

# *Desempenho máximo com custo mínimo em dados*

## *Aplicativos de mineração por meio dos novos dados on-line*

### *Arquitetura de armazém usando área de armazenamento*

### *Rede com malha de canal de fibra*

*M. Saravanamuthu*

*PhD Scholar*

*Anna University Coimbatore Region Chennai,  
Tamilnadu, Índia  
msaravanamuthu@gmail.com*

*Dr. GM Kadhar Nawaz*

*Diretor*

*Departamento de MCA, Sona College of Technology Salem,  
Tamilnadu, Índia  
nawazse@yahoo.co.in*

**Resumo** - Os principais problemas da arquitetura de data warehouse são o custo mais alto e o baixo desempenho. As razões para esses problemas concentram-se no projeto arquitetônico. As soluções são propostas com o novo design de data warehouse modificado através da arquitetura de data warehouse online com o recente suporte tecnológico da computação em nuvem - redes de área de armazenamento através do uso de dispositivos de armazenamento baseados em canal de fibra de alta velocidade com comutadores de canal de fibra. Finalmente, os benefícios e vantagens do projeto proposto foram discutidos com o custo e desempenho.

**Palavras-chave** - Arquitetura de Data Warehouse, Computação em Nuvem, Arquitetura de Storage, Storage Area Network, Network Storage, Fibre Channel Protocol, Fibre Channel Switch/Fabric.

#### I. INTRODUÇÃO

Os problemas nas arquiteturas de data warehouse existentes foram gradualmente resolvidos, estágio por estágio, pelos projetistas de data warehouse. Principalmente, os problemas em aplicativos de inteligência de negócios (BI) executados com ferramentas de mineração de dados (DM) ainda não foram resolvidos. A primeira questão é que o custo do sistema de data warehouse é alto nos modelos existentes devido à enorme quantidade de requisitos de armazenamento de dados e memória para análise de dados. A segunda questão é que o desempenho das ferramentas de BI e DM usadas pelo sistema de data warehouse não é satisfatório devido ao design complexo da arquitetura do data warehouse. Portanto, o design do data warehouse deve ser modificado de acordo com os desenvolvimentos tecnológicos disponíveis na indústria e para atender às necessidades do negócio. Este trabalho de pesquisa propõe uma nova maneira de projetar a arquitetura do data warehouse usando a tecnologia de computação em nuvem com a ajuda do armazenamento em canal de fibra, reduzindo o custo geral com melhoria no desempenho do data warehouse.

#### II. REVISÃO DE LITERATURA

##### A. Armazenagem de dados

###### 1) Componentes

Os principais componentes [1] são armazenamento de dados operacionais, gerenciador de carga, gerenciador de armazém, metadados, data mart e gerenciador de consultas. Estes também podem ser projetados com modelo em camadas [22].

###### 2) Abordagens de desenvolvimento

O modelo Inmon [2] [13] é baseado na abordagem de data warehouse empresarial, ferramentas de banco de dados relacionais, não voláteis e variantes no tempo, orientadas por assunto, integradas, de cima para baixo. O modelo de Kimball [3] [13] é baseado na abordagem de data mart, de baixo para cima, modelagem dimensional, planejar grande, construir pequeno. Modelos de data mart independentes e dependentes também foram usados neste projeto. Data warehousing em tempo real (ativo) e Virtual data warehousing [20] também existem. Outras arquiteturas podem ter modelos de um, dois e três níveis [17]. Alguns podem usar nuvem [25].

As quatro visualizações de uma estrutura de análise de negócios [6] são: visualização de cima para baixo, visualização da fonte de dados, visualização do armazém de dados e visualização de consulta de negócios.

###### 3) Fatores determinantes da arquitetura de armazenamento de dados

Os principais fatores que determinam o sucesso do armazenamento de dados [5] são: função cruzada, suporte em todo o campus, consistência e integração entre áreas temáticas, segurança, controle de acesso, suporte para usuários operacionais e suporte para acesso flexível para tomadores de decisão. Os fatores de influência que determinam a qualidade do data warehouse [4][14] são: integridade, não redundância, aplicação de regras (de negócios), reutilização de dados, estabilidade e flexibilidade, simplicidade e elegância, comunicação e eficácia e desempenho.

Os principais problemas no projeto de data warehouse [1][14] são: Subestimação de recursos para carregamento de dados, Problemas ocultos com sistemas de origem, Dados necessários não capturados,

## Conferência Internacional de Circuitos, Potência e Tecnologias de Computação de 2015 [ICCPCT]

Maior demanda do usuário final, Homogeneização de dados, Alta demanda por recursos, Propriedade de dados, Alta manutenção, Projetos de longa duração e Complexidade de integração.

### B. Computação em Nuvem

#### 1) Computação sob demanda moderna

O modelo de computação sob demanda [11] pode ser utilizado para minimizar o custo de investimento envolvido nos recursos. É simplesmente um tipo de tecnologia de pagamento por uso, sem a compra de nenhum recurso dentro da organização. Portanto, eliminará completamente o custo de manutenção do hardware. A organização pagaria apenas as taxas de aluguel dos recursos usados mensalmente ou anualmente. Isso reduzirá drasticamente os custos de investimento e manutenção dos recursos necessários à organização.

#### 2) Além da área de trabalho

A computação em nuvem [8, 28] nos permite armazenar e executar aplicativos e dados em nuvens dinâmicas para acesso universal. A noção de modelo centrado no desktop usado na computação cliente/servidor será substituída pela computação em nuvem com modelo centrado no usuário, centrado na tarefa e centrado nos dados. É o caminho do futuro. O Cloud BI [19] existe também com vários desafios de nuvem [24].

#### 3) Principais propriedades

As principais propriedades [8] da computação em nuvem são: centrada no usuário, centrada na tarefa, poderosa, acessível, inteligente e programável.

#### 4) Benefícios

Os benefícios [9, 15, 21, 23] da computação em nuvem são: autoatendimento sob demanda, amplo acesso à rede, pooling de recursos, Elasticidade rápida e serviço medido

#### 5) Vantagens

As vantagens [9, 30] da computação em nuvem são: custos mais baixos, facilidade de utilização, qualidade de serviço, confiabilidade, capacidade de gerenciamento, vantagem estratégica, gerenciamento de TI terceirizado, manutenção e atualização simplificadas e baixa barreira à entrada. O novo serviço *Disaster Recovery as a Service-DRaaS* [26] também existe. *High Availability-HA* [27, 29] é um recurso de nível de sistema (arquitetura, design, conjunto de software) que garante o tempo de atividade do sistema.

### C. Rede de Área de Armazenamento

#### 1) A conectividade de armazenamento universal

A computação em nuvem oferece dois modelos de armazenamento [7] *Armazenamento conectado à rede* (NAS) e *rede de área de armazenamento* (SAN). Entre estes, o modelo SAN possui várias variantes com suporte ao *Fibre Channel Protocol* (FCP), que é um dispositivo de armazenamento de alta velocidade com uma enorme taxa de transferência de dados. Uma arquitetura típica de SAN é mostrada na figura 1. O principal recurso de SAN é a conectividade de qualquer para qualquer computador e dispositivos de armazenamento. SAN é conectividade de armazenamento universal.

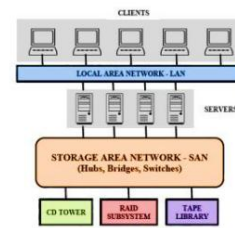


Figura 1: Uma arquitetura SAN típica

#### 2) Componentes

Os principais componentes [10] do SAN são: parte da rede, Parte de hardware, parte de software e parte de conectividade.

##### D. Fibre Channel

#### 1) Interconexão de alta velocidade

A rede de armazenamento vem amadurecendo desde a introdução do *Fibre Channel* (FC) [7] no início dos anos 90. Normalmente os dispositivos Fibre Channel são dispositivos muito rápidos que trabalham com tecnologia Gigabit. Um dispositivo SAN é um bloco de acesso [18] que se conecta a seus clientes usando FC e um protocolo de acesso a dados de bloco, como SCSI.

#### 2) Objetivos do projeto

Por acaso, os objetivos do projeto [12] do FC são contemplados pelos requisitos de uma tecnologia de transmissão para redes de armazenamento como: Transmissão serial para alta velocidade e longas distâncias, Baixa taxa de erros de transmissão e Baixo atraso (latência) dos dados transmitidos. Implementação do Fibre Channel Protocol em hardware nas placas *Host Bus Adapter* (HBA) para liberar as CPUs do servidor.

#### 3) Topologias

As topologias [7] do FC são:

*Ponto a ponto*, no qual dois dispositivos são conectados diretamente entre si por um cabo FC conforme a figura 2(a).

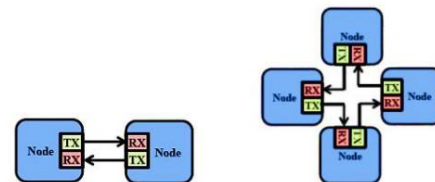


Figura 2: (a) Ponto a ponto, (b) Loop arbitrário

*Loop arbitrário*, no qual vários dispositivos compartilham um único caminho de transmissão. Um loop arbitrário FC pode conectar até 126 dispositivos. É mostrado na figura 2(b).

*Switched ou fabric*, em que cada dispositivo é conectado diretamente a um ponto de conexão central chamado switch. O switch faz e interrompe conexões momentâneas entre pares de dispositivos que precisam se comunicar entre si.

Assim como os switches Ethernet, os switches FC são dispositivos de alto desempenho capazes de intermediar várias comunicações simultâneas entre pares de dispositivos. Isso é mostrado na figura 2(c).

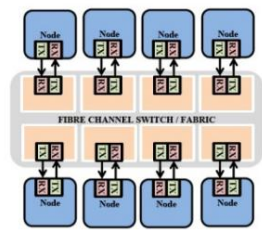


Figura 2(c): Interruptor ou tecido

As características [7] das topologias FC individuais são mostradas na tabela 1 a seguir.

Tabela 1: Características das topologias FC

Attribute	Point-to-Point	Arbitrated Loop	Switched Fabric
Max. ports	2	127	~16777216 (2 <sup>24</sup> )
Address size	N/A	8-bit ALPA	24-bit port ID
Side effect of port failure	Link fails	Loop fails (until port bypassed)	N/A
Frame delivery	In order	In order	Not guaranteed
Mixing different link rates	No	No	Yes
Access to medium	Dedicated	Arbitrated	Dedicated

E. Interruptor/tecido FC

Um comutador FC [7] ou design de malha [16] tem aparência semelhante a um hub. A Figura 3 descreve as principais características de um switch FC. Para simplificar, um switch de quatro portas é representado. Os switches reais contêm mais portas.

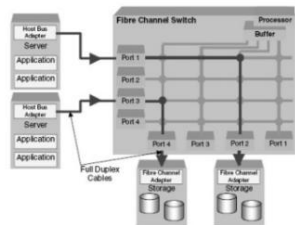


Figura 3: switch/malha FC

A topologia fabric [12] é a mais flexível e escalável das três topologias FC. Uma malha consiste em um ou mais switches FC conectados entre si. Servidores e dispositivos de armazenamento são conectados à malha pelos comutadores FC. Os switches FC podem trocar dados por meio de conexões switch-a-switch (*Inter Switch Links-ISL*). Vários ISLs podem ser instalados entre dois switches para aumentar a largura de banda.

A Figura 4 compara a latência de diferentes componentes FC SAN. Um cabo FC de 10 quilômetros aumenta a latência de uma conexão ponta a ponta. Para componentes de hardware, a regra geral é que um switch FC pode encaminhar um quadro em 2 a 4 microssegundos; um FC HBA requer de 2 a 4 milissegundos para processá-lo.

Uma característica especial da malha é que vários dispositivos podem enviar e receber dados simultaneamente na taxa de dados total. Todos os dispositivos, portanto, têm a largura de banda total disponível para eles ao mesmo tempo.

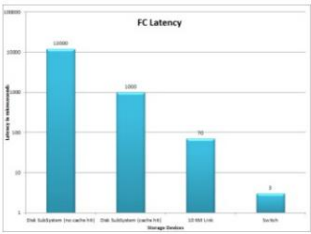


Figura 4: A latência do FC

1) Vantagens

As vantagens [7] do switch FC são: custo mais baixo, desempenho muito alto, suporte de longa distância e menos sobrecarga.

III. O DESENHO DE DADOS PROPOSTO  
ARQUITETURA DO ARMAZÉM

Este trabalho de pesquisa propõe uma nova arquitetura para armazenamento de dados online usando redes de área de armazenamento para resolver os problemas de custo mais alto e menor desempenho enfrentados nos projetos anteriores de armazenamento de dados.

A. As várias gerações de arquiteturas de data warehouse

*Arquitetura de data warehouse de primeira geração:* durante a invenção dos computadores pessoais, todo software desenvolvido era autônomo e não seguro. O projeto de arquitetura de data warehouse de primeira geração também estava no mesmo modelo. Os desenvolvedores de aplicativos desenvolveram software para acessar os dados no data warehouse diretamente por meio de versões simples de ferramentas de mineração de dados sem nenhum tipo de segurança. Essas ferramentas de mineração de dados extraíram conhecimento e geraram várias formas de dados usando técnicas de visualização. A arquitetura do data warehouse de primeira geração foi baseada em um modelo de duas camadas, conforme a figura 5(a).

*Arquitetura de data warehouse de segunda geração:* O rápido desenvolvimento da tecnologia de rede introduziu a computação cliente/servidor. A arquitetura de data warehouse de segunda geração foi baseada neste modelo com segurança sobre os dados. Também permite o acesso a dados em locais geograficamente distribuídos, digamos servidores. As aplicações desenvolvidas para este tipo de sistemas de data warehouse foram executadas em sistemas clientes de forma segura. A arquitetura do data warehouse de segunda geração foi baseada no modelo de três camadas, conforme mostrado na figura 5(b).

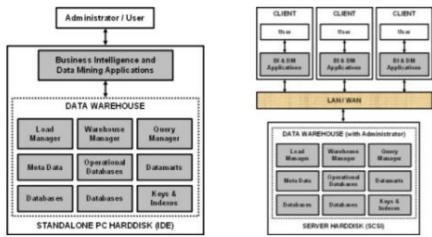


Figura 5 (a): Primeira geração (b): Segunda geração

*Arquitetura de data warehouse de terceira geração:* as melhorias na tecnologia de computação cliente/servidor

Conferência Internacional de Circuitos, Potência e Tecnologias de Computação de 2015 [ICCPCT]

alcançou os aplicativos baseados na web usando a tecnologia da internet. A arquitetura do data warehouse de terceira geração é mostrada na figura 5(c). Porém, aplicações de BI e DM rodando na internet não estavam satisfeitas com o desempenho devido aos protocolos usados e razões de largura de banda. Porque os aplicativos de BI e DM exigem uma grande quantidade de espaço de RAM e espaço de armazenamento em disco rígido ao trabalhar com bancos de dados de data warehouse.

B. A Arquitetura de Data Warehouse Proposta

A necessidade de um projeto de data warehouse modificado e os recursos de armazenamento em rede juntos formam a nova arquitetura de data warehouse online. O projeto proposto vai usar o modelo de rede de área de armazenamento para armazenamento de data warehouse. A estrutura do projeto proposto é mostrada na figura 5(d) como arquitetura de data warehouse de próxima geração (proposta). A malha/switch de canal de fibra desempenha um papel importante entre os servidores e os dispositivos de armazenamento.

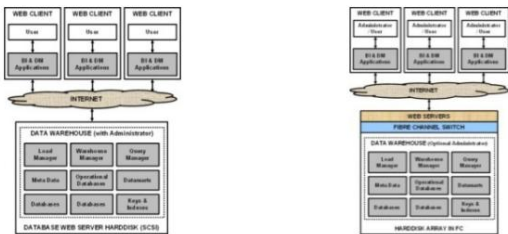


Figura 5 (c): Terceira geração (d): Projeto proposto

C. Comparação de Arquiteturas de Armazém de Dados de Primeira, Segunda e Terceira Geração e Propostas

As várias gerações existentes de projetos de arquitetura de data warehouse são comparadas com o projeto de data warehouse online proposto com vários recursos técnicos conforme a tabela 2.

Tabela 2: Análise de várias gerações de dados arquiteturas de armazéns

Data Warehouse Architecture Features	First Generation	Second Generation	Third Generation	Next Generation (proposed)
1.Number of tiers	Two tier model	Three tier model	Three tier model	Multi-tier model
2.Number of users	Single user model (or multiple users but, one at a time)	Limited multi-user model	Unlimited multi-user model without any bandwidth limitations	Unlimited multi-user model without any bandwidth limitations
3.Data storage technology	Local standalone PC hard disk with space limitations	Shared client/server hard disk with space limitations	Shared database with server hard disk with space limitations	Shared array of hard disk (RAID) without any space limitations
4.Storage technology	IDE or Enhanced IDE	Parallel SCSI	Ultra SCSI	FC (or iSCSI)
5.Maximum data transfer rate	IDE - 3.3 to 25 MBps Enhanced IDE: 16.7 to 133 MBps	Parallel SCSI 5 to 20 MBps	Ultra SCSI 40 to 640 MBps	10GFC to 40GFC/ 100 to 4800 MBps per direction
6.Processor load	Heavy due to multi-tasking in PC	Shared heavily but limited users	Shared heavily but limited bandwidth	Shared lightly with FC management (shared by HBA)
7.Network load	Not applicable	Shared with limited users and traffic	Shared with limited bandwidth	Shared without any restrictions
8.Data accessibility	Fast due to local access	Based on network traffic	Based on network and internet bandwidth	Normally fast due to FC
9.Load balancing	Possible with local buffering only	Possible with C/S and network buffering	Possible with C/S, network and internet buffering	Possible with C/S, network, internet, FC and HBA
10.Response time	Fast due to local processor	Medium because shared	Medium because shared	Fast due to FC
11.Application execution	Local	Client	Client or server	Client or server
12.Overall data warehouse performance	Initially good, but poor when data volume increases due to single user model	Initially good, but not satisfactory when data volume increases	Moderate, but not satisfactory when user volume increases	Excellent due to FC
13.Fault tolerance	Not possible	Possible with backup server	Possible with backup web server	Intelligent FC switch, RAID and HBA
14.Recovery time	Usually long time, but based on backup policy	Moderate, but based on backup policy	Moderate, but based on backup policy	Negligible due to FC, RAID and HBA
15.Security	0%	Not 100% due to C/S policies	Not 100% due to C/S, network and internet	100% assured
16.Scalability	Possible only by connecting limited number of hard disks	Possible only by connecting more hard disks in both C/S	Possible only by connecting more hard disks in Web Server	Default and automatic due to RAID and FC fabric
17.Cost	Low	Medium	High	As per usage (daily / monthly / yearly)
18.Services available	Nil (local services only)	C/S services only	Limited web services	Unlimited web services with cloud model

Os recursos incluem número de camadas, número de usuários, armazenamento de dados, tecnologia de armazenamento, taxa de transferência de dados, carga do processador, carga de rede, acessibilidade de dados, balanceamento de carga, tempo de resposta, execução de aplicativos, desempenho do armazém, tolerância a falhas, tempo de recuperação, segurança, escalabilidade, custo e serviços. De

tabela, é claro que o modelo proposto tem razões mais sólidas do que outros modelos.

D. Benefícios tecnológicos e comerciais da SAN

A rede de armazenamento é um mecanismo para entregar serviços de benefício para a empresa [7]. Uma SAN pode

- Reduzir o custo de capital de armazenamento • Reduzir o custo de gerenciamento de armazenamento • Reduzir os requisitos de largura de banda de I/O • Melhorar a precisão e pontualidade dos dados de negócios • Prolongar a vida útil do equipamento de capital

4. TESTE

O projeto proposto de arquitetura de data warehouse online é testado com interconexão de armazenamento em rede. O dispositivo usado para teste é uma matriz de armazenamento de rede. O nome e o modelo do dispositivo são *Iomega StorCenter px4-300r*. Inicialmente, uma aplicação de mineração de dados é executada e o fluxo de dados da rede e do servidor é analisado sem a conexão do dispositivo. O resultado disso é mostrado na figura 6.

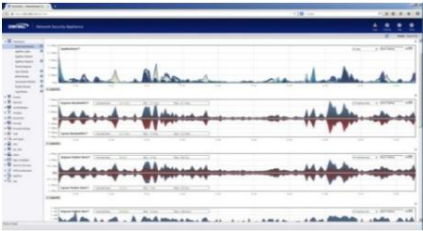


Figura 6: Fluxo de dados sem armazenamento de rede

Observa-se pela figura 6 que, o fluxo de dados é muito pequeno no servidor devido a carga que é dada ao servidor e rede devido ao seu buffer local. A velocidade nativa do dispositivo ainda não foi alcançada neste projeto. Portanto, os dispositivos de armazenamento são subutilizados com enorme capacidade de armazenamento e o servidor e a rede estão sobrecarregados, o que leva à degradação do desempenho. Esse tipo de fluxo de dados não é adequado para aplicativos e gerenciamento de dados on-line porque degradaria o desempenho da rede e do servidor. Além disso, quando o número de usuários aumenta, o desempenho cai novamente devido ao gerenciamento de usuários junto com seus dados.

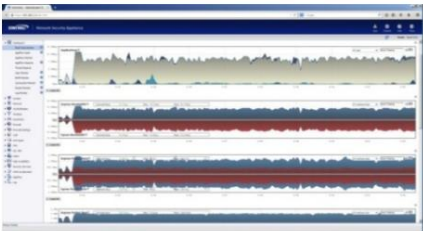


Figura 7: Fluxo de dados com armazenamento em rede

Em seguida, o mesmo aplicativo de mineração de dados é executado junto com a conexão de armazenamento de rede e analisado o desempenho do fluxo de dados. O resultado do fluxo de dados é mostrado na figura 7.

## Conferência Internacional de Circuitos, Potência e Tecnologias de Computação de 2015 [ICCPCT]

Observa-se na figura 7 que o fluxo de dados é muito grande tanto na rede quanto no servidor porque a carga não é fornecida ao servidor e à rede devido ao adaptador de barramento do host. O switch gerenciável está controlando os problemas de buffer com os dados. O switch de canal de fibra nos dará a velocidade do dispositivo nativo sem nenhum atraso na transferência de dados. O resultado mostra que o desempenho é altamente aprimorado tanto no lado da rede quanto no lado do servidor, reduzindo a carga para eles. Quando o volume do usuário aumenta, o dispositivo ainda se comporta da mesma maneira porque os dados são gerenciados pelo switch e os usuários são gerenciados pelo servidor ou pela rede. Assim, os dispositivos de armazenamento são utilizados no máximo e o servidor e a rede são carregados levemente, o que aumenta o desempenho geral da transferência de dados na rede e no servidor, por sua vez, esse design melhora o desempenho geral do aplicativo de mineração de dados devido à arquitetura de armazenamento online proposta projeto com redes de armazenamento com modelo de malha de canal de fibra.

## V. RESULTADOS E DISCUSSÃO

A Figura 8 compara a carga da CPU do tráfego de dados TCP/IP e FC [12]. A razão para a baixa carga de CPU do FC é que uma grande parte da pilha do protocolo FC é realizada no FC HBA.

Por outro lado, nas placas de rede atuais, grande parte da pilha de protocolos TCP/IP é processada na CPU do servidor. A comunicação entre a placa de rede Ethernet e a UCP ocorre por meio de interrupções. Isso custa poder de computação adicional, porque cada interrupção aciona uma mudança de processo cara no sistema operacional. A carga da CPU do FC é baixa porque seu protocolo é processado principalmente no HBA.

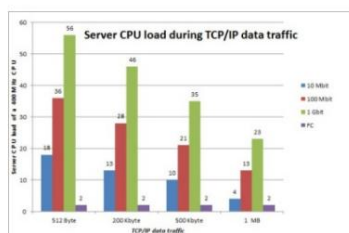


Figura 8: O tráfego de dados TCP/IP sobrecarrega a CPU do servidor.

A Tabela 4 lista o custo da incapacidade de processar informações por vários períodos para duas empresas para as quais o processamento de dados online é essencialmente o negócio [7]. Uma típica operação de compras domésticas processa um grande número de transações de valor relativamente baixo. Com uma corretora de valores, por outro lado, os valores das transações individuais podem ser muito maiores, portanto, o custo por minuto de inatividade é correspondentemente maior. As linhas da tabela são extrapoladas da linha superior por multiplicação. Na realidade, outros fatores, como a incapacidade de cumprir obrigações financeiras, acarretariam consequências ainda maiores do que o custo financeiro sugerido pela tabela 3.

Tabela 3: Custo do tempo de inatividade – Alguns exemplos

Percentage of Uptime	Annual Downtime	Annual Cost (Home Retail)	Annual Cost (Brokerage)
99.9999%	30 seconds	\$950	\$53,750
99.999%	5 minutes	\$9,417	\$537,500
99.99%	52 minutes	\$98,000	\$5,590,000
99.9%	8.75 hours	\$988,750	\$56,000,000
99.5%	43.7 hours	\$5,000,000	\$280,000,000
99%	87.6 hours	\$10,000,000	\$560,000,000
98%	180+ hours	\$20,000,000+	\$1,000,000,000+
95%	450+ hours	\$50,000,000+	\$3,000,000,000+

O valor do tempo de atividade pode diferir para diferentes empresas, mas a mensagem geral da Tabela 3 de que o tempo de inatividade é extremamente caro – é universal. A tecnologia SAN permite novas técnicas de processamento de informações que podem reduzir drasticamente o tempo de inatividade, o que, como pode ser visto na Tabela, pode significar economias de custos muito significativas ou até mesmo a sobrevivência de uma empresa. A Figura 9 ilustra várias técnicas de disponibilidade de dados e aplicativos que aumentam a acessibilidade com custos crescentes[7]. A linha de base de disponibilidade ilustrada na figura é um sistema de computador adequadamente gerenciado, mas normal, cujo backup é feito de forma razoavelmente frequente e regular.

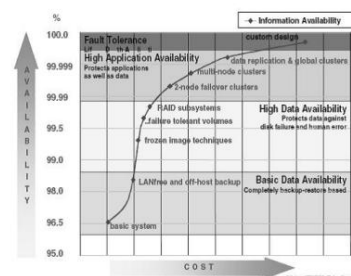


Figura 9: Graus de disponibilidade do sistema.

A figura sugere que o backup sem LAN e fora do host, ambos ativados por SANs, aumentam o tempo de atividade. Ambas as técnicas reduzem o impacto do backup nos aplicativos, o que, por sua vez, permite backups mais frequentes. Backups mais frequentes significam tempos de restauração mais curtos e menos reprodução de diário ou outras formas de reconstrução de dados, o que, por sua vez, significa menos tempo até que os aplicativos voltem a funcionar após uma falha. Como a figura também sugere, essa melhoria tem um custo.

## VI. CONCLUSÃO

O projeto de arquitetura de armazenamento de dados on-line proposto com modelo de rede de área de armazenamento por meio de malha de canal de fibra mostra que o desempenho aprimorado em aplicativos de mineração de dados e o custo de infraestrutura para dispositivos de armazenamento e computação estão sendo reduzidos drasticamente, substituindo-os por aluguel em vez de compra, o que é o núcleo do serviço de computação em nuvem chamado Storage as a Service (SaaS). Se preferirmos a opção de compra, o custo de investimento inicial seria alto e o custo de manutenção seria menor e deveríamos atualizar regularmente o software e o hardware de acordo com a mudança na indústria, que tem várias dores de cabeça como manutenção. Em seguida, o desempenho geral do projeto de arquitetura de data warehouse proposto é mais satisfatório do que os projetos anteriores porque o novo projeto usa alta velocidade



## Conferência Internacional de Circuitos, Potência e Tecnologias de Computação de 2015 [ICCPCT]

dispositivos de armazenamento (RAID) com taxa de transferência de dados muito alta (estrutura de canal de fibra) junto com escalabilidade conforme discutido anteriormente. Finalmente, o acesso aos dados no data warehouse é claramente examinado para que a computação em nuvem e as redes de área de armazenamento forneçam qualquer conectividade como padrão. Assim, as soluções sugeridas neste trabalho de pesquisa podem definitivamente satisfazer as necessidades de negócios em termos de inteligência de negócios e ferramentas de mineração de dados que estão sendo usadas para análise de dados em data warehouse. Este trabalho pode ser estendido para modelos de computação móvel e computação em nuvem pura.

## REFERÊNCIAS

- [1] "Transparências de conceitos de armazenamento de dados", Pearson Education Limited, 1995, &2005.
- [2] William H. Inmon, "Building the Data Warehouse", Quarta Edição, Wiley Publishing, 2005
- [3] Ralph Kimball e Margy Ross, "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", John Wiley & Sons, 3ª ed., 2013.
- [4] Graeme C. Simsion e Graham C. Witt, "Logical Database Design", Elsevier Inc., 2005.
- [5] Peter Cava (UC Berkeley, Gerente de Serviços de Data Warehouse), "Enterprise Data Warehouse", 2006.
- [6] Jiawei Han, MichelineKamber e Jian Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann (Elsevier) Editores, 3ª ed., 2012.
- [7] Richard Barker e Paul Massiglia, "Storage Area Network Essentials-A Complete Guide to Understanding and Implementing SANs", Wiley Computer Publishing, 2010, pp. 4, 29-47, 300-301, 303, 304.
- [8] Michael Miller, "Computação em nuvem: aplicativos baseados na Web que mudam a maneira como você trabalha e colabora online", Que Publishing, 2009, pp. 12-15.
- [9] Barrie Sosinsky, "Bíblia da Computação em Nuvem", Wiley Publishing, Inc., 2011.
- [10] Robert Spalding, "Storage Networks-The Complete Reference", McGraw-Hill/Osborne, 2003, pp. 182-183.
- [11] John W. Rittinghouse e James F. Ransome, "Computação em Nuvem: Implementação, Gerenciamento e Segurança", CRC Press Taylor & Francis Group, 2010, pp. 36-37.
- [12] Ulf Troppens, Wolfgang Muller-Friedt, Rainer Wolafka, Rainer Erkens e Nils Haustein, "Storage Networks Explained-Basics and Application of Fibre Channel SAN,NAS, iSCSI, InfiniBand and FCoE", John Wiley & Sons, 2ª ed. ., 2009, pp. 66-69, 71-72, 89-90, 114-115.
- [13] Ian Abramson, "Data Warehouse-The Choice of Inmon versus Kimball", IAS Inc.
- [14] Sean Kelly, "Data Warehousing in Action", Wiley India, 2007, pp. 112-113.
- [15] Simarpreet Singh Gulati e Sakshi Gupta, "Uma estrutura para aprimorar a segurança e o desempenho em aplicativos multilocatários", International Journal of Information Technology and Knowledge Management, julho-dezembro de 2012, Volume 5, nº 2, pp. 237.
- [16] AN Uwaechia, O. Akinsanmi, "Avaliação de confiabilidade no modelo de desempenho dos repositórios de rede de dados da Universidade Ahmadu Bello para design de rede de área de armazenamento", Jornal internacional de pesquisa inovadora em ciência, engenharia e tecnologia, vol. 2, edição 7, julho de 2013, pp.3311-3315.
- [17] Rajiv Senapati, D.Anil Kumar, "Uma Pesquisa sobre Arquitetura de Data Warehouse", Jornal Internacional de Pesquisa Inovadora em Engenharia de Computação e Comunicação, vol. 2, edição 8, agosto de 2014, pp. 5235-5239.
- [18] David HC Du, "Avanços recentes e desafios futuros dos sistemas de armazenamento", Proc. do IEEE, vol. 96, nº. 11, novembro de 2008, pp. 1875-1886.
- [19] Yuvraj Singh Gurjar, Vijay Singh Rathore, "Cloud Business Intelligence – é o que os negócios precisam hoje", Jornal Internacional de Tecnologia e Engenharia Recentes, Volume-1, Edição-6, janeiro de 2013, pp. 81-86.
- [20]S.Saagari, P.DeviAnusha, Ch.LakshmiPriyanka, VSSNSailaja, "Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Essential Elements to SupportDecision-Making Process in Industries", International Journal of Innovative Technology and Exploring Engineering, Volume -2, Edição-6, maio de 2013, pp. 88-93.
- [21] Dhanamma Jagli e Akanksha Gupta, "Clustering Model for Evaluating SaaS on the Cloud", International Journal of Application or Innovation in Engineering & Management, Volume 2, Edição 12, dezembro de 2013.
- [22] Vaibhav R. Bhedi, Srinivas P. Deshpande e Ujwal A. Lanjewar, "Arquitetura de Data Warehouse para Institutos Financeiros para se Tornar um Sistema Financeiro Básico Integrado Robusto usando BUID", Jornal Internacional de Pesquisa Avançada em Engenharia de Computação e Comunicação, vol. 3, edição 3, março de 2014, pp. 5652-5656.
- [23]Sean Carlin, Kevin Curran, "Cloud Computing Technologies", International Journal of Cloud Computing and Services Science, Vol.1, No.2, junho de 2012, pp. 59-65.
- [24]Kuyoro SO, Ibikunle F. e Awodele O, "Questões e desafios de segurança da computação em nuvem", International Journal of Computer Networks, vol. 3, Edição 5, 2011, pp. 247-255.
- [25] Vikram Jeet Singh, Dharam Pal Singh e Dr. KishoriLalBansal, "Arquitetura proposta: Rede de recuperação de informações médicas baseada em nuvem", Jornal internacional de ciência da computação e tecnologia de engenharia, vol. 4 nº 5, maio de 2013, pp. 485-496.
- [26] Pete Manca. "O futuro da recuperação de desastres como serviço", Disaster Recovery Journal, vol. 28, Número 1, 25 de setembro de 2012. [<http://www.drj.com/articles/online-exclusive/the>

Conferência Internacional de Circuitos, Potência e Tecnologias de Computação de 2015 [ICCPCT]

future-of-disaster-recovery-as-a-service.html] em 31 de janeiro de 2015.

[27] Jim Kaskade, "What Does HA in the Cloud Mean?", 15 de agosto de 2009. [<http://jameskaskade.com/?p=448>] como em 31 Janeiro de 2015.

[28] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica e Matei Zaharia, "A View of Cloud Computing", Comunicações da ACM , abril de 2010, vol. 53 Nº 4, pp. 50-

58.[[http://cacm.acm.org/magazines/2010/4/81493-a-view-of cloud-computing/fulltext](http://cacm.acm.org/magazines/2010/4/81493-a-view-of-cloud-computing/fulltext)] em 31 de janeiro de 2015.

[29] Ohad Flinker, "Welcome to The NinesObserver – Your Guide to Cloud Uptime", 9 de março de 2014.

[<https://www.cloudendure.com/blog/welcome-to-ninesobserver/>] em 31 de janeiro de 2015.

[30] Vantagens e desvantagens da computação em nuvem, [[http://www.levelcloud.net/advantages-and-disadvantages-of cloud-computing/](http://www.levelcloud.net/advantages-and-disadvantages-of-cloud-computing/)] em 31 de janeiro de 2015.