



Redfish PCIe Switch Model & Mockup



Information & Disclaimer

- The information in this presentation represents a snapshot of work in progress within the DMTF.
- This information is subject to change. The Standard Specifications remain the normative reference for all information.
- DMTF Specifications in progress are considered DMTF confidential and are being shown under existing work register agreements.
- Documents cannot be shared outside of the DMTF work group. The documents may be posted on the WG site, but not otherwise distributed without DMTF Board approval.
- Non-member comments, feedback and submissions are subject to the DMTF Patent Policy statement



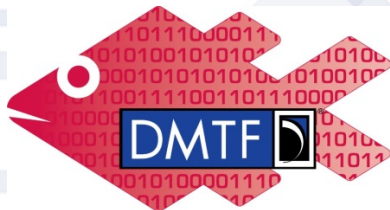
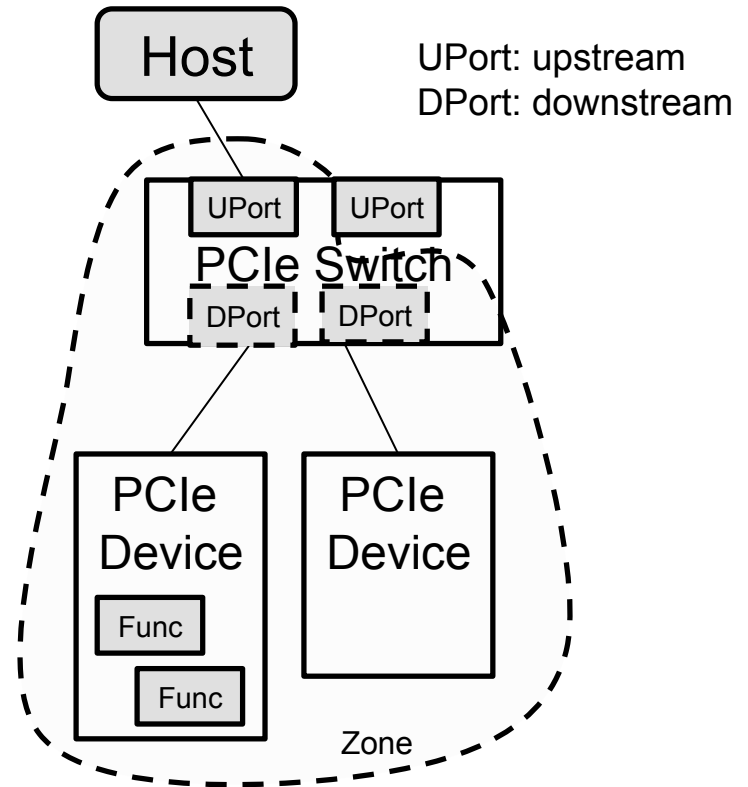
Requirements

PCIe Switch

- Host connected via upstream ports
- PCIe devices connected to downstream ports and contain functions

Requirements

- Discover switches, devices, functions
- Assign devices and functions to hosts (zone)

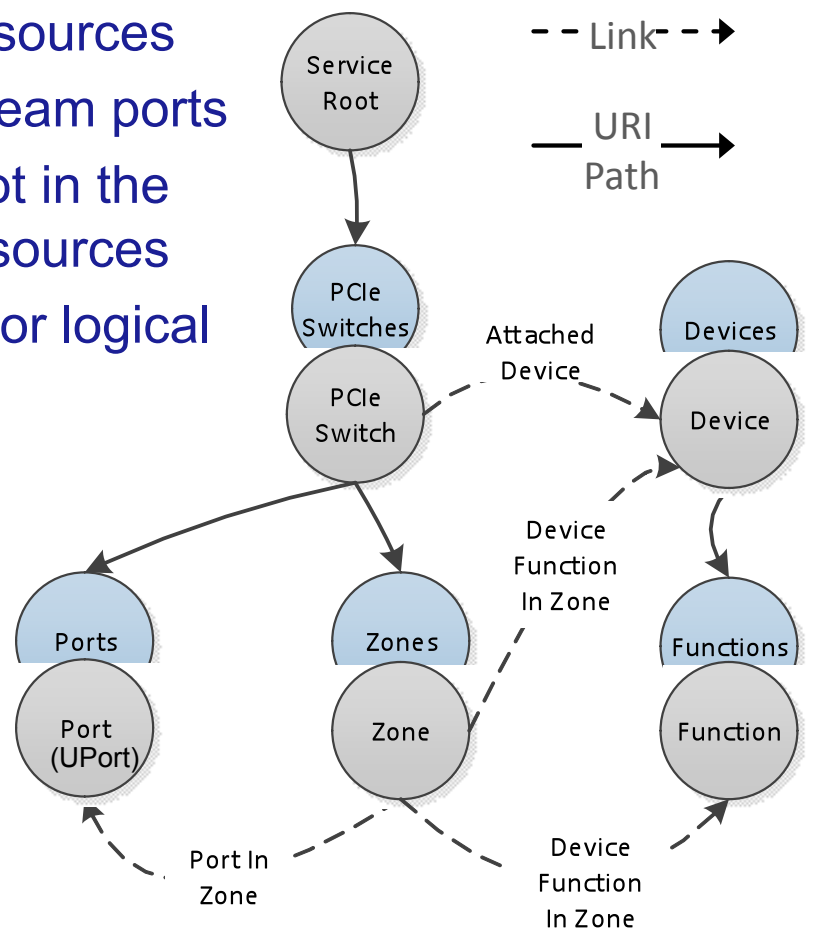
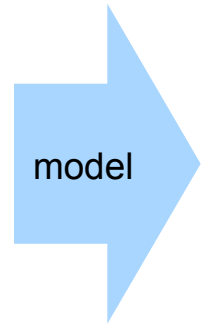
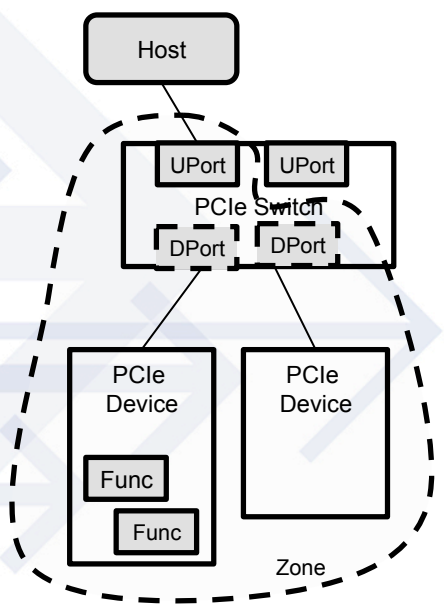


Redfish



PCIe Switch Model

- PCIeSwitch resource contains port resources
- A Port resource represents only upstream ports
- Device resource placed at root, but not in the Service Root; and contain function resources
- A Function resource may be physical or logical

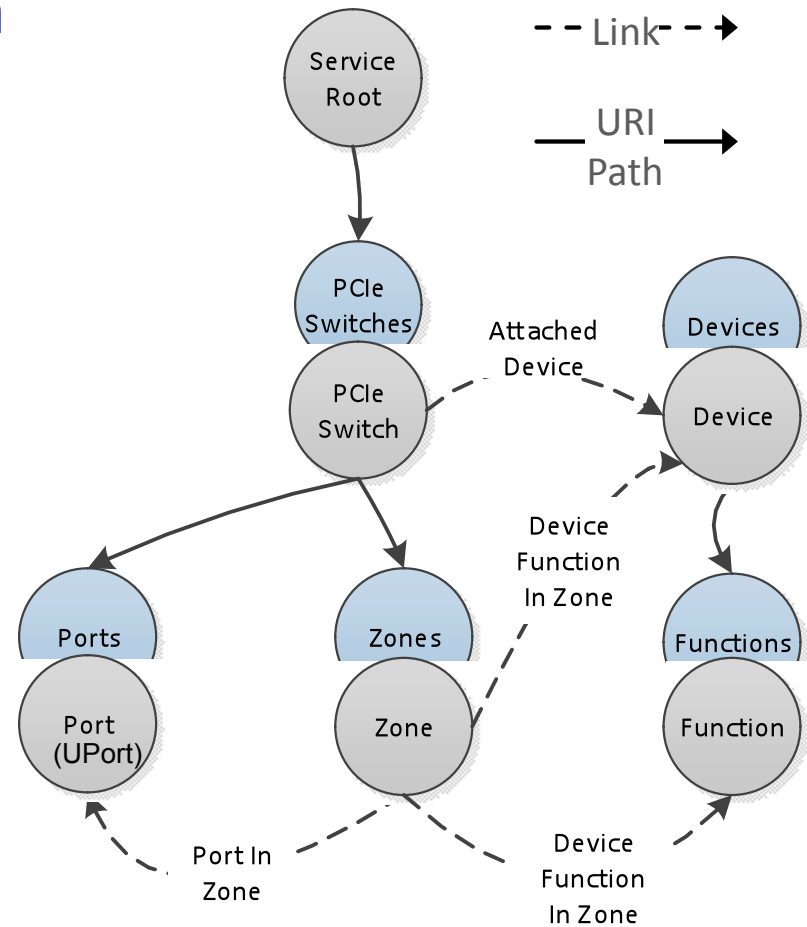




Switch, Port, Device & Function Resources

Discovery process

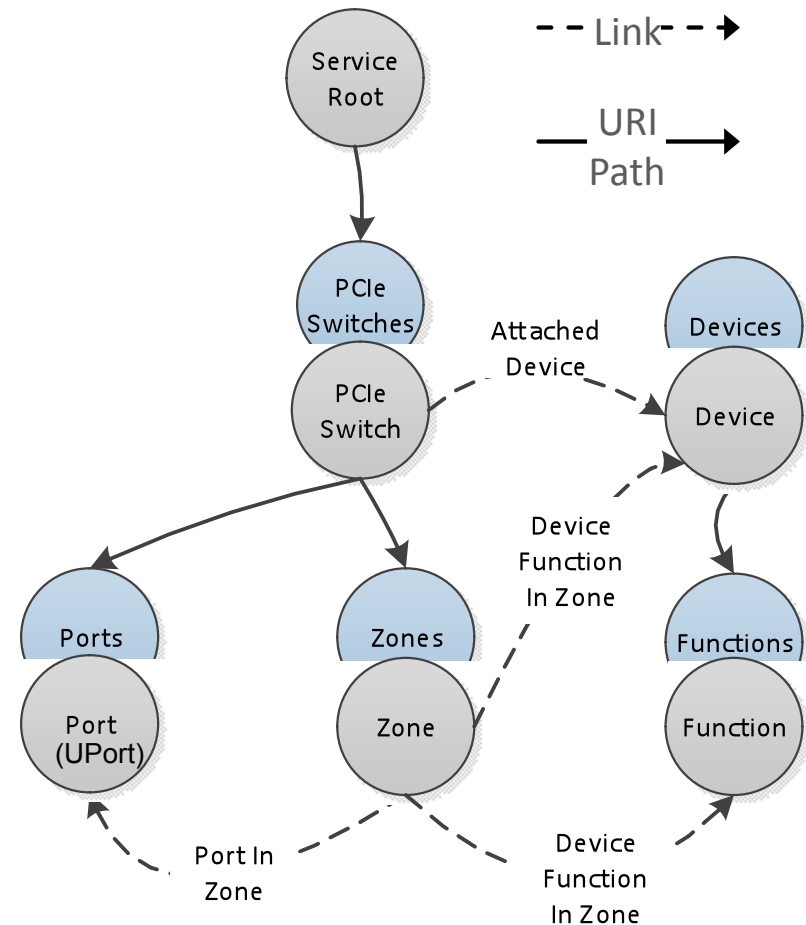
- Discover PCIe Switches from Service Root
- Discover Port resources from PCIe Switch resource
- Discover Device resources by inspecting AttachedDevice structure
- Discover the Functions resources from the Device resource





Zone Resource

- Zone resource specifies the devices and functions assigned to one or more upstream-ports (host ports).
- A Zone resource is created by POST'ing to the Zones collection resource.
- Zone can include
 - One or more upstream port (to support trunking and multi-node)
 - One or more devices or functions





Backup

Examples

- Simple PCIe Switch
- Ganged PCIe Switch
- Hierarchical Switches
- Redundant PCIe Switches



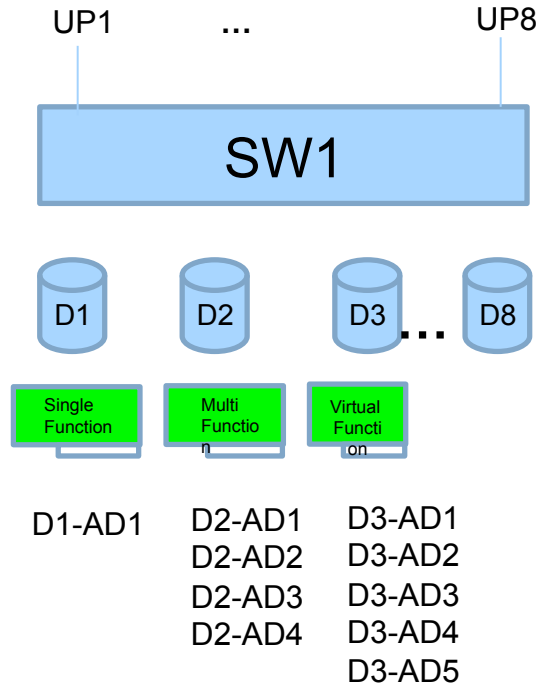
PCIe Switch - Simple

Simple switch with the following attached:

- 1 PCIe device with a single function
- 1 PCIe device with 4 functions
- 1 PCIe SR-IOV device with 5 virtual functions

Mockup

- PCleSwitches/1/Zones/{empty}
- PCleSwitches/1/Ports/UP1, ..., UP8
- Devices/D1/Functions/D1-AD1
- Devices/D2/Functions/D2-AD1, ..., D2-AD4
- Devices/D3/Functions/D3-AD1, ..., D3-AD5

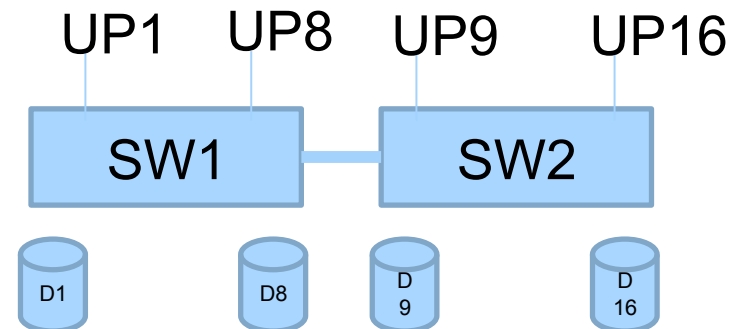




Modeling Ganged PCIe Switch

- Ganged switch is modeled as one giant switch.
- Any device in either switch can be assigned to any device in either switch.
- When the sidelink bandwidth is exhausted, the assignment request may fail.
- Create a concept of preferred upstream ports for each Assignable Device (**not modeled**).

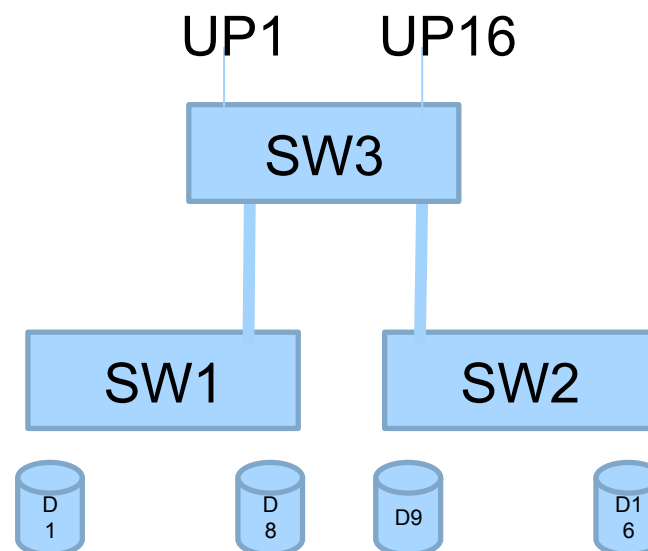
- PCIeSwitches/1/Zones/{empty}
- PCIeSwitches/1/Ports/UP1..UP16
- Devices/D1,...,D8 (prefer UP1-UP8)
- Devices/D9,...,D16 (prefer UP9-UP16)





Modeling PCIe Switch - Hierarchy

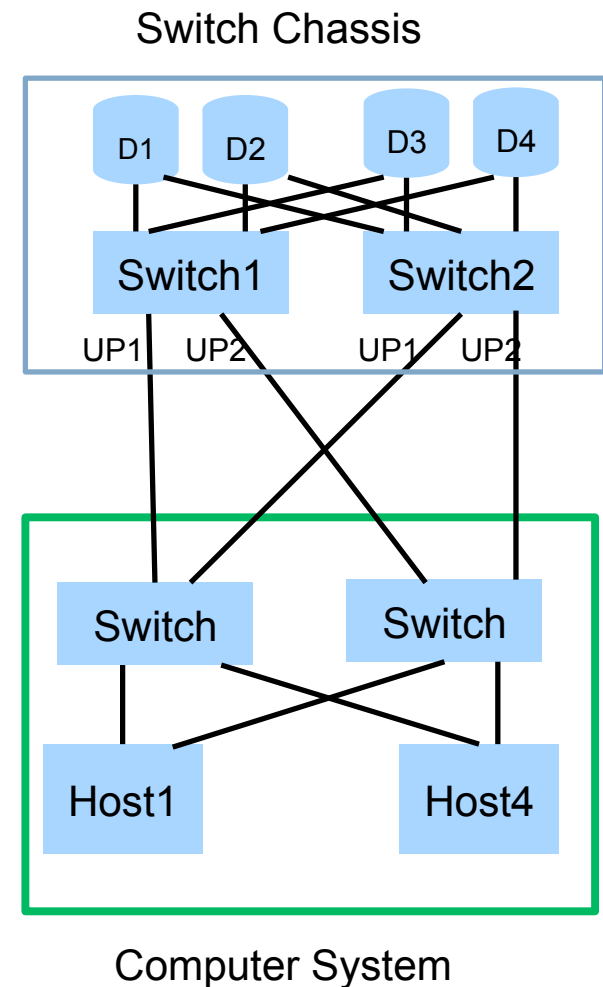
- Hierarchical switch is modeled as one giant switch.
 - Lower level switches are for expansion of SW3.
 - Any device in either lower switch can be assigned to any upstream port.
 - No proximity issues related to switch topology.
- PCIeSwitches/1/Zones/{empty}
 - PCIeSwitches/1/Ports/UP1-UP16
 - Devices/D1,...,D16 (no preferred UP)





Modeling Redundant Switch Config

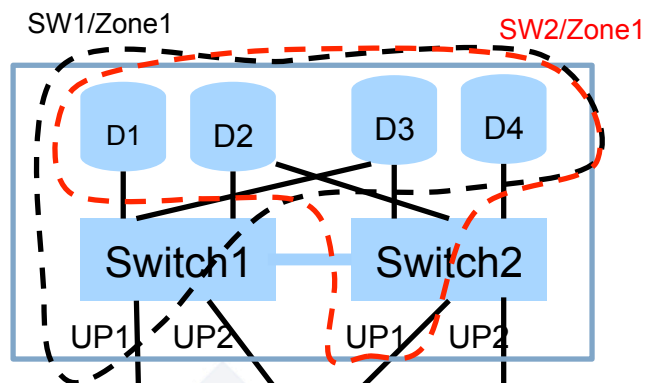
- Switch1 and Switch2 are independent of each other (i.e., they can fail independently).
- From manageability standpoint, they must have same manager so device sharing info is visible across the two switches.
- Use redundancy object from RF (**not modeled**).



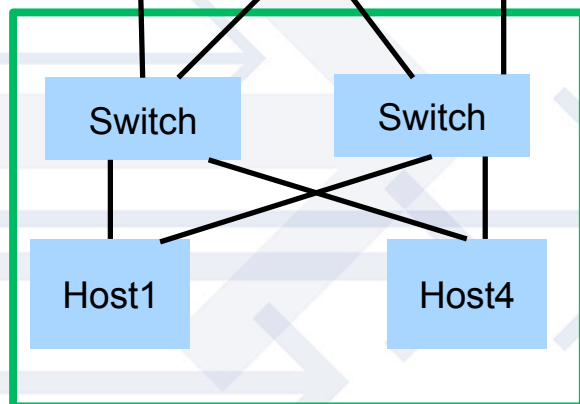


Modeling Redundant Switch Config w/Zones

Switch Chassis



- PCIeSwitches/1/Ports/UP1, UP2
- PCIeSwitches/2/Ports/UP1, UP2
- Devices/D1,...,D4 (no preferred UP)
- PCIeSwitches/1/Zones/1
 - UP1, D1, D2, D3, D4
- PCIeSwitches/2/Zones/1
 - UP1, D1, D2, D3, D4
- Redundancy Set ["PCIeSwitches/1/Zones/1", "PCIeSwitches/2/Zones/1"]



Computer System

With config per the diagram

- If Switch1 fails, we will no longer have access to AD1.
- If Switch2 fails, we will no longer have access to AD4.



No downstream port modeled

- PCIe devices can be assignable or non-assignable.
- DevOps cares about the binding of the upstream ports and the assignable devices/function.
 - Downstream port is transparent.
 - OS representation accuracy is not needed.
 - Avoids showing the downstream port for a non-assignable device.
- For device sharing, the downstream ports are usually synthesized.

