

The background of the slide is a nighttime city skyline, likely Hong Kong, with numerous skyscrapers illuminated. Overlaid on the city are dynamic, glowing blue light trails that curve across the sky, suggesting data flow or digital connectivity. The overall color palette is dominated by deep blues and bright whites from the city lights.

H3C

数字化解决方案领导者

ONEStor模块之tgt详解 (35)

01

tgt 模块的作用

02

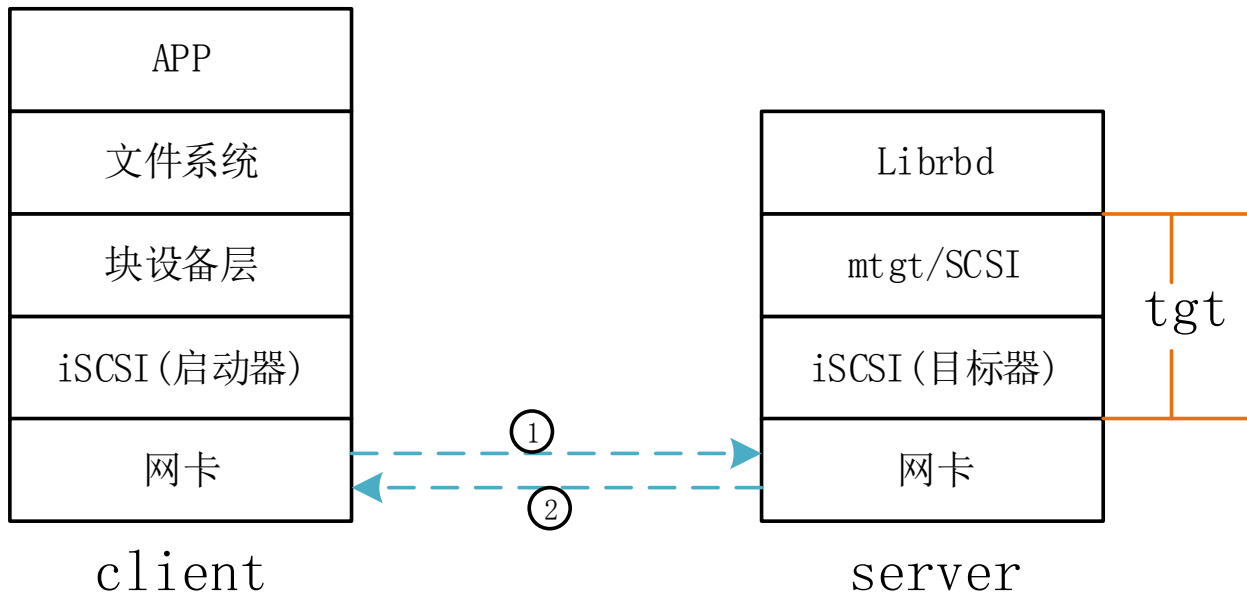
tgt相关模块和协议的介绍

03

tgt问题定位思路

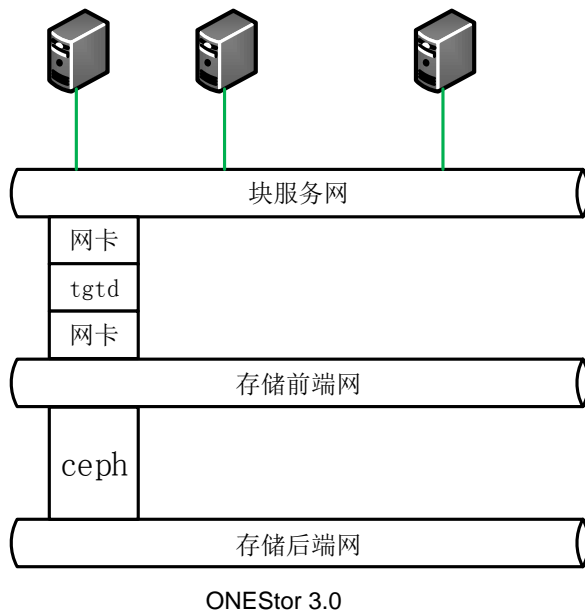
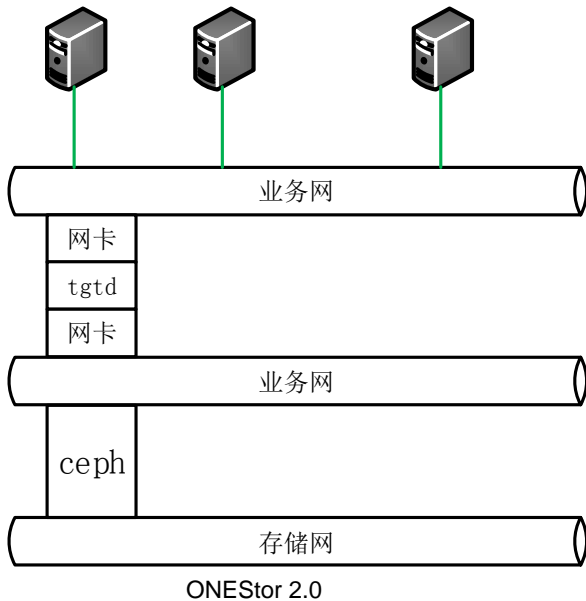
- tgt的架构：应用架构、io网络架构和管理配置架构
- tgt在io读写中的应用

➤ 应用架构

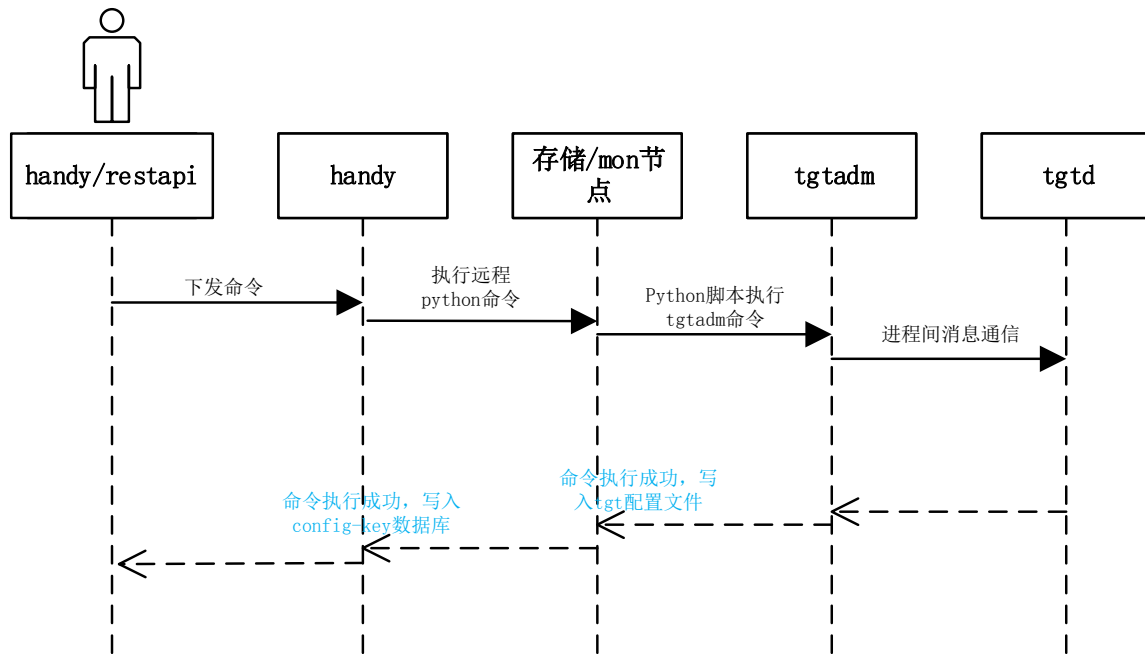


tgt常见应用图

➤ IO网络架构

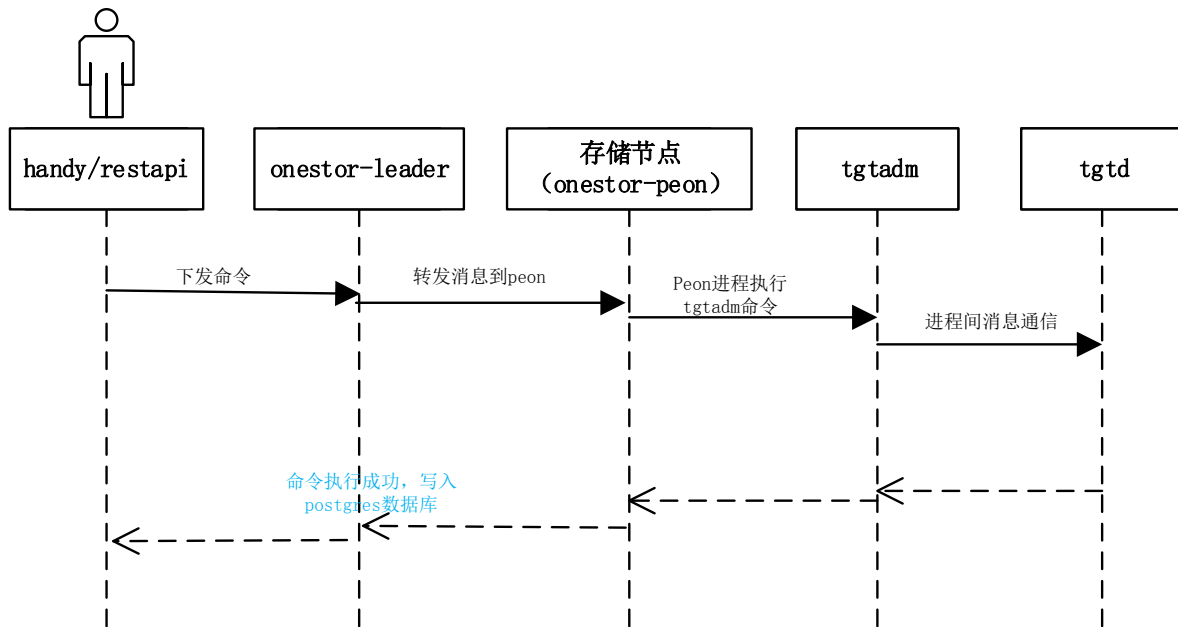


➤ 配置管理架构-2.0

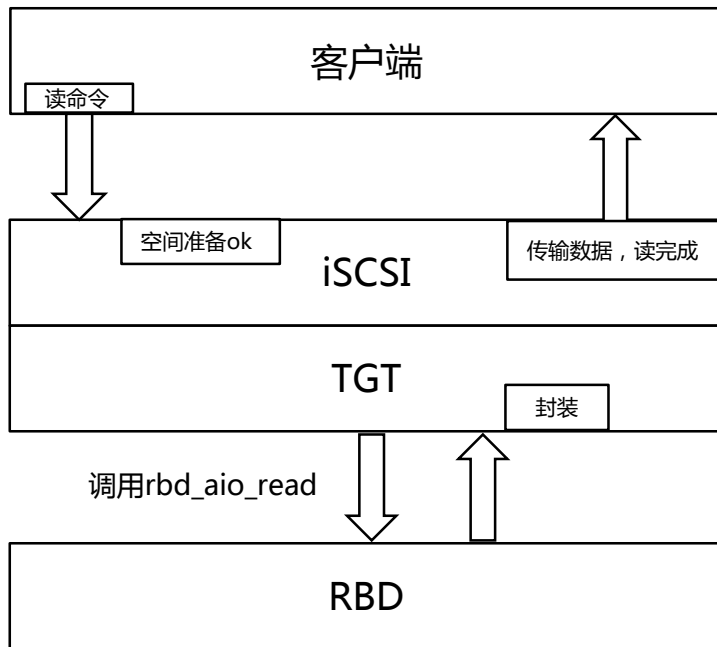
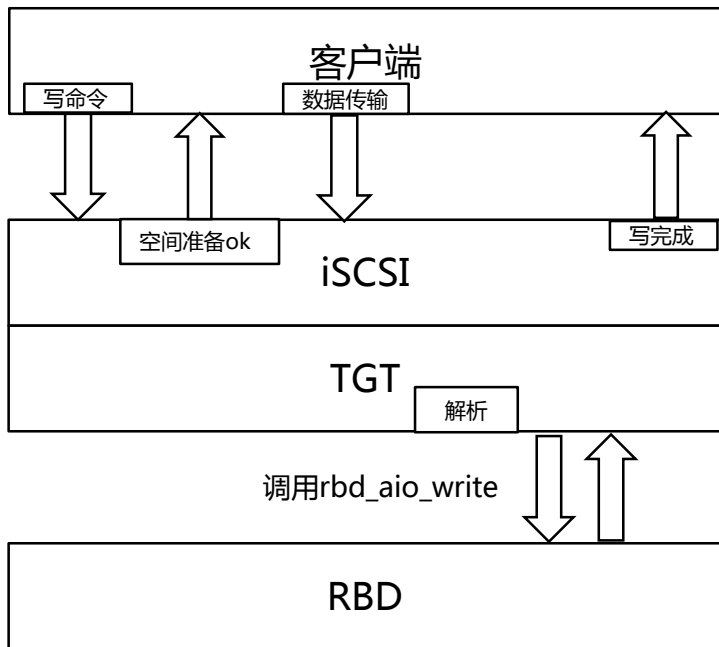


ONESstor 2.0

➤ 配置管理架构-3.0



ONESstor 3.0



01

tgt 模块的作用

02

tgt相关模块和协议的介绍

03

tgt问题的定位思路

ONEStor目标器向外提供的块存储，使用open-iscsi、或者基于iSCSI协议使用存储，使用tcpdump抓包可以看到：

No.	Time	Source	Destination	Protocol	Length	Info
6	2019-11-03 14:01:14.873159	192.169.168.172	192.169.168.170	iSCSI	302	Login Command
8	2019-11-03 14:01:14.873260	192.169.168.170	192.169.168.172	iSCSI	278	Login Response (Success)
11	2019-11-03 14:01:14.873382	192.169.168.172	192.169.168.170	iSCSI	70	Text Command
13	2019-11-03 14:01:14.873419	192.169.168.170	192.169.168.172	iSCSI	214	Text Response
20	2019-11-03 14:01:14.891340	192.169.168.172	192.169.168.170	iSCSI	534	Login Command
22	2019-11-03 14:01:14.891525	192.169.168.170	192.169.168.172	iSCSI	434	Login Response (Success)
24	2019-11-03 14:01:14.892369	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x00
25	2019-11-03 14:01:14.892444	192.169.168.170	192.169.168.172	iSCSI	138	SCSI: Data In LUN: 0x00 (Inquiry Response Data) [SCSI transfer limited due to allocation_length...
26	2019-11-03 14:01:14.892463	192.169.168.170	192.169.168.172	iSCSI	126	SCSI Response (Check Condition) LUN:0x00
28	2019-11-03 14:01:14.892602	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x00
29	2019-11-03 14:01:14.892631	192.169.168.170	192.169.168.172	iSCSI	138	SCSI: Data In LUN: 0x00 (Inquiry Response Data) [SCSI transfer limited due to allocation_length...
30	2019-11-03 14:01:14.892751	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x00
31	2019-11-03 14:01:14.892782	192.169.168.170	192.169.168.172	iSCSI	170	SCSI: Data In LUN: 0x00 (Inquiry Response Data) [SCSI transfer limited due to allocation_length...
32	2019-11-03 14:01:14.892890	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x00 Supported Vital Product Data Pages
33	2019-11-03 14:01:14.892922	192.169.168.170	192.169.168.172	iSCSI	114	SCSI: Data In LUN: 0x00 (Inquiry Response Data) SCSI: Response LUN: 0x00 (Inquiry) (Good)
34	2019-11-03 14:01:14.893011	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x00 Unit Serial Number Page
35	2019-11-03 14:01:14.893037	192.169.168.170	192.169.168.172	iSCSI	142	SCSI: Data In LUN: 0x00 (Inquiry Response Data) SCSI: Response LUN: 0x00 (Inquiry) (Good)
36	2019-11-03 14:01:14.893134	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x00 Device Identification Page
37	2019-11-03 14:01:14.893160	192.169.168.170	192.169.168.172	iSCSI	178	SCSI: Data In LUN: 0x00 (Inquiry Response Data) SCSI: Response LUN: 0x00 (Inquiry) (Good)
38	2019-11-03 14:01:14.894193	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Report LUNs LUN: 0x00
39	2019-11-03 14:01:14.894280	192.169.168.170	192.169.168.172	iSCSI	126	SCSI: Data In LUN: 0x00 (Report LUNs Response Data) SCSI: Response LUN: 0x00 (Report LUNs) (Go...
40	2019-11-03 14:01:14.894699	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x01
41	2019-11-03 14:01:14.894738	192.169.168.170	192.169.168.172	iSCSI	138	SCSI: Data In LUN: 0x01 (Inquiry Response Data) [SCSI transfer limited due to allocation_length...
42	2019-11-03 14:01:14.894833	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x01
43	2019-11-03 14:01:14.894859	192.169.168.170	192.169.168.172	iSCSI	170	SCSI: Data In LUN: 0x01 (Inquiry Response Data) [SCSI transfer limited due to allocation_length...
44	2019-11-03 14:01:14.894958	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x01 Supported Vital Product Data Pages
45	2019-11-03 14:01:14.894983	192.169.168.170	192.169.168.172	iSCSI	114	SCSI: Data In LUN: 0x01 (Inquiry Response Data) SCSI: Response LUN: 0x01 (Inquiry) (Good)
46	2019-11-03 14:01:14.895070	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x01 Unit Serial Number Page
47	2019-11-03 14:01:14.895095	192.169.168.170	192.169.168.172	iSCSI	142	SCSI: Data In LUN: 0x01 (Inquiry Response Data) SCSI: Response LUN: 0x01 (Inquiry) (Good)
48	2019-11-03 14:01:14.895176	192.169.168.172	192.169.168.170	iSCSI	102	SCSI: Inquiry LUN: 0x01 Device Identification Page



iSCSI、SCSI协议区分不开？目标器如何向外提供服务的？基本架构是怎么样的？

iSCSI:

负责iSCSI目标器的生成、目标器针对网卡端口的监听、iSCSI链路的建立销毁、CHAP认证（3.0）、负载均衡（3.0）

SCSI:

协议的解析、各类IO处理

- 负载iSCSI目标器的生成
- 目标器针对网卡端口的监听
- iSCSI链路的建立销毁
- CHAP认证（3.0）
- 负载均衡（3.0）

➤ 负载iSCSI目标器的生成

ONESTor 2.0是通过handy web或者restapi创建的;

ONESTor3.0是在配置块服务网段时, 根据存储集群的fsid生成的默认target。

➤ 目标器针对网卡端口的监听

ONESTor 2.0在tgt进程初始化时, 根据集群配置文件的业务网段, 找到指定的网卡, 进行3260端口的绑定监听;

ONESTor3.0分两部分:

- 1、在tgt主进程初始化配置恢复完成后, 会进行回环网卡的监听。
- 2、在端口的生成, 包括已有块服务网段内网卡扫描网卡, 没有块服务网段内网卡配置块服务网段IP, 找到块服务网段的网卡, 进行3260端口的绑定监听;



两个版本都会有网卡的绑定监听, 所以在部署好集群后, 不要进行网络分配的改变, 如果有改变需要重启tgt进程以恢复监听服务。

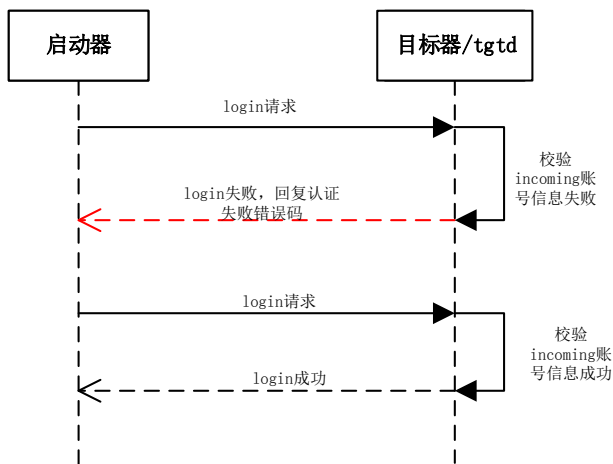
2.0bond卡发生变化, 包括不限于单独的bond卡重启, 需要重启tgt进程以恢复监听服务。

➤ iSCSI链路的建立销毁

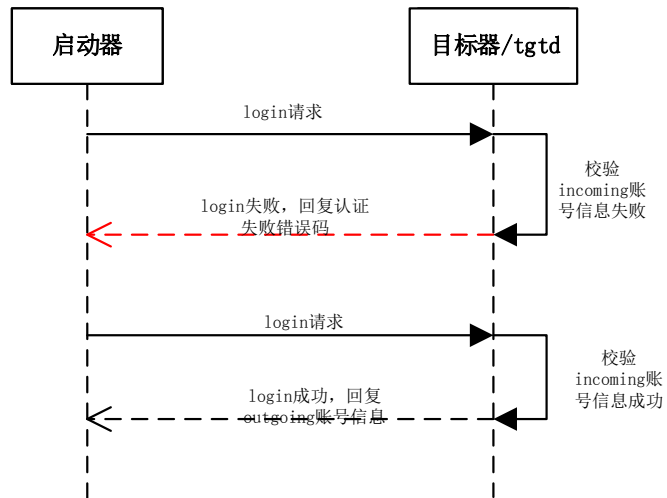
在网络正常，iSCSI监听正常的情况下，进行tcp连接的建立，并协商iSCSI信息。

➤ CHAP认证（3.0）

实现的是在iSCSI login阶段进行认证

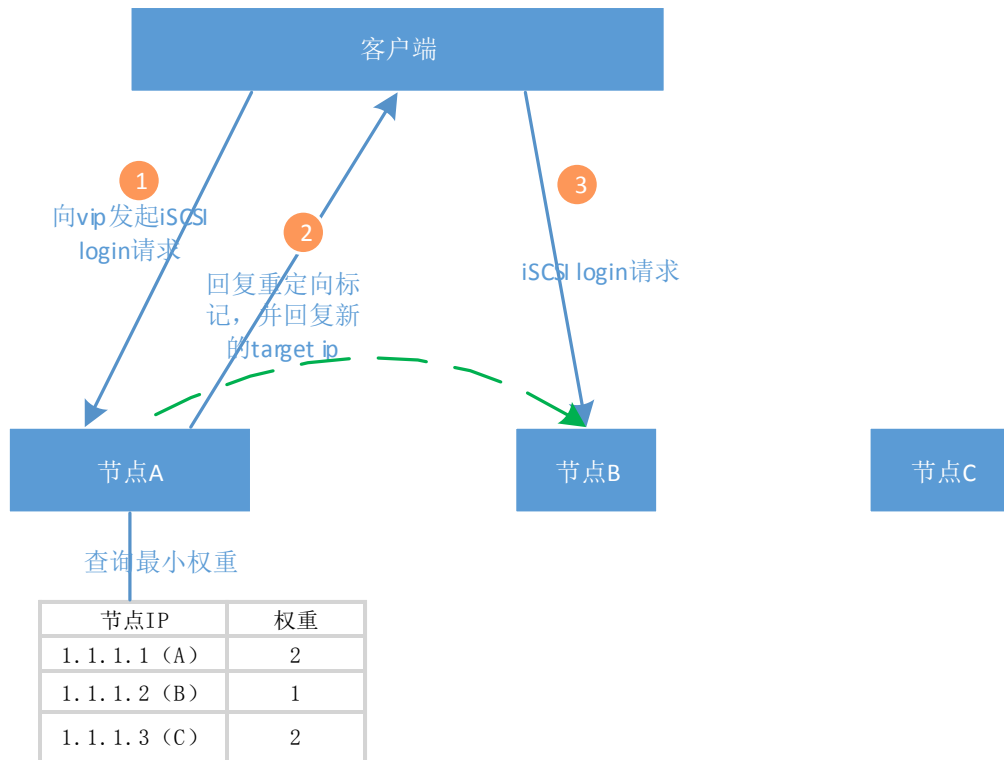


单向认证



双向认证

➤ 负载均衡 (3.0)



至此，启动器和目标器之间的通道连接完成，下一步就是正式的业务运行。

- 协议的解析
- 各类IO处理

➤ 协议的解析

```

24 2019-11-03 14:01:14.892369 192.169.168.172 192.169.168.170 iSCSI 102 SCSI: Inquiry LUN: 0x00
25 2019-11-03 14:01:14.892444 192.169.168.170 192.169.168.172 iSCSI 138 SCSI: Data In LUN: 0x00 (Inquiry Response Data) [S
26 2019-11-03 14:01:14.892463 192.169.168.170 192.169.168.172 iSCSI 126 SCST Response (Check Condition) LUN:0x00

Frame 24: 102 bytes on wire (816 bits), 102 bytes captured (816 bits)
Ethernet II, Src: HuaweiTe_88:c6:9d (28:6e:d4:88:c6:9d), Dst: HuaweiTe_88:c6:98 (28:6e:d4:88:c6:98)
Internet Protocol Version 4, Src: 192.169.168.172, Dst: 192.169.168.170
Transmission Control Protocol, Src Port: 52920, Dst Port: 3260, Seq: 529, Ack: 381, Len: 48
iSCSI (SCSI Command)
  Opcode: SCSI Command (0x01)
  .0.. .... = I: Queued delivery
  TotalAHSLength: 0x00
  DataSegmentLength: 0 (0x00000000)
  LUN
    .00.. .... = Address Mode: Simple logical unit addressing (0x00)
    ..00 0000 0000 0000 = LUN: 0x0000
  InitiatorTaskTag: 0x01000000
  ExpectedDataTransferLength: 0x00000024
  CmdSN: 0x00000000
  ExpStatSN: 0x00000001
  Data In in: 25
  Response in: 26
  Flags: 0xc1, F, R, Attr: Simple
  1... .... = F: Final PDU in sequence
  .1.. .... = R: Data will be read from target
  ..0. .... = W: No data will be written to target
  .... 001 = Attr: Simple (0x1)
SCSI CDB Inquiry

30 1d 90 d2 f4 00 00 01 c1 00 00 00 00 00 00 00 .....
40 00 00 00 00 00 00 01 00 00 00 00 00 24 00 00 .....$.
50 00 00 00 00 00 01 12 00 00 00 24 00 00 00 00 .....$.
60 00 00 00 00 00 00 .....
  
```

从抓包数据就可以看到SCSI协议是包含在iSCSI协议中的，iSCSI协议在TCP协议中，所以，解析到iSCSI协议之后，需要解析出SCSI报文。即iSCSI协议是传输协议，SCSI是业务协议。

➤ 各类IO处理

- ◆ 查询存储信息
- ◆ 扫盘流程
- ◆ 读写类IO
- ◆ VAAI
- ◆ SCSI集群

➤ 各类IO处理

◆ 查询存储信息

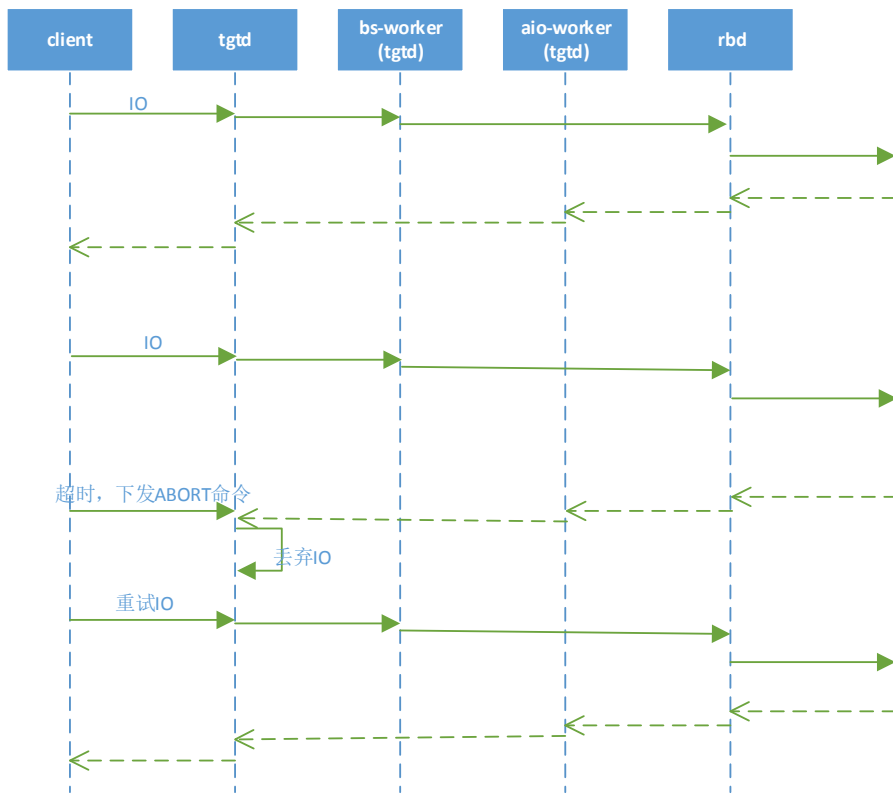
- 实际上针对LUN 0进行命令查询；
- 查询LUN 0的SN等一系列信息生成控制卷；
- 控制卷的作用是承载客户端第一次report luns命令；

◆ 扫盘流程

- INQUIRY: 查询盘的基本属性, VID/PID/SN等；
- Mode Sense: 查询盘的其他基本属性, 写保护开关、写缓存开关等；
- Mgmt Protocol In: 查询盘支持的SCSI命令字；
- TUR: 查询盘是否准备好；
- READ CAPACITY: 查询盘的容量；
- READ: 一般是读盘开始位置, 一般属于盘的元数据区；

➤ 各类IO处理

◆ 读写类IO



01

tgt 模块的作用

02

tgt相关模块和协议的介绍

03

tgt问题的定位思路

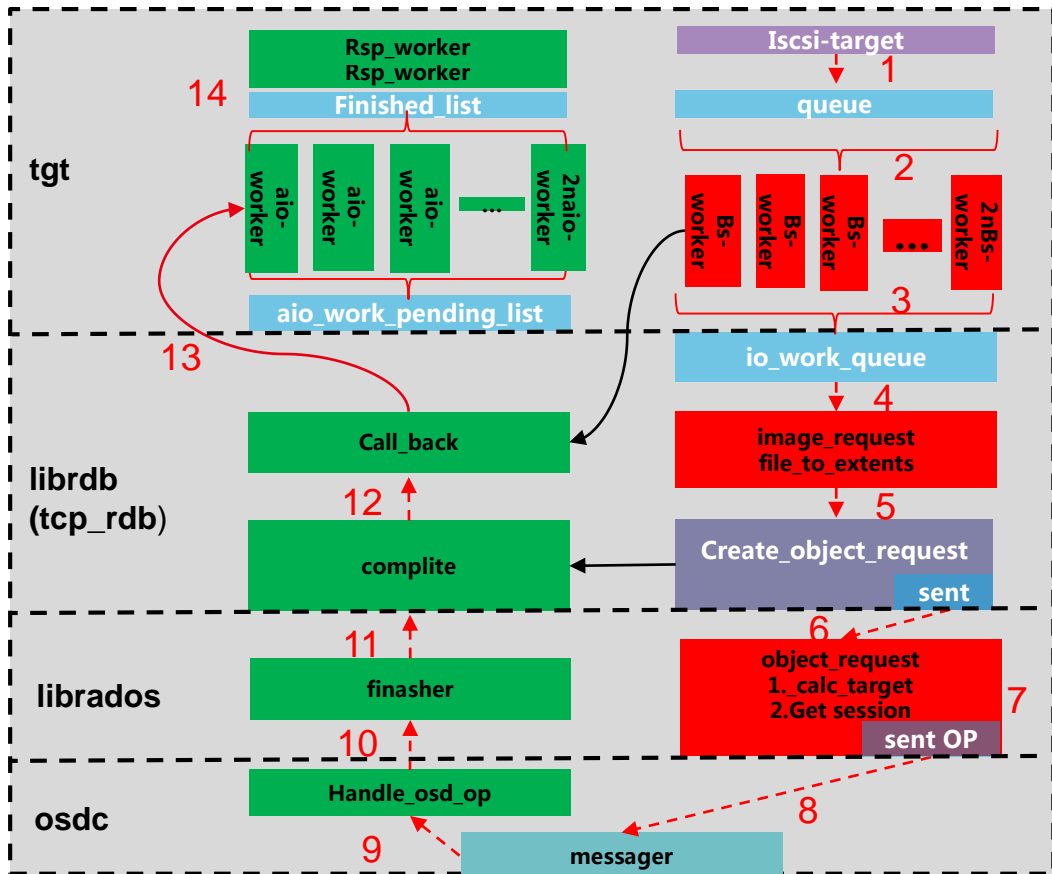
ONEStor 2.0业务对接流程:

1. 登录ONEStor界面，点选存储管理—>块存储—>块设备，点击创建，选择Pool为default;
2. 以iSCSI Target方式创建基于default的存储块，并指定名称为performance，容量为100GB，同时选择新建target，名称为performance;
3. 创建成功后弹出是否开启IQN认证对话框，点击开启，填入客户端iqn（在CAS集群中点击主机，存储适配器查看）；此处也可以选择不开启，对性能测试无影响；
4. 在客户端后台执行iscsiadm -m discovery -t sendtargets -p xx.xx.xx.xx（存储卷业务网IP），可得到预期结果(1)；
5. 在客户端后台执行iscsiadm -m node --targetname iqn.2016-01.com.h3c.onestor.performance -p xx.xx.xx.xx --login（存储卷业务网IP，同上），可得到预期结果(2)；

ONEStor 3.0创建完业务主机组和业务主机后，将卷映射给业务主机组，后进行4-5步操作。

如果无法发现，问题存在哪里？

如果后台可以发现，但是前台无法发现，问题存在哪里？



若在主高可用节点的syslog日志中存在较多io error的打印，可以查看osdc计数连接数

```
tgttd: procallocresp(230) io error 0x7f2082c421d0 8a -110
tgttd: procallocresp(230) io error 0x7f2082c3d3d0 8a -110
tgttd: target_cmd_done(1535) rsp for abort task
```

1. 进入/var/run/ceph目录下有个ceph-tgt.XXXX.asok的文件；
2. 执行 `ceph --admin-daemon ceph-tgt.xxxxxx.asok perf dump > perf.log` 将dump日志保存到perf.log中；
3. 查看perf.log中计数消息，是否超过 2^{32} （400多亿）。

H3C ONEStor 分布式存储

THANKS

— www.h3c.com —

