



Deploying ITD: Server Traffic distribution Using Direct Server Return

Table of Contents:Deploying ITD: Server Traffic distribution Using Direct Server Return	1
1. Introduction :	1
2. Traditional DSR mode using External Load Balancers.....	1
3. DSR mode using Nexus – ITD.	2
4. Deploying ITD for SLB-DSR.....	3
4.1 Server configuration	4
4.2 Nexus ITD configuration	5
4.3 Verification	6
5. Guidelines and Limitations	8
6. References and Further reading.	8

1. Introduction :

Intelligent Traffic Director (ITD) is an ASIC based multi-terabit layer 4 traffic distribution and clustering solution on the Cisco Nexus 5000/6000/7000/9000 Series Switches. ITD provides scalable traffic distribution of client requests to a group of servers or service appliances.

This document describes a typical deployment scenario for distributing traffic with ITD using Direct Server Return (DSR) mode using Nexus switches.

2. Traditional DSR mode using External Load Balancers

With Server Load Balancing (SLB), incoming requests for a particular service are distributed across multiple servers for reasons such as load sharing, capacity and redundancy.

In a traditional DSR mode deployment, the incoming traffic from the client-facing network device is destined to an external Load-balancer appliance via a Virtual IP (VIP). The same IP address is also configured on all the servers (also termed nodes) as a loopback-IP-address, which enables the servers to reply to the client directly using the VIP. This ensures that the traffic bypasses the Load-balancer on the return path thus eliminating any bottlenecks in flow setups or overall throughput. This makes Direct Server Return (DSR) an ideal choice for stateless services like DNS load-balancing and for services with a large amount of server return data to the client like video services.

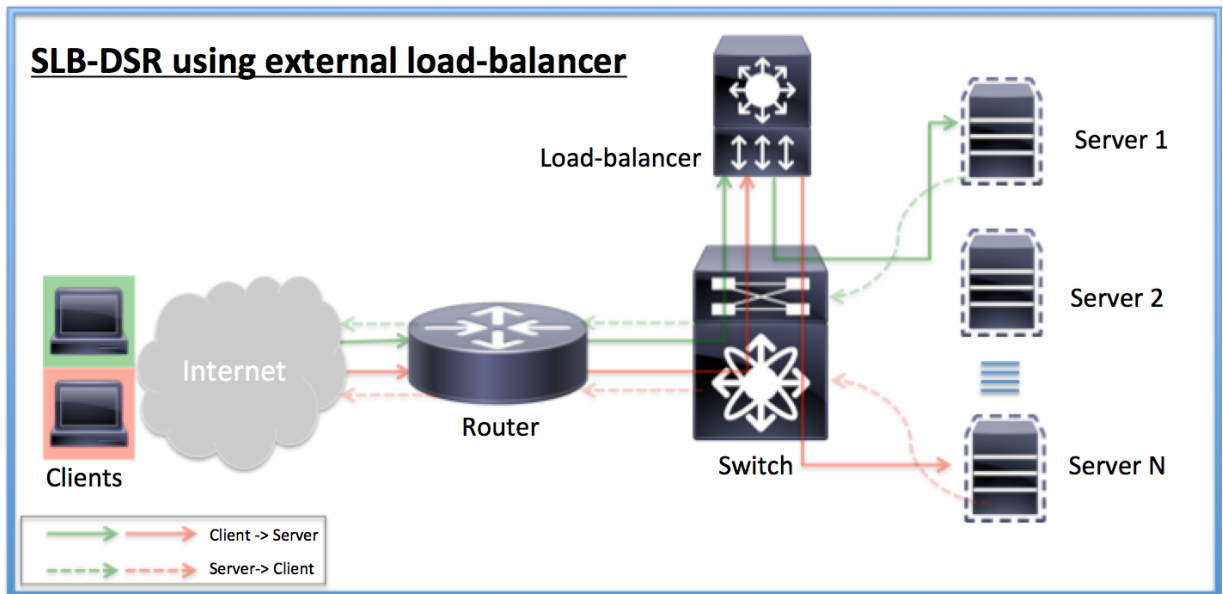


Figure 1 DSR Mode using External Load-Balancer(s).

However this approach requires an external Load Balancer with its separate configuration and management, in addition to the Routed/Switched network devices. Redundancy considerations also require that multiple Load-balancers be deployed for HA or Clustered designs.

3. DSR mode using Nexus – ITD.

With ITD, the Cisco Nexus Series Switches provide traffic distribution to servers, without the need for external load-balancers as seen in figure 2.

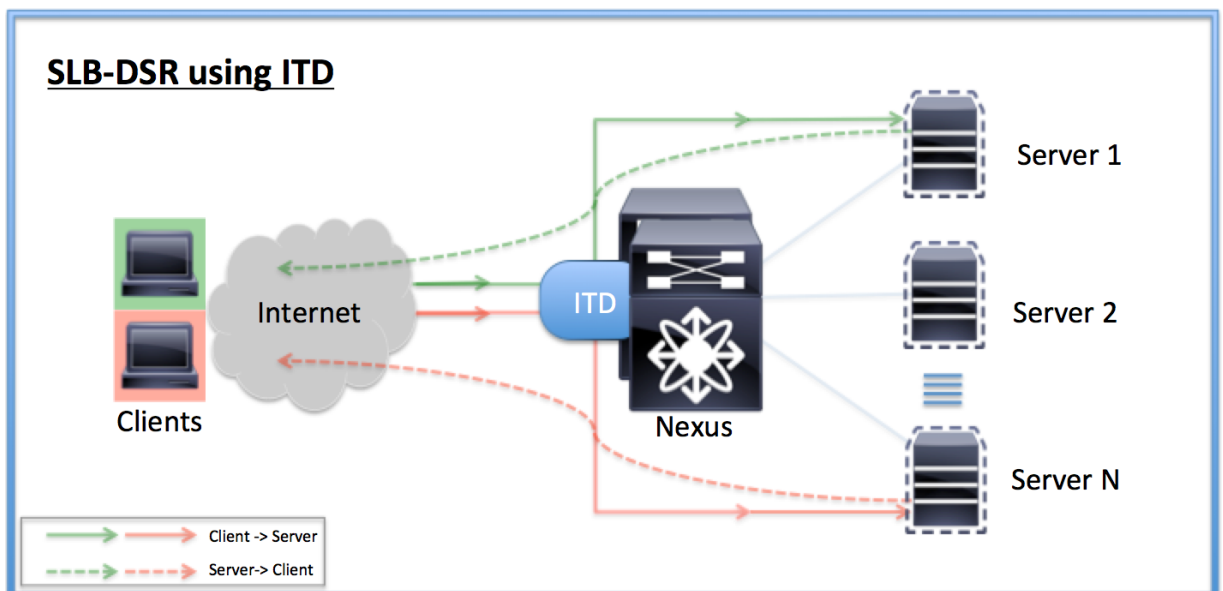


Figure 2 : DSR mode using Nexus-ITD

The Virtual IP, which was on the load-balancer earlier, is now configured within the ITD policy and can also be advertised for routing to the subnet. ITD also provides probes for node failure detection and sophisticated node-failure-actions, which can be customized as per network requirements.

Using ITD provides many benefits such as:

- Reduced Cost: No external SLBs or Application Delivery Controllers (ADCs) are required.
- Unhindered Performance: ITD leverages ASIC based Line-rate traffic distribution to provide Multi-Terabit capacity without increasing latency.
- Scalability: ITD can scale to support a large number of Servers.

4. Deploying ITD for SLB-DSR

The deployment described below involves the following devices:

- 1) Nexus 7700 - Two switches configured in vPC mode running 7.2(0)D1(1).
- 2) 2 Virtual Machines (servers) running Ubuntu Linux 14.04
- 3) 2 Virtual Machines (clients), also running Ubuntu Linux 14.04
- 4) Layer-2 switches (providing connectivity between VM's and Nexus switches)
- 5) The Server nodes host a simple HTTP service for testing/verification .

The same configurations can be applied for Server Load-balancing with DSR mode using ITD on Nexus 9000 series switches as well. The Nexus 5000/6000 switches currently do not support ITD probes.

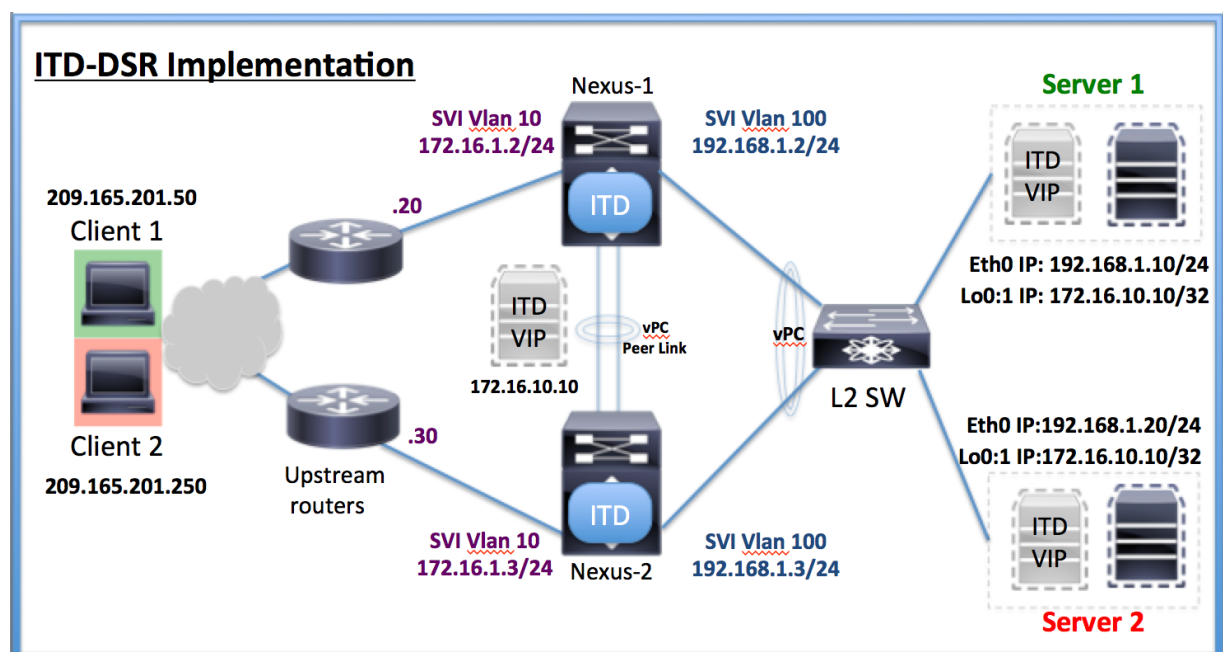


Figure 3: Direct Server return Implementation with ITD on Nexus switches.

Note: Here the clients cannot be on the same VLAN as the Servers. Also, the Virtual IP is on a different subnet, thus preventing ARP issues. This subnet can be advertised through routing protocols using ITD VIP "advertise" feature.

4.1 Server configuration

Direct Server Return requires the server to respond to clients directly by using the Virtual IP address as the source IP address. This necessitates that the server be configured with the VIP in addition to the regular IP address of the server. On typical Linux machines, this can be done using a Loopback address.

```
# sudo ifconfig lo:1 172.16.10.10 netmask 255.255.255.255 -arp up
```

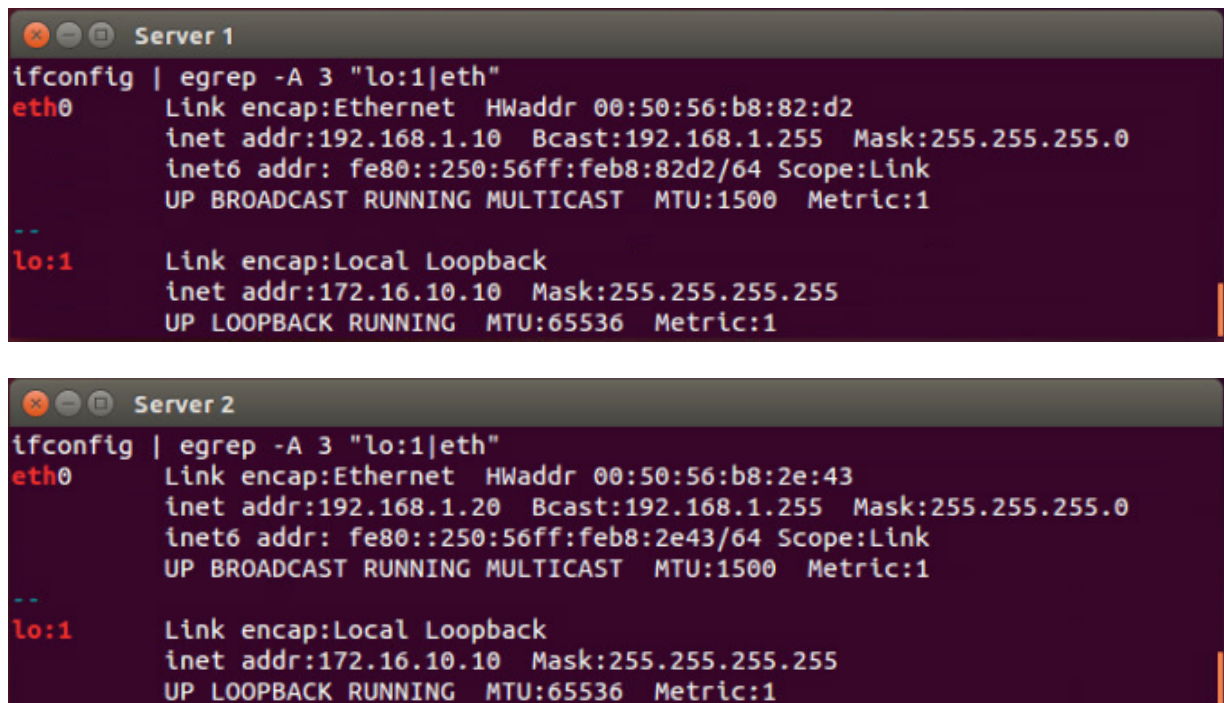
This command configures a new virtual loopback interface “lo:1” with the IP address 172.16.10.10. Since the same Virtual IP is configured on multiple devices, there is a possibility of duplicate IP detection if the Servers also respond to ARP for the VIP. Hence the configuration uses ‘-arp’ to disable ARP on this interface.

However, Linux also responds to ARP on wrong and unassociated interfaces by default. Hence the below configuration is required to ensure that the server does not respond to ARP for this IP on any other interface as well:

```
Append and save the following configuration to the file “/etc/sysctl.conf”  
net.ipv4.conf.all.arp_ignore=1  
net.ipv4.conf.all.arp_announce=2
```

Every server in the setup needs to be configured with the same Virtual IP address with the corresponding ARP changes using the steps above.

The VIP configuration on the server can be verified using ‘ifconfig’ from the terminal.



```
Server 1  
ifconfig | egrep -A 3 "lo:1|eth"  
eth0      Link encap:Ethernet  HWaddr 00:50:56:b8:82:d2  
          inet addr:192.168.1.10  Bcast:192.168.1.255  Mask:255.255.255.0  
          inet6 addr: fe80::250:56ff:feb8:82d2/64 Scope:Link  
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1  
--  
lo:1      Link encap:Local Loopback  
          inet addr:172.16.10.10  Mask:255.255.255.255  
          UP LOOPBACK RUNNING  MTU:65536  Metric:1  
--  
Server 2  
ifconfig | egrep -A 3 "lo:1|eth"  
eth0      Link encap:Ethernet  HWaddr 00:50:56:b8:2e:43  
          inet addr:192.168.1.20  Bcast:192.168.1.255  Mask:255.255.255.0  
          inet6 addr: fe80::250:56ff:feb8:2e43/64 Scope:Link  
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1  
--  
lo:1      Link encap:Local Loopback  
          inet addr:172.16.10.10  Mask:255.255.255.255  
          UP LOOPBACK RUNNING  MTU:65536  Metric:1
```

Figure 4: Server verification.

4.2 Nexus ITD configuration

This deployment example uses vPC's on the Nexus switches for connectivity towards the Server side. vPC is expected to be already configured on the Nexus switches and is not covered here. The configuration of switch 'Nexus-1' is shown below. The vPC peer 'Nexus-2' needs to be configured identically except for the SVI IP addresses.

The following features are prerequisites for ITD and need to be enabled in addition to other features already in use:

```
feature pbr           !Enables redirection functionality required for ITD
feature sla sender    !Enables probing mechanism used by ITD
feature sla responder !Enables probing mechanism used by ITD
feature interface-vlan !Enables users to create SVIs
feature itd           !Enables the ITD functionality itself
```

The ingress-interface SVI Vlan 10 and the 'server-group' SVI Vlan100 are created.

```
interface Vlan10
  description ITD-DSR Ingress interface
  no shutdown
  ip address 172.16.1.2/24
interface Vlan100
  description ITD-DSR Server VLAN
  no shutdown
  ip address 192.168.1.2/24
```

A device group 'server-group100' is created and the physical-interface IP addresses of the servers are added to it. ICMP probes are also configured for this device group with default values. The probe timers can be tweaked as required (see guidelines and limitations section for probes on Nexus 5000/Nexus 9000)

```
itd device-group server-group100
  probe icmp
  node ip 192.168.1.10
  node ip 192.168.1.20
```

An ITD Service "vip-dsr100" is configured to use the created device-group "server-group100". The Virtual IP 172.16.10.10 is configured for the ITD Service with advertisement enabled to advertise the VIP to upstream routing protocols. The load balance method of 'Source IP' is selected. Note that the command - "Failaction node reassign" needs to be configured explicitly to enable reassigning traffic buckets of a failed node(see references for details). Finally, statistics need to be enabled in order to view ITD traffic distribution statistics.

```
itd vip-dsr100
  device-group server-group100
  virtual ip 172.16.10.10 255.255.255.255 advertise enable
  ingress interface Vlan10
  failaction node reassign
  load-balance method src ip
  no shut
itd statistics DSRService
```

4.3 Verification

The configured ITD Service can be verified as below:

```
PSK_N7700_1-ITD-DSR(config-itd)# sh itd
Legend:
ST(Status): ST-Standby,LF-Link Failed,PF-Probe Failed,PD-Peer Down,IA-Inactive
Name          LB Scheme  Status  Buckets
-----
vip-dsr100    src-ip     ACTIVE  2
Exclude ACL
-----
Device Group          Probe  Port
-----
server-group100      ICMP
Pool                  Interface  Status  Track_id
-----
vip-dsr100_itd_pool  Vlan10    UP      3
Virtual IP            Netmask/Prefix  Protocol  Port
-----
172.16.10.10 / 255.255.255.255          IP        0
Node  IP          Cfg-S  WGT  Probe  Port  Probe-IP  STS  Trk#  Sla_id
-----
1     192.168.1.10  Active  1    ICMP
      Bucket List
      vip-dsr100_itd_vip_1_bucket_1
Node  IP          Cfg-S  WGT  Probe  Port  Probe-IP  STS  Trk#  Sla_id
-----
2     192.168.1.20  Active  1    ICMP
      Bucket List
      vip-dsr100_itd_vip_1_bucket_2
```

Figure 5: Nexus ITD verification

From the Client VMs, the Virtual IP’s reachability and the HTTP service can be verified as shown below. For differentiation, a folder named with the Server number and IP address has been created on each Server.

Since the ITD by default uses the last octet (or least significant bits (LSBs)) for bucketing, skewed IP addresses .50 and .250 have been used for the clients to show the load-distribution in action.

Client 1: Ping and HTTP test to the VIP.

```
Client 1
64 bytes from 172.16.10.10: icmp_seq=142 ttl=63 time=0.447 ms
64 bytes from 172.16.10.10: icmp_seq=143 ttl=63 time=0.504 ms
64 bytes from 172.16.10.10: icmp_seq=144 ttl=63 time=0.525 ms
64 bytes from 172.16.10.10: icmp_seq=145 ttl=63 time=0.409 ms
64 bytes from 172.16.10.10: icmp_seq=146 ttl=63 time=0.405 ms
64 bytes from 172.16.10.10: icmp_seq=147 ttl=63 time=0.535 ms
64 bytes from 172.16.10.10: icmp_seq=148 ttl=63 time=0.339 ms
64 bytes from 172.16.10.10: icmp_seq=149 ttl=63 time=0.480 ms
```



Client 2: Ping and HTTP test to the VIP.

```
Client 2
64 bytes from 172.16.10.10: icmp_seq=214 ttl=63 time=0.660 ms
64 bytes from 172.16.10.10: icmp_seq=215 ttl=63 time=0.390 ms
64 bytes from 172.16.10.10: icmp_seq=216 ttl=63 time=0.436 ms
64 bytes from 172.16.10.10: icmp_seq=217 ttl=63 time=0.599 ms
64 bytes from 172.16.10.10: icmp_seq=218 ttl=63 time=0.383 ms
64 bytes from 172.16.10.10: icmp_seq=219 ttl=63 time=0.365 ms
64 bytes from 172.16.10.10: icmp_seq=220 ttl=63 time=0.589 ms
64 bytes from 172.16.10.10: icmp_seq=221 ttl=63 time=0.392 ms
```

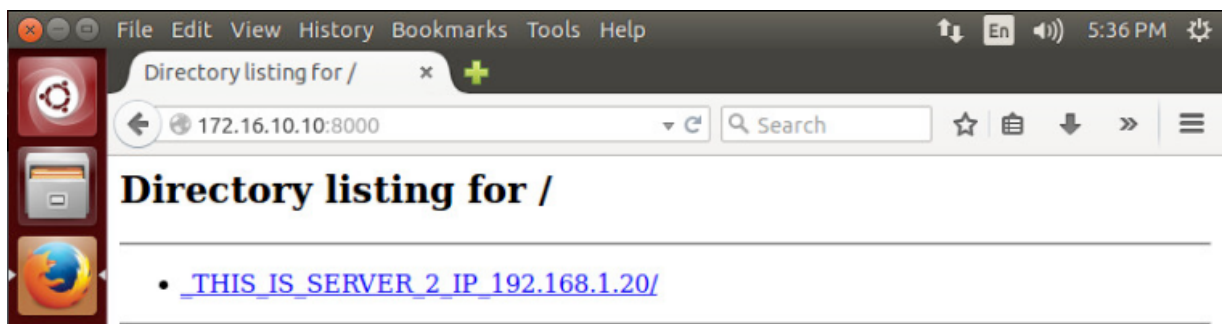


Figure 6: Client service verification.



Using the CLI on the Nexus7700, statistics can be verified for ITD redirected traffic:

```
PSK_N7700_1-ITD-DSR# show itd vip-dsr100 statistics
```

Service	Device Group	VIP/mask	#Packets	
vip-dsr100	server-group100	172.16.10.10 / 255.255.255.255	98 (100.00%)	
Traffic Bucket	Assigned to	Mode	Original Node	#Packets
vip-dsr100_itd_vip_1_bucket_1	192.168.1.10	Redirect	192.168.1.10	49 (50.00%)
Traffic Bucket	Assigned to	Mode	Original Node	#Packets
vip-dsr100_itd_vip_1_bucket_2	192.168.1.20	Redirect	192.168.1.20	49 (50.00%)

Figure 7: ITD statistics

5. Guidelines and Limitations

- Configuration of Loopback on the Server nodes is mandatory for DSR mode to work even if the servers run other operating systems(Windows for eg.).
- Nexus vPC peer devices need to have identical configuration of the ITD service to operate correctly. This includes the number, order, configuration of the nodes as well as other parameters like load-balance method, probe specifications etc.
- If failaction reassign is not configured with no standby, the traffic will get routed normally without ITD redirection upon failure of a Node.
- If the Server nodes support multiple NICs with Port-channel, the Nexus switches can connect directly to the servers via vPC.
- Using weights for the nodes, traffic can be distributed unequally (when using servers of different capacities).
- Starting release 7.2(0)D1(1), probes can be configured per-node if required.
- For ITD-ICMP Probes on the Nexus 9000, the features “SLA responder/sender” are currently not required as prerequisites.
- Nexus 5000/6000 series switches currently do not support ITD Probes.

6. References and Further reading.

Creating Virtual interfaces in Linux:

<http://linuxconfig.org/configuring-virtual-network-interfaces-in-linux>

Linux ARP announce/ARP ignore:

http://kb.linuxvirtualserver.org/wiki/Using_arp_announce/arp_ignore_to_disable_ARP

Linux ARP Flux considerations: <http://linux-ip.net/html/ether-arp.html> - ether-arp-flux

Nexus 7000 ITD configuration Guide: [Nexus 7000 / 7.x / ITD Config Guide](#)

Nexus 9000 ITD configuration guide: [Nexus 9000 / 7.x / ITD Config Guide](#)

Nexus 5500 ITD config guide: [Nexus 5500 ITD Config Guide](#)

Nexus 5600 ITD config guide: [Nexus 5600 ITD Config Guide](#)