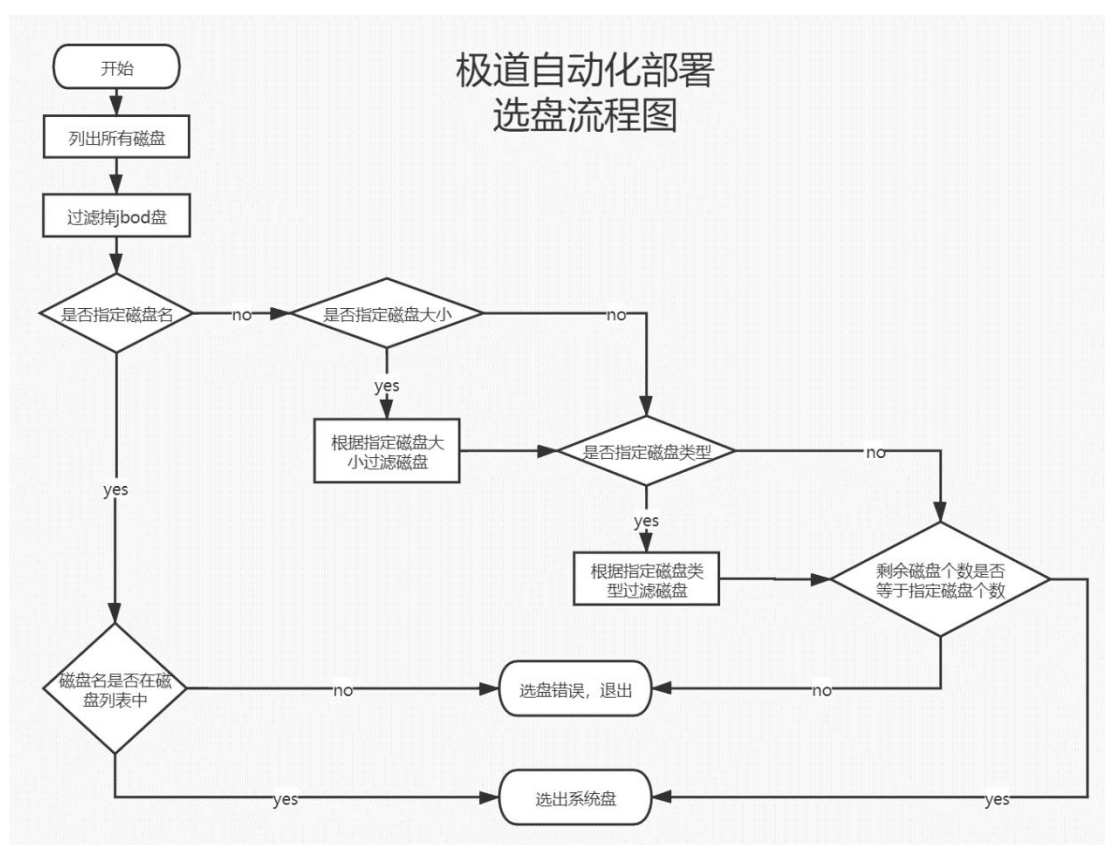
    storage_controller: 1-2
```

8、配置文件说明

8.1 集群配置部分（cluster_config）

配置	默认值	说明
add_node	false	此次部署是否为添加节点
cluster_type	compute	支持集群类型 有:compute, anna alamo or alamo_d
arch_type	x86_64	架构类型，目前支持 x86_64 和 arm64
cluster_name	xtao	集群名字
cluster_node_number	3	此次安装节点个数
existing_cluster_node_number	0	已存在节点个数，默认为 0，如果是添加节点，需要填写。
existing_large_cluster_node_number	0	
ipmi_user	ADMIN	
ipmi_passwd	ADMIN	
mgmt_nic	enp0s8	管理网络网卡名，通常不需要修改
mgmt_subnet	192.168.200.0/24	管理网络网段，将通过管理网推送服务
pxe_nic	enp0s3	PXE 网络网卡名，通常不需要修改
pxe_subnet	192.168.20.0/24	PXE 网络网段，将通过 PXE 网络安装系统
system_disk		指定系统盘
disk_name		系统盘名字
disk_num		系统盘个数
disk_size		系统盘大小，单位 GB
disk_type		系统盘类型，SSD 或 HDD

通过 PXE 给目标节点安装系统是全自动化的，包括将系统安装到哪个磁盘上也是自动化选择的。为了能够准确的将系统安装到指定磁盘上，需要用户输入一些提示性信息，将根据用户指定的信息选出 1 块或 2 块磁盘作为系统盘；如果选出 2 块系统盘，将做 raid1。关于选盘逻辑如下图所示：



8.2 网络配置部分 (nic_config)

用于配置目标节点网络，根据目标节点网络规划章节规划的网络填写配置

配置	说明
subnet	该类型网络的子网
begin	从 subnet 网段的那个地址开始分配地址，默认是从该网段 第一个地址开始分配
interface	配置网卡

nodes	节点索引。说明： 1. 1-3 表示配置 1 到 3 号节点；1, 3 表示配置 1 号和 3 号节点； 2. 节点索引不能超过节点总数； 3. 如果物理机目标节点网卡名不统一，可以一个节点一个节点的配置； 4. 可以网卡复用，即一个网卡上配置多个类型的网络；
nic	网卡名字。如果填写两个网卡名，就表明要做网卡 bond，此时可以同时指定 bond_name
bond_name	做网卡 bond 则必须填写此项配置，否则忽略此项配置

8.3 服务配置部分 (services_config)

用于配置目标节点的服务，并将使用 ansible 将这些服务安装到目标节点上。

配置	说明
compute	配置 ACHELOUS 计算集群相关的服务
compute_base	计算系统的基础服务，每个节点都必须安装
compute_controller	计算节点的控制器，要求奇数个节点安装，通常为前 3 个或前 5 个节点
compute_login	计算系统用户登录相关软件，通常安装在控制器上
anna	配置 ANAN 存储集群相关的服务
anna_base	ANNA 存储系统的基础服务，每个节点都必须安装
storage_controller	ANNA 存储系统的控制器，要求个数为奇数
alamo	配置 ALAMO 存储集群相关的服务
cold_base	ALAMO 存储系统的基础服务，每个节点都必须安装
storage_controller	ALAMO 存储系统的控制器，要求个数为奇数
alamo_d	配置 ALAMO 双控存储集群相关的服务
coldD_base	ALAMO 存储系统的基础服务，每个节点都必须安装
storage_controller	ALAMO 双控存储系统的控制器，要求个数为奇数

8.4 补充说明

- 1、 所有系统的控制器都要求个数为奇数，当目标节点总个数是 2 个及以下时，将不再检查控制器是否为奇数。
- 2、 填写配置文件时可以将不相关配置删除，比如配置 ALAMO 集群，可以将

计算和 ANNA 相关的配置删除，以保持配置文件简洁。

- 3、 由于 Achelous 计算集群内部会使用 ANNA 作为内部存储，所以配置计算时必须同时配置 ANNA 存储，内部 ANNA 存储可以和计算控制器重合。
- 4、 填写配置文件的时候可能由于目标节点还没有安装系统，无法取得目标节点的网卡名字，可以先使用默认网卡名，待配置程序执行到配置网卡那一步发现目标节点没有填写的网卡名，会报错并退出。此时可以修改配置文件，并重新执行配置程序。

9、配置过程简介

部署过程分为四个阶段：

- 1、 配置部署服务器；
- 2、 目标节点安装系统；
- 3、 生成目标节点网络和服务配置文件；
- 4、 目标节点安装服务；

! Tips : 执行 `xt-deploy -h` 查看部署程序的使用说明

1. -f 参数指定配置文件
2. --skip_install_os 如果目标节点系统已经安装完成，可以使用此参数跳过安装系统；目标节点未安装系统则不要使用该参数。

usage: xt-deploy [-h] [-f FILE] [--skip_install_os] [--commit]

optional arguments:

-h, --help	show this help message and exit
-f FILE, --file FILE	the path of config file
--skip_install_os	skip install os
--commit	commit exist configuration

10、配置步骤

步骤 1、在部署服务器中执行如下命令，自动化部署过程开始执行。

```
xt-deploy -f /usr/local/deploy/alamo.yml
```

! Tips: 此时目标节点应为关机状态。

步骤 2、当输出如下图所示信息，表明步骤 1 已经完成，此时逐个开启目标节点安装系统；部署服务器会不断扫描目标节点的开启状态，发现目标节点启动，便自动执行安装。

```
-----  
load xtao system (x86_64)  
-----  
create private docker hub  
-----  
Start to boot nodes...  
-----  
█
```

! Tips: 目标节点开机前，请确认节点 boot 顺序配置正确，并确保 VirtualBox 自带 DHCP 未启用。



目标节点的屏幕显示

步骤 3、当输出如下提示信息表明步骤 2(目标节点安装系统)已经完成，将自动开始步骤 3(生成目标节点网络和服务配置文件)，

```
.....  
All nodes(['192.168.200.1', '192.168.200.2']) is ready.  
['192.168.200.1', '192.168.200.2']  
install os success, will config services
```

! Tip : 如果配置文件有误，配置程序将提示相关错误信息并退出，此时需要修改配置文件并重新执行配置程序。

步骤 4、当输出如下提示信息表明步骤 3(生成目标节点网络和服务配置文件)已经完成，正在执行步骤 4(目标节点安装服务)。此时需要 license 才能继续执行。

```
TASK: [prepare | ensure /etc/ansible/xtao-product-key is exist] *****
ok: [localhost]

TASK: [prepare | Wait for put docker_all.tar.gz to /etc/ansible/] *****
skipping: [localhost]

TASK: [prepare | Wait for apply xtao-license and put it to /etc/ansible/] *****
```

11、申请 license

11.1 申请临时 license 方法

临时 license 有效期为 3 天。申领步骤为：

步骤 1、复制/etc/ansible/xtao-product-key 文件中的内容

步骤 2、开启浏览器，登陆临时 license 申请地址

<http://www.xtaotech.com:88>，输入步骤 1 中复制的内容，注意前后不要留空

格，空行。

步骤 3、点击【申请 license】按钮，xtao-license 文件将被自动下载到本地。

步骤 4、通过 ftp 工具连接部署服务器，将 xtao-license 文件放入部署服务器的

/etc/ansible 目录下。部署服务器将自动识别，并开始后续服务的安装。

！ Tips：注意不要将许可文件改名，如果本机系统自动为文件加了后缀，如.txt，请手动将后缀删除，否则部署服务器不会识别该文件。

11.2 申请正式 license 方法

请将/etc/ansible/xtao-product-key 文件通过邮件，发送给极道工作人员申请。

12、附录

申请临时 license 网址: www.xtaotech.com:88 常见问题，请访问如下链接寻求帮助：

<http://note.youdao.com/noteshare?id=e7844784b11225dda41883f92e>
[adbd2&sub=AC7B2A068D2F4CBE90FFB6FCC04E36AC](http://note.youdao.com/noteshare?id=e7844784b11225dda41883f92e)