



Transwarp Data Hub Version 4.6

安装手册

星环信息科技（上海）有限公司

版本号 T00146x-01-010, 2016-08-08

目录

1. 简介	2
1.1. Transwarp Data Hub介绍	2
1.2. 集群结构	2
2. 安装准备	4
2.1. 硬件环境要求	4
2.2. 操作系统要求	4
2.3. Java环境要求	5
2.4. 支持的浏览器	5
3. 安装前的检查	6
3.1. 系统磁盘分区要求	6
3.2. 内存容量要求	7
3.3. 网络设置	8
3.4. 集群和网络拓扑要求	8
3.5. NTP服务设置	8
3.6. 安全设置	8
3.7. 系统的推荐设置	9
4. TDH服务的角色	10
4.1. Zookeeper角色	10
4.2. HDFS角色	10
4.3. YARN角色	10
4.4. Hyperbase角色	10
4.5. Inceptor角色	11
4.6. Stream角色	11
4.7. Sqoop角色	11
4.8. Kafka角色	11
4.9. Oozie角色	11
4.10. Discover 角色	12
4.11. HUE角色	12
4.12. Elasticsearch角色	12
5. 操作系统安装	13
6. Transwarp Manager的安装	14
6.1. 安装前系统配置改动	14
6.2. 安装介质	14
6.3. 安装Transwarp Manager	14
7. 用Transwarp Manager配置集群	22
7.1. 添加集群服务器	24
7.2. 删除集群中的服务器	27
7.3. 其他配置	27
8. 用Transwarp Manager安装服务	32
8.1. Guardian的安装和配置	32
8.2. 选择服务	35
8.3. 服务角色的分配	37

8.4. 服务的配置	38
8.5. 服务安装	44
9. 用Transwarp Manager安装新增服务	46
10. 许可证的激活与认证	51
10.1. 许可证管理	51
10.2. 许可证激活和使用	52
10.3. 查看许可证授权管理	52
10.4. 许可证过期预警	53
11. 卸载TDH	54
11.1. 使用方法	54
11.2. 常见问题	54
12. 安装和设置问题及回答	56
附录 A: 集群服务使用的网络端口	58
A.1. Zookeeper相关端口	58
A.2. HDFS相关端口	58
A.3. YARN相关端口	58
A.4. Hyperbase相关端口	59
A.5. Inceptor相关端口	59
A.6. Discover相关端口	59
A.7. Stream相关端口	59
A.8. Kafka相关端口	59
A.9. Oozie相关端口	59
A.10. Sqoop相关端口	60
A.11. HUE相关端口	60
A.12. Elasticsearch相关端口	60
A.13. Transwarp Manager相关端口	60
A.14. Kerberos相关端口	60
客户服务	62

免责声明

本说明书依据现有信息制作, 其内容如有更改, 恕不另行通知。星环信息科技(上海)有限公司在编写该说明书的时候已尽最大努力保证期内容准确可靠, 但星环信息科技(上海)有限公司不对本说明书中的遗漏、 不准确或印刷错误导致的损失和损害承担责任。具体产品使用请以实际使用为准。

注释: Hadoop® 和 SPARK® 是ApacheTM 软件基金会在美国和其他国家的商标或注册的商标。 Java® 是 Oracle公司在美国和其他国家的商标或注册的商标。 Intel® 和Xeon® 是英特尔公司在美国、 中国和其他国 家的商标或注册的商标。

版权所有 © 2013年-2016年星环信息科技(上海)有限公司。保留所有权利。

©星环信息科技(上海)有限公司版权所有, 并保留对本说明书及本声明的最终解释权和修改权。本说明书 的版权归星环信息科技(上海)有限公司所有。未得到星环信息科技(上海)有限公司的书面许可, 任何人 不得以任何方式或形式对本说明书内的任何部分进行复制、摘录、备份、修改、传播、翻译成其他语言、或 将其全部或部分用于商业用途。

手册版本信息

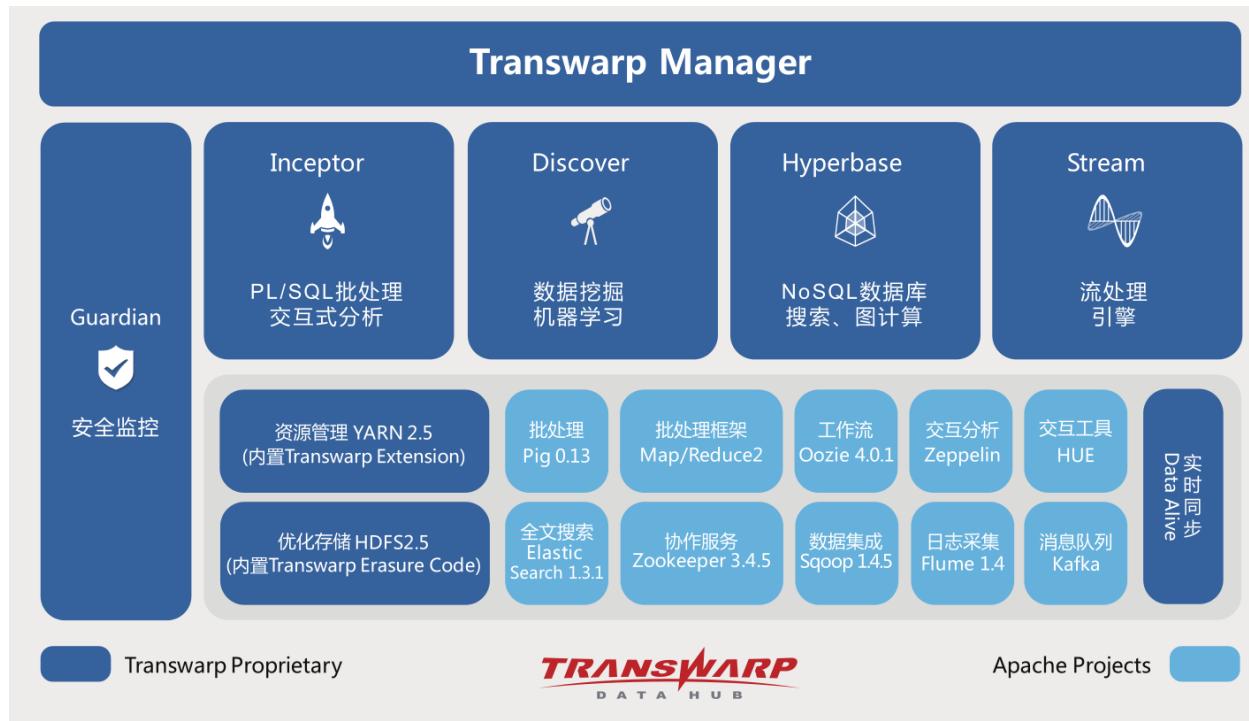
版本号: T00146x-01-010

发布日期: 2016-08-08

1. 简介

1.1. Transwarp Data Hub介绍

Transwarp Data Hub（简称TDH）是国内首个全面支持Spark的Hadoop发行版，也是国内落地案例最多的商业版本，是国内外领先的高性能平台，比开源基于Hadoop MapReduce计算框架的版本快10x~100x倍。TDH应用范围覆盖各种规模和不同数据量的企业，通过内存计算、高效索引、执行优化和高度容错的技术，使得一个平台能够处理GB级到PB级的数据，并且在每个数量级上，都能比现有技术提供更快的性能；企业客户不再需要混合架构，TDH可以伴随企业客户的数据增长，动态不停机扩容，避免MPP或混合架构数据迁移的棘手问题。



星环科技的大数据平台TDH包含五个组成部分：Hadoop企业版，Inceptor分布式内存分析引擎，Hyperbase分布式实时在线数据处理引擎，Stream流处理引擎和Discover机器学习引擎，五个产品组件构成了完整的大数据平台。

在统一的分布式存储之上数据平台上通过YARN提供统一的资源管理调度，结合LDAP与Kerberos提供完备的权限管理控制，不同的部门以及使用租户可以按需创建计算集群访问其授权数据，包括基于Spark计算框架的SQL类统计分析应用与数据挖掘类应用，或者基于MapReduce计算框架的应用。同时，平台通过使用Hyperbase结合Inceptor，为用户提供基于SQL的高并发的查询以及分析能力。在数据导入与交换方面，数据平台提供多样的数据导入与交换形式，包括Flume提供海量数据文件的聚合汇总到HDFS的功能，FTP Over HDFS提供文件通过FTP传入HDFS的通道，Sqoop提供与关系型数据库的数据交换以及Kafka消息队列集群接收实时流数据。此外，TDH数据平台提供实时的流处理能力，通过消息队列Kafka接收实时数据流，做到数据不丢不重，通过Stream基于Spark Streaming提供类似与批处理系统的计算能力、健壮性、扩展性的同时，将数据延时降低至秒级甚至毫秒级。

1.2. 集群结构

创建一个集群前，首先必须对整个集群进行规划，规划包括集群节点、网络、软件模块分布等部分。运维人员必须保证将用于构造集群的服务器满足Transwarp Data Hub系统的基本要求。这些要求包括硬件要求、软件要求、内存要求、网络要求等。第二章将对这些要求进行详细介绍。Transwarp Data Hub通过网络连接所

有服务器并实现整个集群的服务，因此架构师需要仔细计算应用服务对网络带宽和延迟的需求并规划整个网络拓扑，如您需要规划帮助可以咨询星环科技的售前人员和架构师。在应用规划中您需要根据应用的特点决定哪些软件模块和组件需要安装。我们将列出安装前的硬件和软件要求，在满足这些要求的基础上，您才可以开始进行系统安装。系统安装包括以下步骤：

1. 在集群中所有的节点上安装所需要的操作系统。
2. 在管理节点上安装Transwarp Manager。
3. 通过Transwarp Manager来把所有的节点加到集群中、对这些节点部署相关软件和配置，并运行服务。

2. 安装准备

2.1. 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，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

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Stream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

2.2. 操作系统要求

所有集群中的节点，必须运行于同一操作系统。Transwarp Data Hub支持以下操作系统见。

操作系统	版本
Red Hat Enterprise Linux	6.4、6.5、6.6
CentOS	6.4、6.5、6.6
SUSE Linux Enterprise Server	11 SP3、11 SP4
Asianux Server 4 (Hiranya)	SP4

2.3. Java环境要求

如果您的集群的机器上还没有安装JDK，您无需自行安装JDK——我们提供的Transwarp Data Hub安装工具会帮您安装。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
```

如果版本和下面列出的不一样，请和我们联系。

TDH目前支持以下JDK版本：

- Oracle JDK 1.7

2.4. 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)	36.0.1985.143及以上
FireFox	36.0及以上
Safari	7.0.2及以上
Internet Explorer (不推荐)	不支持

3. 安装前的检查

3.1. 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200GB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100GB~300GB，并将该分区挂载到/目录。
- 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I/O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I/O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。

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，文件类型推荐为ext4，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	100GB	/	ext4
/dev/sda2	32GB		swap
/dev/sda3	100GB	/var/log	ext4
/dev/sda4	368GB	/mnt/disk1	ext4
/dev/sdb1	600GB	/mnt/disk2	ext4

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，文件类型推荐为ext4，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
/dev/sdc1	600GB	/mnt/disk2	ext4
/dev/sdd1	600GB	/mnt/disk3	ext4
/dev/sde1	600GB	/mnt/disk4	ext4
/dev/sdf1	600GB	/mnt/disk5	ext4

3.2. 内存容量要求

每个节点必须至少有64GB的RAM。根据节点所安装的Transwarp Data Hub服务，节点可能需要超过64GB的RAM。下表列出在节点上运行不同服务时，该节点所需的额外内存。

服务	要求
Management Server	8GB
HDFS NameNode	32GB
HDFS NameNode	32GB
HDFS Standby NameNode	32GB
HDFS DataNode	4GB
Inceptor Server	4GB
Inceptor executor	32GB
YARN ResourceManager	4GB
YARN NodeManager	4GB
NodeManager分配给Container的计算资源数	用户指定
ZooKeeper	4GB
HBase Master	4GB

具体节点需要内存的计算步骤如下所示：1. 确认所有会在节点上运行的TDH服务。

1. 确认每个服务要求的内存容量。
2. 将所有内存要求相加。
3. 如果相加后内存要求小于64B，则最低内存要求为64GB。如果相加后内存要求大于64GB，则最低内存要求为相加后的和。

比如，如果节点上运行以下服务：

- HDFS DataNode
- YARN ResourceManager
- HBase RegionServer
- YARN NodeManager分配给Inceptor executor为32G

则节点的内存容量要求如下（生产环境实际使用内存要结合具体的应用场景）： 4GB+4GB+32GB+32GB=72GB

3.3. 网络设置

安装Transwarp Data Hub需要最低的网络为千兆以太网。当一台机器上有多个网络适配器时，用户可以在安装Transwarp Data Hub之前对其进行绑定配置。

3.4. 集群和网络拓扑要求

- 决定集群中的节点数目。
- 决定集群中的机柜数目及每个机柜的名称。
- 决定每个机柜中的节点数目。
- 决定每个节点所在的子网（或多个子网）。
- 决定每个节点的主机名和IP地址。
- 决定哪个机器是管理节点。
- 决定哪些机器是NameNode。
- 决定哪些机器是客户端，哪些机器运行TDH服务，或二者兼是。
- 一旦主机名分配给NameNode，主机名则不能再更改。
- 确定你知道要加入到TDH集群中的每个节点的root密码。
- 管理节点必须和集群中的其他节点属于同一子网。
- 决定在集群中使用哪些组件。
- 决定网络带宽和交换机背板带宽。决定交换机型号。
- 决定如何连接到交换机。必须知道需要用到哪些以太网端口和是否需要绑定。
- 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。

3.5. NTP服务设置

决定如何进行时间同步。管理节点将负责所有服务器上的时间的同步，但您需要决定是否使用外部的NTP服务。如果不使用外部NTP服务，集群中所有服务器的时间是相同的，但这个时间有可能不是标准时间，这有可能导致集群与外部连接时产生错误。

3.6. 安全设置

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

3.7. 系统的推荐设置

以下推荐配置可帮助确保TDH集群的性能优化和可管理性。

- 节点的主机名解析。注意，主机名只能由英文、数字和“-”组成，否则之后的安装会出现问题。
- 要同时添加一组节点到集群中。
- 要减少网络延迟，集群中的所有节点都必须属于同一子网。
- 每个节点应配置一块10GE的网卡，用于节点间的通信和执行集群中需要网络连接的任务。
- 如果节点没有使用10GE的网卡，则可使用网卡绑定以便将多个网卡组合在一起以提升网络流量。绑定的网卡必须使用工作模式6。
- 每个节点推荐最小的系统分区，至少有300GB的磁盘空间。
- 每个节点应至少有6T的可用磁盘空间用于HDFS。
- 如果可能，避免将物理磁盘分为多个逻辑分区。除了系统分区外，每个物理磁盘应当仅有一个分区，且该分区包含整个物理磁盘。
- 仅使用物理机器，不要使用虚拟机器。虚拟机可能会明显导致HDFS I/O的缓慢。
- 节点所在的单个或多个子网不允许有其他机器。
- 集群中不能同时有物理机器和虚拟机器。
- 要确保集群中的机器不成为性能和I/O的瓶颈，所有机器必须有相似的硬件和软件配置，包括RAM、CPU和磁盘空间。
- 每个节点应至少有64GB的内存。
- 由于服务可能生成大量日志，推荐将/var/log放置在其他逻辑分区。这可保证日志不会占满根分区的空间。
- 要加快对本地文件系统的读取，可使用noatime选项挂载磁盘，这表示文件访问次数不会被写回。

4. TDH服务的角色

要在集群中运行服务，需要指定集群中的一个或多个节点执行该服务的特定功能。这些功能被归类为功能性的角色，角色用于定义节点能为集群中的服务提供哪些功能。角色分配是必须的，没有角色集群将无法正常工作。在分配角色前，需要了解这些角色的含义，以及对应于服务节点需要哪些角色。

4.1. Zookeeper角色

角色	描述
Zookeeper	一个或多个运行Zookeeper服务的节点。Zookeeper服务是指包含一个或多个节点的集群提供服务框架用于集群管理。对于集群，Zookeeper服务提供的功能包括维护配置信息、命名、提供Hyperbase的分布式同步，以及当HMaster停止时触发master选择。Hyperbase需要有一个Zookeeper集群才能工作。推荐在Zookeeper集群中至少有3个节点。

4.2. HDFS角色

角色	描述
NameNode	HDFS系统中的节点用于维护文件系统中所有文件的目录结构并跟踪文件数据存储于哪些数据节点。当客户端需要从HDFS文件系统中获得文件时，它通过和NameNode通讯来知道客户端哪个数据节点上有客户端需要的文件。 一个集群中只能有一个NameNode。NameNode不能被赋予其他角色。
DataNode	在HDFS中，DataNode是用来存储数据块的节点。
Standby NameNode	Standby模式的NameNode元数据（Namespcae information 和 Block locations map）都是和Active NameNode中的元数据是同步的，一旦切换成Active模式，马上就可以提供NameNode服务。
JournalNode	Standby NameName和Active NameNode通过JournalNode通信，保持信息同步。

4.3. YARN角色

角色	描述
ResourceManager	ResourceManager负责将各个资源部分（计算、内存、带宽等）安排给基础NodeManager，并与NodeManager一起启动和监视它们的基础应用程序。
TimelineServer	提供对YARN的作业历史日志信息的展现服务。
NodeManager	NodeManager管理一个YARN集群中的每个节点。提供针对集群中每个节点的服务，从监督对一个容器的终生管理到监视资源和跟踪节点的状态。

4.4. Hyperbase角色

角色	描述
Master	节点实现Hyperbase数据库以下功能：配区域给RegionServers，平衡整个集群。 1. 确认RegionServer在运行中。 2. 如果没找到参考值，删除parents。 3. 管理数据库表。 4. 在RegionServers中分发消息。 5. 监控RegionServers以决定是否有必要执行恢复，如果是，则执行恢复。可以分配HMaster角色给一个或多个节点以进行备份切换。如果你分配角色给多个节点，这将创建一个active-standby状态的HMaster节点的集群，即一个节点处于active状态而集群中的另一个节点处于standby状态。如果active状态的HMaster停止，Zookeeper集群将选出一个inactive状态的HMaster来作为active状态的HMaster。
RegionServer	负责服务和管理Hyperbase区域的节点。
Hyperbase Thrift	Thrift Client API开放的节点，客户端可通过Thrift和Hyperbase通讯。
HBase-rest	Hyperbase的restAPI。

4.5. Inceptor角色

角色	描述
MetaStore	存储表结构及其元数据信息
Inceptor Server	Inceptor运行的节点，帮助查询和管理HDFS的大型数据集。Inceptor Server提供工具从文件系统中提取、转换和装载数据，实施数据格式结构，以及提供HDFS或Hyperbase的直接文件访问。

4.6. Stream角色

角色	描述
JobServer	Stream的任务管理工具

4.7. Sqoop角色

角色	描述
Sqoop	Sqoop是一款用于在(Inceptor, Hyperbase)与传统的数据库(MySql、Postgresql...)间进行数据的传递的工具

4.8. Kafka角色

角色	描述
Kafka Server	Kafka的任务管理工具。

4.9. Oozie角色

角色	描述
Oozie server	Oozie的任务管理工具。

4.10. Discover 角色

角色	描述
RStudio	RStudio是一个R的集成开发环境(IDE)，它包括一个控制台和语法编辑器，支持直接代码执行，提供绘图、历史、调试和工作区管理等工具。

4.11. HUE角色

角色	描述
HUE server	HUE是一组可与您的集群交互的网络应用程序。HUE应用能让您浏览HDFS和工作，管理metastore，运行Inceptor，浏览Sqoop出口数据，提交MapReduce程序，构建自定义的搜索引擎与Solr，与Oozie一起调度重复性的工作流。

4.12. Elasticsearch角色

角色	描述
Elasticsearch server	上的Elasticsearch是一个开源、独立、子包含的小型库，可与上的工作进行交互。

5. 操作系统安装

在安装Transwarp Data Hub之前，集群中的所有节点必须满足[安装准备](#)和[安装前的检查](#)中所列举的所有要求。

可以使用两种方式来安装集群中的服务器的操作系统，单独安装方式和PXE安装方式。

其中单独安装方式是使用Red Hat Enterprise Linux for Servers、CentOS系列或SUSE Linux Enterprise Server的安装光盘在每台服务器上独立安装操作系统。

注意：

对于Red Hat Enterprise Linux for Servers、CentOS 系列的操作系统，需要使用visudo命令打开/etc/sudoers配置文件：

```
# visudo
```

将该文件中的“Defaults requiretty”注释掉 如要使用非root用户安装产品，使用的用户必须拥有sudo权限。所以，管理员需要向/etc/sudoers文件中添加sudo用户。方法为在该文件以下部分最后一行的内容：

i

```
## Next comes the main part: which users can run what software on
## which machines (the sudoers file can be shared between multiple
## systems).
## Syntax:
##
##       user      MACHINE=COMMANDS
##
## The COMMANDS section may have other options added to it.
##
## Allow root to run any commands anywhere
root      ALL=(ALL)          ALL
sudo      ALL=(ALL)          NOPASSWD: ALL
```

6. Transwarp Manager的安装

本章节将详细描述如何在管理节点上安装Transwarp Manager。

6.1. 安装前系统配置改动

如果您有DNS，那么您可以直接跳到下一节。如果您没有DNS，在安装前，请打开管理节点下的/etc/hosts文件，查看该文件是否包含管理节点的hostname和它IP地址的对应，例如：

```
172.16.2.125 tw-node2125
```

如果/etc/hosts文件不包含这一行，请添加。

登入到管理节点使用 **hostname** 命令检查管理节点的名字并保证名字和上述/etc/hosts里设置的名字相同：

```
#hostname  
#tw-node2125
```

注意，/etc/hosts文件的第一行必须为127.0.0.1的记录，不能将此行注释掉，不能将当前主机名写在该行中。通常如下：



```
127.0.0.1 localhost
```

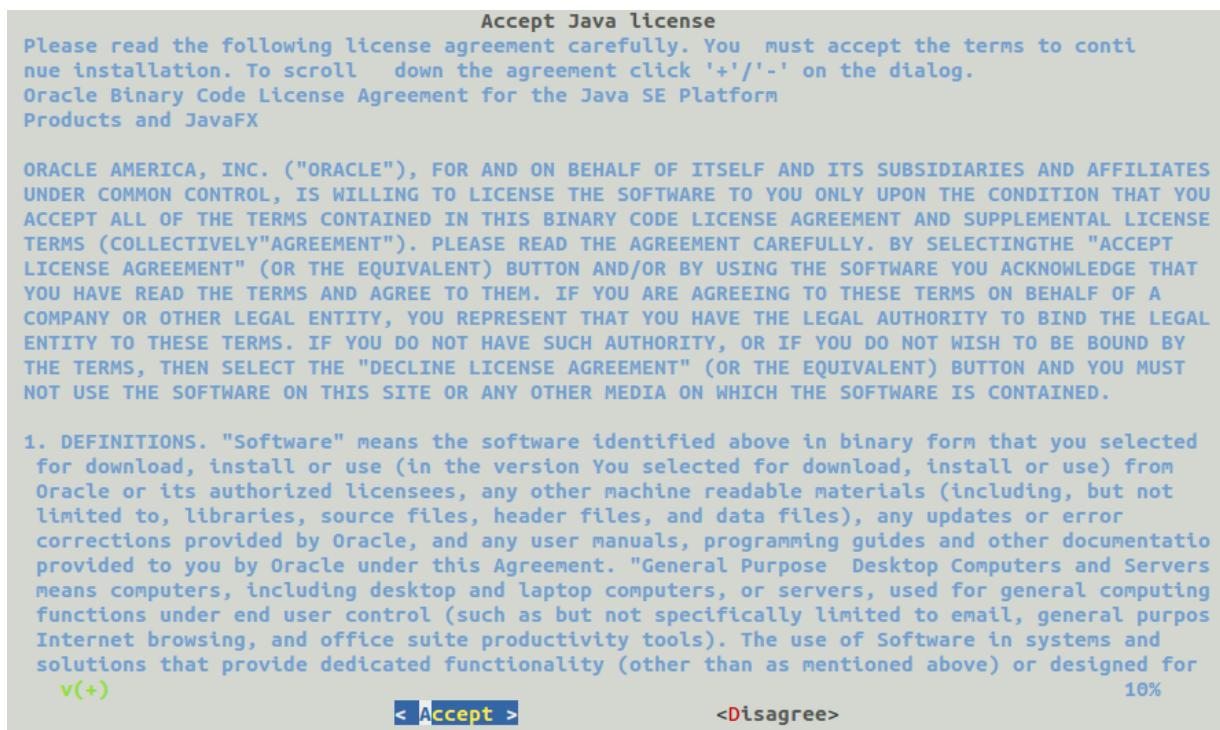
6.2. 安装介质

如果有安装光盘可将安装光盘插入服务器CD/DVD ROM中直接运行。由于安装过程中会使用到操作系统的安装介质，建议用户将OS的ISO文件事先拷贝到管理服务器上。如客户服务代表或技术支持提供的是压缩包，请将压缩文件导入服务器中，解压后 **cd** 进入目录运行下面的 **install**，例如：

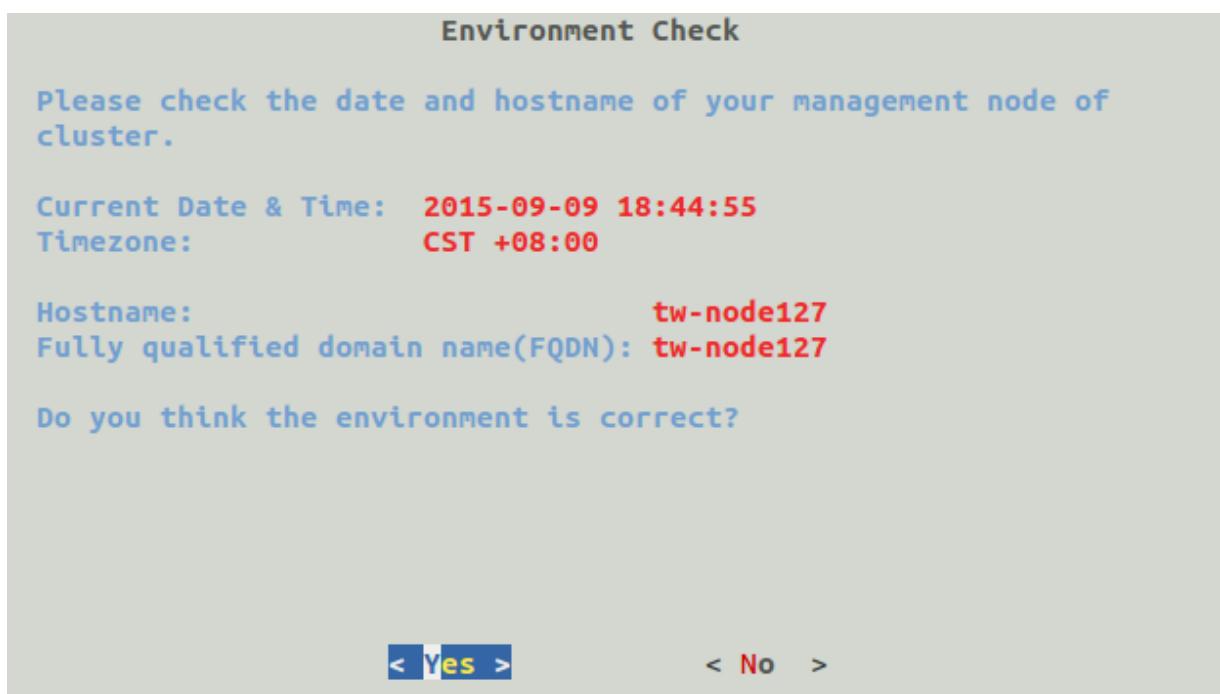
```
$ cd ${package} #进入解压后的目录  
$ ./install #执行install二进制文件
```

6.3. 安装Transwarp Manager

1. 系统首先需要您阅读Java许可。阅读完毕，点击“Accept”进入下一步。

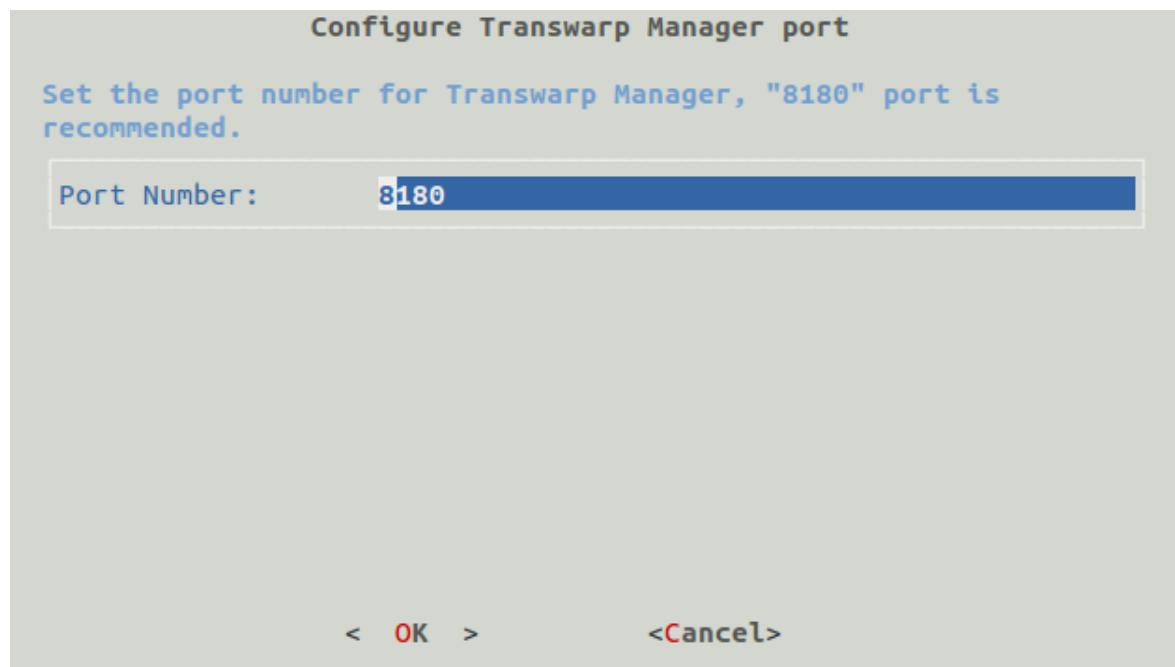


- 系统将自动检查管理节点的环境配置，主要包括时间、日期、时区及主机名信息并显示在屏幕上，请确认。

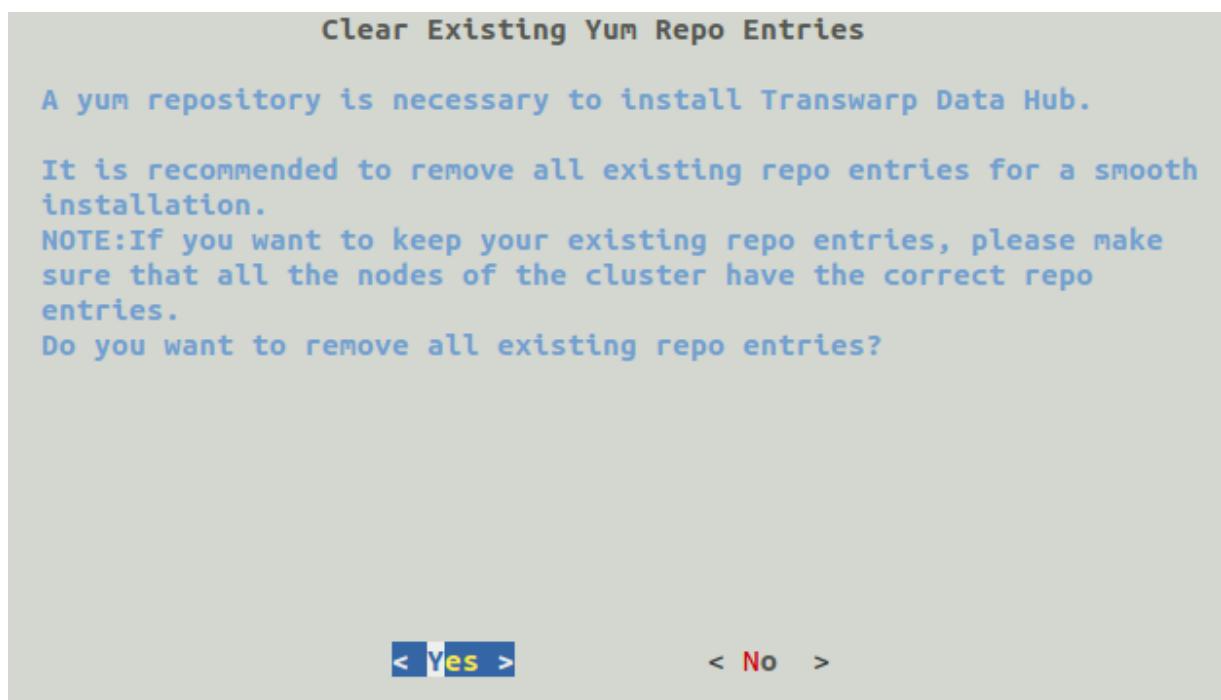


- 取决于管理节点上的网卡数量，您需要进行如下操作：

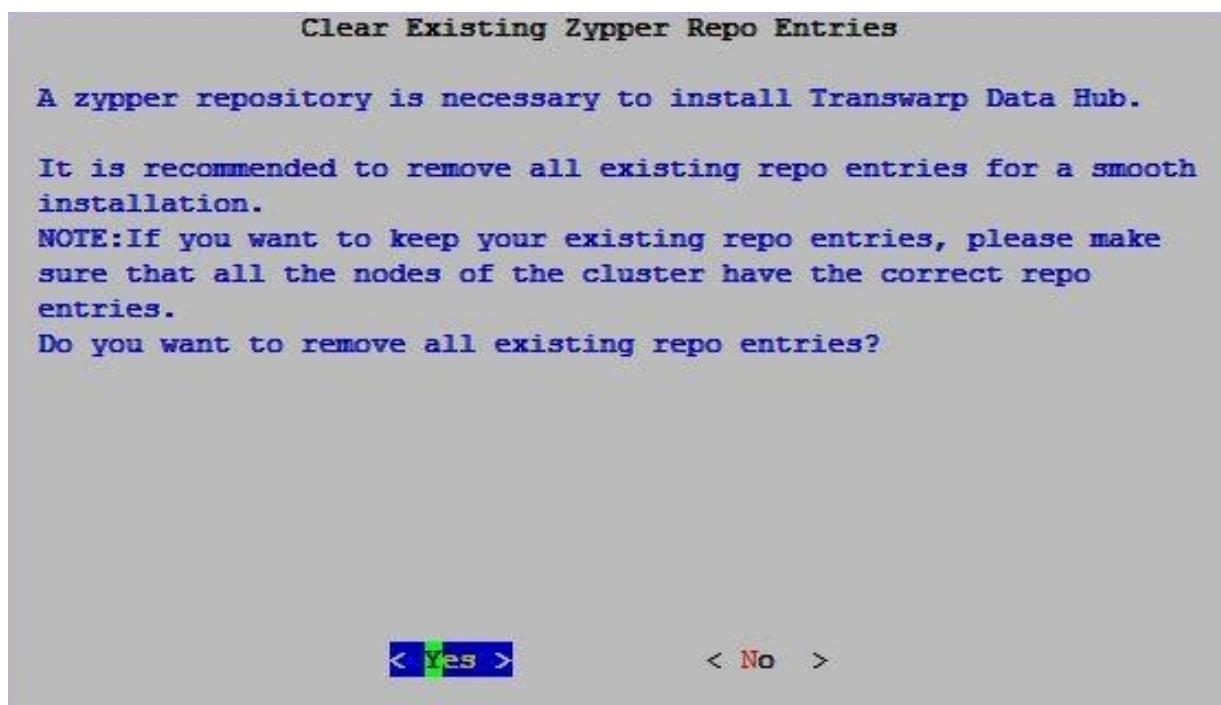
- 如果管理节点上只有一块网卡，安装结束后，您会被要求设置Transwarp Manager端口，推荐默认端口“8180”。



- 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。
4. 安装Transwarp Manager需要一个包含对应版本操作系统的资源库(repo)。这一步中，您需要决定是否清除当前集群上已有的资源库配置。如果您的操作系统为CentOS或者Red Hat，您会看到如下提示：



如果您的操作系统为SUSE，您会看到如下提示：

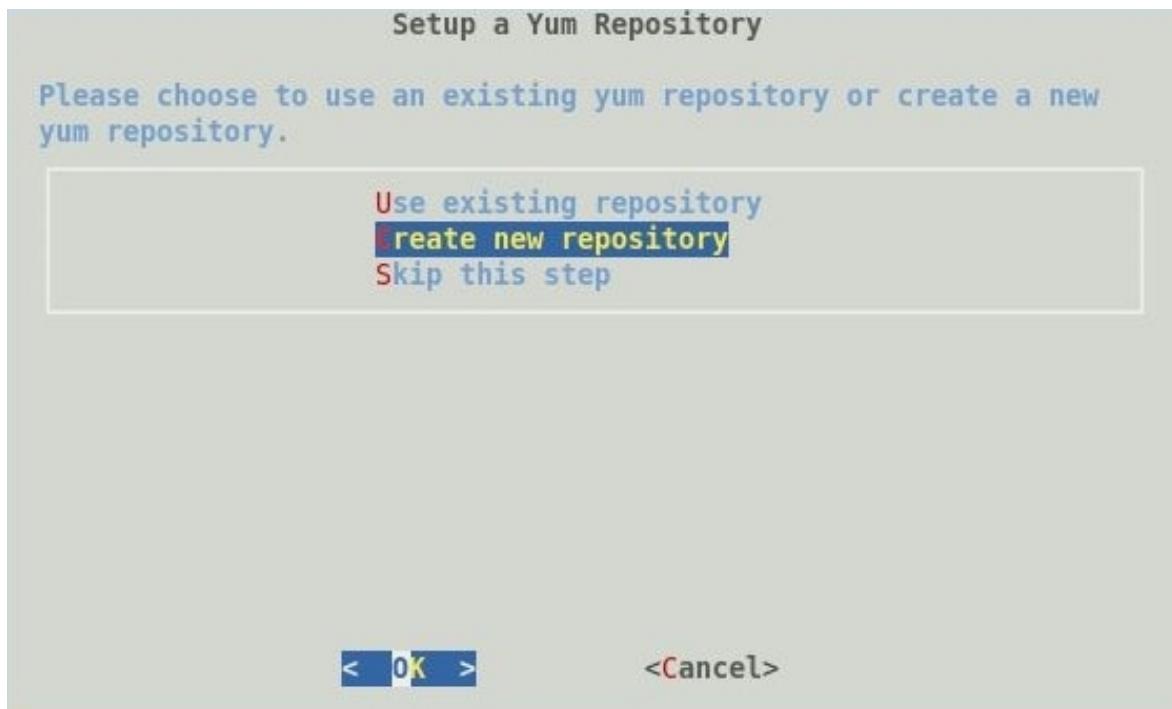


在这一步，取决于您的集群是否部署在公有云上，您的操作会有所不同。

- 如果您的集群 不是部署在公有云上，推荐 选择“Yes”将集群上原有的资源库配置清除，在下一步重新配置，以保证平滑的安装。
 - 如果您的集群 部署在公有云上，您 不能 清除已有的资源库配置，必须选择“No”。
5. (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。取决于您在上一步的操作，这里您看到的提示将会略有不同：
- 如果您在上一步中选择了“Yes”清除了原有的资源库配置，那么您在这一步 必须 添加一个资源库。您会看到下面两个选项：



- Use existing repository: 使用已经创建好的资源库
- Create new repository: 创建一个新资源库
- 如果您在上一步中选择了“No”保留了原有的资源库配置，您会看到三个选项：



- Use existing repository: 添加已经创建好的资源库
- Create new repository: 创建并添加一个新资源库
- Skip this step: 跳过该步骤，不添加资源库

因为您在上一步保留了原有的资源库配置，您在这一步可以添加资源库（“Use existing repository”或者“Create new repository”），也可以选择不添加资源库（“Skip this step”）。

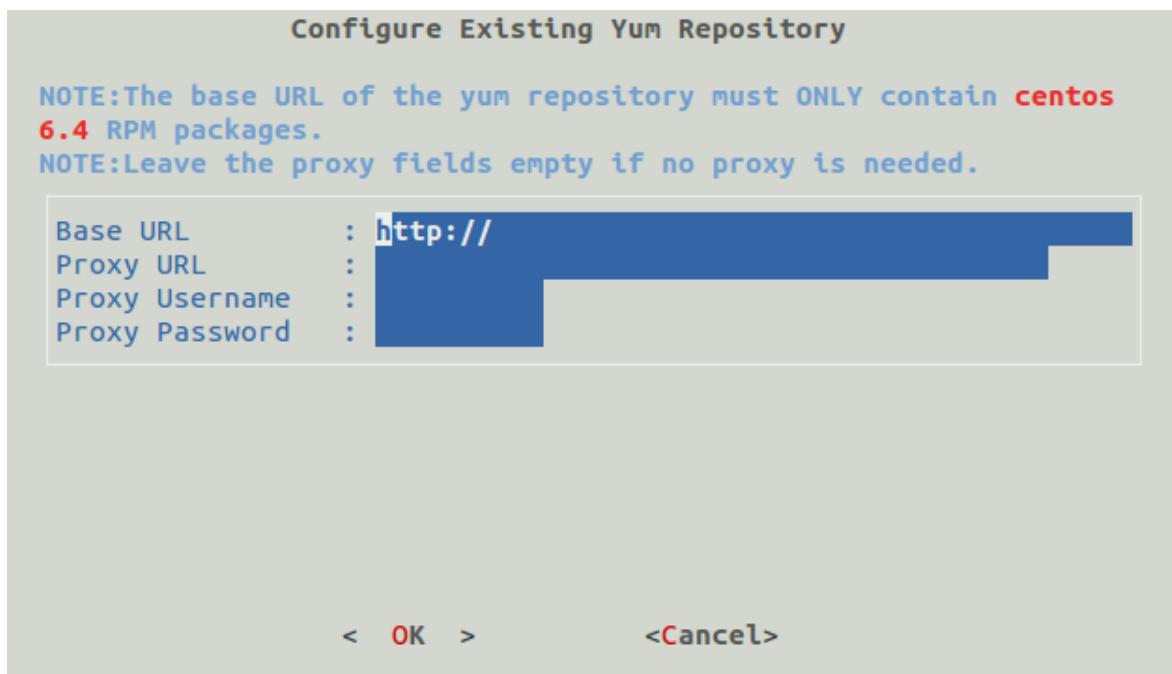
版本信息



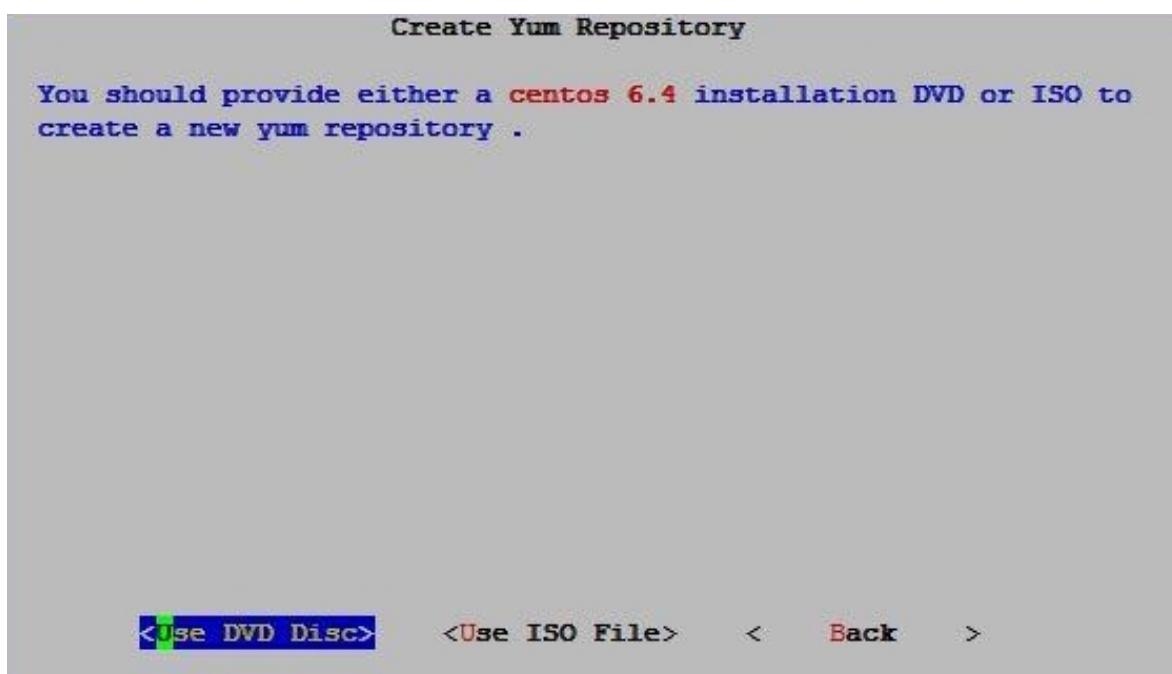
如果您使用的Transwarp Data Hub是4.3.7之前的4.3.x版本或者4.5.3之前的4.5.x版本，您不能选择“Skip this step”，必须选择另外两个选项。这是一个已知问题，在4.3.7, 4.5.3和4.6.0之后的版本中已经修复。

取决于您的选择，您需要进行下面的操作：

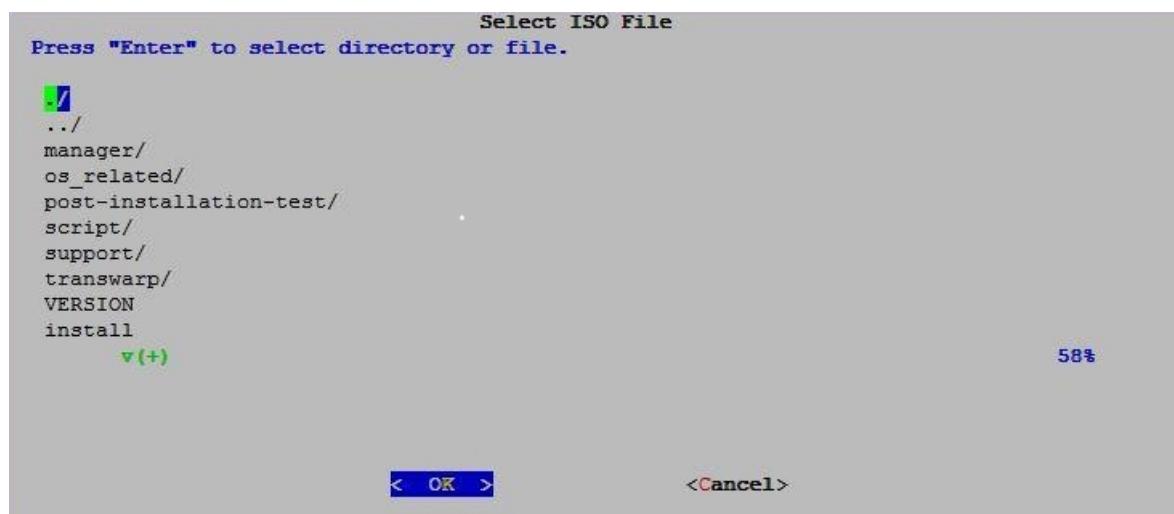
- 如果您选择“Use existing repository”，您需要输入已经创建好的资源库的URL：



- 如果您选择“Create new repository”，您需要准备好对应版本操作系统的ISO文件或光盘。我们建议您采用ISO镜像方法生成资源库库包。使用镜像生成时，您需要将OS的ISO文件copy至服务器用户目录中，然后点击“Use ISO File”：

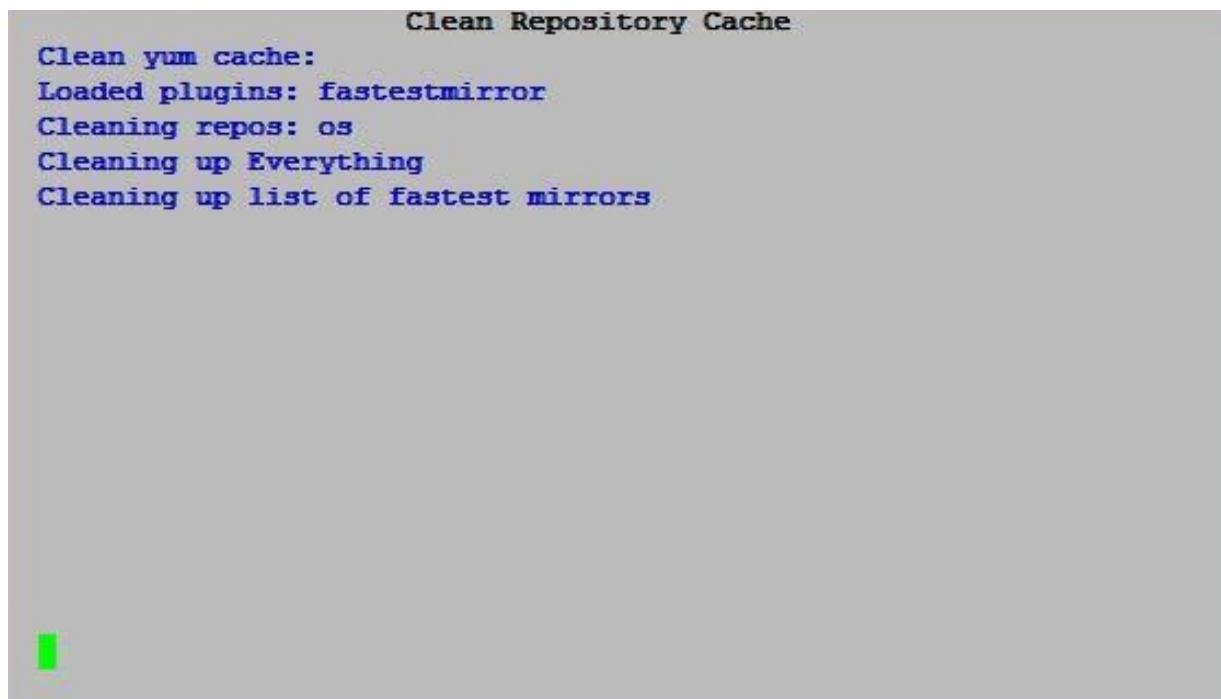


并在“Select ISO File”中选择ISO文件所在的目录：



- 如果您选择“Skip this step”，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：



- 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。
- 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。

Installation Completed

Transwarp Manager has been installed successfully.

You can visit Transwarp Manager at

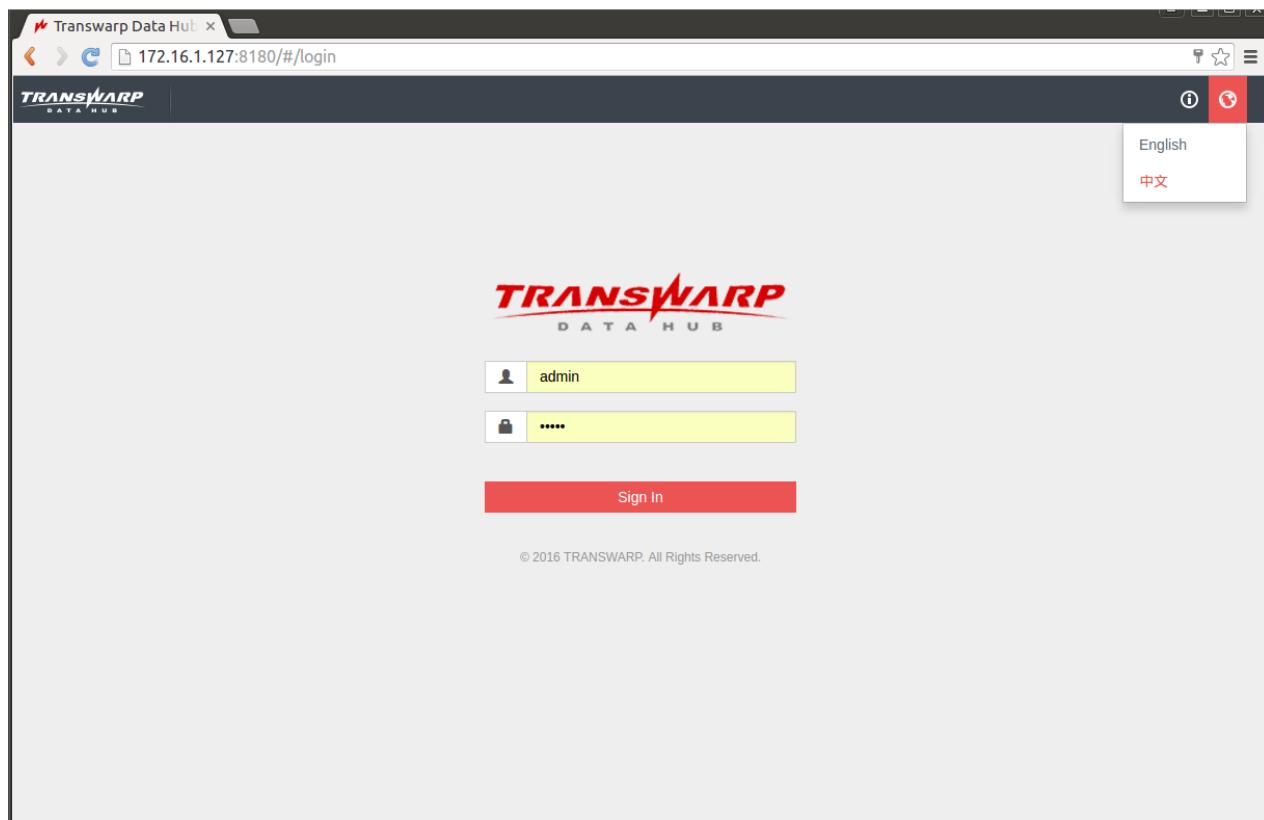
<http://tw-node127:8180>

by username/password(**admin/admin**).

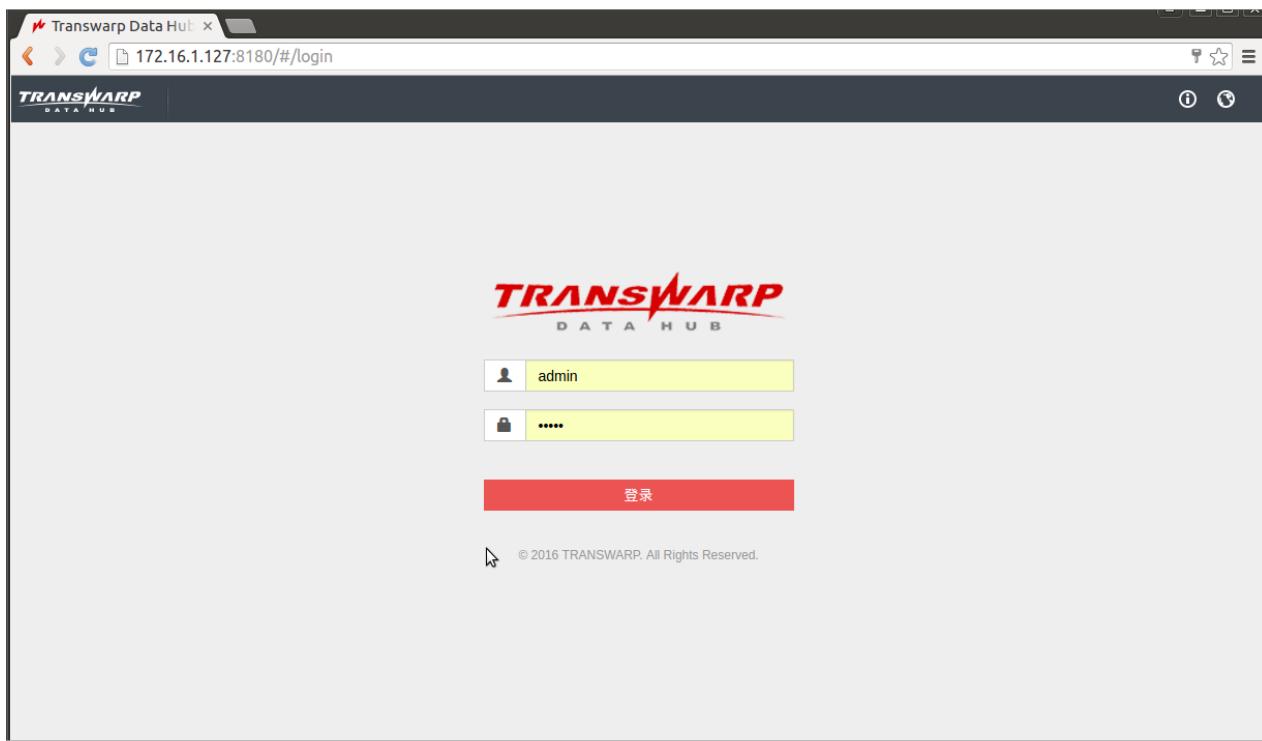
<**Finish**>

7. 用Transwarp Manager配置集群

打开客户端浏览器（推荐使用Google Chrome浏览器），输入安装好的管理节点IP或DNS地址，比如`http://192.168.1.10:8180/`（192.168.1.10是管理节点的IP地址）。访问这个地址，您会看到下面的登陆页面。Transwarp Manager会试着根据您的系统判断您使用的语言，如果当前显示的语言和您想要使用的不同，可以点击右上角的(C)来选择语言。目前Transwarp Manager支持中文和英文：



以admin的身份登录，密码也是admin。



登陆后，您需要先接受最终用户协议才可以进行进一步操作。

Transwarp Data Hub最终用户协议

请仔细阅读下面的协议

最终用户许可协议

重要须知—请认真阅读：本《最终用户许可协议》（以下称“协议”）是您（个人或单一实体）与星环信息科技（上海）有限公司之间有关Transwarp Data Hub软件产品的法律协议。本“软件产品”包括计算机软件，并可能包括相关媒体、印刷材料，和“联机”或电子文档（“软件产品”）。本“软件产品”还包括对星环信息科技（上海）有限公司提供的您的原“软件产品”的任何更新和补充资料。任何与本“软件产品”一同提供给您的并与单独一份最终用户许可证相关的软件产品是根据那份许可协议中的条款而授予您。您一旦安装、复制、下载、访问或以其它方式使用“软件产品”，即表示您同意接受本《协议》各项条款的约束。如您不同意本《协议》中的条款，请不要安装或使用“软件产品”；但您可将其退回原购买处，并获得全额退款。

软件产品许可证
本“软件产品”受著作权法及国际条约款和其它知识产权法及条约的保护。本“软件产品”只许可使用，而不出售。

一、许可证的授予。
本《协议》授予您下列权利：
1. 应用软件。您可在一组计算机或工作站组成的单个集群上安装、使用、访问、显示、运行或以其它方式互相作用于（“运行”）本“软件产品”的一份副本。
2. 储存／网络用途。您还可以在一个仅用于通过内部网络在您的其它计算机上运行“软件产品”的储存设备（如网络服务器）上存放或安装一份“软件产品”副本；但是，您必须为从储存设备运行“软件产品”的每一个集群获得一份许可证。一份“软件产品”许可证不得在不同的计算机集群中共同或同时使用。
3. 保留权利。未明示授予的一切其它权利均为星环信息科技（上海）有限公司所有。

二、其它权利和限制的说明。
1. 对反向工程(Reverse Engineering)、反向编译(Decompilation)、反汇编(Disassembly)的限制。您不得对本“软件产品”进行反向工程、反向编译(decompile)或反汇编(disassemble)，但尽管有这项限制，如适用法律明示允许上述活动，则不在此限。
2. 商标。本《协议》不授予您有任何星环信息科技（上海）有限公司商标或服务商标的任何权利。
3. 出租。不得出租、租赁或出借本“软件产品”。
4. 支持服务。星环信息科技（上海）有限公司可能为您提供与“软件产品”有关的支持服务（“支持服务”）。支持服务的使用受用户手册、“联机”文档和/或其它星环信息科技（上海）有限公司提供的材料中所述的各项政策和计划的制约。提供给您作为支持服务的一部分的任何附件代码应被视为本“软件产品”的一部分，并须符合本《协议》中的

拒绝 接受

下一步 >

7.1. 添加集群服务器

您需要给您的集群设置一个名字。输入集群名字后点击“下一步”。

感谢选用 Transwarp Data Hub

本向导将帮助通过创建集群，选择节点和安装服务来使用Transwarp Data Hub.

- Transwarp Hadoop
- Transwarp Inceptor
- Transwarp Hyperbase
- Transwarp Stream

首先请给集群取一个名字

Transwarp Data Hub管理工具能同时管理多个集群，请给要创建的集群取一个名字以便区分，集群创建后仍可重命名.

集群名称

下一步 ▶

您在这个页面添加/删除/修改机柜。将鼠标移到机柜的名字和描述上可以对机柜的名字和描述进行修改。完成后点击“下一步”。

集群名称 添加节点 添加服务

1.编辑机柜 **2.分配节点** **3.其他配置**

机柜的概念在Hadoop中十分重要，每个节点的机柜信息可以帮助Hadoop决定数据分配。可在此页添加/修改/删除机柜。

机柜名称	节点数	描述
/default-rack	0	This is the default rack

+添加机柜

下一步 ▶

在这一步给您的集群分配节点。

点击“添加节点”会弹出下面窗口：

在搜索框中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。

<input type="text" value="172.16.2.[25-27]"/>	SSH端口号	22	<input type="button" value="搜索"/>
IP地址	已被管理	已配置SSH访问	机柜
<input checked="" type="checkbox"/> 172.16.2.25	NO	NO	
<input checked="" type="checkbox"/> 172.16.2.26	NO	NO	
<input checked="" type="checkbox"/> 172.16.2.27	NO	NO	

点击“继续”。并提供SSH访问权限并输入您管理节点的密码：

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

* 输入密码：

* 确认密码：

* SSH端口号：

输入密码后，点击“完成”。节点便添加完成。如果我们只有一个机柜，Transwarp Manager自动将所有的节点分配给了默认机柜。如果您有多个机柜，您可以点击“分配机柜”来分配机柜。（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）

集群名称 **添加节点** **添加服务**

1.编辑机柜 2.分配节点 3.其他配置

以下节点将被添加到集群中，如果有需要可以将其删除。在前往下一步之前，请对每一个节点分配机柜以及配置SSH访问。

检测到只有一个机柜，已自动为新添加的节点分配。可以在选中节点后，重新分配机柜。

<input checked="" type="checkbox"/>	IP地址	已被管理	已配置SSH访问	机柜
<input checked="" type="checkbox"/>	172.16.2.25	NO	YES	/default-rack
<input checked="" type="checkbox"/>	172.16.2.26	NO	YES	/default-rack
<input checked="" type="checkbox"/>	172.16.2.27	NO	YES	/default-rack

+ 添加节点 **分配机柜** **删除**

◀ 上一步 **下一步 ▶**

7.2. 删除集群中的服务器

如果您需要删除刚刚添加的节点，您可以先选中需要删除的节点，然后点击“删除”。

集群名称 **添加节点** **添加服务**

1.编辑机柜 2.分配节点 3.其他配置

以下节点将被添加到集群中，如果有需要可以将其删除。在前往下一步之前，请对每一个节点分配机柜以及配置SSH访问。

检测到只有一个机柜，已自动为新添加的节点分配。可以在选中节点后，重新分配机柜。

<input type="checkbox"/>	IP地址	已被管理	已配置SSH访问	机柜
<input type="checkbox"/>	172.16.2.25	NO	YES	/default-rack
<input checked="" type="checkbox"/>	172.16.2.26	NO	YES	/default-rack
<input type="checkbox"/>	172.16.2.27	NO	YES	/default-rack

+ 添加节点 **分配机柜** **删除**

◀ 上一步 **下一步 ▶**

节点添加工作完成后，点击“下一步”进入集群的“其他配置”。

7.3. 其他配置

7.3.1. 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。否则请选择不能通过主机名访问，Transwarp Manager会相应为您配置/etc/hosts文件，然后添加节点。

The screenshot shows the 'DNS Configuration' step of the Transwarp Manager setup. At the top, there's a header bar with the Transwarp logo and a user icon labeled 'admin'. Below the header, there are three tabs: '1.编辑机柜' (Edit Cabinet), '2.分配节点' (Assign Nodes), and '3.其他配置' (Other Configuration). The third tab is currently selected. A note below the tabs states: '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), 管理工具会在安装包的时候自动配置/etc/hosts.' (Node access: If nodes cannot be accessed by host name (whether DNS or /etc/hosts is configured), the management tool will automatically configure /etc/hosts during the package installation.) There are two radio button options for node access: '节点间已经可以通过主机名访问' (Nodes can already be accessed by host name) and '需要管理工具配置/etc/hosts' (Management tool needs to configure /etc/hosts). The second option is selected. Another note below says: 'NTP服务器: TDH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Transwarp管理工具自动选择一台节点作为内部NTP服务器' (NTP server: TDH uses an NTP server to ensure node clock synchronization in the cluster, you can specify an external NTP server or let the Transwarp management tool automatically select one node as the internal NTP server). There are two radio button options for NTP servers: '使用外部NTP服务器' (Use external NTP server) and '不使用外部NTP服务器' (Do not use external NTP server). The second option is selected. At the bottom of the screen are '上一步' (Previous Step) and '下一步' (Next Step) buttons.

7.3.2. 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部NTP服务器。

如果您需要使用外部NTP服务器，请选择“使用外部NTP服务器”，然后点击“下一步”：

● 集群名称 ○ 添加节点 添加服务

1.编辑机柜 2.分配节点 3.其他配置

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), 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 需要管理工具配置/etc/hosts

NTP服务器: TDH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Transswap管理工具自动选择一台节点作为内部NTP服务器

NTP服务器: 使用外部NTP服务器 不使用外部NTP服务器

+ 添加NTP服务器 **○ 测试连接**

◀ 上一步 **下一步 ▶**

点击“添加NTP服务器”，输入您想要添加的NTP服务器:

● 集群名称 ○ 添加节点 添加服务

1.编辑机柜 2.分配节点 3.其他配置

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), 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 需要管理工具配置/etc/hosts

NTP服务器: TDH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Transswap管理工具自动选择一台节点作为内部NTP服务器

NTP服务器: 使用外部NTP服务器 不使用外部NTP服务器

+ 添加NTP服务器 **○ 测试连接**

X

◀ 上一步 **下一步 ▶**

多次点击“添加NTP服务器”可以让您添加多台外部NTP服务器:

集群名称 添加节点 添加服务

1. 编辑机架 2. 分配节点 3. 其他配置

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), 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 需要管理工具配置/etc/hosts

NTP服务器: TDH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Transwarp管理工具自动选择一台节点作为内部NTP服务器

NTP服务器: 使用外部NTP服务器 不使用外部NTP服务器

[+ 添加NTP服务器](#) [测试连接](#)

s1a.time.edu.cn	x	ntp.sjtu.edu.cn	x
<input checked="" type="checkbox"/> s1a.time.edu.cn		<input checked="" type="checkbox"/> ntp.sjtu.edu.cn	

[上一步](#) [下一步](#)

点击“测试链接”，您可以测试添加的NTP服务器的连接。NTP服务器选择完成后，点击“下一步”继续。系统会开始自动添加节点：

集群名称 添加节点 添加服务

安装正在进行中

0 / 3 节点安装成功

搜索节点...

节点名称	IP地址	进度	状态	操作
tw-node2125	172.16.2.125	<div style="width: 50%;"></div>	安装 Transwarp Manager agent	详情
tw-node2126	172.16.2.126	<div style="width: 50%;"></div>	安装 JDK	详情
tw-node2127	172.16.2.127	<div style="width: 50%;"></div>	安装 JDK	详情

添加节点的过程会进行一段时间，成功后您会看到：

◎ 添加节点 分配角色 配置服务

Congratulations! All Completed!

3 / 3 节点安装成功

搜索节点...

节点名称	IP地址	进度	状态	操作
sles11sp3-node1	172.16.2.25	<div style="width: 100%; height: 10px; background-color: #2e6b2e;"></div>	Node installation completed	详情
sles11sp3-node2	172.16.2.26	<div style="width: 100%; height: 10px; background-color: #2e6b2e;"></div>	Node installation completed	详情
sles11sp3-node3	172.16.2.27	<div style="width: 100%; height: 10px; background-color: #2e6b2e;"></div>	Node installation completed	详情

[增加更多节点](#) [完成](#) [添加角色 ▶](#)

您可以点击“增加更多节点”来向集群添加更多节点，并重复[添加集群服务器](#)和其他配置中的步骤。您也可以点击“下一步”开始向集群[安装服务](#)。

8. 用Transwarp Manager安装服务

接下来，我们介绍如何使用Transwarp Manager向集群安装Transwarp Data Hub服务。

8.1. Guardian的安装和配置

Guardian服务为您的集群提供安全认证和授权机制，它并不是必须的。您可以在安装集群其他服务前就通过Transwarp Manager安装和配置Guardian服务，您也可以在集群安装完成后通过Transwarp Manager安装Guardian。本节我们将演示在安装其他服务前安装和配置Guardian。

- 集群配置完成后，Transwarp Manager会带您进入“添加服务”页面，该页面顶部会出现Guardian安装的提示。点击页面底部的“添加Guardian服务”进行添加：

The screenshot shows the 'Add Service' page of the Transwarp Manager interface. At the top, there is a message: '如果需要启用安全，请单击左下角的“添加Guardian服务”按钮添加Guardian服务。Guardian服务并不是必需的。' (If you need to enable security, click the 'Add Guardian Service' button at the bottom-left to add the Guardian service. The Guardian service is not required.) Below this, there are three tabs: '1.选择服务' (Select Service), '2.分配角色' (Assign Roles), and '3.配置服务' (Configure Services). The '1.选择服务' tab is selected. On the left, under '选择需要安装的产品或者服务:' (Select products or services to install), several services are listed with their icons and descriptions: Transwarp Hadoop, Transwarp Inceptor, Transwarp Hyperbase, Transwarp Stream, and Transwarp Discover. On the right, under '定制服务' (Custom Services), several services are listed with their icons and descriptions: ZooKeeper, HDFS, YARN, Hyperbase, Inceptor-SQL, Discover, Stream, Kafka, Oozie, Sqoop, Hue, and Elastic Search. The 'Stream' service is currently selected. At the bottom of the page, there are two buttons: '< 增加更多节点' (Add More Nodes) and a red-bordered '添加Guardian服务' (Add Guardian Service) button.

Transwarp Manager会向您请求安装确认：

Add Guardian Service

To enable Kerberos, a Guardian service is required, would you like add one now?

取消 **确认**

点击“确认”继续。

2. 下面，您需要对Guardian服务分配角色，包括：

- 分配Guardian server所在节点：

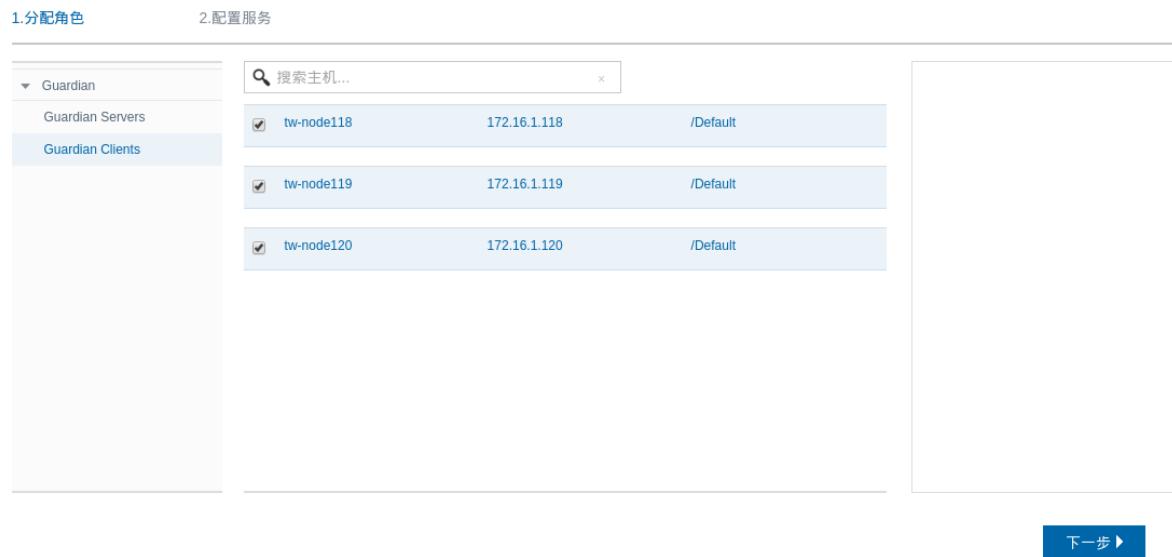
1.分配角色 2.配置服务

Guardian	搜索主机...		
Guardian Servers	<input checked="" type="checkbox"/> tw-node118	172.16.1.118	/Default
Guardian Clients	<input checked="" type="checkbox"/> tw-node119	172.16.1.119	/Default
	<input type="checkbox"/> tw-node120	172.16.1.120	/Default

下一步▶

Guardian服务支持高可用性（HA），所以您可以选择启用多台节点作为Guardian server。

- 分配Guardian client所在节点：



我们建议您为集群的每一台节点都安装Guardian client。

配置完成后点击“下一步”继续。

3. 这里，您需要设置Guardian服务中的两个重要的密码，

2. 配置服务	
属性	基础参数
服务名称	<input type="text" value="Guardian1"/>
Guardian管理员用户名	<input type="text" value="admin"/>
Guardian管理员密码	<input type="password"/>
OpenLDAP超级密码	<input type="password"/>

下一步▶

包括：

- Guardian管理员密码。Guardian管理员将管理集群上所有用户在Guardian中的用户名和密码。
- OpenLDAP超级密码。OpenLDAP超级密码用于管理OpenLDAP自身。

设置完成后点击“下一步”继续。

4. 此时，Transwarp Manager会开始安装Guardian:

节点名称	描述	开始时间	完成时间	状态
	Install Guardian	2016/1/18 21:12:55	2016/1/18 21:15:40	SUCCESSFUL
▶ configure OpenLDAP Server (tw-node118)				
▶ configure OpenLDAP Server (tw-node119)				
▶ configure Kerberos Server (tw-node118)				
▶ configure Kerberos Server (tw-node119)				
▶ configure Kadmin Server (tw-node118)				
▶ configure Kadmin Server (tw-node119)				

安装完成后点击“Close”退出。

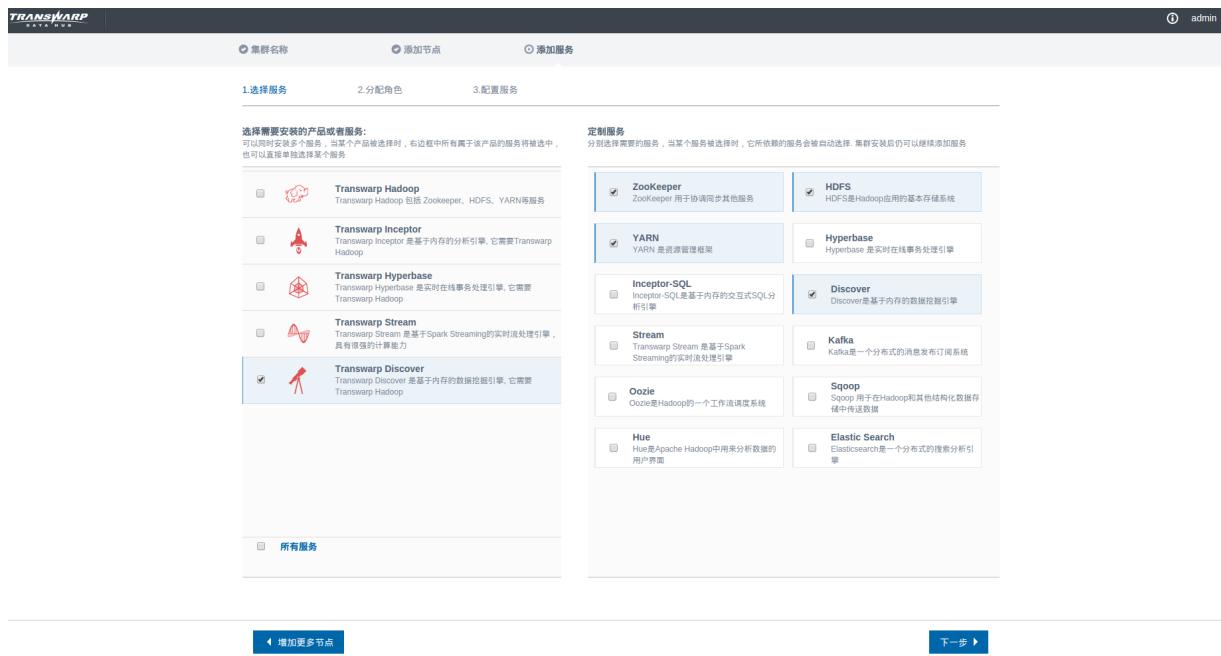
5. 此时，您已经完成Guardian服务在集群上的安装和配置，您可以开始选择安装其他服务。

8.2. 选择服务

管理界面的左侧列表显示Transwarp Data Hub中的 **产品**，包括：Hadoop, Inceptor, Hyperbase, Stream 和Discover。右侧列表显示的则是Transwarp Data Hub中的 **服务**，包括：Zookeeper, HDFS, YARN, HBase, Inceptor, Discover, Stream, Kafka, Oozie, Sqoop, HUE和Elasticsearch。Transwarp Data Hub中的产品由一个或多个服务组成。为集群安装服务时，您可以通过选择产品来安装属于该产品的服务，您也可以直接通过选择服务来进行安装：

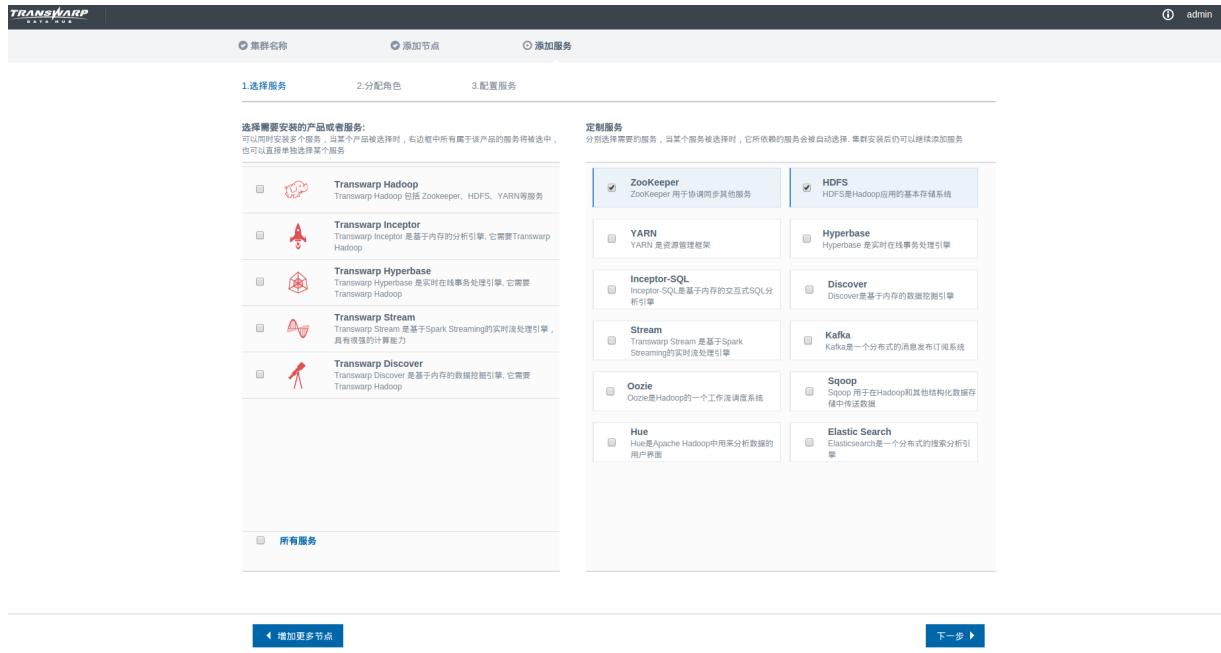
- 当某个产品被选择时，右边框中所有属于该产品的服务将被选中。

例如假设您选择Discover产品，那么ZooKeeper, HDFS, YARN, Discover这四个服务都会被同时选择

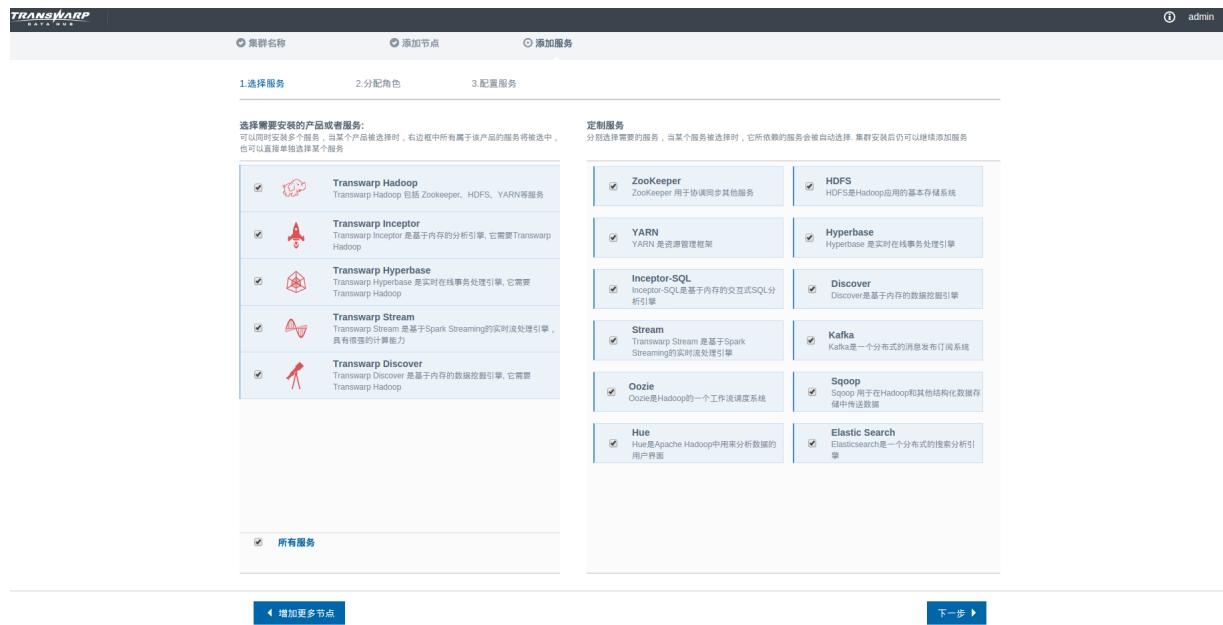


- 直接选择服务安装时，被选服务所依赖的服务会被自动选择。

如选择HDFS服务，则ZooKeeper服务也会被自动选择

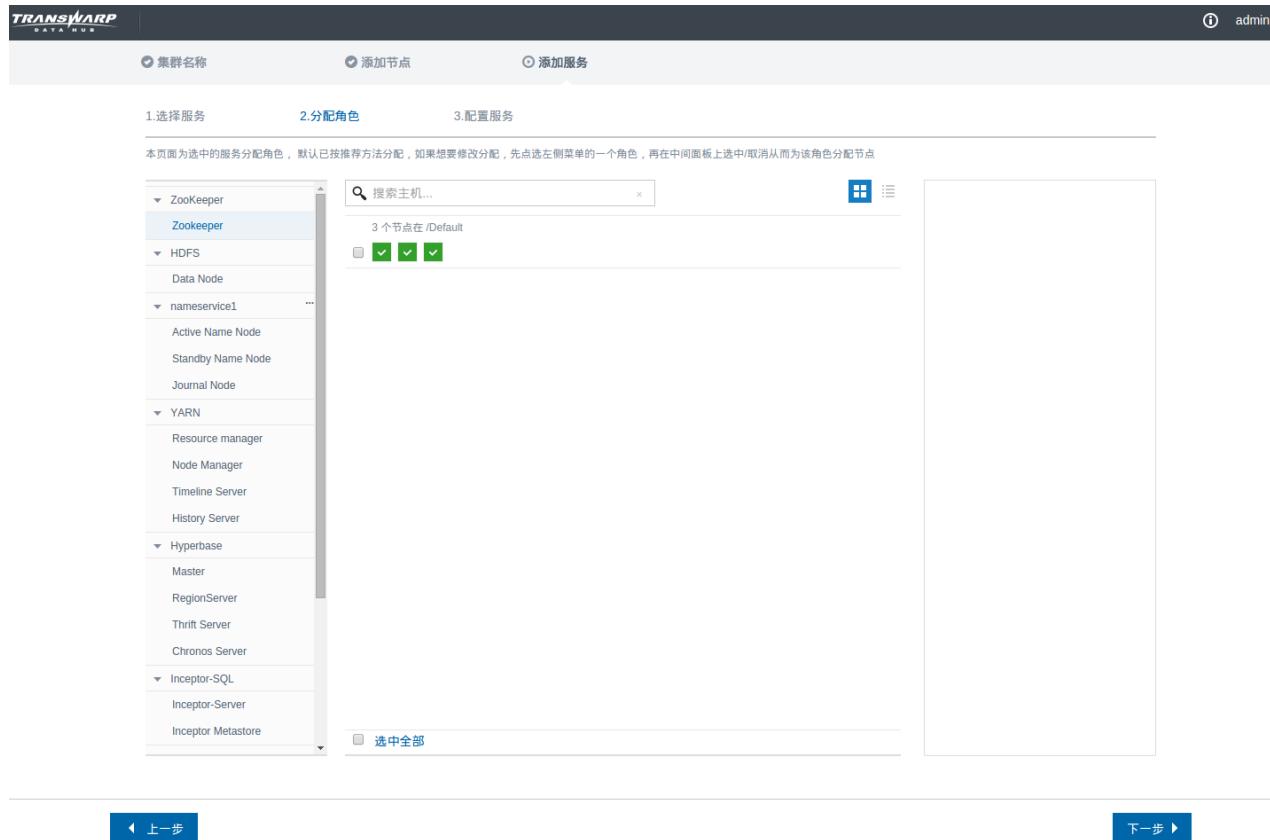


- 您还可以直接选择安装“所有服务”：



8.3. 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角色下的节点。服务角色下被选中的节点将承担该服务角色。例如下图中，Zookeeper下的三个勾选框全部选中代表集群中的三个节点都将承担Zookeeper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”直接进行服务的配置。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：

- 推荐 将Zookeeper角色分配给 奇数 个节点，建议在集群大小在10个节点以下配置3个，20个节点以下配置5-7个，20个节点以上配置9-11个或以上。
- 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HBase Master角色分配给 奇数 个节点。
- 分配了HBase RegionServer的节点 必须 同时拥有HDFS DataNode角色。

8.4. 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。

The screenshot shows the Transwarp Data Hub configuration interface. At the top, there are tabs for '1.选择服务', '2.分配角色', and '3.配置服务'. The '3.配置服务' tab is active. On the left, a sidebar lists services: ZOOKEEPER (ZooKeeper1), HDFS (HDFS1), YARN (YARN1), HYPERBASE (Hyperbase1), INCEPTOR_SQL (InceptorSQL1), DISCOVER (Discover1), STREAM (Stream1), KAFKA (Kafka1), and OOZIE (Oozie1). The 'ZOOKEEPER (ZooKeeper1)' service is selected. In the main area, there are two tabs: '属性' (Properties) and '基础参数' (Basic Parameters). The '基础参数' tab is selected. A table shows configuration parameters:

属性	值
服务名称	ZooKeeper1

At the bottom, there are '上一步' (Previous Step) and '下一步' (Next Step) buttons.

8.4.1. 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:

The screenshot shows the Transwarp Data Hub configuration interface. At the top, there are three tabs: '1. 选择服务' (Select Service), '2. 分配角色' (Assign Role), and '3. 配置服务' (Configure Service). The '3. 配置服务' tab is active. On the left, a sidebar lists various services: ZOOKEEPER (ZooKeeper1), HDFS (HDFS1), YARN (YARN1), HYPERBASE (Hyperbase1), INCEPTOR_SQL (InceptorSQL1), DISCOVER (Discover1), STREAM (Stream1), KAFKA (Kafka1), and OOZIE (Oozie1). The main area displays a configuration table for the selected service. The table has two tabs at the top: '属性' (Properties) and '基础参数' (Basic Parameters). The '属性' tab is selected. The table has two columns: '属性' (Property) and '值' (Value). One row shows '服务名称' (Service Name) with the value 'ZooKeeper1'. At the bottom of the page are two buttons: '上一步' (Previous Step) and '下一步' (Next Step).

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”，例如图中显示的Zookeeper服务名为Zookeeper1。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。例如集群中可以有两个Inceptor服务实例，分别名为Inceptor1和Inceptor2。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：

The screenshot shows the Transwarp Data Hub configuration interface. At the top, there are three tabs: '1. 选择服务' (Select Service), '2. 分配角色' (Assign Role), and '3. 配置服务' (Configure Service). The '3. 配置服务' tab is active. On the left, a sidebar lists various services: ZOOKEEPER (ZooKeeper1), HDFS (HDFS1), YARN (YARN1), HYPERBASE (Hyperbase1), INCEPTOR_SQL (InceptorSQL1), DISCOVER (Discover1), STREAM (Stream1), KAFKA (Kafka1), and OOZIE (Oozie1). The main area displays a configuration table for the selected service. The table has three tabs at the top: '属性' (Properties), '基础参数' (Basic Parameters), and '自定义参数' (Custom Parameters). The '基础参数' tab is selected. The table has two columns: '基础参数' (Basic Parameter) and '值' (Value). Several rows are listed: 'tickTime' with value '9000', 'initLimit' with value '10', 'syncLimit' with value '5', 'zookeeper.client.port' with value '<节点特定，请点击以分别修改>' (Node-specific, click to modify individually), 'zookeeper.peer.communicate.port' with value '<节点特定，请点击以分别修改>' (Node-specific, click to modify individually), and 'zookeeper.leader.elect.port' with value '<节点特定，请点击以分别修改>' (Node-specific, click to modify individually). At the bottom of the page are two buttons: '上一步' (Previous Step) and '下一步' (Next Step).

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。基础参数数值

不仅在这一步可以设置，在安装完成后还可以修改。

基础参数分为节点特定参数（例如上图中的 `zookeeper.client.port`, `zookeeper.peer.communicate.port`, `zookeeper.leader.elect.port`）和非节点特定参数（例如上图中的 `tickTime`, `init.Limit` 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：

The screenshot shows the '3. 配置服务' (Configure Service) step of the Transwarp Manager setup. On the left, a sidebar lists services: ZOOKEEPER (ZooKeeper1), HDFS (HDFS1), YARN (YARN1), HYPERBASE (Hyperbase1), INCEPTOR_SQL (InceptorSQL1), DISCOVER (Discover1), STREAM (Stream1), KAFKA (Kafka1), and OOZIE (Oozie1). The 'ZOOKEEPER (ZooKeeper1)' service is selected. The main area has tabs: 属性 (Properties), 基础参数 (Basic Parameters), and 自定义参数 (Custom Parameters). The '基础参数' tab is active, showing parameters like tickTime (9000), initLimit (10), syncLimit (5), and three 'node-specific' parameters: zookeeper.client.port, zookeeper.peer.communicate.port, and zookeeper.leader.elect.port, each with a placeholder text: '<节点特定，请点击以分别修改>'.

- 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>”

This screenshot shows the same configuration interface as the previous one, but with a red box highlighting the 'zookeeper.client.port' input field. This indicates that the user is about to click on it to edit the value.

来打开为节点分别配置的窗口：



然后点击节点对应的值进行修改。

- **自定义参数:**

属性	基础参数	自定义参数
ZOOKEEPER (ZooKeeper1)		值
HDFS (HDFS1)		
YARN (YARN1)		
HYPERBASE (Hyperbase1)		
INCEPTOR_SQL (InceptorSQL1)		
DISCOVER (Discover1)		
STREAM (Stream1)		
KAFKA (Kafka1)		
OOZIE (Oozie1)		

除了HUE和Sqoop以外的服务都可以添加自定义参数。这里显示的是Zookeeper服务的自定义参数配置页。

除了以上的通用可配置项，集群中的一些服务有其特殊的可配置项，包括：

- **YARN的资源配置设置:**

1.选择服务 2.分配角色 3.配置服务

配置选中的服务

	属性	基础参数	自定义参数	资源配置设置
ZOOKEEPER (ZooKeeper1)				
HDFS (HDFS1)				
YARN (YARN1)				
HYPERBASE (Hyperbase1)				
INCEPTOR_SQL (InceptorSQL1)				
DISCOVER (Discover1)				
STREAM (Stream1)				
KAFKA (Kafka1)				
OOZIE (Oozie1)				

Configure YARN service Scheduler

Use Scheduler: Capacity Scheduler Fair Scheduler

全局配置

Resource Calculator: Dominant Default

默认情况下, Scheduler只基于内存调度,也可以运用Dominant Resource的概念,配置成基于内存和CPU的调度方式。

Queue配置

	容量	最大Application数量	状态	Queue配置
root	100	10000	Running	详情 添加
default	100	10000	Running	详情 删除 添加

[上一步](#) [下一步](#)

YARN负责集群中的任务调度，在“资源配置设置”中，您可以设置YARN如何将计算资源配给它调度的任务。YARN的资源配置设置可以在这一步进行设置，也可以在这一步使用默认设置，之后再根据实际情况调整。

• Inceptor的高级参数

1.选择服务 2.分配角色 3.配置服务

配置选中的服务

	属性	基础参数	高级参数	自定义参数	资源配置
ZOOKEEPER (ZooKeeper1)					
HDFS (HDFS1)					
YARN (YARN1)					
HYPERBASE (Hyperbase1)					
INCEPTOR_SQL (InceptorSQL1)					
DISCOVER (Discover1)					
STREAM (Stream1)					
KAFKA (Kafka1)					
OOZIE (Oozie1)					
SQOOP (Scoop_Ser...)					
HUE (Hue1)					
ELASTICSEARCH (Elasticse...)					

高级参数

参数	值
hive.server2.enabled	TRUE
hive.server2.authentication	NONE

[上一步](#) [下一步](#)

在这里可以设置两个Inceptor高级参数：

- **hive.server2.enabled:** 是否使用InceptorServer 2作为Inceptor server。TRUE代表使用，FALSE代表不使用。不使用InceptorServer 2即使用InceptorServer 1作为Inceptor

server。InceptorServer 2支持安全认证和授权，是默认选择，也是我们推荐的选择。

- **hive.server2.authentication:** 在hive.server2.enabled为TRUE的情况下，为InceptorServer 2选择哪一种安全认证方式。认证方式有NONE, KERBEROS和LDAP。使用LDAP的情况下您会看到两个LDAP相关参数：

属性	基础参数	高级参数	自定义参数	资源分配
高级参数				
hive.server2.enabled	值			
hive.server2.authentication	LDAP			
hive.server2.authentication.ldap.baseDN	ou=People,dc=tdh			
hive.server2.authentication.ldap.url	ldap://tw-node2022:389 ldap://tw-node2023:389			

这两个参数Transwarp Manager已经为您配置好，不需要改动。

- **基于Spark的服务的资源分配设置:** Transwarp Data Hub中基于Spark的服务包括Inceptor, Stream和Discover，可以分别对它们进行资源分配设置。
 - **Discover的资源分配设置:**

运行Application的Queue:	<input type="text" value="default"/>
Executor 内核数:	<input type="text" value="1"/>
每个executor所使用的core的数量.	
Executor 内存:	<input type="text" value="512"/>
每一个executor使用的内存(MB).	
Worker数量:	<input type="text" value="2"/>

对Discover服务，您可以设置的有：

- 运行Application的Queue: Discover 应用运行的任务队列。
- Executor内核数：每个Executor分配的CPU数量。
- Executor内存：每个Executor分配的内存。
- Worker数量：集群中总共启用多少个运行Discover任务的节点。
- **Inceptor和Stream的资源分配设置:**

运行Application的Queue:

Application Master内存: application master使用的内存(MB).

Executor资源: Fixed Ratio

Memory : core比例: 内存(GB)与CPU core数的比例。例如, 如果这个比例为2, 而CPU core数为1, 那么内存使用量就为2GB.

core百分比: 异构系统中每个executor使用的core数量占所在节点core总量的比例。例如, 如果这个比例为0.1, 而一个节点上有10个core, 那么该节点上的executor将使用1个core.

Executor 数量: 每个worker上只运行一个executor 总共运行指定数目的executor

Executor 分布: 每个worker上都运行executor executor运行在指定数量的worker上

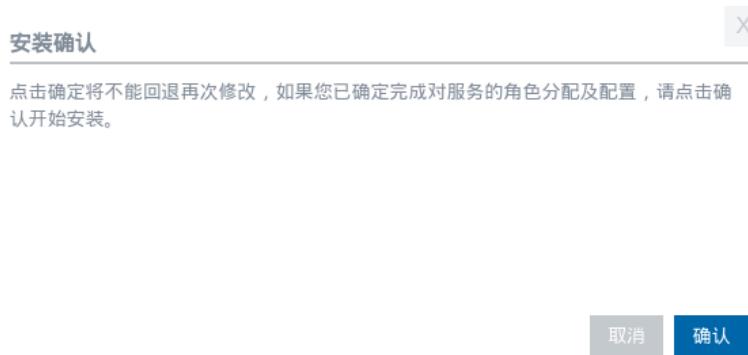
Executors运行在指定的worker列表

对Inceptor和Stream服务, 您可以设置的有:

- 运行Application的Queue: Inceptor或Stream运行的任务队列。
- Application Master内存: 分配给Inceptor或Stream的Application Master的内存。
- Executor资源: 选择给Executor分配固定的内存和CPU (“Fixed”) 还是分配一定比例的内存和CPU (“Ratio”) 。
 - 如果选择 “Fixed”, 那您需要设置 “Executor内核数” (每个Executor分配的CPU数量) 和 “Executor内存” (每个Executor分配的内存) 。
 - 如果选择 “Ratio”, 那您需要设置 “Memory:Core比例” (内存(GB)与CPU核数的比例) 和 “Core百分比” (每个Executor使用的CPU核数占所在节点总CPU核数的比例) 。
- Executor数量: 在每个每个Worker上运行一个Executor还是在整个集群中运行指定数量的Executor。
- Executor分布: Executor在Worker上的分布。

8.5. 服务安装

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



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, 您将看到以下页面:

ZOOKEEPER initialize ZooKeeper : ZooKeeper1

2/3 步已完成

- ▶ install packages of ZooKeeper1
- ▶ update configuration of ZooKeeper1
- ▶ start ZooKeeper service: ZooKeeper1

点击“下一步”按键代表安装完成。如出现下面界面表示安装成功。

恭喜!

所有的服务均已安装配置并运行在集群上。

点击“完成”进入Transwarp Manager的控制面板。

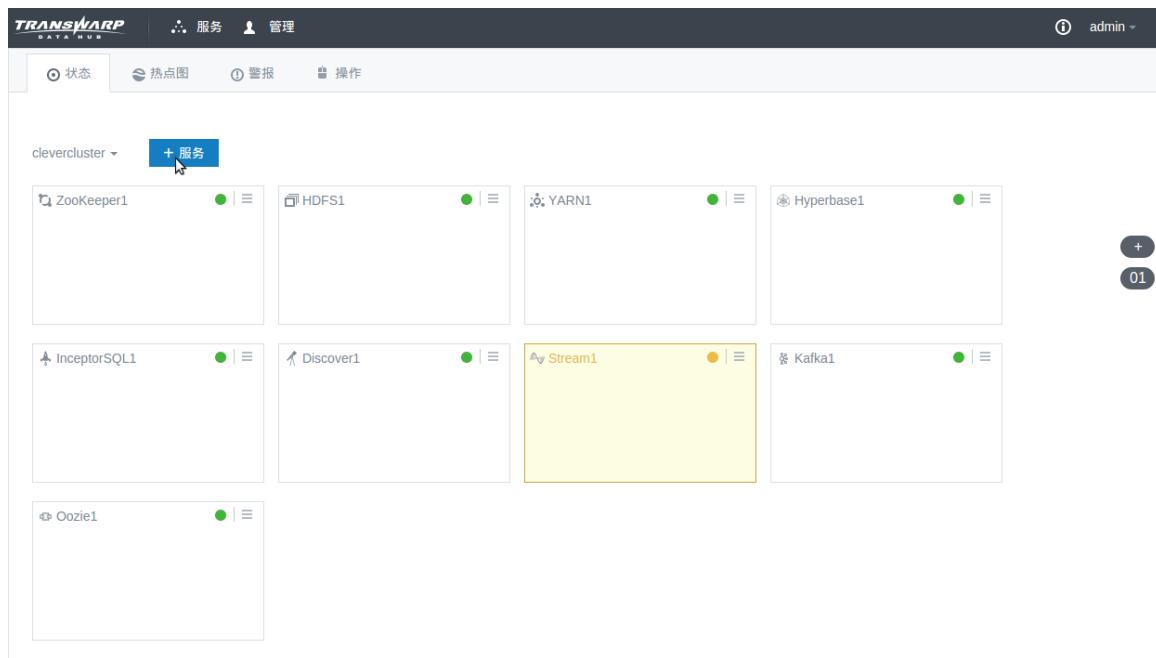
TEST

ZooKeeper1	HDFS1	YARN1	Hyperbase1	InceptorSQL1	Discover1
Stream1	Kafka1	Oozie1	Sqoop_Service1	Hue1	Elasticsearch1

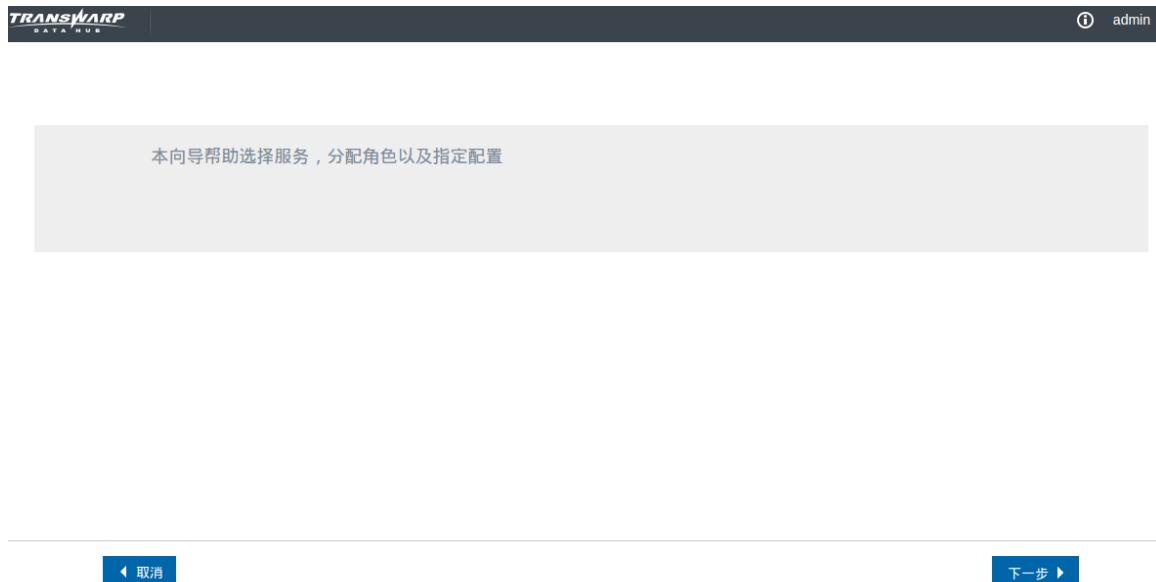
9. 用Transwarp Manager安装新增服务

在安装完您所选择的服务后，您还可通过Transwarp Manager主界面向集群新增服务。以下，我们将以Sqoop服务为例，演示如何在Transwarp Manager主界面添加并安装服务。

1. 将光标放在Transwarp Manager主页面上集群名旁的“+”上，点击弹出的“+服务”。



2. 进入向导，点击“下一步”。



3. 进入“选择服务”页面，选择Sqoop，点击“下一步”。

TRANSWARP DATA HUB

① admin

1.选择服务 2.分配角色 3.配置服务 4.安装

选择想要添加的服务，可能存在某些由于依赖服务未添加而无法添加的服务，请先添加其依赖的服务。鼠标悬停在一个服务上可以看到它依赖的服务

ZooKeeper ZooKeeper 用于协调同步其他服务	HDFS HDFS是Hadoop应用的基本存储系统	YARN YARN 是资源管理框架
Hyperbase Hyperbase 是实时在线事务处理引擎	Inceptor-SQL Inceptor-SQL是基于内存的交互式SQL分析引擎	Discover Discover是基于内存的数据挖掘引擎
Stream Transwarp Stream 是基于Spark Streaming的实时流处理引擎	Kafka Kafka是一个分布式的消息发布订阅系统	Oozie Oozie是Hadoop的一个工作流调度系统
Sqoop Sqoop 用于在Hadoop和其他结构化数据存储中传递数据	Hue Hue是Apache Hadoop中用来分析数据的用户界面	Elastic Search Elasticsearch是一个分布式的搜索分析引擎

◀ 上一步 **下一步 ▶**

4. 新增服务往往需要依赖集群中已有的服务，您在这一步可以设置新增的服务对集群中已有服务的依赖。

TRANSWARP DATA HUB

172.16.2.56:8180/#/addservice/addspecifydependency?installClusterId=1¤tServiceId=18

① admin

1.选择服务 2.指定依赖 3.分配角色 4.安全模式 5.配置服务 6.安装

为选择的服务指定其依赖的服务

HDFS	YARN
<input checked="" type="radio"/> HDFS1	YARN1

◀ 上一步 **下一步 ▶**

5. 接下来，您需要为Sqoop服务**分配角色**，完成后点击“**下一步**”继续。

1.选择服务 2.分配角色 3.配置服务 4.安装

本页面为选中的服务分配角色， 默认已按推荐方法分配，如果想要修改分配，先点选左侧菜单的一个角色，再在中间面板上选中/取消从而为该角色分配节点

Sqoop	搜索主机...	3个节点在 /Default
<input checked="" type="checkbox"/> S	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

选中全部

[◀ 上一步](#) [下一步 ▶](#)

6. 设置新增服务的安全模式，安全模式在安装完成后可以更改：

1.选择服务 2.指定依赖 3.分配角色 4.安全模式 5.配置服务 6.安装

指定安全认证模式

简单认证模式 Kerberos认证模式

[◀ 上一步](#) [下一步 ▶](#)

7. 这一步，您需要为Sqoop服务进行配置：

配置选中的服务

属性	值
服务名称	Sqoop_Service1

◀ 上一步 下一步 ▶

配置完成后点击“下一步”。

8. 确认配置，确定无误后点击“下一步”：

SQOOP(Sqoop_Service2)

基础参数	
sqoop.http.port	12000
sqoop.admin.port	12001

◀ 上一步 下一步 ▶

弹出“安装确认”弹框，点击“确认”，开始安装：

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](#)

9. 安装过程一般需要几分钟时间。安装成功后点击“继续”进入“服务安装完成”页面，点击“完成”，完成Sqoop安装。

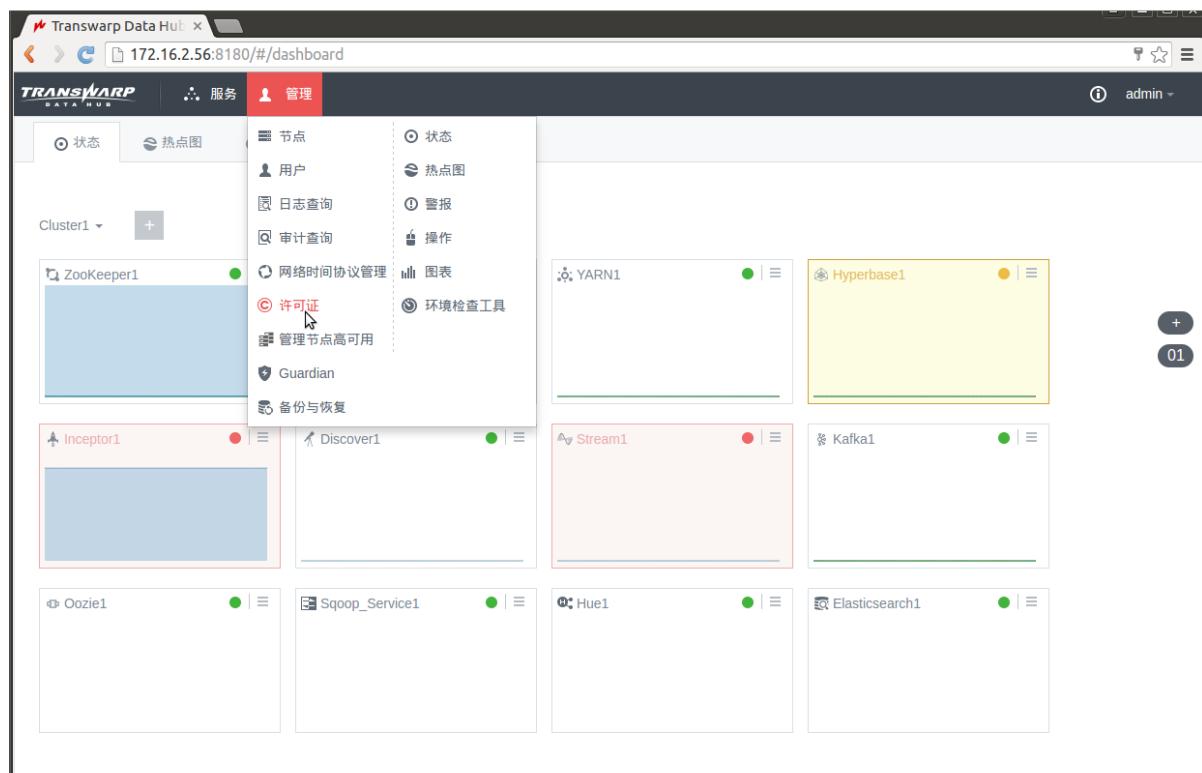
The screenshot shows the Transwarp Manager interface for service installation. At the top, there's a navigation bar with the Transwarp logo and a user icon labeled "admin". Below the navigation bar, a progress bar indicates "4. 安装" (Installation) is the current step. The main content area displays the "SQOOP initialize Sqoop : Sqoop_Service1" section, showing a completed step: "3/3 步已完成" (3/3 steps completed). A list of three tasks is shown, each with a green checkmark indicating completion: "Install related packages of sqoop server", "Update configuration of sqoop server", and "Start sqoop server". At the bottom right of the content area, there is a blue button labeled "继续 >" (Continue).

10. 许可证的激活与认证

Transwarp Data Hub采用许可证授权方式提供相关服务，因此在安装过程中需要输入许可证并激活以启用所有服务。

10.1. 许可证管理

将鼠标移至Transwarp Manager主界面上方的“管理”菜单，在弹出的菜单栏中点击“许可证”进入许可证管理页面。



许可证管理页面分为“概要”和“详情”两个板块。“概要”板块显示Server Key、序列号、许可证服务的健康状况，点击“启动”和“停止”可以启动或者停止许可证集群，还可以上传License和查看License Key。“详情”板块显示当前每一个已安装服务的License类型、节点容量、License到期时间。

The screenshot shows the Transwarp Data Hub license management interface. At the top, there's a header bar with the Transwarp logo, a search bar containing '172.16.2.56:8180/#/license/summary', and a user dropdown for 'admin'. Below the header is a navigation menu with '服务' (Services) and '管理' (Management). The main content area has a sidebar with a '许可证' (License) icon. The main panel displays a '概要' (Summary) section with a 'Server Key' field containing 'fowF1u6ID5YGSi/0qdbGmPZi09Ni.IIURY8lwNPzksvG/LXY GyOaccnKc'. It also includes a '序列号(可选)' (Optional Serial Number) field and a status indicator for the '许可证服务' (License Service) showing 'HEALTHY' with a green dot. There are '启动' (Start) and '停止' (Stop) buttons next to it. Below this are two buttons: '上传License' (Upload License) and '查看License Key' (View License Key). The bottom section is titled '详情' (Details) and contains a table listing various products and their license details:

产品	License类型	节点容量	截止日期
Transwarp Manager Enterprise	EVALUATION	50	Sep 29, 2016
Transwarp Hadoop Enterprise	EVALUATION	50	Sep 29, 2016
Transwarp Hyperbase Enterprise	EVALUATION	50	Sep 29, 2016
Transwarp Inceptor Enterprise	EVALUATION	50	Sep 29, 2016
Transwarp Streaming Enterprise	EVALUATION	50	Sep 29, 2016
Transwarp Discover Enterprise	EVALUATION	50	Sep 29, 2016
Transwarp Holodesk Enterprise	EVALUATION	50	Sep 29, 2016

10.2. 许可证激活和使用

点击“上传 License”可以上传一个许可证，系统会弹出以下窗口。

The dialog box is titled '上传License'. It contains two input fields: '序列号.' (Serial Number) and 'License Key'. Below these fields is a note: '直接粘贴到上面的文本框或者通过文件上传' (Paste directly into the text box or upload from file) followed by a '选择文件...' (Select File...) button. At the bottom are two buttons: '取消' (Cancel) and '保存' (Save).

如果您还没有购买商业版的Transwarp Data Hub, 您使用的将是拥有30天有效期的试用版Transwarp Data Hub。您将不会有序列号, 但会有一个试用版的License Key。输入License Key后选择保存您便可以开始使用试用版的Transwarp Data Hub了。

如果您购买了商业版的Transwarp Data Hub, 我们将为您提供序列号和License Key。将两条信息都输入后点击“保存”，您便可以开始使用商业版的Transwarp Data Hub。

10.3. 查看许可证授权管理

如果您已经上传过License, 您可以点击“查看License Key”来查看您的License Key。

10.4. 许可证过期预警

当许可证离过期还有7天时，控制面板上会显示告警信息，及时提醒用户更换许可证。

11. 卸载TDH

卸载需要使用uninstall.sh 及 uninstall_agent_template.sh这两个脚本。执行这两个脚本会卸载TDH的服务相关RPM包及删除数据目录，具体包括：

- 卸载目标节点上所有安装来自transwarp源的RPM（不包括krb5-1ib 及 openldap）
- 停止所有TDH相关服务，并删除服务启动脚本
- 删除/etc/\${sid}, /var/log/\${sid}, /etc/\${sid}, /var/run/\${sid}, /var/lock/subsys/\${sid} 目录
- 删除数据目录，删除所有定义在hdfs-site.xml, yarn-site.xml, zoo.cfg及ngmr-env.sh文件中定义的本地数据目录
- 删除/usr/lib/ngmr目录
- 删除transwarp-manager master/agent节点上/etc/transwarp-manager, /var/log/transwarp-manager/, /usr/lib/transwarp-manager, /var/lib/transwarp-manager目录，卸载transwarp-manager 及 transwarp-manager-common RPM。

11.1. 使用方法

在管理节点的安装目录下运行uninstall.sh进行卸载。该脚本支持4个参数：

- **--keepdata**: 在执行卸载动作时不删除数据目录。默认（不指定改参数时）删除节点数据目录。
- **--nodes**: 指定目标节点执行卸载。目标节点的值以空格分隔。比如 **--nodes="172.16.1.1 172.16.1.2"**。不指定该参数时，卸载程序会自动解析Transwarp Manager的元数据，如果元数据存在会自动获得节点信息。
- **--user**: 指定目标节点执行卸载程序的用户。默认卸载程序的执行用户为root。当指定其它用户时，需要保证其它用户可以无密码方式执行 **sudo**
- **-y**: 默认执

说明

- 卸载脚本的运行需要在执行过程中确认（如果没有使用 **-y** 参数），请在确认卸载程序解析节点和服务信息后输入y确认。输入其它字符，卸载程序将被取消。
- 卸载程序会在当前目录生成，生成卸载日志uninstall.log，具体卸载信息可以从该日志文件查看。
- 卸载程序会自动解析元数据中服务信息。如果无法解析元数据，或者元数据不存在。可以通过在当前目录创建service_list文件，将需要卸载的服务的sid写到该文件中，一个服务一行。通过这样的方式指定需要卸载的服务，否则卸载程序将会按照“Zookeeper1, HDFS1, YARN1, Hyperbase1, Inceptor1, Discover1, Stream1, Kafka1, Sqoop1, Oozie1, HUE1, Elasticsearch1”的设置进行卸载。

11.2. 常见问题

- uninstall.sh及uninstall_agent_template.sh支持哪个TDH版本？

一般而言，所有版本都支持。对于TDH4.2之前的版本，Transwarp Manager的元数据不存放在文件中，无法自动解析节点和服务信息，需要手动通过 **--nodes** 及service_list文件进行指定。对于4.2之后的版本可以自动解析信息。

- 可否重复执行卸载？

可以重复执行，但是在一次执行后，Transwarp Manager节点的信息随着元数据信息一同被删除，卸载程序无法解析需要的节点信息，需要通过—nodes指定需要执行的节点。服务信息在第一次成功解析后，会保存在service_list文件中，不需要重新指定，如果需要更改可直接更改该文件。

- 如何处理错误？

通过控制台及log输出可以查看错误信息。程序仅在与目标节点发生通信错误时及卸载RPM不成功时会停止当前节点过程，其它节点继续执行。遇到其它错误时会忽略继续执行。请注意查看log，确认卸载程序是否到达预期的效果。远程登录卸载程序通过/etc/transwarp/下的ssh identity文件进行远程登录，一般情况在Transwarp Manager安装节点过程中该文件被生成，同时目标节点完成相应配置。可以使用该文件进行认证登录，如果该文件丢失或损坏，在执行过程中会要求用户输入密码以登录远程系统。

12. 安装和设置问题及回答

问题: 如何修改CentOS和Redat 操作系统中的hostname?

第一, 执行hostname xxx指令(暂时生效)。第二, 修改/etc/sysconfig/network中的hostname(重启后能够生效)。第三, 修改/etc/hosts文件, 确保集群中的所有节点的hostname都正确。注意: 安装完成后尽量不要修改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
::1 localhost localhost.localdomain
```

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。或在管理节点上的命令行方式执行#hostname (机器名字), 比如: #hostname TDH-Transwarp Manager。

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。

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。

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: 第一次加节点失败?

在TDH 4.2v以前的版本中首次在Transwarp Manager中加节点必须3台以上才可以完成节点的配置, 因此推荐用户首次添加3台以上数据节点。

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:

HDFS initialize HDFS : HDFS1

4/5 steps are completed

Node Name	Description	Start Time	Complete Time	Status
transwarp-13	start journal node	2016/1/12 18:08:20	2016/1/12 18:08:30	Successful
transwarp-14	start journal node	2016/1/12 18:08:20	2016/1/12 18:08:29	Successful
transwarp-15	start journal node	2016/1/12 18:08:20	2016/1/12 18:08:29	Successful
transwarp-13	stop active name node	2016/1/12 18:08:30	2016/1/12 18:08:31	Successful
transwarp-13	format name node	2016/1/12 18:08:31	2016/1/12 18:08:34	Failed
transwarp-13	start active name node			Waiting
transwarp-14	bootstrap standby name node			Waiting
transwarp-13	initialize HA state			Waiting
▶ start HDFS service: HDFS1				

Buttons: ReEdit Service | Skip Current Service

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
```

- 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。）
- 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。

附录 A: 集群服务使用的网络端口

集群中使用的网络端口如下所列。在系统部署之前，确保这些端口在网络系统中保持打开状态。

A.1. Zookeeper相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
2291	License Cluster Client Port	不能修改	否
2988	License Cluster Peer Port	不能修改	否
3988	License Cluster Leader Port	不能修改	否
2181	Zookeeper Client Port	不能修改	否
2888	Zookeeper Peer Port	不能修改	否
3888	Zookeeper Leader Port	不能修改	否

A.2. HDFS相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
8020	NameNode RPC Address	namenode. rpc-port	是
50070	NameNode HTTP Address	namenode. http-port	是
8485	JournalNode RPC Address	journalnode. rpc-port	否
8480	JournalNode HTTP Address	journalnode. http-port	否
50010	DataNode Address	datanode. port	是
50075	DataNode HTTP Address	datanode. http-port	是
50020	DataNode IPC Address	datanode. ipc-port	否

A.3. YARN相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
8032	ResourceManager Address	resourcemanager. port	是
8031	ResourceManager Resource-Tracker Address	resourcemanager. resource-tracker. port	是
8030	ResourceManager Scheduler Address	resourcemanager. scheduler. port	否
8033	ResourceManager Admin Address	resourcemanager. admin. port	否
8088	ResourceManager Web Port	resourcemanager. webapp. port	是

A.4. Hyperbase相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
60000	HBase Master Port	master.port	是
60010	HBase Master Web Port	master.info.port	是
60020	HBase RegionServer Port	regionserver.port	是
60030	HBase RegionServer Web Port	regionserver.info.port	是
9090	Thrift Port	不能修改	是
9095	Thrift Info Port	不能修改	是

A.5. Inceptor相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
4040	Inceptor UI Port	是	是
3306	MySQL JDBC Port	不能修改	否
9083	Metastore Thrift Port	是	否
10000	Hive Port	是	是

A.6. Discover相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
8787	Rstudio Server address	/etc/rstudio/rserver.conf	www-port
4088	Spark context UI	discover.ui.port	是

A.7. Stream相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
4044	Spark UI port	ngmr.ui.port	是
8090	JobServer UI/RestAPI Port	transwarp.jobserver.port	是

A.8. Kafka相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
9092	Port Kafka Broker Listens On	不能修改	是

A.9. Oozie相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
11000	Oozie Listening Port	oozie.http.port	是

A.10. Sqoop相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
12000	Sqoop HTTP Port	sqoop.http.port	是
12001	Sqoop Admin Port	sqoop.admin.port	是

A.11. HUE相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
8000	Web Service Port	不能修改	是

A.12. Elasticsearch相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
9200	ES HTTP Port	不能修改	是
9300	ES Transport Port	不能修改	是

A.13. Transwarp Manager相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
8180	Transwarp Manager Port	不能修改	是
2551	Transwarp Manager Master RPC Port	不能修改	否
10208	Transwarp Manager Agent RPC Port	不能修改	否
22	Transwarp Manager Repo Port	不能修改	否
48121	Transwarp Manager Metric Port	不能修改	否
8649	Ganglia gmond Port (UDP/TCP)	不能修改	否
8651	Ganglia gmetad Port	不能修改	否
123	NTP Port (UDP)	不能修改	否

A.14. Kerberos相关端口

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
88	Kerberos KDC Port (TCP, UDP)	不能修改	是

端口号	端口用途	Transwarp Manager配置项	是否被外部访问
464	Kerberos kpasswd Port (TCP/UDP)	不能修改	是
749	Kerberos admin Port (TCP/UDP)	不能修改	是

客户服务

技术支持

感谢你使用星环信息科技（上海）有限公司的产品和服务。如您在产品使用或服务中有任何技术问题，可以通过以下途径找到我们的技术人员给予解答。

email: support@transwarp.io

技术支持热线电话: 4008 079 976

技术支持QQ专线: 3221723229, 3344341586

官方网址 : www.transwarp.io

意见反馈

如果你在系统安装，配置和使用中发现任何产品问题，可以通过以下方式反馈:

email: support@transwarp.io

感谢你的支持和反馈，我们一直在努力！