

崔炳华

狗币 (DogeCoin) 地址 : DNY1rCigWdqbSFe9oi84MA5DeDEvwQ7tyF




☰ 目录视图

☰ 摘要视图

RSS 订阅


个人资料




i_chips

+ 加关注

✉ 发私信





访问： 9665次

积分： 291分

排名： 千里之外

.....

原创： 19篇 转载： 5篇

译文： 0篇 评论： 6条

文章搜索

Q

文章分类

OpenStack (15)

Cinder (6)

云计算 (6)

GlusterFS (2)

iSCSI (1)

云存储 (5)

Swift (6)

Ubuntu (4)

Quorum (1)

Hadoop (1)

HDFS (1)

Keystone (3)

算法 (1)

LDAP (2)

JXplorer (1)

samba (1)

Windows (1)

💡 有奖征资源，博文分享有内涵 社区问答： 半嵴iOS测试指南 专访阿里陶辉 2014 CSDN博文大赛 10月微软MVP申请

原 [置顶] Ubuntu下Ceph单节点和多节点安装小结

分类： Ceph 2014-02-26 16:39 🔍 1115人阅读 💬 评论(0) 收藏 举报

Ceph Ubuntu 单节点 多节点 安装

目录(?)	[+]
目录(?)	[+]

版权声明：本博客欢迎转载，转载时请以超链接形式标明文章原始出处！谢谢！

博客地址：http://blog.csdn.net/i_chips

Ubuntu下Ceph单节点和多节点安装小结

崔炳华

2014年2月26日

1 概述

Ceph是一个分布式文件系统，在维持POSIX兼容性的同时加入了复制和容错功能。Ceph最大的特点是分布式的元数据服务器，通过CRUSH（Controlled Replication Under Scalable Hashing）这种拟算法来分配文件的location。Ceph的核心是RADOS（ReliableAutonomic Distributed Object Store），一个对象集群存储，本身提供对象的高可用、错误检测和修复功能。

Ceph生态系统架构可以划分为四部分：

- client：客户端（数据用户）。client向外export出一个POSIX文件系统接口，供应用程序调用，并连接mon/mds/osd，进行元数据及数据交互；最原始的client使用FUSE来实现的，现在写到内核里面了，需要编译一个ceph.ko内核模块才能使用。
- mon：集群监视器，其对应的daemon程序为cmon（Ceph Monitor）。mon监视和管理整个集群，对客户端export出一个网络文件系统，客户端可以通过mount -t ceph monitor_ip:/ mount_point命令来挂载Ceph文件系统。根据官方的说法，3个mon可以保证集群的可靠性。
- mds：元数据服务器，其对应的daemon程序为cmds（Ceph Metadata Server）。Ceph里可以有多个MDS组成分布式元数据服务器集群，就会涉及到Ceph中动态目录分割来进行负载均衡。
- osd：对象存储集群，其对应的daemon程序为cosd（Ceph Object StorageDevice）。osd将本地文件系统封装一层，对外提供对象存储的接口，将数据和元数据作为对象存储。这里本地的文件系统可以是ext2/3，但Ceph认为这些文件系统并不能适应osd特殊的访问模式，它们之前自己实现了ebofs，而现在Ceph转用btrfs。

Ceph支持成百上千甚至更多的节点，以上四个部分最好分布在不同的节点上。当然，对于基本的测试，可以把

Ceph (2)

CentOS (1)

内核 (1)

文章存档

2014年06月 (3)

2014年05月 (1)

2014年02月 (2)

2014年01月 (3)

2013年12月 (4)

展开

阅读排行

Ubuntu下Ceph单节点和: (1114)

OpenStack Swift学习笔记 (1053)

Ubuntu下Swift异地多节点部署 (977)

Ubuntu下Swift、Keystor (895)

GlusterFS常用命令小结 (855)

OpenStack常用命令小结 (825)

OpenStack Keystone概念 (478)

Ubuntu下Swift多节点安装 (471)

Ubuntu下Swift单节点安装 (457)

Ubuntu下LDAP和OpenS (400)

评论排行

OpenStack常用命令小结 (3)

Ubuntu下Swift异地多节点部署 (2)

Hadoop分布式文件系统 (1)

OpenStack Swift学习笔记 (0)

LDAP概念小结 (0)

OpenStack Keystone概念 (0)

Ubuntu下配置支持Windc (0)

Ubuntu下Ceph单节点和: (0)

Gluster vs Ceph：开源有 (0)

CentOS 6.4升级内核小结 (0)

推荐文章

最新评论

Hadoop分布式文件系统（HDFS yebai: 记录一下。

OpenStack常用命令小结 i_chips: @ytreel:cinder --version 查看下版本号呢？应该还是没有安装好Cinder吧？

OpenStack常用命令小结 ytreel: 楼主，向您求教一下！我的系统中找不到cinder backup一些列的命令（create delet...

Ubuntu下Swift异地多节点部署指 i_chips: @jiakon:你好，部署到两个以上的区域的话，光纤通道一定要给力，否则确实有可能导致服务质量问题。...

Ubuntu下Swift异地多节点部署指 jiakon: 你好，请教一个问题：你这种部署方式是否会因为两个区域间网络延时导致服务质量问题呢？因为proxy，Z...

OpenStack常用命令小结 waterfall_bo: 这一堆命令。。。真多

mon和mds装在一个节点上，也可以把四个部分全都部署在同一个节点上。

2 环境准备

2.1 版本

- Linux系统版本：Ubuntu Server 12.04.1 LTS；
- Ceph版本：0.72.2（稍后安装）；

2.2 更新source（可选步骤）

“网速较慢”或者“安装软件失败”的情况下，可以考虑替换成国内的镜像：

```
# sudo sed -i's#us.archive.ubuntu.com#mirrors.163.com#g' /etc/apt/sources.list
```

```
# sudo apt-get update
```

Ubuntu 12.04默认的Ceph版本为0.41， 如果希望安装较新的Ceph版本，可以添加key到APT中，更新sources.list：

```
# sudo wget -q -O- 'https://ceph.com/git/?p=ceph.git;a=blob_plain;f=keys/release.asc'| sudo apt-key add -
```

```
# sudo echo deb http://ceph.com/debian/ $(lsb_release -sc) main | sudo tee/etc/apt/sources.list.d/ceph.list
```

```
# sudo apt-get update
```

2.3 系统时间

```
# date # 查看系统时间是否正确，正确的话则忽略下面两步
```

```
# sudo date -s "2013-11-04 15:05:57" # 设置系统时间
```

```
# sudo hwclock -w # 写入硬件时间
```

2.4 关闭防火墙

请确保已关闭SELinux（Ubuntu默认未开启）。

另外，建议关闭防火墙：

```
# sudo ufw disable # 关闭防火墙
```

3 Ceph单节点安装

3.1 节点IP

- 192.168.73.129（hostname为ceph1， 有两个分区/dev/sdb1和/dev/sdb2提供给osd， 另外还安装了client/mon/mds）；

3.2 安装Ceph库

```
# apt-get install ceph ceph-common ceph-mds
```

```
# ceph -v # 将显示ceph的版本和key信息
```

3.3 创建Ceph配置文件

```
# vim /etc/ceph/ceph.conf
```

```
[global]
```

```
max open files = 131072
```

```
#For version 0.55 and beyond, you must explicitly enable
```

```
#or disable authentication with "auth" entries in [global].
```

```
auth cluster required = none
```


auth service required = none

auth client required = none

[osd]

osd journal size = 1000

#The following assumes ext4 filesystem.

filestore xattruse omap = true

#For Bobtail (v 0.56) and subsequent versions, you may

#add settings for mkcephfs so that it will create and mount

#the file system on a particular OSD for you. Remove the comment `#`

#character for the following settings and replace the values

#in braces with appropriate values, or leave the following settings

#commented out to accept the default values. You must specify the

#--mkfs option with mkcephfs in order for the deployment script to

#utilize the following settings, and you must define the 'devs'

#option for each osd instance; see below.

osd mkfs type = xfs

osd mkfs options xfs = -f #default for xfs is "-f"

osd mount options xfs = rw,noatime # default mount option is"rw,noatime"

#For example, for ext4, the mount option might look like this:

#osd mkfs options ext4 = user_xattr,rw,noatime

#Execute \$ hostname to retrieve the name of your host,

#and replace {hostname} with the name of your host.

#For the monitor, replace {ip-address} with the IP

#address of your host.

[mon.a]

host = ceph1

mon addr = 192.168.73.129:6789

[osd.0]

```
host = ceph1
```

```
#For Bobtail (v 0.56) and subsequent versions, you may
```

```
#add settings for mkcephfs so that it will create and mount
```

```
#the file system on a particular OSD for you. Remove the comment `#`
```

```
#character for the following setting for each OSD and specify
```

```
#a path to the device if you use mkcephfs with the --mkfs option.
```

```
devs = /dev/sdb1
```

```
[osd.1]
```

```
host= ceph1
```

```
devs= /dev/sdb2
```

```
[mds.a]
```

```
host= ceph1
```

注意，对于较低的Ceph版本（例如0.42），需要在[mon]项下添加一行内容：`mondata = /data/$name`，以及在[osd]项下添加一行内容：`osd data = /data/$name`，以作为后续的数据目录；相应的，后续针对数据目录的步骤也需要调整。

3.4 创建数据目录

```
# mkdir -p /var/lib/ceph/osd/ceph-0
```

```
# mkdir -p /var/lib/ceph/osd/ceph-1
```

```
# mkdir -p /var/lib/ceph/mon/ceph-a
```

```
# mkdir -p /var/lib/ceph/mds/ceph-a
```

3.5 为osd创建分区和挂载

对新分区进行xfs或者btrfs的格式化：

```
# mkfs.xfs -f /dev/sdb1
```

```
# mkfs.xfs -f /dev/sdb2
```

第一次必须先挂载分区写入初始化数据：

```
# mount /dev/sdb1 /var/lib/ceph/osd/ceph-0
```

```
# mount /dev/sdb2 /var/lib/ceph/osd/ceph-1
```

3.6 执行初始化

注意，每次执行初始化之前，都需要先停止Ceph服务，并清空原有数据目录：

```
# /etc/init.d/ceph stop
```

```
# rm -rf /var/lib/ceph/*/ceph-*/*
```

然后，就可以在mon所在的节点上执行初始化了：

```
# sudo mkcephfs -a -c /etc/ceph/ceph.conf -k /etc/ceph/ceph1.keyring
```

注意，一旦配置文件ceph.conf发生改变，初始化最好重新执行一遍。

3.7 启动Ceph服务

在mon所在的节点上执行：

```
# sudo service ceph -a start
```

注意，执行上面这步时，可能会遇到如下提示：

```
=== osd.0 ===
```

```
Mounting xfs onceph4:/var/lib/ceph/osd/ceph-0
```

```
Error ENOENT: osd.0 does not exist. create it before updating the crush map
```

执行如下命令后，再重复执行上面那条启动服务的命令，就可以解决：

```
# ceph osd create
```

3.8 执行健康检查

```
# sudo ceph health # 也可以使用ceph -s命令查看状态
```

如果返回的是HEALTH_OK，则代表成功！

注意，如果遇到如下提示：

```
HEALTH_WARN 576 pgs stuckinactive; 576 pgs stuck unclean; no osds
```

或者遇到如下提示：

```
HEALTH_WARN 178 pgs peering; 178pgs stuck inactive; 429 pgs stuck unclean; recovery 2/24 objects degraded(8.333%)
```

执行如下命令，就可以解决：

```
# ceph pg dump_stuck stale && cephpq dump_stuck inactive && ceph pg dump_stuck unclean
```

如果遇到如下提示：

```
HEALTH_WARN 384 pgs degraded; 384 pgs stuck unclean; recovery 21/42degraded (50.000%)
```

则说明osd数量不够，Ceph默认至少需要提供两个osd。

3.9 Ceph测试

客户端挂载mon所在的节点：

```
# sudo mkdir /mnt/mycephfs
```

```
# sudo mount -t ceph 192.168.73.129:6789:/ /mnt/mycephfs
```

客户端验证：

```
# df -h #如果能查看到/mnt/mycephfs的使用情况，则说明Ceph安装成功。
```

4 Ceph多节点安装

对于多节点的情况，Ceph有如下要求：

- 修改各自的hostname，并能够通过hostname来互相访问。
- 各节点能够ssh互相访问而不输入密码（通过ssh-keygen命令）。

4.1 节点IP

- 192.168.73.129（hostname为ceph1，有一个分区/dev/sdb1提供给osd）；
- 192.168.73.130（hostname为ceph2，有一个分区/dev/sdb1提供给osd）；
- 192.168.73.131（hostname为ceph3，安装了client/mon/mds）；

4.2 配置主机名

在每个节点上设置相应的主机名，例如：

```
# vim /etc/hostname
```

```
ceph1
```

修改/etc/hosts，增加如下几行：

```
192.168.73.129 ceph1
```

```
192.168.73.130 ceph2
```

```
192.168.73.131 ceph3
```

4.3 配置免密码访问

在每个节点上创建RSA密钥：

```
# ssh-keygen -t rsa # 一直按确定键即可
```

```
# touch /root/.ssh/authorized_keys
```

先配置ceph1，这样ceph1就可以无密码访问ceph2和ceph3了：

```
ceph1# scp /root/.ssh/id_rsa.pub ceph2:/root/.ssh/id_rsa.pub_ceph1
```

```
ceph1# scp /root/.ssh/id_rsa.pub ceph3:/root/.ssh/id_rsa.pub_ceph1
```

```
ceph1# ssh ceph2 "cat /root/.ssh/id_rsa.pub_ceph1>> /root/.ssh/authorized_keys"
```

```
ceph1# ssh ceph3 "cat /root/.ssh/id_rsa.pub_ceph1 >> /root/.ssh/authorized_keys"
```

节点ceph2和ceph3也需要参照上面的命令进行配置。

4.4 安装Ceph库

在每个节点上安装Ceph库：

```
# apt-get install ceph ceph-common ceph-mds
```

```
# ceph -v # 将显示ceph的版本和key信息
```

4.5 创建Ceph配置文件

```
# vim /etc/ceph/ceph.conf
```

```
[global]
```

```
max open files = 131072
```

```
#For version 0.55 and beyond, you must explicitly enable
```

```
#or disable authentication with "auth" entries in [global].
```

```
auth cluster required = none
```

```
auth service required = none
```

```
auth client required = none
```


[osd]

osd journal size = 1000

#The following assumes ext4 filesystem.

filestore xattruse omap = true

#For Bobtail (v 0.56) and subsequent versions, you may

#add settings for mkcephfs so that it will create and mount

#the file system on a particular OSD for you. Remove the comment `#`

#character for the following settings and replace the values

#in braces with appropriate values, or leave the following settings

#commented out to accept the default values. You must specify the

#--mkfs option with mkcephfs in order for the deployment script to

#utilize the following settings, and you must define the 'devs'

#option for each osd instance; see below.

osd mkfs type = xfs

osd mkfs options xfs = -f #default for xfs is "-f"

osd mount options xfs = rw,noatime # default mount option is"rw,noatime"

#For example, for ext4, the mount option might look like this:

#osd mkfs options ext4 = user_xattr,rw,noatime

#Execute \$ hostname to retrieve the name of your host,

#and replace {hostname} with the name of your host.

#For the monitor, replace {ip-address} with the IP

#address of your host.

[mon.a]

host = ceph3

mon addr = 192.168.73.131:6789

[osd.0]

host = ceph1

#For Bobtail (v 0.56) and subsequent versions, you may

```
#add settings for mkcephfs so that it will create and mount
```

```
#the file system on a particular OSD for you. Remove the comment `#`
```

```
#character for the following setting for each OSD and specify
```

```
#a path to the device if you use mkcephfs with the --mkfs option.
```

```
devs = /dev/sdb1
```

```
[osd.1]
```

```
host = ceph2
```

```
devs= /dev/sdb1
```

```
[mds.a]
```

```
host = ceph3
```

配置文件创建成功之后，还需要拷贝到除纯客户端之外的每个节点中（并且后续也要始终保持一致）：

```
ceph1# scp /etc/ceph/ceph.conf ceph2:/etc/ceph/ceph.conf
```

```
ceph1# scp /etc/ceph/ceph.conf ceph3:/etc/ceph/ceph.conf
```

4.6 创建数据目录

在每个节点上创建数据目录：

```
# mkdir -p /var/lib/ceph/osd/ceph-0
```

```
# mkdir -p /var/lib/ceph/osd/ceph-1
```

```
# mkdir -p /var/lib/ceph/mon/ceph-a
```

```
# mkdir -p /var/lib/ceph/mds/ceph-a
```

4.7 为osd创建分区和挂载

对于osd所在的节点ceph1和ceph2，需要对新分区进行xfs或者btrfs的格式化：

```
# mkfs.xfs -f /dev/sdb1
```

对于节点ceph1和ceph2，第一次必须先分别挂载分区写入初始化数据：

```
ceph1# mount /dev/sdb1 /var/lib/ceph/osd/ceph-0
```

```
ceph2# mount /dev/sdb1 /var/lib/ceph/osd/ceph-1
```

4.8 执行初始化

注意，每次执行初始化之前，都需要在每个节点上先停止Ceph服务，并清空原有数据目录：

```
# /etc/init.d/ceph stop
```

```
# rm -rf /var/lib/ceph/*/ceph-*/*
```

然后，就可以在mon所在的节点ceph3上执行初始化了：

```
# sudo mkcephfs -a -c /etc/ceph/ceph.conf -k /etc/ceph/ceph3.keyring
```

注意，一旦配置文件ceph.conf发生改变，初始化最好重新执行一遍。

4.9 启动Ceph服务

在mon所在的节点ceph3上执行：

```
# sudo service ceph -a start
```

注意，执行上面这步时，可能会遇到如下提示：

```
=== osd.0 ===
```

```
Mounting xfs onceph4:/var/lib/ceph/osd/ceph-0
```

```
Error ENOENT: osd.0 does not exist. create it before updating the crush map
```

执行如下命令后，再重复执行上面那条启动服务的命令，就可以解决：

```
# ceph osd create
```

4.10 执行健康检查

```
# sudo ceph health # 也可以使用ceph -s命令查看状态
```

如果返回的是HEALTH_OK，则代表成功！

注意，如果遇到如下提示：

```
HEALTH_WARN 576 pgs stuckinactive; 576 pgs stuck unclean; no osds
```

或者遇到如下提示：

```
HEALTH_WARN 178 pgs peering; 178pgs stuck inactive; 429 pgs stuck unclean; recovery 2/24 objects degraded(8.333%)
```

执行如下命令，就可以解决：

```
# ceph pg dump_stuck stale && cephpq dump_stuck inactive && ceph pg dump_stuck unclean
```

如果遇到如下提示：

```
HEALTH_WARN 384 pgs degraded; 384 pgs stuck unclean; recovery 21/42degraded (50.000%)
```

则说明osd数量不够，Ceph默认至少需要提供两个osd。

4.11 Ceph测试

客户端（节点ceph3）挂载mon所在的节点（节点ceph3）：

```
# sudo mkdir /mnt/mycephfs
```

```
# sudo mount -t ceph 192.168.73.131:6789:/ /mnt/mycephfs
```

客户端验证：

```
# df -h #如果能查看到/mnt/mycephfs的使用情况，则说明Ceph安装成功。
```

5 参考资料

- 1) 《Ceph：一个 Linux PB 级分布式文件系统》，<http://www.ibm.com/developerworks/cn/linux/l-ceph/>
- 2) 《Ubuntu 12.04 Ceph分布式文件系统之部署》，<http://hobbylinux.blog.51cto.com/2895352/1175932>
- 3) 《How To Install CephOn FC12, FC上安装Ceph分布式文件系
统》，<http://blog.csdn.net/polisan/article/details/5624207>
- 4) 《Ceph配置说明书》，<http://www.doc88.com/p-9062058419946.html>

主题推荐

- ubuntu

分布式文件系统

文件系统

负载均衡

应用程序

猜你在找

- centos 6.4下ceph-0.72的安装

Openstack cinder配置iscsi存储专用网络

分布式文件系统选型

OSDI，SOSP与美国著名计算机系的调查

一位老码农的编程简史

centos 6.4下ceph-0.72的安装

Openstack cinder配置iscsi存储专用网络

分布式文件系统选型

OSDI，SOSP与美国著名计算机系的调查

一位老码农的编程简史

OpenStack Cinder源码分析之六

openstack nova resize API 详解

OpenStack Nova：虚拟机初始化user-data & Cloud-

OpenStack Hacker养成指南

【NPC】12、汉密尔顿回路规约到汉密尔顿路径

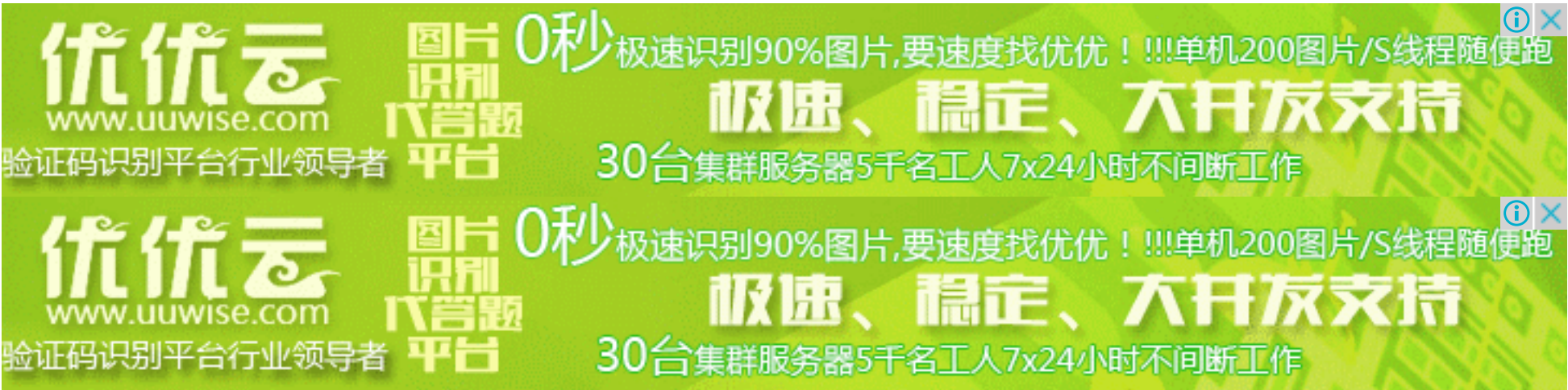
OpenStack Cinder源码分析之六

openstack nova resize API 详解

OpenStack Nova：虚拟机初始化user-data & Cloud-

OpenStack Hacker养成指南

【NPC】12、汉密尔顿回路规约到汉密尔顿路径



查看评论

暂无评论

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

核心技术类目

全部主题

Java

VPN

Android

iOS

ERP

IE10

Eclipse

CRM

JavaScript

Ubuntu

NFC

WAP

jQuery

数据库

BI

HTML5

Spring

Apache

Hadoop

.NET

API

HTML

SDK

IIS

Fedora

XML

LBS

Unity

Splashtop

UML

components

Windows Mobile

Rails

QEMU

KDE

Cassandra

CloudStack

FTC

coremail

OPhone

CouchBase

云计算

iOS6

Rackspace

Web App

SpringSide

Maemo

Compuware

大数据

aptech

Perl

Tornado

Ruby

Hibernate

ThinkPHP

Spark

HBase

Pure

Solr

Angular

Cloud Foundry

Redis

Scala

Django

Bootstrap

公司简介 | 招贤纳士 | 广告服务 | 银行汇款帐号 | 联系方式 | 版权声明 | 法律顾问 | 问题报告 | 合作伙伴 | 论坛反馈

网站客服
 杂志客服
 微博客服
 webmaster@csdn.net
 400-600-2320

京 ICP 证 070598 号

北京创新乐知信息技术有限公司 版权所有

江苏乐知网络技术有限公司 提供商务支持

Copyright © 1999-2014, CSDN.NET, All Rights Reserved



公司简介 | 招贤纳士 | 广告服务 | 银行汇款帐号 | 联系方式 | 版权声明 | 法律顾问 | 问题报告 | 合作伙伴 | 论坛反馈

