



## PRINCIPAIS MEIOS DE ARMAZENAMENTO DE DADOS UTILIZADOS EM DATACENTERS MODERNOS<sup>1</sup>

Gilnei Ferraz

**Resumo:** A necessidade por armazenamento de dados cresceu rapidamente nos últimos anos, se tornando um fator crítico nos datacenters modernos. Este artigo buscou conhecer as principais tecnologias e arquiteturas empregadas no armazenamento de dados nesses ambientes de missão crítica. Dada a imensa variedade de necessidades de negócios, é natural que os datacenters modernos empreguem diferentes arquiteturas e meios de armazenamento. Entre as arquiteturas mais utilizadas encontram-se, a já consolidada *Storage Area Network* (SAN), e mais recentemente, a hiperconvergência. Essas arquiteturas podem empregar arranjos de discos magnéticos, discos de estado sólido, ou mesmo arranjos híbridos, com os dois tipos de discos formando assim uma hierarquia de memória, onde os dados mais acessados são armazenados em discos mais rápidos e dados menos requisitados ficam em discos mais lentos.

**Palavras-chave:** Datacenter moderno. Armazenamento de dados. Storage.

### 1 INTRODUÇÃO

É notório que a demanda por armazenamento de dados aumentou nos últimos anos, com a crescente utilização de computação em nuvem a demanda tende a continuar aumentando consideravelmente. Nos datacenters são utilizados diversos meios de armazenamento, normalmente dimensionados considerando-se a quantidade de acessos e criticidade de tais dados para o negócio.

Os gestores de datacenters devem considerar a disponibilidade, segurança e escalabilidade dos recursos de armazenamento de dados. A depender da informação armazenada, existirão diferentes meios físicos de armazenamento recomendados. Os meios de armazenamento mais rápidos também são mais caros, então deve-se balancear a utilização de meios de armazenamento buscando uma melhor relação custo-benefício.

---

<sup>1</sup> Artigo apresentado como Trabalho de Conclusão do Curso de Especialização em Datacenter: Projeto, Operação e Serviços, da Universidade do Sul de Santa Catarina, como requisito parcial para a obtenção do título de Especialista em Datacenter: Projeto, Operação e Serviços.



A transmissão de dados em um datacenter exige muita atenção, pois os volumes de dados comumente são gigantescos, onerando muito a rede de dados. A forma mais utilizada de tratar tal volume é a segregação em uma rede dedicada para armazenamento, denominada *Storage Area Network* (SAN). A SAN apresenta método de acesso em baixo nível, ou seja, nível de bloco de dados. Por se tratar de uma rede segregada, dependendo do meio físico de transmissão e das tecnologias utilizadas o custo pode ser elevado. Uma solução normalmente empregada é a utilização do protocolo *Internet Small Computer System Interface* (iSCSI) ou o *Fibre Channel over Ethernet* (FCoE), que permite que ativos utilizados em redes locais comuns (LANs), como switches, possam ser utilizados na rede segregada para armazenamento.

O meio de armazenamento mais comum atualmente ainda é o disco rígido (*Hard Disk Drive* ou HDD). Padrões como *Serial ATA* (SATA) e principalmente o *Serial Attached SCSI* (SAS) dominam o ambiente do datacenter. Aos poucos, os Discos de Estado Sólido (*Solid State Disk* ou SSD) vão sendo utilizados, mas um fator ainda limitante é o elevado valor dessa tecnologia. No tocante a mídias utilizadas para backup, as fitas magnéticas ainda são utilizadas, principalmente devido ao seu baixo custo e boa capacidade de armazenamento.

Diante do exposto, percebe-se que o armazenamento de dados é um fator crítico em datacenters. Atualmente existem diversas tecnologias que permitem um balanceamento objetivando um melhor custo-benefício, considerando aspectos de velocidades de acesso, disponibilidade, segurança e escalabilidade do armazenamento de dados.

## **2 Fundamentação teórica**

O armazenamento de dados é um fator crítico nos datacenters modernos. Um datacenter é um conjunto de componentes de alta tecnologia que permite fornecer serviços e infraestrutura de valor agregado, tipicamente processamento e armazenamento de dados, em larga escala, para qualquer tipo de organização (Veras, 2009). O surgimento de serviços populares tais como *e-mail* baseado na *Web*, busca (*searching*) e redes sociais, associados ao aumento da conectividade através de banda larga e redes ópticas, impulsionaram o modelo de comunicação centrado no servidor (*server-side*). Cada vez mais o processamento e o armazenamento estão sendo movidos dos PCs para grandes



provedores de serviços (Verdi et al. 2010). Isso faz com que o *storage* deva ser examinado com atenção, já que em muitos casos, os recursos virtuais acessarão uma *storage area network* (SAN) compartilhada (Moda et al. 2008). Tal diversidade de tecnologias e fabricantes elevam a complexidade do gerenciamento de um datacenter.

O datacenter é um ambiente que consome muita energia, os discos ficam em segundo lugar no consumo de energia em servidores (de Carvalho Junior, 2014). Isso torna a escolha de meios de armazenamento ainda mais delicada, pois deve-se acrescentar o consumo de energia na escolha da solução.

Uma arquitetura que está sendo amplamente discutida atualmente é a *Hyper-Converged Infrastructure* (HCI). Os responsáveis pelo planejamento de TI contam cada vez mais com as soluções de HCI para simplificar e acelerar as implementações de infraestrutura, facilitar o gerenciamento operacional diário, reduzir os custos e aumentar a velocidade e a agilidade da TI (Colm Keegan, 2016).

## **2.1 Análise Crítica dos Dados Coletados**

O armazenamento de dados é um fator crítico no sucesso do datacenter. As necessidades atuais das empresas têm demandado cada vez mais dessa área. Desde o surgimento dos primeiros discos rígidos, com pouca capacidade e elevado tempo de acesso, muito se evoluiu nas tecnologias de armazenamento de dados, assim como nas arquiteturas computacionais. O *storage* do datacenter sempre foi dominado pelo uso de discos rígidos (HDD). Os padrões SATA e principalmente o SAS, com elevado número de rotações por minuto (RPM), predominavam nas *Network Attached Storage* (NASs) e SANs.

A memória *flash* surgiu na década de 1980, porém inicialmente ficou muito restrita a nichos específicos. Com o barateamento desse tipo de memória secundária as empresas de tecnologia começaram a desenvolver soluções comerciais baseadas em discos de estado sólido (*SSDs*). O *SSD* possui diversas vantagens comparado ao disco magnético tradicional, dentre elas a maior largura de banda, maior resistência a choques mecânicos, menor consumo de energia, menor peso e menor tamanho.

No disco *SSD* o tempo de busca de um dado é zero, esse disco pode acessar dados de forma duas a três vezes mais veloz que um disco magnético de alta performance.



O custo dos discos *SSDs* foi por um longo tempo um impeditivo para que as empresas o adotassem em soluções de alta performance. Ocorre que a indústria já percebe que, apesar do custo superior ao disco magnético, o *SSD* pode ser mais econômico no médio a longo prazo, desde que seja adotada uma estratégia correta para o negócio.

Para Tanenbaum (2013), a solução para o armazenamento de grandes quantidades de dados é uma hierarquia de memória, pois a velocidade de acesso sobe exponencialmente com o custo do disco. Sendo assim, a combinação de diferentes tecnologias de discos pode propiciar uma melhor relação custo-benefício para o *storage* do datacenter.

Os *players* do mercado já oferecem soluções *all-flash* e híbridas de armazenamento. Devido a essas características de alta performance, baixo consumo de energia, conseqüentemente menos calor e necessidade de refrigeração, aliados a economia de espaço no datacenter, o *SSD* conquista cada vez mais o mercado de *storage*.

Usualmente os dados que estão em uso ativo ou devem ser acessados (lidos e gravados) com frequência, são chamados de dados quentes e devem ser armazenados nos discos com melhores performances, *SSDs*, com um custo maior. Já os dados menos acessados, pouco requisitados, são chamados de dados frios, estes devem ser armazenados em discos magnéticos comuns, com menor performance, mas a um custo menor. Nesse contexto, as soluções híbridas, com discos magnéticos e *SSDs*, são as mais utilizadas, essas devem ser capazes de promover dados frios para quentes, quando a frequência e velocidade de acesso exigirem. Assim como demoverem dados quentes para frios, quando esses tiverem suas frequência e velocidade de acesso diminuídas.

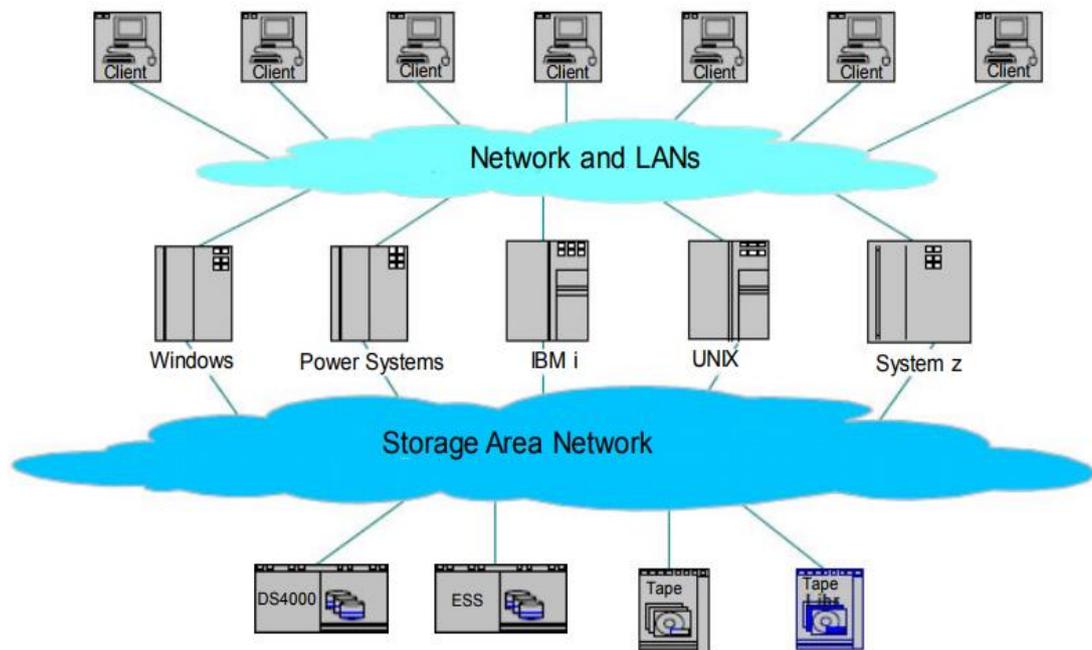
Com o uso cada vez mais frequente de armazenamento flash e com a conseqüente diminuição no tempo de acesso aos dados, o gargalo se tornou à camada de rede. A arquitetura consolidada no ambiente de datacenter para armazenamento em grande escala é a *Storage Area Network (SAN)*.

Segundo a *Storage Network Industry Association (SNIA)*, uma *Storage Area Network (SAN)* é uma rede especializada de alta velocidade que fornece acesso de rede em nível de bloco ao armazenamento. As *SANs* geralmente são compostas por *hosts*,

switches, elementos de armazenamento e dispositivos de armazenamento que estão interligados usando uma variedade de tecnologias, topologias e protocolos.

Nessa arquitetura os *arrays* de discos, chamados de *targets*, são ligados aos servidores, chamados de *initiators*, por uma rede segregada da LAN. A SAN possui protocolos e hardwares específicos para transferência de dados, o principal e mais caro protocolo é o *Fibre Channel (FC)*, mas também são usados os protocolos FCoE e iSCSI.

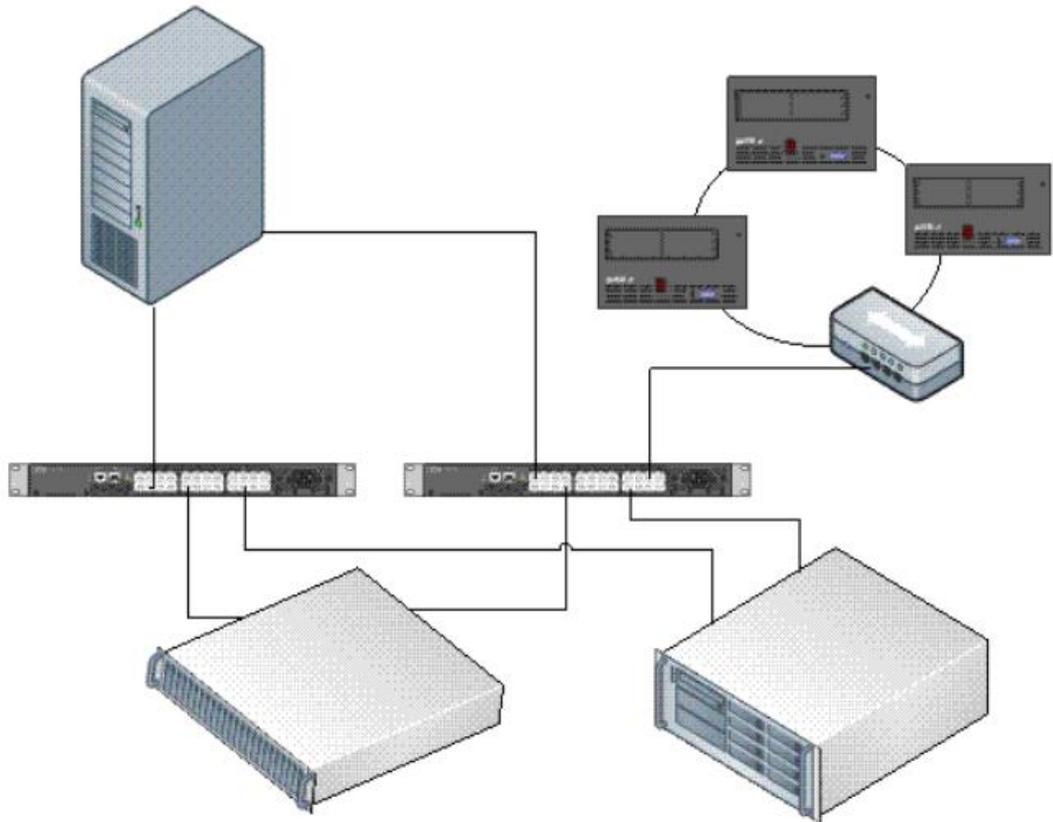
Figura 1.0 - Arquitetura SAN



Fonte: IBM – *Red Book - Introduction to Storage Area Networks*, 2016.

A principal topologia usada na SAN é a *Switched Fabric*, onde os *hosts*, com adaptadores *Host Bus Adapter (HBA)* são conectados a switches *Fibre Channel* que fazem a comutação dos dados até o *array* de discos. A topologia *Fibre Channel Switched Fabric (FC-SW)* permite múltiplas e coexistente conexões ponto a ponto via comutação em camada 2, provendo maior desempenho e escalabilidade.

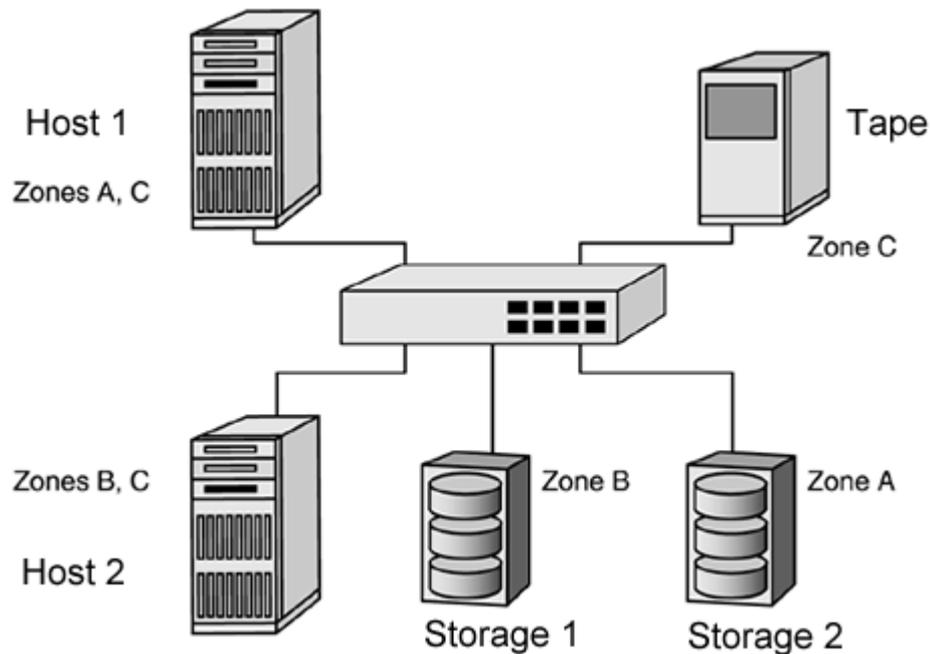
Figura 2.0 - Topologia *Fibre Channel Switched Fabric*



Fonte: IBM – *Red Book - Introduction to Storage Area Networks*, 2016.

Em conjunto com a topologia *switched fabric* é implementado o conceito de zoneamento. O zoneamento (*Fabric zoning*) permite a separação de dispositivos baseados: na função, separação de departamentos, ou potenciais conflitos entre sistemas operacionais. De forma simplificada, o zoneamento restringe o acesso a certos nós/recursos.

Figura 3.0 - Zoneamento (Fabric zoning)

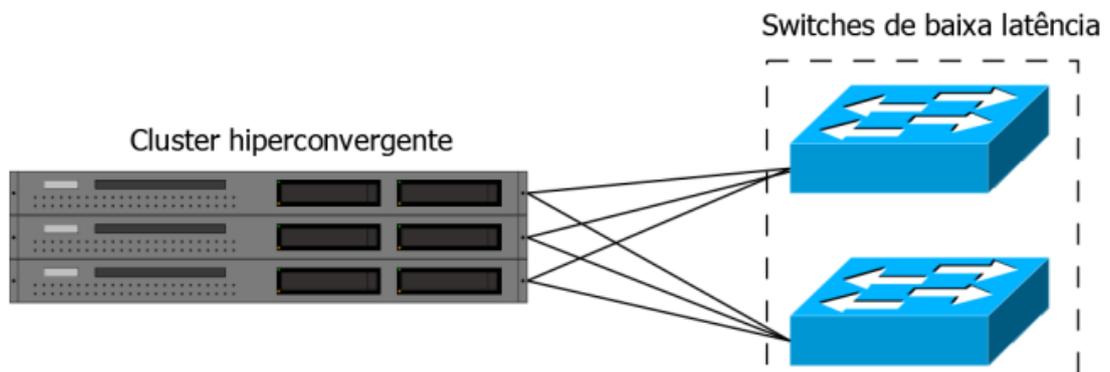


Fonte: *Designing Storage Area Networks: A Practical Reference for Implementing Fibre Channel and IP SANs*, 2003.

O zoneamento pode ser configurado de duas maneiras, por porta ou *World Wide Name* (WWN). O zoneamento por porta é mais seguro que zoneamento baseado em WWN, pois o intruso não consegue enganar a porta manipulando as informações do quadro *Fibre Channel*. As zonas podem ser definidas usando o *ID* do domínio do *switch* e o número da porta para identificar exclusivamente os membros da zona. A vantagem é a facilidade de configuração e a definição de zoneamento permanece intacta mesmo se um HBA ou controlador de destino for substituído. A desvantagem é a flexibilidade reduzida, a zona permanece com a porta e não com o dispositivo. Se houver uma mudança de um servidor de uma porta para outra, o servidor não mais pertencerá a zona que era designado. Já no zoneamento baseado em WWN, a zona é designada ao dispositivo. A vantagem é manter a definição de zoneamento mesmo quando o dispositivo é movido para uma porta ou comutador diferente. A desvantagem é que as definições de zoneamento devem ser alteradas sempre que um HBA é substituído por outro, tendo um WWN diferente. Isso facilita o traslado dos dispositivos, mas é mais vulnerável a intrusões.

Por essa tecnologia ter custo elevado, necessitar de gerenciamento especializado, acrescido do fato da rede se tornar o gargalo das soluções atuais, surgiu a hiperconvergência, que busca diminuir a distância entre os dados e o processador. Na arquitetura hiperconvergente os discos, *SSDs* e magnéticos, ficam no *host*. Para garantir a alta disponibilidade são montados arranjos de servidores, que trabalham como uma entidade única, um cluster. Esses servidores são interconectados por switches de baixa latência, garantindo assim a replicação de dados necessária ao ambiente de alta disponibilidade do datacenter.

Figura 4.0 – Topologia hiperconvergente



Fonte: Facilidades e Desafios de uma Solução Hiperconvergente, 2017.

No armazenamento também são necessários discos de alta performance como os *SSDs*. Segundo Souza Junior (2017), com o armazenamento de dados local, elimina-se o gargalo de rede já que os dados da VM não precisam mais trafegar da *storage* até o hipervisor. O armazenamento na hiperconvergência possui memória flash (discos *SSDs*) para garantir alta taxa de transferência e discos rígidos comuns para maior volume de armazenamento com menor custo financeiro, além disso é possível categorizar os dados, onde informações mais acessadas sejam mantidas em locais de armazenamento com uma taxa de leitura e escrita muito mais rápidas que os tradicionais discos rígidos.



### 3 CONCLUSÕES

Nesse artigo pode-se concluir que diferentes arquiteturas e meios são empregados no armazenamento de dados do datacenter. As diferentes necessidades de negócio ditam as principais características dessas soluções.

A *Storage Area Network* (SAN) é uma arquitetura consolidada no ambiente do datacenter. Nessa arquitetura os *arrays* de discos, chamados de *targets*, são ligados aos servidores, chamados de *initiators*, por protocolos e hardwares específicos para transferência de dados, sendo o protocolo proprietário *Fibre Channel*(FC) o mais utilizado, mas também são usados os protocolos FCoE e iSCSI.

Com o aumento da velocidade dos discos a rede passou a ser exigida cada vez mais. Nesse contexto surge a arquitetura hiperconvergente, a qual busca a aproximação do processamento e armazenamento dos dados. Nessa arquitetura os discos, SSDs e magnéticos, são conectados no *host*. A alta disponibilidade e a tolerância a falhas são garantidas, pois os servidores trabalham com uma entidade única, um cluster. Os nós do cluster são interconectados por switches de baixa latência, garantindo assim a replicação de dados necessária ao ambiente de alta disponibilidade do datacenter.

Conclui-se com esse trabalho que o armazenamento de dados é uma característica muito importante no ambiente de missão crítica do datacenter, sendo diversas as necessidades de negócios e tecnologias disponíveis no mercado. Como sugestão para futuras pesquisas pode-se relacionar os seguintes estudos: percentual de adoção de soluções hiperconvergentes em datacenters modernos; modelo matemático que possibilite identificar qual solução é mais adequada a determinados requisitos de negócio; custo médio das diferentes arquiteturas de armazenamento para pequenas e médias empresas.



## REFERÊNCIAS

VERAS, M. **Datacenter - Componente Central da Infraestrutura de TI**. Brasport, 2009.

VERDI, Fábio Luciano; Rothenberg, Christian Esteve.; Pasquini, Rafael.; Magalhães, Maurício F. **Novas Arquiteturas de Data Center para Cloud Computing**. 2010. Disponível em: <http://www.dca.fee.unicamp.br/~chesteve/pubs/MC-DATA-CENTER-NETWORKS-SBRC2010.pdf>. Acesso em: 12 de agosto de 2017.

MODA, Cássio.; Cremonin, Fabiano Loverbeck.; Cremonin, Rodrigo Marassi. **Virtualização e alta disponibilidade em ambiente corporativo**. 2008. Disponível em: <http://rodrigomarassi.com/wp-content/themes/minicard/images/artigo-cientifico-virtualizacao-alta-disponibiliade.pdf>. Acesso em: 12 de agosto de 2017.

de Carvalho Junior, Osvaldo Adilson. **GreenMACC - Uma arquitetura para metaescalamento verde com provisão de QoS em uma nuvem privada - ICMC-USP.2014** - Disponível em: [http://www.teses.usp.br/teses/disponiveis/55/55134/tde-08042015-161656/publico/Tese\\_Osvaldo\\_Versao\\_Revisada.pdf](http://www.teses.usp.br/teses/disponiveis/55/55134/tde-08042015-161656/publico/Tese_Osvaldo_Versao_Revisada.pdf). Acesso em: 9 de agosto de 2017.

Colm Keegan. **Transformação hiperconvergente: noções básicas sobre o SDDC (Software Defined Data Center)**. White paper. 2016. Disponível em: <https://brazil.emc.com/collateral/white-papers/esg-whitepaper-hyperconverged-transformation-sddc.pdf>. Acesso em: 9 de agosto de 2017

TANENBAUM, Andrew S. **Organização estruturada de computadores** / Andrew S. Tanenbaum, Todd Austin; tradução Daniel Vieira; revisão técnica Wagner Luiz Zucchi – São Paulo: Pearson Prentice Hall, 2013.

VELHO, Jesus Antonio et all. **Tratado de Computação Forense**. p43-44 Millennium Editora. 2016.

ROSÁRIO, Djan de Almeida do. **Disponibilidade e qualidade operacional de datacenters**. Palhoça: Unisul virtual, 2016.

FACCIONI FILHO, Mauro. **Conceitos e infraestrutura de datacenters**. Palhoça: Unisul virtual, 2016.



**Armazenamento em flash no caminho certo do datacenter** – Disponível em:

[http://i.dell.com/sites/doccontent/shared-content/data-sheets/pt/Documents/Dell\\_SD\\_Flash\\_Storage\\_Hits\\_Its\\_Stride\\_in\\_the\\_Data\\_Center\\_br.pdf](http://i.dell.com/sites/doccontent/shared-content/data-sheets/pt/Documents/Dell_SD_Flash_Storage_Hits_Its_Stride_in_the_Data_Center_br.pdf).

Acesso em: 22 de setembro de 2017.

**Introduction to Storage Area Networks** - Disponível em:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg245470.pdf>. Acesso em: 14 de dezembro de 2017.

**Fibre Channel e Storage Area Network** – Disponível em:

[https://www.gta.ufrj.br/grad/08\\_1/san/index.html](https://www.gta.ufrj.br/grad/08_1/san/index.html). Acesso em: 15 de dezembro de 2017.

SOUZA JUNIOR, Zilmar. **Facilidades e desafios de uma solução hiperconvergente**.

Unisul, 2017.

CLARK, Tom. **Designing Storage Area Networks: A Practical Reference for Implementing Fibre Channel and IPSANs.2.ed.** Addison-Wesley, 2003.