

# 星环 Transwarp Data Hub（TDH）6

## 大数据平台部署文档

深圳迅策科技有限公司

2020 年 2 月 20 日

# 目 录

1. 环境要求.....	3
1.1 硬件环境要求.....	3
1.2 操作系统要求.....	3
1.3 Java 环境要求.....	4
1.4 支持的浏览器.....	4
2. 操作系统配置.....	4
2.1 系统磁盘分区要求.....	4
2.2 主机名配置.....	7
2.3 安全设置.....	7
2.4 安装介质准备.....	7
3. 安装 Transwarp Manager .....	8
4. 配置集群.....	13
4.1 DNS 配置.....	13
4.2 NTP 配置.....	13
4.3 添加/删除集群服务器.....	14
5. 安装服务.....	20
5.1 上传产品包.....	20
5.2 选择服务.....	23
5.3 服务角色的分配.....	25
5.4 服务的配置.....	26
5.5 服务的安装.....	32
5.6 安装 TDH 客户端.....	34
6. 安装和设置问题及回答 .....	36

## 1. 环境要求

### 1.1 硬件环境要求

Transwarp Data Hub(简称 TDH)提供集群服务来保证服务的高可用性，因此集群系统必须是 3 台以上物理服务器组成。

官方建议生产环境每台服务器必须具备以下最低配置：

2 颗 6 核心或以上带超线程 x86 指令集 CPU 的服务器

64GB 以上内存

2 个 300G 以上的硬盘做 RAID1，作为系统盘

4 个以上的 300GB 容量以上的 硬盘作为数据存放硬盘

2 个千兆以上网卡

为实现较好的性能并实现最高的性价比，TDH 针对生产环境集群内不同的模块有不同的推荐配置，Namenode 和 Transwarp Manager 推荐使用以下配置：

2 颗 6 核带超线程的 x86 指令集 CPU 服务器(比如 Intel(r) Xeon(r) E5-2620 v2 CPU)

128GB 以上内存

6 个 600G 以上的硬盘做 RAID1 作为系统盘和管理数据

2 个千兆以上网口的网卡

生产环境数据节点 (Datanode) 的硬件配置应该根据不同应用的特点进行选配，使用 Inceptor, Discover 和 Slipstream 配置的用户需要比较高的计算能力的服务器，其数据节点的推荐配置如下：

2 颗 8 核心或以上带超线程 x86 指令集 CPU 的服务器(比如 Intel® Xeon® E5-2650 v2)

128GB 以上内存

2 个 300G 以上的硬盘做 RAID1，作为系统盘

10 个以上的 2TB 容量以上的硬盘作为数据存放硬盘

2 个千兆以上网卡

针对 TDH 测试环境，至少三台服务器，每台配置必须具备以下最低配置：

4 核心 CPU 以上

24G 以上内存

至少两块硬盘，一块至少 500G 以上，一块至少 100G 以上。

1 块千兆以上网卡

### 1.2 操作系统要求

自 5.2 版本起，TDH 支持集群的节点运行在不同的操作系统上。您只需要在添加集群服务器时，添加不同的 RPM 库配置，并为节点分配相应的 RPM 库即可。

以下为 TDH 支持的操作系统：

操作系统	版本
Red Hat Enterprise Linux	7.2, 7.3, 7.4, 7.5, 7.6
CentOS	7.2, 7.3, 7.4, 7.5
SUSE Linux Enterprise Server	12 SP2, 12 SP3

### 1.3 Java 环境要求

注：如果集群的机器上还没有安装 JDK，无需自行安装 JDK——Transwarp Data Hub 安装工具会帮自动安装。如果您的机器上有不同版本的 JDK，可能会和安装工具安装的 JDK 发生冲突。请在安装前运行下面指令来查看您的 JDK 版本。

```
java -version
```

TDH 目前支持以下 JDK 版本：

Oracle JDK 1.8

### 1.4 支持的浏览器

Transwarp Data Hub 采用 Web 页面作为平台管理界面，管理平台支持以下浏览器：

浏览器	版本
Google Chrome（推荐）	71 及以上
FireFox	64 及以上
Safari	11.1.2 及以上，或 12.0.2 及以上

## 2. 操作系统配置

### 2.1 系统磁盘分区要求

系统安装和运行需要占用硬盘空间，在安装前操作系统硬盘必须留出 300GB 空间。对磁盘进行分区时需要遵守以下几点要求：

- 至少要分出 swap 和加载于 “/” 的系统分区。
- 推荐系统分区大小为 200GB~300GB，并将该分区挂载到/目录。
- 请在某数据盘上为 TxSQL 预留不小于 200GB 的空间，并将 TxSQL 的 data dir 设置为该数据盘的某个目录（例如 /mnt/disk1/txsqldata/）。
- 推荐把每个物理磁盘挂载在 /mnt/disknn(nn 为 1 至 2 位的数字) 上不同的挂载点。建议使用 EXT4 文件系统。每个这样的目录会被管理节点自动配置为 HDFS DataNode 的数据目录。
- HDFS DataNode 的数据目录不能放在系统分区，以避免空间不足和 IO 竞争。同时也建议不要将数据分区和系统分区放在同一块磁盘上以避免 IO 竞争。除非整个 HDFS 规划空间不足，否则不要在系统分区所在磁盘上创建数据分区。
- 为了保证 Docker 运行的稳定性，需要专门给 Docker 分配一个磁盘分区，推荐分区大小为 100~300GB。同时为了保障高可用，**推荐** 在非 Docker 所在磁盘预留一个同样大小的空分区作为备用分区，以便在 Docker 所在磁盘发生故障时快速恢复服务。
- 为了保证流畅运行，请尽量把 Docker 卷组划分在数据盘下。如果数据盘较多且存储空间过剩，建议使用单独一块数据盘作为 Docker 分区。否则，建议将一块数据盘的一个分区作为 Docker 分区。

**注：**如果您使用的是 RedHat 或者 CentOS 系统，Docker 分区 必须 使用 XFS 文件系统。其他磁盘或分区 推荐 使用 EXT4 文件系统。更多详细信息请参考 Docker 官方文档。

- 在安装之前需要对 docker 分区进行格式化处理：

- Redhat/CentOS

在 Redhat/CentOS 上，docker 分区必须采用 XFS 格式，实现的步骤如下：

1. 创建目录/var/lib/docker

```
mkdir -p /var/lib/docker
```

2. 对 docker 分区进行 xfs 格式化

```
mkfs.xfs -f -n ftype=1 /dev/<p_name>
```

3. 挂载分区

```
mount /dev/<p_name> /var/lib/docker
```

4. 进行验证，检查是否格式化成功

```
xfs_info /dev/<p_name> | grep ftype=1
```

如果该语句返回结果中有 ftype=1 字样，则说明格式化成功。

5. 配置/etc/fstab

执行语句下述命令查看 UUID：

```
blkid /dev/<p_name>
```

将查到的 UUID 值<UUID>添加在/etc/fstab 中：

UUID=<UUID> /var/lib/docker xfs defaults,uquota,pquota 0 0

#### 例 1. 磁盘资源不富余时的规划

某台机器有两块硬盘，容量都为 600GB，分区及挂载目录将如下，其中/dev/sda1 用作系统分区：

文件系统	大小	挂载目录	文件系统类型
/dev/sda1	368GB	/	EXT4
/dev/sda2	32GB		swap
/dev/sda3	100GB	/var/log	EXT4
/dev/sda4	100GB	（空，Docker 备用分区）	XFS
/dev/sdb1	250GB	/mnt/disk1	EXT4
/dev/sdb2	250GB	/mnt/disk2	EXT4
/dev/sdb3	100GB	/var/lib/docker	XFS

**注意**，这是磁盘资源不富余情形下的规划。如果磁盘资源比较富余，建议操作系统安装在单独磁盘上，防止数据分区与系统分区造成数据读写竞争，如下一个例子。

#### 例 2. 磁盘资源富余时的规划

某台机器有 6 块硬盘，容量都为 600GB，分区及挂载目录将如下，其中一块磁盘用作系统分区：

文件系统	大小	挂载目录	文件系统类型
/dev/sda1	400GB	/	EXT4
/dev/sda2	32GB		swap
/dev/sda3	168GB	/var/log	EXT4
/dev/sdb1	600GB	/mnt/disk1	EXT4

文件系统	大小	挂载目录	文件系统类型
/dev/sdc1	600GB	/mnt/disk2	EXT4
/dev/sdd1	400GB	/mnt/disk3	EXT4
/dev/sdd2	200GB	/var/lib/docker	XFS
/dev/sde1	400GB	/mnt/disk4	EXT4
/dev/sde2	200GB	（空，Docker 备用分区）	XFS

## 2.2 主机名配置

确定每台机器的 IP 地址和主机名。决定如何分配 IP(使用 DHCP 或静态分配)。决定如何解析主机名（使用 DNS 或/etc/hosts）。如果使用/etc/hosts，管理节点将负责更新集群中每台机器的/etc/hosts。

## 2.3 安全设置

禁掉 SELinux 和 iptables(Transwarp Manager 会自动禁掉 SELinux 和 iptables)。

手动关闭防火墙，命令：

```
systemctl stop firewalld
```

禁用系统防火墙，命令：

```
systemctl disable firewalld
```

## 2.4 安装介质准备

注：CDH 安装过程中，会使用 YUM 工具自动安装系统依赖包，为加快安装速度，建议配置本地 YUM 源或国内 YUM 镜像。

CDH 安装介质网络下载地址：<https://nj.transwarp.cn:8180/?p=370>

# Manager 安装包

MANAGER-Basic-Component-Transwarp-6.0.1905a-final.tar.gz

# TDH 除 Discover 外所有产品的产品包

TDH-Image-Registry-Transwarp-6.0.2-final.tar.gz

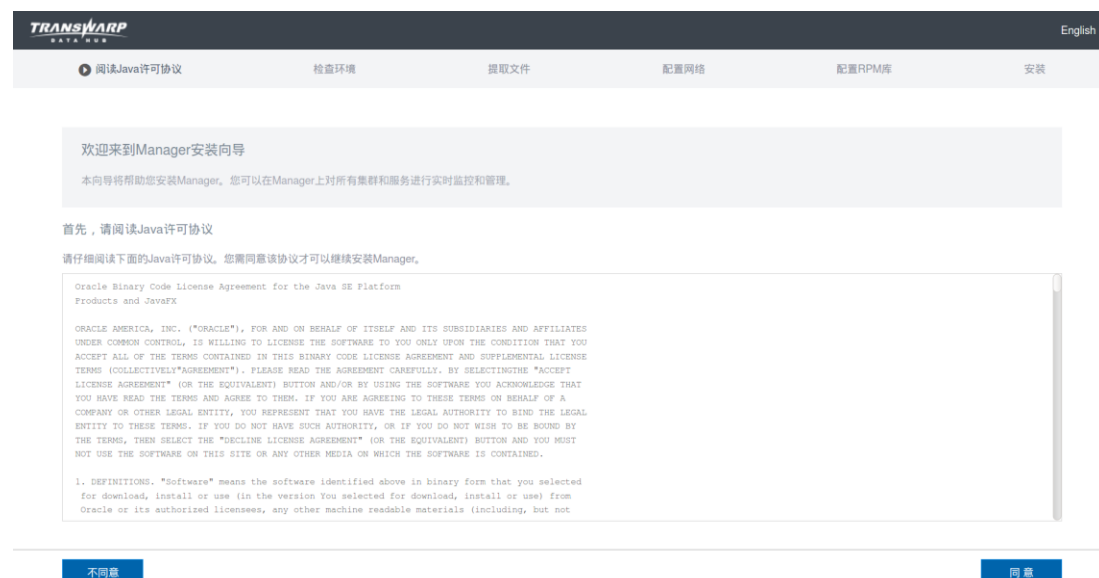
## # Discover 产品包

TDH-Machine-Learning-Transwarp-6.0.2-final.tar.gz

### 3. 安装 Transwarp Manager

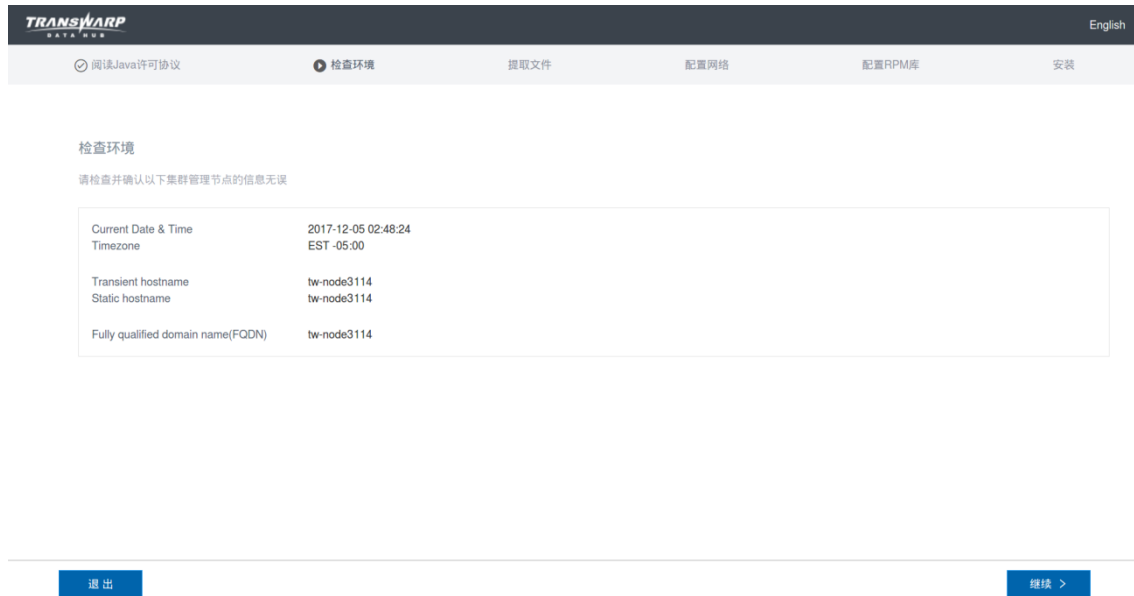
Transwarp Manager 的安装需要先解压安装包，然后运行 Web Installer 使用图形化界面安装。

1. 解压 Transwarp Manager 安装包。在安装包所在的服务器上运行以下命令：
2. *# 解压出安装目录*
3. `tar xvzf <manager-install-pkg>`
- 4.
5. *# 进入解压后的 transwarp 目录*
6. `cd transwarp`
- 7.
8. *# 执行 install 二进制文件*  
`./install`  
<manager-install-pkg>为 Transwarp Manager 安装包的名称，通常为 MANAGER-Basic-Component-Transwarp-<version>.tar.gz，具体名称请参阅“2.4 安装介质准备”小节。
9. 通过浏览器访问管理节点，进入 Web Installer 界面。
10. 系统首先需要您阅读 Java 许可。阅读完毕，点击“同意”进入下一步。

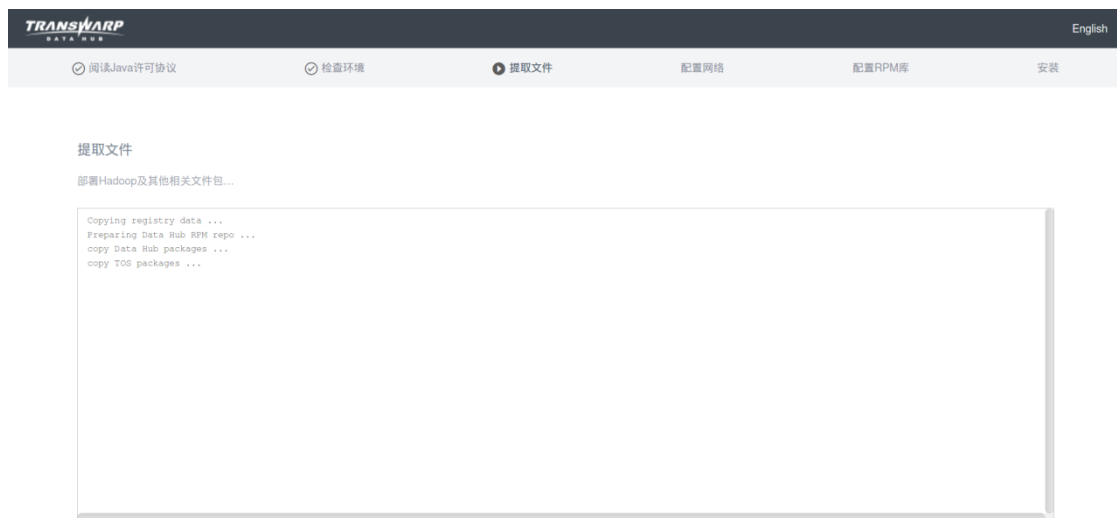


11. 系统将自动检查管理节点的环境配置，主要包括时间、日期、时区及主机名信息并显示在屏幕上，请确认。





## 12. 等待安装



## 13. 取决于管理节点上的网卡数量，您需要进行如下操作：

- 如果管理节点上只有一块网卡，安装结束后，您会被要求设置 Transwarp Manager 端口，推荐默认端口“8180”。

TRANSWARP  
DATA FLOW

English

阅读Java许可协议

检查环境

提取文件

配置网络

配置RPM库

安装

配置网络

请设置部署Manager的端口号，并选择用于集群间通讯的网卡

设置端口号：

请设定部署Manager的端口号 (推荐使用8180)，即您可以访问到Manager的端口号。Manager将监听在所有网卡上的此端口

8180

选择网卡：

请选择用于集群间通讯的网卡。Manager将监听在此网卡IP

退出

继续

- 如果管理节点上有多块网卡，系统会要求您从中选择一块网卡，用于 Transwarp Manager 和集群中其他节点通信。在这一步，您应该选择用于集群内部通信的网卡。

14. 安装 Transwarp Manager 需要一个包含对应版本操作系统的资源库（repo）。如果您的操作系统为 CentOS 或者 Red Hat，您会看到如下提示：

TRANSWARP  
DATA FLOW

English

阅读Java许可协议

检查环境

提取文件

配置网络

配置RPM库

安装

配置RPM库

请选择RPM库的配置方式并填写相应的配置信息。填写完成后点击“应用配置”，应用对RPM库的配置。（注：如果选择“使用远程RPM库”或“创建新的RPM库”，当前的配置将会被清除）

☒ 使用远程RPM库

☐ 创建新的RPM库(DVD)

☐ 创建新的RPM库 (ISO)

☐ 后台手工配置

RPM库的基准地址必须只包含CentOS Linux release 7.2.1511 (Core)RPM包（注：代理信息可为空）

基准地址 \*

代理地址

代理用户名

代理密码

退出

应用配置 >

15. （该步骤中所有操作以 CentOS/Red Hat 系统为例,SUSE 系统的操作完全相同，区别在于如果您使用的是 SUSE 系统，示例步骤中显示“yum”的地方将会在您的操作中显示“zypper”）这一步中，您可以选择是否添加一个新的操作系统资源库。取决于您的选择，您需要进行下面的操作：

- 如果您选择“使用远程 RPM 库”，您需要输入已经创建好的资源库的 URL：

配置RPM库

请选择RPM库的配置方式并填写相应的配置信息。填写完成后点击“应用配置”，应用对RPM库的配置。（注：如果选择“使用远程RPM库”或“创建新的RPM库”，当前的配置将会被清除）

☒ 使用远程RPM库    ☐ 创建新的RPM库(DVD)    ☐ 创建新的RPM库 (ISO)    ☐ 后台手工配置

RPM库的基准地址必须只包含CentOS Linux release 7.2.1511 (Core)RPM包（注：代理信息可为空）

基准地址 \*

代理地址

代理用户名

代理密码

退出

应用配置 >

- 如果您选择“创建新的 RPM 库 (DVD)”，您需要准备好对应版本操作系统的光盘。

配置RPM库

请选择RPM库的配置方式并填写相应的配置信息。填写完成后点击“应用配置”，应用对RPM库的配置。（注：如果选择“使用远程RPM库”或“创建新的RPM库”，当前的配置将会被清除）

☐ 使用远程RPM库    ☒ 创建新的RPM库(DVD)    ☐ 创建新的RPM库 (ISO)    ☐ 后台手工配置

请插入CentOS Linux release 7.2.1511 (Core) Installation Disc 1于DVD驱动器中

退出

应用配置 >

- 如果您选择“创建新的 RPM 库 (ISO)”，您需要准备好对应版本操作系统的 ISO 文件。我们 **建议** 您采用 ISO 镜像方法生成资源库库包。

TRANSWARP  
DATA HUB

English

阅读Java许可协议

检查环境

提取文件

配置网络

配置RPM库

安装

### 配置RPM库

请选择RPM库的配置方式并填写相应的配置信息。填写完成后点击“应用配置”，应用对RPM库的配置。（注：如果选择“使用远程RPM库”或“创建新的RPM库”，当前的配置将会被清除）

☐ 使用远程RPM库

☐ 创建新的RPM库(DVD)

☒ 创建新的RPM库 (ISO)

☐ 后台手工配置

请选择CentOS Linux release 7.2.1511 (Core)安装ISO文件

ISO文件

浏览

退出

应用配置 >

- 。如果您选择“后台手工配置”，您在该步骤无需进行其他操作。您进行选择后，系统会清理资源库缓存：

TRANSWARP  
DATA HUB

English

阅读Java许可协议

检查环境

提取文件

配置网络

配置RPM库

安装

### 安装

正在安装Manager...

```
Creating Repo ...
Loaded plugins: fastestmirror
Cleaning repos: os transwarp
Cleaning up everything
```

16. 资源库缓存清理完毕后，系统会自动开始安装和配置 Transwarp Manager。安装程序会自动安装必需的软件包，全程静默安装，安装配置完成后自动跳转到下一步。
17. Transwarp Manager 安装完成，可以访问提示的安装地址并使用默认的用户名/密码（admin/admin）去登录管理界面继续接下来的配置。

#### 安装完成

您已经成功安装Manager。

接下来，可以在此地址访问Manager：<http://hw-node3114:8180/>，用户名/密码信息为 (admin/admin)

## 4. 配置集群

### 4.1 DNS 配置

如果配置集群中的节点可以通过用主机名互相访问，用户必须配置有效的 DNS 服务器或/etc/hosts 文件。否则请选择需要管理工具配置/etc/hosts，Transwarp Manager 会相应为您配置/etc/hosts 文件，然后添加节点。

TRANSWARP  
DATA HUB

商业版 ⓘ ⓘ admin

集群名称

添加节点

添加服务

1.全局设定

2.编辑机柜

3.编辑RPM或配置

4.节点分配

节点访问: 如果节点间无法通过主机名访问(无论是配置了DNS或者/etc/hosts), Transwarp管理工具会在安装包的时候自动配置/etc/hosts.

节点访问: 

节点间已经可以通过主机名访问

需要Transwarp管理工具配置/etc/hosts

NTP服务器: 使用NTP服务器确保集群中节点时钟同步, 可以指定外部的NTP服务器或者依赖Manager自动选择一台节点作为内部NTP服务器

NTP服务器: 

Manager不管理NTP配置

内部NTP服务器作为独立NTP服务器, 其它节点向内部NTP服务器同步

所有节点向外部NTP同步

内部NTP服务器向外部NTP同步, 其它节点向内部NTP服务器同步

下一步 ▶

### 4.2 NTP 配置

使用 NTP 服务器可以确保集群中节点时钟同步，您可以指定外部的 NTP 服务器或者依赖 Transwarp Manager 自动选择一台节点作为内部 NTP 服务器。

要使用外部 NTP 服务器为您的集群保证时钟同步,您需要确保您的集群可以连接到该外

如果您需要使用外部 NTP 服务器,选择“所有节点向外部 NTP 同步”或“内部 NTP 服务器向外部 NTP 同步,其它节点向内部 NTP 服务器同步”,如下图所示。点击“添加 NTP 服务器”,输入您想要添加的 NTP 服务器,并点击“测试链接”来测试添加的 NTP 服务器的连接。多次点击“添加 NTP 服务器”可以让您添加多台外部 NTP 服务器。

The screenshot shows the 'Add Service' (添加服务) tab in the Transwarp Data Hub configuration tool. The interface is in Chinese and includes a progress bar at the top with four steps: 1. Global Settings (全局设定), 2. Edit Rack (编辑机柜), 3. Edit RPM Repository (编辑RPM库配置), and 4. Node Allocation (节点分配). The current step is 1. Global Settings. Below the progress bar, there is a section for 'Node Access' (节点访问) with two options: 'Nodes can be accessed via hostnames' (节点间已经可以通过主机名访问) and 'Need Transwarp management tool configuration /etc/hosts' (需要Transwarp管理工具配置/etc/hosts). The 'NTP Settings' (NTP服务器) section explains that NTP servers are used to ensure clock synchronization across nodes and can be either external or internal. It provides four radio button options: 'Manager-managed NTP configuration' (Manager-管理NTP配置), 'Internal NTP server as independent NTP server, other nodes sync to internal NTP server' (内部NTP服务器作为独立NTP服务器,其它节点向内部NTP服务器同步), 'All nodes sync to external NTP server' (所有节点向外部NTP同步), and 'Internal NTP server sync to external NTP server, other nodes sync to internal NTP server' (内部NTP服务器向外部NTP同步,其它节点向内部NTP服务器同步). The last option is selected. Below these options are buttons for 'Add NTP Server' (添加NTP服务器) and 'Test Connection' (测试连接). A text input field contains 'cn.ntp.org.cn', and a green checkmark indicates the connection is successful. At the bottom right, there is a 'Next Step' (下一步) button.

全局设定完成后,点击“下一步”。

### 4.3 添加/删除集群服务器

您可以在此步骤添加/删除/修改机柜,添加节点,以及配置 RPM 库。

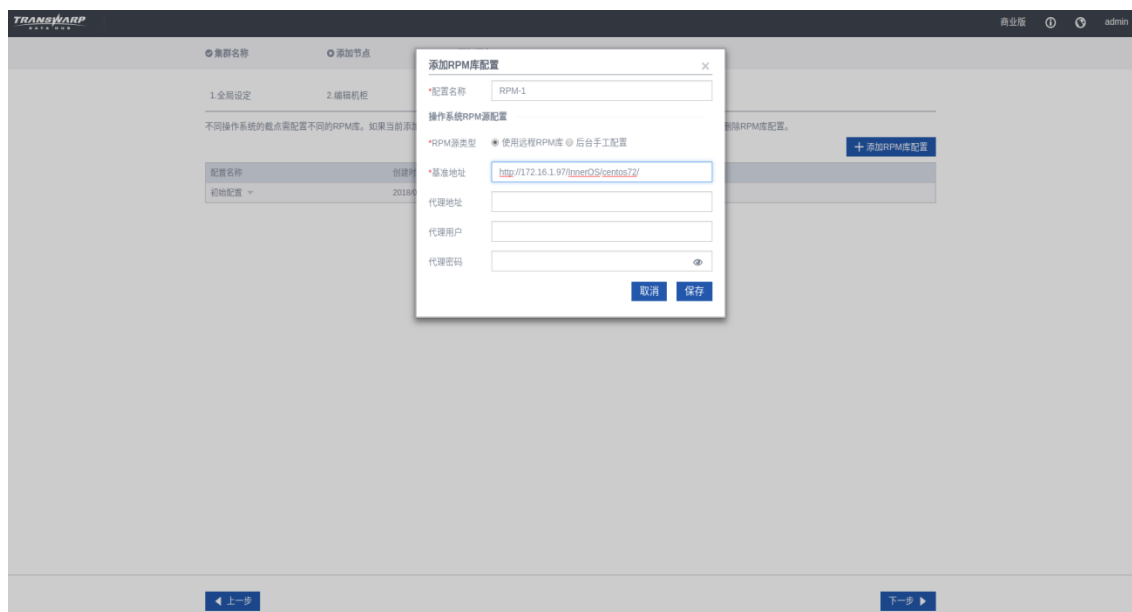
1. 点击机柜的名字和描述可以对它们进行编辑。节点数会随着您之后的配置自动改变。完成后点击“下一步”。



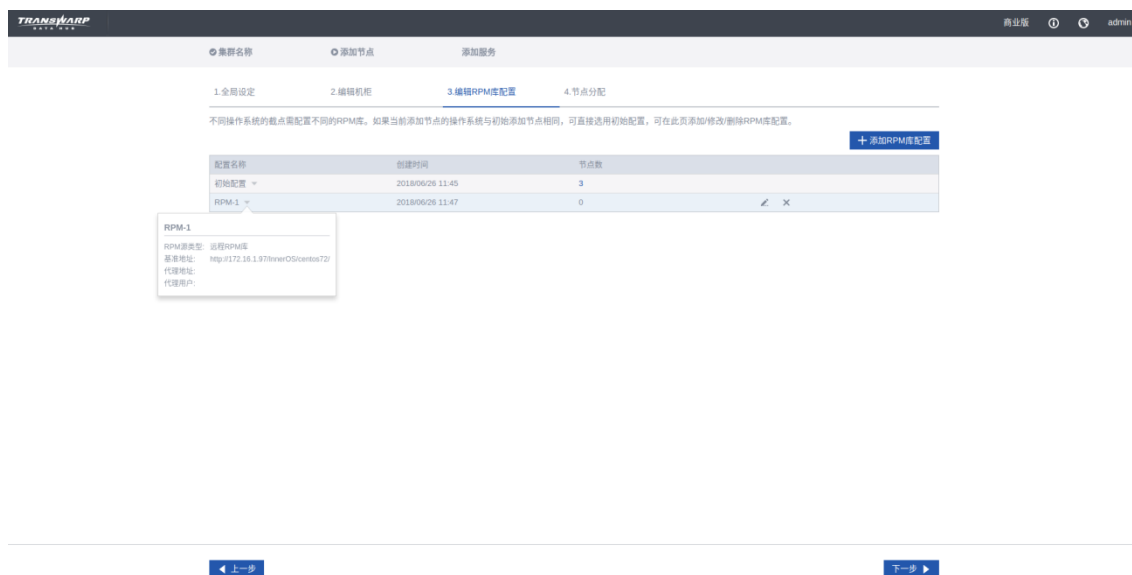
您可以在这一步添加新的 RPM 库。自 5.2 版本起，TDH 支持集群的节点运行在不同的操作系统上，您只需要添加 RPM 库配置，并为节点配置相应的 RPM 库即可。



2. 点击“添加 RPM 库配置”，并填写相关配置：



- 完成后，点击“保存”来保存您的配置。您可以点击相应 RPM 的下拉箭头看到详细配置信息：



- 点击“下一步”，为您的集群分配节点。





5. 点击“添加节点”，在弹出窗口中按照提示输入符合格式的节点 IP，然后点击“搜索”，Transwarp Manager 会搜索您输入的节点：

1.搜索节点

2.提供SSH访问权限

×

请用IP地址搜索节点，IP可以使用如下模式：172.16.1.[110-115], 172.16.1.134, 172.16.1.[156-170]。节点可以多次添加，每次请添加具有相同访问方法的节点。

Q

10.0.16.15, 10.0.16.[17-18]

SSH端口号

22

搜索

<input checked="" type="checkbox"/>	IP地址	已配置SSH访问	已被管理	集群
<input checked="" type="checkbox"/>	10.0.16.15	NO	NO	
<input checked="" type="checkbox"/>	10.0.16.17	NO	NO	
<input checked="" type="checkbox"/>	10.0.16.18	NO	NO	

继续 ▶

6. 点击“继续”，将出现进行 SSH 访问权限配置的界面，在该节点上您可以选择访问节点使用身份，并且提供了两种认证方法：
- 所有节点使用同一密码
- 如果您选择此选项，请在对应的输入框中输入您管理节点的密码、确认密码、输入 ssh 端口号：

1. 搜索节点

2. 提供SSH访问权限

×

安装 Data Hub需要root或sudo访问权限，管理工具使用root或者sudo的权限通过SSH访问节点。

访问节点使用身份：

root

其它用户

可以为已经选择的用户选择使用密码或者密钥来进行访问。

认证方法：

所有节点使用同一密码

所有节点使用同一密钥

\*输入密码：

\*确认密码：

\*SSH端口号：

22

×

完成

。 所有节点使用同一密钥

如果您选择此选项，请点击 **选择文件** 按钮，在目录窗口中选择待上传的密钥文件，并输入 ssh 端口号；也可以输入密钥口令（可选）：

1. 搜索节点

2. 提供SSH访问权限

×

安装 Data Hub需要root或sudo访问权限，管理工具使用root或者sudo的权限通过SSH访问节点。

访问节点使用身份：

root

其他用户

可以为已经选择的用户选择使用密码或者密钥来进行访问。

认证方法：

所有节点使用同一密码

所有节点使用同一密钥

\*私有密钥文件：

选择文件

输入密钥口令：

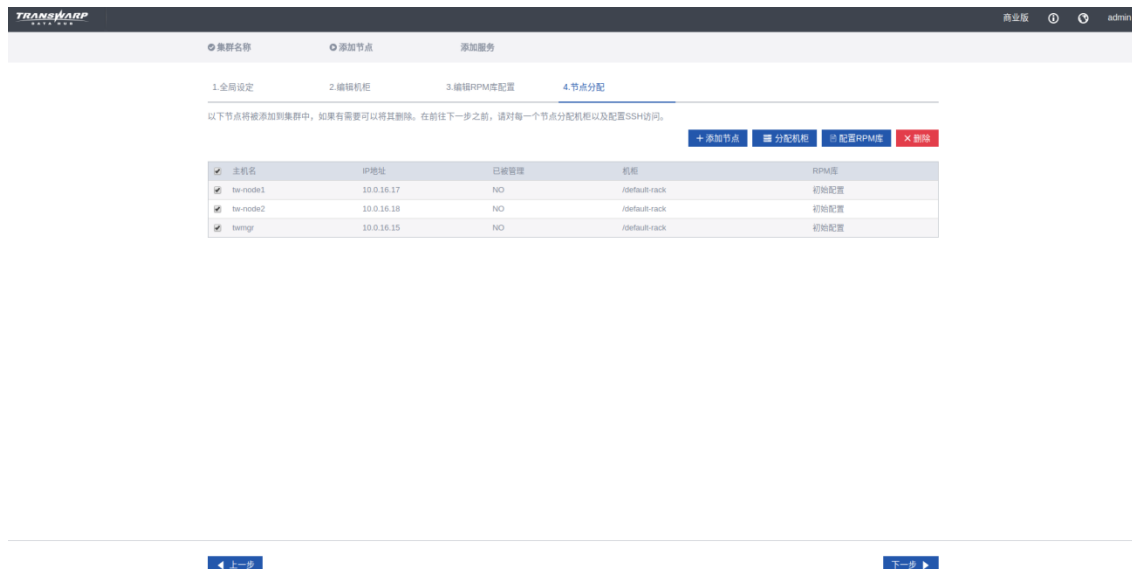
确认密钥口令：

\*SSH端口号：

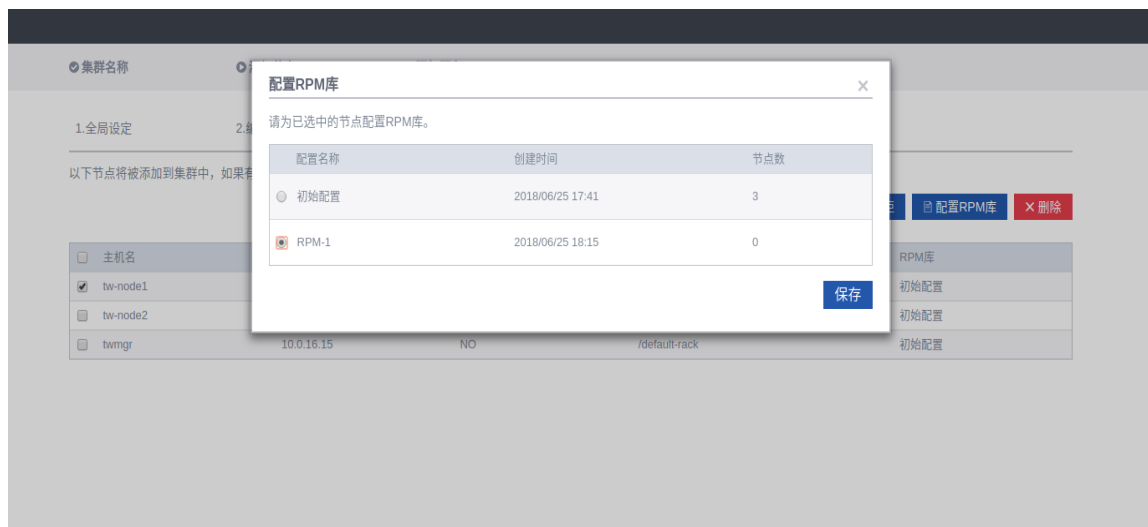
22

完成

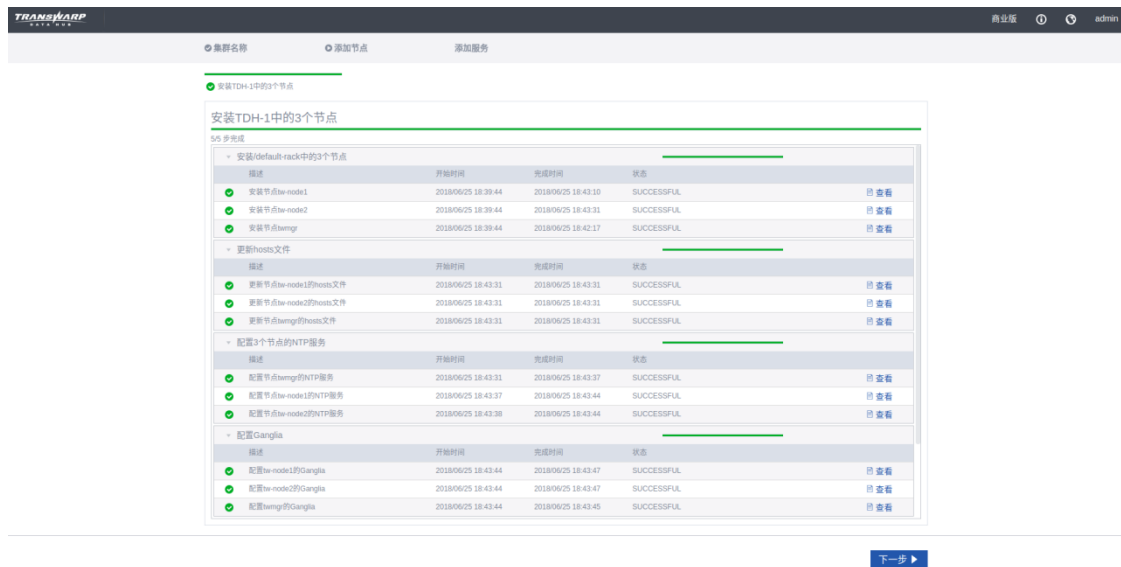
7. 上述配置完成后，点击“完成”。节点便添加完成。如果只有一个机柜，Transwarp Manager 会自动将所有节点分配给默认机柜。如果您有多个机柜，您可以通过点击“分配机柜”来指明节点和机柜之间的分配关系。如果您的集群中有 6 个以上的节点，那么您必须配置机柜信息，方便以后的运维和服务配置。



8. 在这一步您可以为节点分配 RPM 库，点击“配置 RPM 库”，选择一个 RPM 库。默认使用初始配置。节点添加工作完成后，点击“下一步”继续。



9. 系统会开始自动添加节点，添加节点的过程会进行一段时间，成功后您会看到成功页面。



10. 点击“下一步”开始向集群安装服务。

## 5. 安装服务

### 5.1 上传产品包

自 6.0 版本开始，Transwarp Manager 作为一个单独的产品安装，各产品线的产品需要在 Transwarp Manager 的应用市场里先上传产品包，而后 Transwarp Manager 会自动完成解压，取代了之前版本的手动解压步骤，使安装服务更加便捷，并且更加灵活，例如，您可以上传不同版本的产品包，来安装不同版本的服务。

如果您刚刚完成第一个集群的节点安装，您可能会进入以下界面：



在此界面您可以直接点击**上传**来选择产品包上传。您也可以点击**取消**，回到 Transwarp Manager 主界面，稍后再按照以下步骤上传产品包：

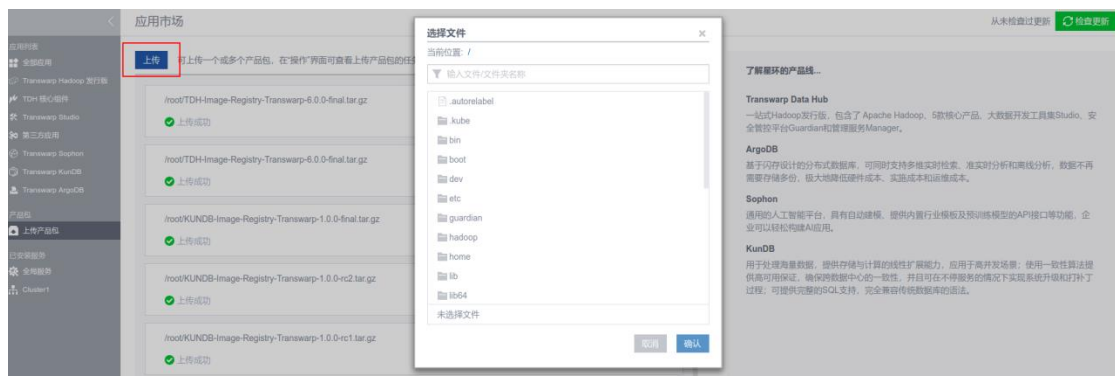
1. 从 Transwarp Manager 菜单点击 **应用市场**，打开应用市场页面：



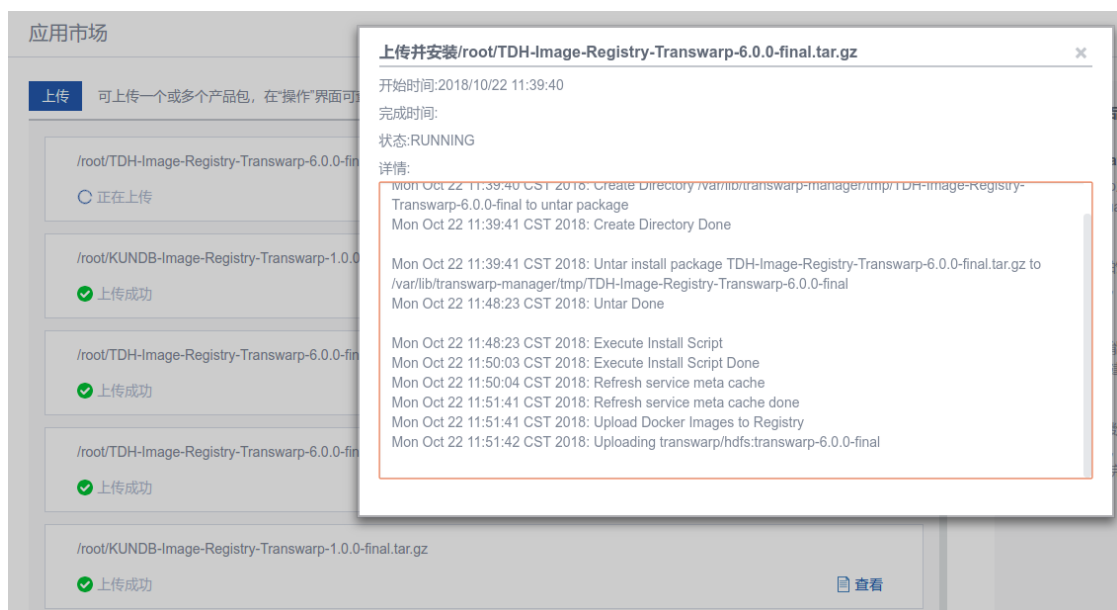
2. 进入应用市场默认页面，从左侧导航栏点击 **上传产品包**。



3. 在上传产品包页面，点击左上角的 **上传**，并在以下弹出窗口中选择产品包的位置：



4. 上传需要一段时间，完成后可以看到状态图标。您也可以点击产品包右侧的 **查看** 来查看上传进程。



5. 上传成功后，相应的服务将会出现在安装服务时的选择服务页面。

## 5.2 选择服务

在安装服务之前请再次确认已按照安装前的检查的要求分配资源，并已成功上传产品包。**只有产品包上传成功的产品才会出现在选择服务页面。**

TDH 中，各个服务将被包装为容器，安装在 TCOS 上。TCOS 是基于 Docker 和 Kubernetes 的云操作系统，在部署安装 TDH 服务时不会显示在服务列表中，默认安装，无需您手动添加。但您可以在后续步骤 **服务角色的分配** 和 **配置服务页面** 为 TCOS 服务分配角色和配置参数、属性等。

- 左侧为服务所属产品，右侧为具体的服务。当某个产品被选择时，属于该产品的所有服务将被选中。

例如假设您选择 TDH 核心组件，那么右侧 Inceptor、Search 等服务将被选中。

集群名称

添加节点

添加服务

1. 上传产品包

2. 选择服务

3. 选择服务版本

4. 分配角色

5. 配置服务

6. 服务总览

选择需要安装的产品或服务

可以同时安装多个服务，当某个产品被选择时，右边框中所有属于该产品的服务将被选中，也可以直接单独选择某个服务

HADOOP

Transwarp Hadoop 包括 ZooKeeper, HDFS, YARN, Kafka 等服务

CORE\_PRODUCT

TDH 核心组件包括 Inceptor, Slipstream, Hyperbase, Search, Discover 等服务

STUDIO

Transwarp Studio 是 tdh 中的大数据开发工具集

THIRD\_PARTY

第三方应用包含 Redis, Codis 和 Hbase 等开源组件

所有服务

定制服务

分别选择需要的服务，当某个服务被选择时，它所依赖的服务会被自动选择。集群安装后仍可以继续添加服务

ZOOKEEPER

ZooKeeper 用于协调同步其他服务

SEARCH

Search 是一个分布式的搜索分析引擎

HDFS

HDFS 是 Hadoop 应用的基本存储系统

YARN

YARN 是资源管理框架

SHIVA

Shiva 是一个通用的分布式存储服务

SLIPSTREAM

Slipstream 是支持 SQL 的实时复杂流处理引擎

SLIPSTREAM\_STUDIO

Slipstream Studio 是图形化的流上任务设计、调试、监控和分析的服务平台

KAFKA

Kafka 是一个分布式的消息发布订阅系统

MILANO

日志收集和检索

TXSQL

TxSQL 是一个分布式的关系型数据库

HYPERBASE

Hyperbase 是实时在线事务处理引擎

INCEPTOR

Inceptor 是基于内存的交互式 SQL 分析引擎

INCEPTOR\_GATEWAY

InceptorGateway 是多 Inceptor 的连接管理以及中转站

PILOT

Pilot 是一个可视化的、直观的、可交互的数据挖掘工具

上一步

下一步

- 如果您选择的服务有依赖的服务，那么会出现提示说明被依赖的服务也会被安装。

## 确认

以下服务被已选服务所依赖，也将被安装: HDFS, ZOOKEEPER, YARN, TXSQL

取消

确认

- 如果您上传了多个版本的产品包，您可以在下一步选择要安装的版本。



本页面为选中的服务确认安装版本，默认已为各服务选中了推荐的版本，您也可以根据需求修改。


SEARCH	transwarp-6.0.0-final
HDFS	transwarp-6.0.0-final
INCEPTOR_GATEWAY	transwarp-6.0.0-final
HYPERBASE	transwarp-6.0.0-final
INCEPTOR	transwarp-6.0.0-final
SHIVA	transwarp-6.0.0-final
STELLARDB	transwarp-6.0.0-final
ZOOKEEPER	transwarp-6.0.0-final
YARN	transwarp-6.0.0-final
SLIPSTREAM	transwarp-6.0.0-final
SLIPSTREAM_STUDIO	transwarp-6.0.0-final

- 您还可以选择“所有服务”，直接安装全部服务。

选择需要安装的产品或者服务


可以同时安装多个服务，当某个产品被选择时，右边框中所有属于该产品的服务将被选中，也可以直接单独选择某个服务

☒




**HADOOP**  
Transwarp Hadoop 包括 ZooKeeper, HDFS, YARN, Kafka 等服务

☒




**CORE\_PRODUCT**  
TDH 核心组件包括 Inceptor, Slipstream, Hyperbase, Search, Discover 等服务

☒




**STUDIO**  
Transwarp Studio 是 tdh 中的大数据开发工具集

☒



**THIRD\_PARTY**  
第三方应用包含 Redis, Codis和Hbase等开源组件

☒



**KUNDB**  
Transwarp KunDB 是一款分布式的交易型数据库

☒

**所有服务**

定制服务

分别选择需要的服务，当某个服务被选择时，它所依赖的服务会被自动选择。集群安装后仍可以继续添加服务

☒

**SHIVA**  
Shiva是一个通用的分布式存储服务

☒

**MILANO**  
日志收集和检索

☒

**HBASE**  
HBase是一个开源的NoSQL数据库

☒

**INCEPTOR**  
Inceptor是基于内存的交互式SQL分析引擎

☒

**PILOT**  
Pilot是一个可视化的，直观的，可交互的数据挖掘工具

☒

**GOVERNOR**  
Governor是元数据管理工具

☒

**CODIS**  
Codis 是一种基于代理机制的Redis集群解决方案

☒

**STELLARDB**  
StellarDB是一款实时的图数据库

☒

**NOTIFICATION**  
Notification是集成的消息接收，存储和转发的中间件

☒

**REDIS**  
Redis是一个开源的高性能内存数据库

☒

**HYPERBASE**  
Hyperbase是实时在线事务处理引擎

☒

**TRANSPORTER**  
TDT是数据整合工具

☒

**WORKFLOW**  
Workflow是图形化的工作流设计、调试、调度和分析的服务平台

☒

**RUBIK**  
Rubik是OLAP Cube的可视化设计工具

☒

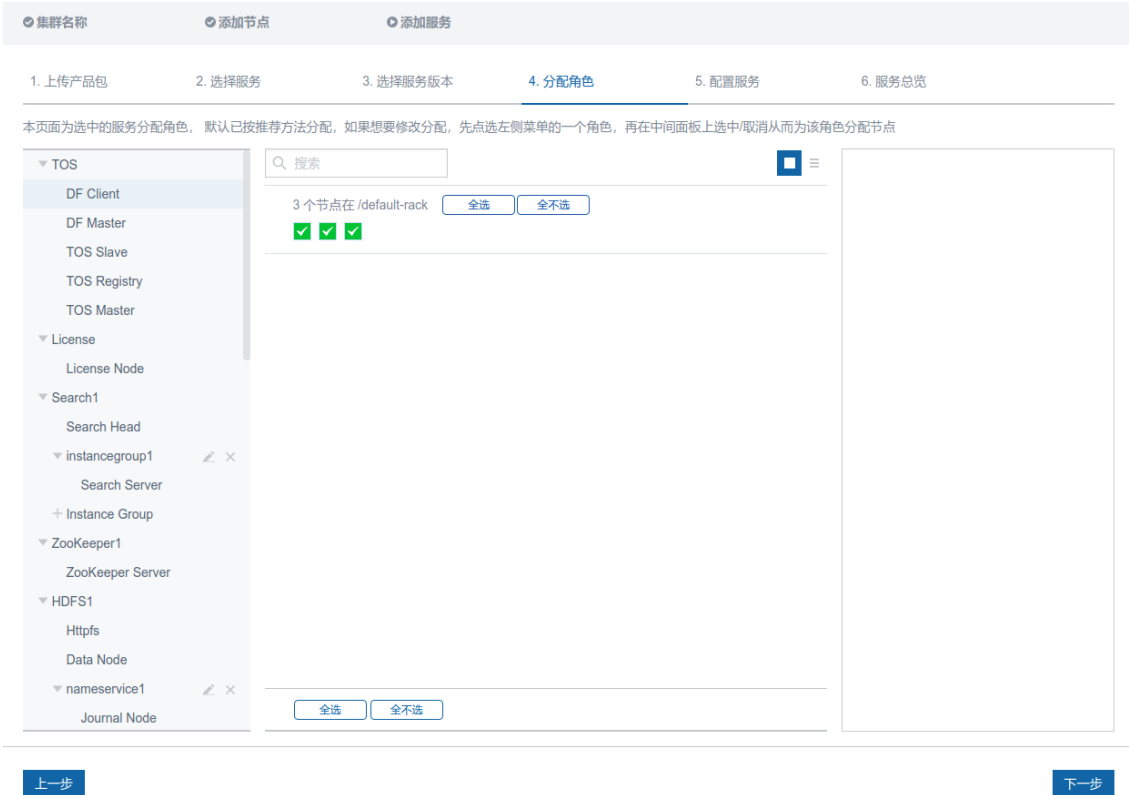
**SLIPSTREAM**  
Slipstream是支持SQL的实时复杂流处理引擎

☒

**KUNDB**  
KUNDB是一个交易型分布式数据库

### 5.3 服务角色的分配

选择想要安装的服务之后，接下来分配各项服务的角色。各服务角色都对应若干个勾选框，每个勾选框代表集群中的一个节点。分配角色的方法为勾选或取消勾选服务右侧对应的节点，被选中的节点将承担该服务角色。例如下图中，三个勾选框被选中代表集群中的这三个节点将承担 DF Client 角色。



Transwarp Manager 对服务的角色提供默认分配，如果您想要使用默认的分配，可以点击“下一步”直接进行配置。但是我们推荐您根据实际的生产场景进行规划，同时我们推荐遵守下面的一些分配原则：

- 首次安装集群时需要安装 TCOS 角色和 License 角色。TOS Registry 角色 **必须** 安装在 Transwarp Manager 所在节点。TOS Slave 角色和 DF Client 角色需要安装在每个节点上。
- 推荐** 将 Zookeeper 角色分配给 **奇数** 个节点，建议集群大小在 20 个节点以下配置 3 个，20 个节点以上配置 5-7 个，不超过 7 个。
- 在集群资源宽裕的情况下，将 HDFS NameNode 和 DataNode 配置给不同的节点，也就是说避免同一节点既是 NameNode 又是 DataNode 的情况。
- 推荐** 将 Hyperbase Master 角色分配给 **奇数** 个节点。
- 分配了 Hyperbase RegionServer 的节点 **必须** 同时拥有 HDFS DataNode 角色。

## 5.4 服务的配置

服务的配置非常重要，请您务必在这一步仔细检查每一项服务的配置。在“配置服务”下，您可以为所有的服务进行配置。您可以通过直接点击属性和参数的值对它们进行修改。

● 集群名称

● 添加节点

● 添加服务

1. 上传产品包

2. 选择服务

3. 选择服务版本

4. 分配角色

5. 配置服务

6. 服务总览

配置选中的服务

TOS

License

Search1

ZooKeeper1

HDFS1

InceptorGateway1

YARN1

Hyperbase1

TxSQL1

Shiva1

Inceptor1

StellarDB1

Slipstream1

SlipstreamStudio1

属性

基础参数

自定义参数

外部配置文件

属性	值
*服务名称	<div>TOS</div>

上一步

下一步

5. 4. 1. 服务的可配置项

Transwarp Data Hub 中的每个服务都自有一系列的可配置项。

通用的可配置项包括（以 Zookeeper 服务为例）：

- 属性：

● 集群名称

● 添加节点

● 添加服务

1. 上传产品包

2. 选择服务

3. 选择服务版本

4. 分配角色

5. 配置服务

6. 服务总览

配置选中的服务

属性

基础参数

自定义参数

外部配置文件

属性	值
*服务名称	<input type="text" value="TOS"/>

TOS

License

Search1

ZooKeeper1

HDFS1

InceptorGateway1

YARN1

Hyperbase1

TxSQL1

Shiva1

Inceptor1

StellarDB1

Slipstream1

SlipstreamStudio1

上一步

下一步

- Transwarp Data Hub 中每个服务都可以配置属性，这里显示的是 Zookeeper 服务的属性配置页。在“属性”下可以设置该服务在集群中的服务名，用于在集群中指代该服务。如果不设置，默认服务名为“服务+整数”，这是因为在集群计算资源允许的情况下，一些服务可以在集群中有多个实例，这些实例工作时需要通过服务名互相区分（如 Zookeeper1 和 Zookeeper2）。在第一次安装集群时，每个服务都只能安装一个实例，额外的服务实例可以向安装好的集群添加。
- 基础参数：

属性	基础参数	自定义参数	外部配置文件
基础参数		值	
tickTime		9000	
initLimit		10	
syncLimit		5	
zookeeper.client.port		2181	
zookeeper.peer.communicate.port		<节点特定, 请点击分别修改>	
zookeeper.leader.elect.port		<节点特定, 请点击分别修改>	
zookeeper.server.memory		1024	
zookeeper.container.limits.memory		-1	
zookeeper.container.limits.cpu		-1	
zookeeper.container.requests.memory		-1	
zookeeper.container.requests.cpu		-1	
zookeeper.memory.ratio		-1	

Transwarp Data Hub 中每个服务都有一系列基础参数可以配置，这里显示的是 Zookeeper 服务的基础参数配置页。如果没有特殊要求，您可以直接使用 Transwarp Manager 对基础参数提供的默认值。基础参数值不仅在这一步可以设置，安装完成后还可以在服务配置页面进行修改。

基础参数分为节点特定参数（例如上图中的 `zookeeper.peer.communicate.port`，`zookeeper.leader.elect.port`）和非节点特定参数（例如上图中的 `tickTime`，`initLimit` 和 `syncLimit`）。

- 对于非节点特定参数，直接点击参数值便可以进行修改。
- 对于节点特定参数，需要点击红框中的“<节点特定, 请点击进行修改>”：



所有服务都可以添加自定义参数。这里显示的是 TCOS 服务的自定义参数配置页。

- 外部配置文件:

您也可以通过外部配置文件的方式添加自定义配置, 点击 **添加** 来选择文件, 并选择需要应用的 Host。

属性

基础参数

自定义参数

外部配置文件

请添加使用的外部服务所需的配置文件

添加

配置文件

操作

请选择文件更新Hosts

选择

5. 4. 2. 特殊配置要求

- Shiva 参数配置

请注意 `tabletserver.store.datadirs` 参数在部署时会默认选择本机所有容量超过 200GB 的挂载点作为数据目录。如本机不存在容量超过 200GB 的挂载点, 请手动指定数据目录。

- 特定角色的配置要求

由于 NameNode, JournalNode, Ganglia, TCOS Master 的 etcd, TxSQL, Guardian ApacheDS, ZooKeeper 等服务或角色对磁盘 IO 要求比较高, 所以要求这些服务或角色目录不能在同一块磁盘上, 并且不能全部在系统盘上, 如果可以, 应尽量都不安装在系统根目录。

上述服务或角色所使用的磁盘配置或路径如下。请检查相关配置项的值是否为正确路径。

服务	角色	配置项/路径
TOS	TOS Master (etcd)	/var/etcd/data/
Ganglia	Gmetad	/var/lib/ganglia/rrds/

服务	角色	配置项/路径
ZooKeeper	ZooKeeper Server	/var/<service_id>/, 如 /var/zookeeper1/
HDFS	Name Node	dfs.namenode.name.dir
	Data Node	dfs.datanode.data.dir
	Journal Node	hdfs-site.xml 中的配置项 dfs.journalnode.edits.dir, 默认为 /hadoop/journal
TxSQL	TxSQL Server	data.dir 和 log.dir
Guardian	ApacheDS	guardian.apacheds.data.dir
	TxSQL Server	data.dir
Transwarp Manager (开启 HA 时)	TxSQL Server	/var/lib/transwarp-manager/master/data/txsql/

## 5.5 服务的安装

配置好参数后，点击“下一步”，在出现的如下所示窗口中可以看到配置信息总览，如果需要修改，可点击“上一步”返回修改。



● 集群名称

● 添加节点

● 添加服务

1. 上传产品包

2. 选择服务

3. 选择服务版本

4. 分配角色

5. 配置服务

6. 服务总览

TOS

基础参数

tos.master.etcd.port	4001
tos.master.etcd.election.timeout	1250
tos.master.etcd.heartbeat.interval	250
tos.master.leader.elect.port	4002
tos.master.apiserver.port	8080
tos.master.apiserver.secure.port	8553
tos.master.scheduler.port	10251
tos.master.controller.port	10252
tos.slave.kubelet.port	10250
tos.slave.kubelet.healthzport	10248
tos.master.dashboard.username	dashboard
tos.master.dashboard.password	password
tos.registry.ui.port	8081
tos.registry.port	5000
service.cluster.ip.range	10.10.10.0/24
cluster.dns.ip	10.10.10.10
dragonfly.enabled	false
dragonfly.supernode.tomcat.http.port	8451
dragonfly.supernode.noinx.peer.port	8452

上一步

下一步

确认配置信息后，点击 下一步 准备执行安装。

此时，Transwarp Manager 会弹出一个“安装确认”窗口确认开始安装。点击“确认”继续。

安装确认

X

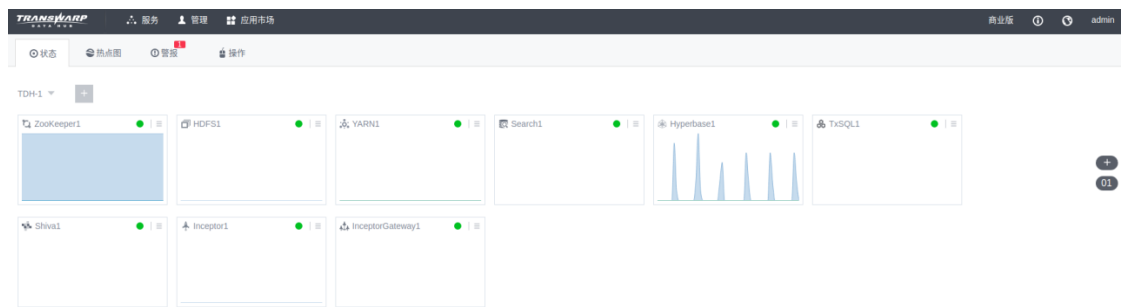
点击确定将不能回退再次修改，如果您已确定完成对服务的角色分配及配置，请点击确认开始安装。

取消

确认

安装开始，系统将花费一些时间安装您所选中的服务，请耐心等待。

安装完成后，点击“完成”进入 Transwarp Manager 的控制面板。



## 5.6 安装 TDH 客户端

服务安装完成后，用户必须安装 TDH 客户端，才可以在交互界面上执行各个服务的命令行，如 `beeline`、`hdfs`、`dfs` 等。

安装过程如下。

### 1. 获取安装包

在 Transwarp Manager 界面上点击 **管理** → **下载客户端**，下载 TDH 客户端。



### 2. 解压文件

将下载的 `tdh-client.tar` 文件放在集群中的任意机器上，然后于该机器上执行如下命令解压文件：

```
tar -xvf tdh-client.tar
```

执行成功后文件会被解压至目录 `TDH-Client` 中。

3. 执行脚本

请执行位于目录 `TDH-Client` 内的脚本 `init.sh`。

`source TDH-Client/init.sh`  
您可以以任何用户执行该脚本，但是我们建议以 `root` 用户身份执行。当以其他用户身份执行时，需要输入 `root` 密码。

**注意** ，用户必须使用 `source` 命令执行该脚本。`bash init.sh` 和 `./init.sh` 都不起作用。

该脚本接受两个参数，默认都是 `y`。第一个参数表示是否修改 `/etc/hosts` 文件（`y` 表示修改，`n` 表示不修改），第二个参数表示是否安装 `Kerberos Client`（`y` 表示安装，`n` 表示不安装）。

因此有如下表格所示的组合方式：

命令	含义
<code>source TDH-Client/init.sh</code>	修改 <code>/etc/hosts</code> 文件，并且安装 <code>Kerberos client</code>
<code>source TDH-Client/init.sh y y</code>	修改 <code>/etc/hosts</code> 文件，并且安装 <code>Kerberos client</code>
<code>source TDH-Client/init.sh y n</code>	修改 <code>/etc/hosts</code> 文件，不安装 <code>Kerberos client</code>
<code>source TDH-Client/init.sh n y</code>	不修改 <code>/etc/hosts</code> 文件，安装 <code>Kerberos client</code>
<code>source TDH-Client/init.sh n n</code>	不修改 <code>/etc/hosts</code> 文件，并且不安装 <code>Kerberos client</code>

当使用 `source` 命令执行完脚本后，如果输出结中没有出现 `ERROR`，则表示成功。此时可以直接执行相关服务的命令。

## 6. 安装和设置问题及回答

### 1. 问题：如何修改 CentOS 和 Redhat 操作系统中的 hostname？

使用 `hostnamectl set-hostname <hostname>` 命令。 **注意：** 安装完成之后尽量不要修改 hostname，很多参数和配置都与 hostname 挂钩；在 `/etc/hostname` 文件中 localhost 必须在 127.0.0.1 的第一个，例如：

```
127.0.0.1 localhost localhost.localdomain localhost4
localhost4.localdomain4
::1          localhost localhost.localdomain localhost6
localhost6.localdomain6
```

### 2. 问题：在安装第一步 hostname 检查中就失败退出 TDH Transwarp Manager 安装，如何解决？

登入到管理节点上执行 hostname 指令和 `/etc/hosts` 里的 IP 后面的名字比较，看名字是否匹配。如不匹配，修改 `/etc/hosts` 文件，确保集群中的所有节点的 hostname 都正确。

### 3. 问题：为 Docker 创建逻辑卷组时，在执行 `pvccreate /dev/xxx` 后出现 `Can't open /dev/vdb exclusively. Mounted filesystem?` 如何解决？

该情况表示分区被占用，需要查看是否挂载文件系统。如果已挂载需要 `umount` 掉。

### 4. 问题：为 Docker 创建逻辑卷组时，在执行 `pvccreate /dev/xxx` 后出现 `"Can't initialize physical volume "/dev/vdb" of volume group "docker" without -ff"` 如何解决？

该情况表示物理卷已经创建，不需要再重复创建。可以直接执行后面步骤。

### 5. 问题：安装后访问 `http://hostname:8180` 出错可能有哪些原因？

1. 可能是安装出现问题：请仔细按照安装文档进行安装。
2. 确保访问方式是 http。
3. 确保 http 的 8180 端口没有被防火墙阻止。

### 6. 问题：节点断开可能有哪些原因？

1. 可能是硬件原因（网线，网口，交换机等问题）。
2. 可能是机器防火墙屏蔽了 ssh。

### 7. 问题：如何检查防火墙是否开启？

1. 在命令行状态下执行 `setup` 命令，按菜单指引防火墙配置→不选防火墙开启，然后保存设置。
2. 在命令行下面执行 `service iptables stop` 关闭防火墙服务。

**问题：**Transwarp Data Hub 对 Hostname 是否有要求？

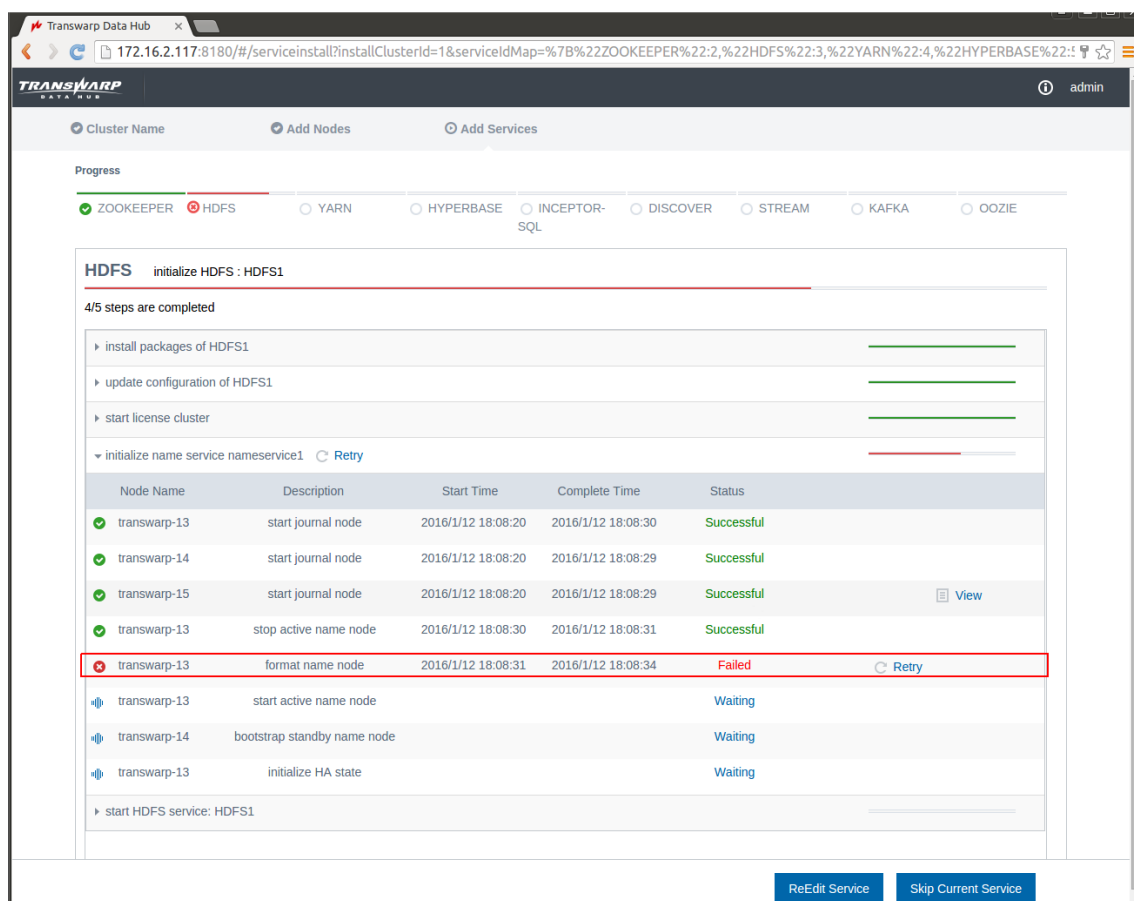
Transwarp Data Hub 对需要添加或安装的节点名有一定的要求，必须是数字，字母或符号组成的，不支持使用\_(下划线)和.(点)的 hostname，比如 TDH1.cluster.com, TDH\_2.cluster.com 等。 用户需要编辑 /etc/sysconfig/network 文件里的 hostname 名字。

### 问题：Transwarp Manager 对网络延迟是否有要求？

Transwarp Manager 对节点和服务添加中对网络的延迟有一定要求，如果客户端浏览器到 Transwarp Manager 的延迟过长会导致添加失败，在客户端使用 ping 程序 ping 一下安装 Transwarp Manager 服务器的延迟，如果超过 2ms，可能会发生安装失败的问题。

### 问题：HDFS Format NameNode 失败怎么办？

该问题表现为安装服务进行到 HDFS 的安装时出现。



解决方法为：确定该集群上是否安装过 HDFS。如果安装过 HDFS，且需要保留其中的数据，请不要继续下面的步骤，而应进行备份或升级操作。如果该集群上的没有需要保留数据的 HDFS，可继续下面的步骤，清空 HDFS 中的残存数据。

- 在所有 Journal Node 上，清空（清空指删除目录中的所有内容，但保留目录，下同）/hadoop/journal 中的所有内容，然后执行：

```
service hadoop-hdfs-journalnode-hdfs1 restart
```

- b. 在所有 Name Node 上，清空 `dfs.namenode.name.dir` 配置的相应目录的所有内容，`dfs.namenode.name.dir` 参数值在 `/etc/<hdfs-service-name>/conf/hdfs-site.xml` 文件中（`<hdfs-service-name>` 为集群上之前安装的 HDFS 服务的服务名，下同。）
- c. 在所有 Data Node 上，清空 `dfs.datanode.data.dir` 配置的相应目录的所有内容，`dfs.datanode.data.dir` 参数值在 `/etc/<hdfs-service-name>/conf/hdfs-site.xml` 文件中。

**问题：Docker 分区所在磁盘发生故障，如何快速恢复服务？**

1. 运行 `mount /dev/<p_name> /var/lib/docker`，将分区要求里预留的 Docker 备用分区 `/dev/<p_name>` 挂载到 `/var/lib/docker`；
2. 运行 `systemctl restart docker` 重启 Docker 服务；
3. 运行 `docker info` 验证重启后的 Docker 是否正常。如果没有报错，则说明 Docker 正常运行。