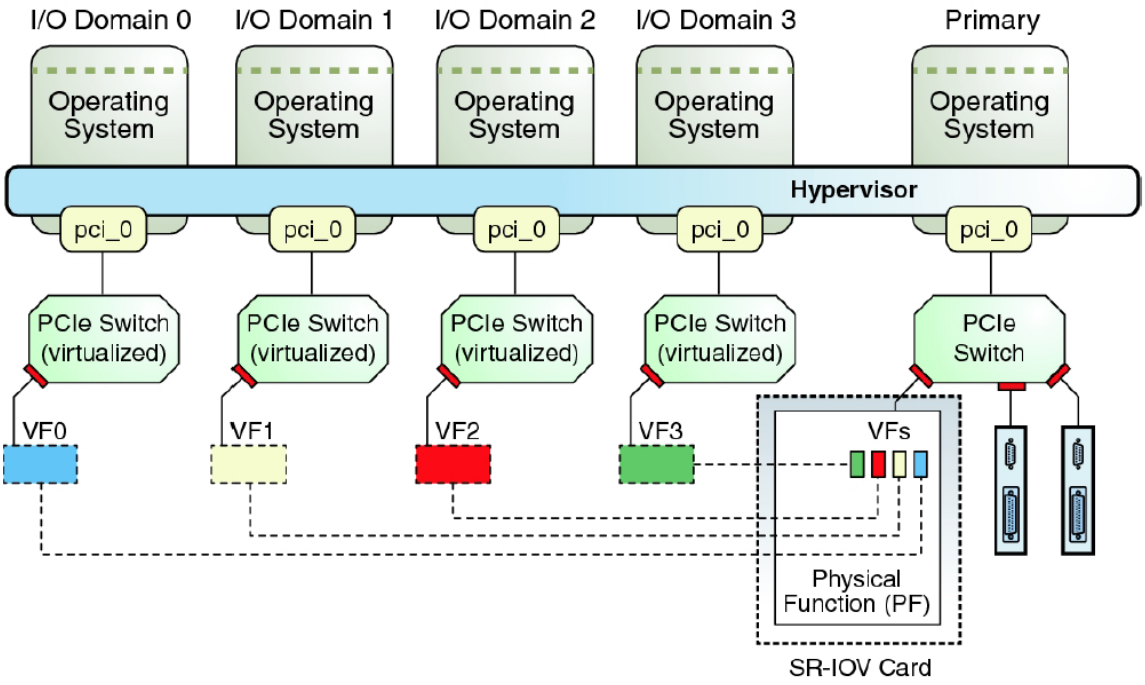
SR-IOV Teknolojisine Genel Bakış

SR-IOV, "Single root I/O Virtualization" nın baş harflerinden oluşur. PCIe ("Peripheral Component Interconnect Express") SR-IOV implementasyonu, http://www.pcisig.com adresinde tanımlanan standartların 1.1 sürümüne dayalı yapılandırılmaktadır. SR-IOV standartları, PCIe cihazların/aygıtların ("devices"), sanal makineler arasında verimli bir şekilde paylaşımını sağlar. Bu standartlar donanım içerisinde yerine getirilmektedir. SR-IOV'nin, fiziksel kartın sanallaştırılmadan sağladığı performans ile karşılaştırılabilir bir performans sağlaması bu teknolojiyi buluta "Cloud" doğru evrilen IT dünyasında oldukça çekici kılmaktadır.

Fiziksel fonksiyon ("physical function") olarak tanımlanan tek bir I/O kaynağı pek çok sanal makine tarafından paylaşılabilir. Paylaşılan cihaz/aygıt, sanal makineye adanmış bir kaynak sağlar. Bu şekilde yapılan paylaşımda, her bir sanal makine tek bir kaynağa erişim sağlar. Bu nedenle, SR-IOV fonksiyonu kullanılabilir hale getirilen bir PCIe aygıtı, örneğin bir Ethernet portu, her birinin kendine ait PCIe konfigürasyonu olan ayrı ve çoklu yapılar olarak görünürler.

Aşağıda olan şekilde fiziksel ve sanal fonksiyonların farkı net olarak görülmektedir. Şekilde, fiziksel PCIe switch üzerindeki SR-IOV özelliği kullanılabilir hale getirilmiş kartın (fiziksel fonksiyon olarak tanımlanır), sanallaştırılmış PCIe swtich'ler üzerinde sanal fonksiyonlar ("virtual function") olarak paylaşımı görülmektedir. Fiziksel kartın, "Primary domain" üzerinde olduğu ve paylaştırılan sanal fonksiyonların, I/O domain'ler üzerine atanmış olduğu görülmektedir.



SR-IOV'nin sunduğu fonksiyon türlerini aşağıdaki şekilde daha detaylı açıklayabiliriz.

Fiziksel fonksiyon ("Physical function"): SR-IOV spesifikasyonları tarafından tanımlanan SR-IOV yeteneklerini destekler. Fiziksel fonksiyon, SR-IOV'nin yeteneklerini kapsadığı gibi sunulan bu fonksiyonelitiyi de yönetir. Herhangi bir PCIe aygıtı nasıl sistem tarafından algılanabiliyorsa, yönetilebiliyorsa, üzerinde manipülasyon yapılabiliyorsa aynı şekilde fiziksel fonksiyon da bu özellikleri sağlar. Fiziksel fonksiyon, PCIe cihazının konfigürasyonunu yapmak ve yönetmek için kullanılır.

Sanal fonksiyon ("Virtual function"): Fiziksel fonksiyon ile ilişkilendirilmiş bir fonksiyon olarak tanımlanabilir. Fiziksel fonksiyonu aksine, sadece kendi davranışını kontrol edebilir.

Her bir SR-IOV cihazı/aygıtı, bir fiziksel fonksiyona sahiptir ve her fiziksel fonksiyon kendisiyle ilişkilendirilen 256 tane sanal fonksiyona sahiptir. Açıkcası bu sayı, SR-IOV cihazının ne olduğuna bağlıdır. Sanal fonksiyonlar, fiziksel fonksiyonlar tarafından oluşturulur.

Fiziksel fonksiyon içerisinde SR-IOV kullanılabilir şekle getirildikten sonra ("SR-IOV enable"), her bir sanal fonksiyonun PCI konfigürasyon tanım alanına "bus", aygıt ve fiziksel fonksiyonun fonksiyon numarası ulaşabilir. Her bir sanal fonksiyon, PCI bellek alanına sahiptir. Bu alan "register set" lerin haritalanması için kullanılır. Sanal fonksiyonun cihaz sürücüsü ("device driver"), "register set" üzerinde çalışır ve kendi fonksiyonelitesini bu "register set" üzerinde kullanılabilir hale getirir. Bu noktadan sonra sanal fonksiyon, gerçek bir PCI cihazı gibi görünür. Bu işlemin ardından, sanal fonksiyonu I/O domain üzerine atayabilirsiniz. Bu özellik sayesinde, sanal fonksiyonu, fiziksel cihazı paylaşmak için kullanabilmekteyiz. Ek olarak bu sayede herhangi bir CPU ve hipervizor yükü getirmeden I/O işlemleri yapabilmekteyiz.

SR-IOV sayesinde, daha yüksek performans alınabilmekte ve gecikmeler ("latency") azalmaktadır çünkü sanal sistem ortamından donanıma doğrudan ulaşılabilmektedir. Güç kazancı sağladığı, adaptör sayısında azalmaya neden olduğu, daha az kablolama ihtiyacı doğurduğu, daha az switch portu gerektirdiği için maliyet tarafında da avantaj sağlamaktadır.

Oracle VM for SPARC'ın SR-IOV implementasyonu, statik ve dinamik konfigürasyon yöntemlerini içerir. Oracle VM for SPARC SR-IOV özelliğiyle aşağıda özetlenen operasyonları yapabiliriz:

-Spesifik bir fiziksel fonksiyon üzerinde sanal fonksiyon oluşturulabilir.

-Fiziksel fonksiyon üzerinde, spesifik bir sanal fonksiyon silinebilir.

-Bir domain'e sanal fonksiyon atanabilir.

-Bir domain'den sanal fonksiyon silinebilir.

SR-IOV fiziksel fonksiyon cihazları üzerinde sanal fonksiyon oluşturabilmek veya fiziksel fonksiyon cihazları üzerinden sanal fonksiyonu silebilmek için, PCIe bus üzerinde I/O sanallaştırmasını kullanılabilir şekle (enable) getirmeniz gerekmektedir. *"ldm set-io"* veya *"ldm add-io"* komutlarını, *"iov"* özelliğini konfigüre etmek için kullanabilirsiniz.

SPARC M7 serisi sunucularda, SPARC T7 serisi sunucularda ve Fujitsu M10 sunucularda, PCIe bus üzerinde I/O sanallaştırma "default" olarak açıktır.

Dinamik ve statik PCIe SR-IOV özellikleri, SPARC T4, SPARC T5, SPARC T7, SPARC M5, SPARC M6, SPARC M7 serisi sunucularda desteklenmektedir. Fujitsu M10 sadece Ethernet aygıtları için dinamik özellikleri içerir. Sistem üzerinde ki diğer aygıtlar, statik metodun kullanılmasını gerektirir. SPARC T3 platformu sadece statik PCIe SR-IOV özelliğini destekler.

SR-IOV için donanım gereksinimleri:

-Ethernet SR-IOV: SR-IOV özelliğini, On-board üzerindeki PCIe SR-IOV aygıtlarında ve PCIe SR-IOV plug-in kartlarında kullanabilirsiniz. Yukarda ismi ifade edilen platformlar içim on-board üzerindeki tüm SR-IOV aygıtları desteklenir.

-InfiniBand SR-IOV: InfiniBand aygıtlar, SPARC T4, SPARC T5, SPARC T7, SPARC M5, SPARC M6, SPARC M7 ve Fujitsu M10 sunucularda desteklenir.

Fibre Channel SR-IOV: Fibre Channel aygıtlar, SPARC T4, SPARC T5, SPARC T7, SPARC M5, SPARC M6, SPARC M7 ve Fujitsu M10 sunucularda desteklenir.

SR-IOV'nin kullanılması için sistem mikrokodunun da belli bir sürüm üzerinde olması gerekmektedir. Ethernet SR-IOV özelliğini kullanmak için, tüm "domain" lerde en azından Oracle Solaris 11.1 SRU 10 işletim sistemi olmalıdır. Infiniband SR-IOV için primary domain, I/O domain ve non-primary root domain üzerinde en azından Solaris 11.1 SRU 10 işletim sistemi olmalıdır. Sanal fonksiyonu konfigüre etmeyi planladığınız InfiniBand SR-IOV'ye sahip tüm root domain'lerde */etc/system* dosyası *"set ldc:ldc_maptable_entries = 0x20000"* satırını içermelidir. Sanal fonksiyonu eklemek istediğiniz her I/O domain üzerinde */etc/system* dosyası *"set rdsv3:rdsv3_fmr_pool_size = 16384"* satırını içermelidir.

Mevcut teknolojiyle en azından şu an için aşağıda olan limitasyonlara sahibiz:

-Kendileriyle ilişkili herhangi bir root domain çalışmıyorsa, I/O domain başlatılamaz.

-Üzerine bir veya birden fazla sanal fonksiyon atananan domain'ler üzerinde "migration" özelliği kullanılamaz.

-Eger SR-IOV kartı, domain'e Direct I/O (DIO) özelliği ile atanmış ise, kart üzerindeki SR-IOV özelliği kullanılabilir olmayacaktır.

-SPARC T7 ve SPARC M7 sunucularında, PCIe bus üzerinden PCIe endpoint aygıtlarını ve SR-IOV sanal fonksiyonlarını maksimum 31 domain'e atayabilirsiniz.

-PCIe bus'ın sahibi "root" domain'dir. Dolayısıyla, bus'ın başlatılmasından, yönetilmesinden kendisi sorumludur. Bu nedenle root domain'in, SR-IOV özelliğini destekleyen OS sürümünde olması gerektiği gibi aynı zamanda aktifte olmalıdır. Root domain üzerinde halt, shutdown, reboot işlemleri, domain'in bus'a erişimini etkileyeceği için, elbette PCIe bus üzerindeki aygıtlar da etkilenecektir. Sonuç olarak, I/O domain çalışırken eger root domain reboot olursa, PCIe SR-IOV sanal fonksiyonların kullanıldığı I/O domain'in davranışı kararsız olacaktır. Örneğin I/O domain bu durumda panik alabilir veya sonrasında reboot edebilir kendisini. Root domain'in reboot'u sonrası, her bir domain'i manual olarak durdurmanız ve yeniden başlatmanız gerekebilir. Önemli not, eger *"I/O Domain Resiliency"* özelliğini kullanırsanız, PCIe bus'ın sahibi olan root domain çalışamaz durumda olsa bile, I/O domain çalışmaya devam eder.

Static SR-IOV

Statik SR-IOV özelliği, root domain'in geciktirilmiş re-konfigürasyonda ("delayed reconfiguration") olmasını gerektirir veya I/O domain'in durdurulmasını gerekir. Bu yöntemi, eger ilgili OS/platform dinamik SR-IOV'yi desteklemiyorsa kullanmanız gerekir.

Dinamik SR-IOV

Dinamik SR-IOV sayesinde, root domain'de herhangi bir geciktirilmiş yeniden konfigürasyona gerek kalmadan ("delayed reconfiguration") sanal fonksiyonlar oluşturabilir ve silinebilir. Aynı şekilde, I/O domain'i durdurmadan, sanal fonksiyon I/O domain'e eklenebilir veya I/O domain'den çıkarılabilir. *ldm*, "Logical Domain" ajanı ve Oracle I/O sanallaştırma çatısı ("virtualization framework") ile bağlantı kurar ve bu değişikliklerin dinamik olarak yapılmasını sağlar.

SR-IOV sanal fonksiyonların konfigürasyonunu yapmadan önce, PCIe bus üzerinde I/O sanallaştırma özelliğini kullanılabilir şekle getirmemiz gerektiğini ifade etmiştik. SPARC M7, SPARC T7 ve Fujitsu M10 sunucularda, PCIe bus üzerinde I/O sanallaştırması kendiliğinde "default" açıktır. Ama eger I/O sanallaştırma özelliğinin kullanılabilir şekle getirilmesi gereken bir platform üzerinde çalışıyorsanız, "root domain", geciktirilmiş konfigürasyon ("delayed reconfiguration") durumundayken, I/O sanallaştırma, PCIe bus üzerinde kullanılabilir hale getirilmelidir.

PCIe SR-IOV Sanal Fonksiyon Kullanımının Planlanması

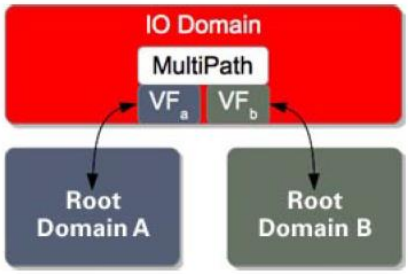
Konfigürasyonda, sanal fonksiyonların nasıl kullanılacağı öncesinde detaylı bir şekilde planlanmalıdır, tasarlanmalıdır. SR-IOV cihazlarının hangi sanal fonksiyonlarının mevcut ve gelecekteki ihtiyaçlarınızı karşılayıp karşılayamayacağını iyi analiz etmeniz gerekmektedir. Dinamik SR-IOV kullanılabilir olsa bile, önerimiz,

mümkün ise tüm sanal fonksiyonların bir kerede oluşturulmasıdır. Mutlaka SR-IOV'nin limitasyonlarını çok iyi anlayarak yapınızı oluşturmanız gerekmektedir.

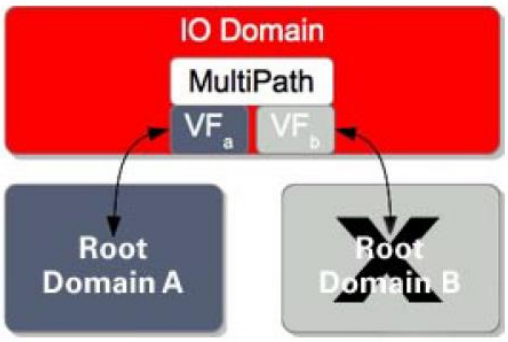
I/O Domain Resiliency

Bu özellik sayesinde, I/O domain, kendisiyle ilişkili root domain çalışamaz durumda olsa bile, çalışmaya devam edecektir. Root domain bir şekilde kesintiye uğradıysa, I/O domain, etkilenen cihazları alternate I/O path'e aktaracak ve bu sayede servisler/uygulamalar çalışmaya devam edecektir.

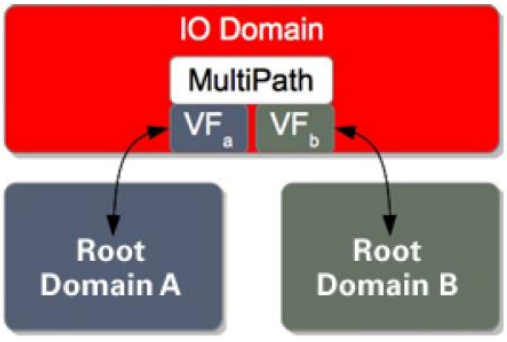
Aşağıda olan şekilde bu akışı açıklayıcı şekilde resmetmektedir.



Şekilde her bir root domain, I/O domain'e bir sanal fonksiyon sağlar. I/O domain'de multipathing teknolojilerini (IPMP-Network için, MpxIO-Fibre Channel Cihazlar) kullanarak root domain'lerin yedekli çalışmasına imkan sağlar.



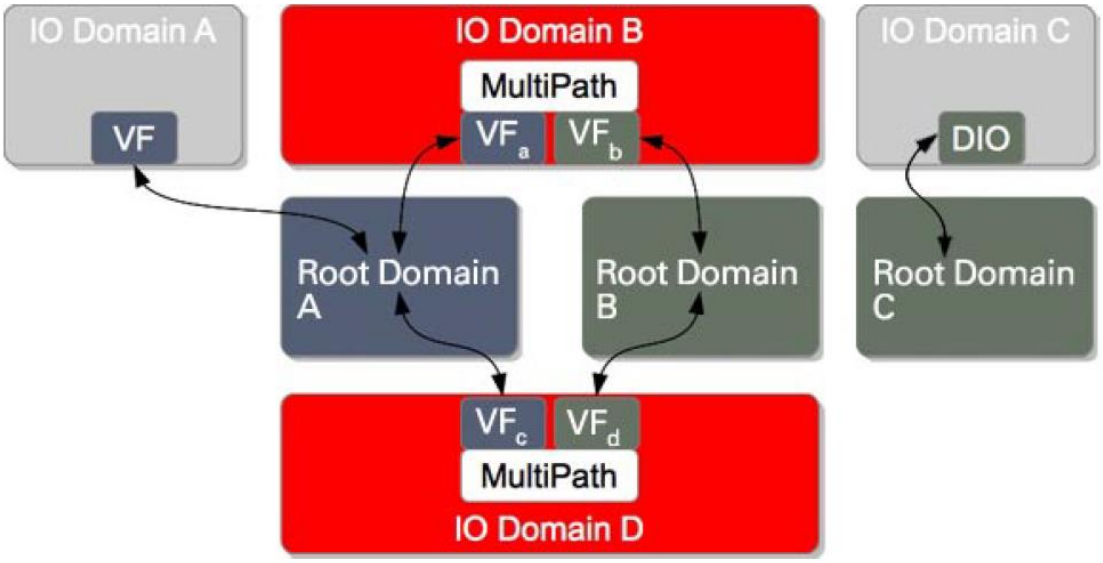
Root domain B'nin panik veya reboot aldığını varsayalım, I/O domain'de sanal fonksiyon B geçici olarak durdurulur, multipathing yazılımı, tüm I/O yu root domain A üzerine yönlendirir.



Root domain B, tekrar servis verebilir duruma geldiğinde, sanal fonksiyon B operasyonu yeniden başlatır. Multipath grup, tekrar tam yedeklilik için hazırda bekler.

Bu konfigürasyonda sanal fonksiyon, sanal network aygıtı veya sanal depolama aygıtı olabilir. Yani, I/O domain sanal fonksiyonların veya sanal aygıtların herhangi bir kombinasyonunu içerecek şekilde yapılandırılabilir.

Bu yapıyı biraz daha anlaşılır yapmak için aşağıdaki şekil üzerinden açıklama yapmak istiyorum.



Şekilde, I/O domain A ve I/O domain C, “resilient” özelliğine sahip değildir çünkü üzerlerinde bir multipathing yapılandırması yoktur. I/O doman A, kendisine atanan bir sanal fonksiyona sahiptir. Domain C ise I/O cihazına/aygıtına doğrudan erişime sahiptir.

I/O domain B ve I/O domain D, “resilient” özelliğine sahiptir. I/O domain A, B ve D’nin, root domain A’ya bağlı olduğu görüyoruz. I/O domain B ve D, root domain B’ye bağlıdır. I/O domain C, root domain C’ye bağlıdır. Örneğin root domain A’da bir kesinti olursa, IO domain A’da da kesinti olacaktır. Buna rağmen, I/O domain B ve D, alternate path’lere aktarım yaparak çalışmaya devam edeceklerdir. Eğer root domain C bir kesinti alırsa, I/O domain C’de hata alarak *failure-policy* özelliğinde tanımlanan değere göre davranış gösterecektir.

SR-IOV özellliğini özetleyecek olursak, iyi bir tasarım planlaması ile tam bir yedekli yapı sağlayacağı gibi daha az kablo, daha az güç, daha az switch port gerektirdiği için de maliyet avantajı sağlayacak ve yönetim kolaylığı getirecektir.

Asiye Yiğit - 21 Temmuz 2016

Kaynakça:

Oracle®VMServer for SPARC 3.3 Administration Guide