Análise de um sistema de backup/recovery para grandes volumes de dados

João Messias Alves da Silva
joao.messias@cptec.inpe.br
INPE/CPTEC

Sylvio Villas Boas Neto
sylvio.neto@cptec.inpe.br
INPE/CPTEC

Eugênio Sper de Almeida
eugenio.almeida@cptec.inpe.br
INPE/CPTEC

Resumo: Um sistema de backup/recovery é fundamental para qualquer instituição, principalmente aquelas que recebem e processam grandes volumes de dados. Este trabalho teve como objetivo analisar um sistema de backup/recovery para grandes volumes da dados. Iniciou-se revendo os conceitos de backup e em seguida apresentou-se a implementação do sistema a ser analisado. Analisando a implementação, verificou-se que os negócios da instituição podem ser afetados devido a falhas. Esta análise tem a importância de permitir que este projeto seja revisto e aprimorado.

Palavras Chave: backup - recovery - volume de dados - meteorologia - EMC Networker

1. INTRODUÇÃO

O grande volume de dados produzido diariamente por instrumentos e por modelos computacionais tendem a ter vida longa e estar publicamente acessíveis com o objetivo de análise contínua (HEY et al., 2009). Um dos grandes produtores de dados são os centros meteorológico, cujo objetivo é melhorar a observação, a compreensão e a previsão do comportamento da atmosfera.

Disseminar esse conhecimento de forma rápida e ampla, e orientar a comunidade na sua utilização permite que um grande número de decisões diárias sensíveis as condições meteorológicas possam ser feitas no melhor interesse do bem-estar econômico e social do país, em todas as esferas da vida (MASON, 1996).

Os efeitos dos fenômenos meteorológicos observados ou previstos devem ser melhor explorados em seus aspectos favoráveis, e atenuados ou evitados em seus piores efeitos. Atualmente diversos setores da sociedade usufruem das informações meteorológicas disponibilizadas a um custo insignificante e sua aplicabilidade inclui: comércio, turismo, agricultura, transporte, setores relacionados a energia, desastres naturais, etc..

A rotina operacional de um centro meteorológico consiste da coleta e assimilação de dados observacionais meteorológicos; da recepção, processamento e distribuição de imagens provenientes de satélites meteorológicos; da execução de modelos atmosféricos em clusters de *High Performance Computing* (HPC) ou supercomputadores; e da geração e disseminação de previsões meteorológicas. Esses dados meteorológicos também são utilizados por pesquisadores no estudo do comportamento da atmosfera e na melhoria dos modelos atmosféricos.

O tamanho dos dados recebidos e produzidos varia de alguns KB até dezenas de GB. A grande maioria dos dados é não estruturada, no entanto existem informações importantes armazenadas de forma estruturada. Esses dados são de extrema importância nos processos operacionais e em pesquisas na área de meteorologia.

Normalmente um centro meteorológico possui um Datacenter onde encontra-se sua infraestrutura computacional. No entanto, os dados podem ser destruídos, alterados ou perdidos devido a erros das pessoas, falhas de software, empregados descontentes, hackers e destruição de equipamentos devido a falta de segurança física.

Qualquer uma dessas ameaças impacta na Segurança da Informação de qualquer instituição, afetando a Confidencialidade, a Integridade e a Disponibilidade da informação (ISO/IEC, 2005). Preservar e assegurar esta informação requer tecnologias para salvaguardá-la, e restaurá-la em caso de danos acidentais ou propositais.

Sistemas de *backup/recovery* são implementados primordialmente para salvaguardar informações e fornecer uma recuperação efetiva em caso de desastres. A implementação de sistema de *backup/recovery* resulta em um baixo impacto na continuidade dos negócios.

A motivação de salvar estes dados encontra se justamente na necessidade de preservar a qualidade dos informações coletadas, produzidas e distribuídas pelo centro meteorológico, visando garantir que os serviços e informações estejam sempre disponíveis a toda sociedade. Desta forma este trabalho tem como objetivo analisar e avaliar o sistema de *backup/recovery* de um Datacenter meteorológico

2. CONCEITOS DE BACKUP

O objetivo do *backup* é armazenar a informação de uma instituição ou empresa (com processo de negócios baseados em Tecnologia da Informação (TI)), gerando uma cópia fiel de seus dados essenciais, devendo ser criada e retida visando recuperação em casos de perdas ou

desastres. Conceitos e estratégias previamente definidas devem ser seguidos em todas as operações de backup.

Define-se como *Recovery Time Objetive* (RTO) a quantidade máxima de tempo que um processo de negócios baseados em TI pode estar indisponível antes do início de consequências inaceitáveis a uma organização (perdas financeiras, impacto na satisfação do cliente, reputação, etc.). *Recovery Point Objetive* (RPO) é a quantidade máxima de dados que se pode perder, antes de causar danos prejudiciais para a organização (PING et al., 2010).

O tempo de retenção é o período de tempo em que os dados devem ser mantidos intactos antes da exclusão ou passar para outro nível de armazenamento para fins de arquivamento. A janela de backup é o período de tempo adequado para executar um procedimento de backup, sem prejuízo da aplicação (ISMAIL et al., 2013).

O backup total (*full backup*) consiste na cópia de todos os dados especificados (área) em um determinado momento. Possui duas desvantagens: a leitura e escrita de todo o sistema de arquivos é lenta, e o armazenamento de uma cópia do sistema de arquivos consome significativa capacidade da mídia de backup (CHEVERNAK et al., 1998).

O backup incremental (*incremental backup*) copia os dados que mudaram após o último backup, sendo mais rápidos e menores. Sua vantagem é reduzir o tamanho dos backups, uma vez que apenas uma pequena porcentagem de arquivos mudam em um determinado dia (CHEVERNAK et al., 1998).

No backup diferencial ou acumulativo apenas os dados modificados desde o último backup total são copiados. O processo de recuperação dos dados inicia restaurando o último backup total e em seguida a restauração do backup diferencial apropriado (CHEVANANCE, 2004).

O backup sintético (*synthetic backup)* visa diminuir o tempo de *recovery e reduzir o* impacto no tráfego de dados na rede e no cliente durante a realização do *backup*. Ele é composto pelo último *backup full* e pelos incrementais subsequentes que já existem em fita ou disco (VERAS, 2010).

O backup pode ser classificado com relação ao estado da aplicação: *hot* (quente) *e cold* (frio). No processo de *cold backup* (*off-line backup)* a aplicação encontra-se inativa, enquanto que no *hot backup* (on-line backup), a aplicação está em execução com os usuários acessando os dados (MCDOWALL, 2011). A disponibilidade do serviço não é comprometida no *hot backup*, no entanto existe uma degradação do desempenho (VERAS, 2010).

Um backup pode ser realizado por imagem (ou bloco), arquivo ou aplicação. O backup por Imagem consiste do *backup* bloco por bloco do conteúdo do disco, sendo mais rápido para recuperar um sistema em caso de desastre. O backup por arquivo visa arquivos e diretórios, permitindo a recuperação de arquivos individualmente. No caso de backup por aplicação, existem APIs personalizadas para a realização do backup de determinados aplicativos (VERAS, 2010).

Segundo Furtado et al. (2002), a política de backup é um documento que informa quais dados são salvos, sua periodicidade, a forma de realização dos backups e como os testes são executados periodicamente. As cópias de segurança dos dados permitem que as empresas retornem à operação normal de suas atividades com maior rapidez após uma catástrofe ou ataque de alguma ameaça do mundo virtual.

Na política de backup deve estar definido o RPO, RTO, janela de backup e tempo de retenção de dados. Adicionalmente, a granularidade do backup (total, incremental, diferencial e sintético) devem ser estabelecidas para os dados que se deseja salvar, assim como o método utilizado (*hot* e/ou *cold)* e os níveis de backup (imagem, arquivo ou aplicação). As

configurações e estratégias devem ser decididas em função das necessidades do negócio e do RPO/RTO definidos.

A ausência de planejamento ou a incorreta definição de políticas e procedimentos de backup podem ter efeitos indesejados para todos os agentes dos sistema, como por exemplo o desempenho da rede. As janelas de backup devem sempre respeitar uma margem de segurança entre as operações, permitindo que imprevistos como uma operação de recuperação possa ser executada sem afetar outros processos.

Existem três estratégias de backup: Disk to Tape (D2T), Disk to Disk (D2D) e Disk to Disk to Tape (D2D2T). Na estratégia D2T, os dados são copiados do disco para fita de forma sequencial, degradando a eficiência e causando impacto no RPO e RTO. A estratégia D2D copia os dados de disco para disco e sua vantagem é produzir RPO e RTO próximos a zero. Na estratégia D2D2T os dados são copiados de disco para disco e posteriormente para fita (MAGRYŚ et al., 2011).

2.1. Arquiteturas de Backup e Deduplicação

Segundo Dharma et al. (2013), As arquiteturas de backup podem ser classificadas em tradicional e topologia de backup em Storage Area Network (SAN). As tradicionais são *Direct-Attached Backup* (DAB) e *LAN Based Backup* (LBB). As baseadas em SAN são *LAN-Free, Serverless* e Network Attached Storage (NAS) backup.

No *DAB* os dispositivos de armazenamento são conectados diretamente ao servidor. Sua vantagem é a rapidez do backup e o funcionamento simplificado. As desvantagens são a desorganização no armazenamento, custo elevado ao utilizar vários servidores com necessidade de várias unidades de fitas, dificuldade de compartilhar unidades de mídia e duplicação de dados semelhantes quando utilizado múltiplos servidores (NAIK, 2003).

No LBB, ou NAS *backup*, os servidores e os dispositivos de backup utilizam a LAN de forma compartilhada. Possui a vantagem da redução de custos. No entanto, as operaçoes de backup elevam o volume de dados na LAN, necessitando segregar este tráfego de backup em segmento de LAN separado (NAIK, 2003).

Na *LAN-Free backup, ou SAN Based Backup (*SBB*),* o tráfego de dados utiliza a SAN. Desta forma, alivia o tráfego na LAN e a concorrência entre os aplicativos (NAIK, 2003). Recomendado para backups de aplicações críticas (24x7x365) com grandes quantidades de dados e exige uma SAN centralizada. Para reduzir de custos com licenciamentos podem ser reduzidos realizando parte do backup pela SAN e parte pela LAN (VERAS, 2010).

O *Serverless Backup,* ou *Server-Free Backup ,* utiliza os recursos da SAN para mover os dados do disco para o dispositivo de backup. São chamados *servless* por não necessitarem recursos dos servidores para realizar o movimento dos dados de *backup*. geralmente um software especifico realiza a cópia dos dados para um dispositivo de backup (VERAS, 2010).

O NAS backup realiza o backup dos dados de um dispositivo NAS, que é um dispositivo de armazenamento conectado diretamente na LAN. O Network Data Management Protocol (NDMP) é um protocolo projetado para realizar backup de dispositivos NAS. Possui um gerenciamento centralizado e permite a separação do fluxo de dados do gerenciamento dos dados.

A deduplicação é uma técnica de compressão de dados onde os dados duplicados são eliminados, mantendo uma referência única ao invés de se armazenar todos os dados idênticos (JUNIOR, 2011). Sua vantagem é a economia no armazenamento, menos dependência de backup em fita, recuperação mais rápida após um desastre e maior rapidez na operação de backup (VERAS, 2010). Hoje a tecnologia de deduplicação é bem aceita no ambiente de

datacenters, backups de máquinas virtuais e escritórios remotos (DUBOIS E AATRUDA, 2010).

3. EMC Networker

O *EMC Networker* é um software que executa operações de *backup*/recuperação dados em fitas e discos, replicação contínua, e deduplicação em ambientes físicos e virtuais (DUBOIS E AMATRUDA, 2010). Ele é composto principalmente por cinco componentes: NetWorker Client, NetWorker Storage Node, NetWorker Server, NetWorker Management Console e NetWorker License Manager (EMC CORPORATION, 2007a).

Os *Networker Clients* (clientes de backup) são usualmente servidores de dados ou de aplicações. Sua função é coletar os dados a serem salvos, enviá-los ao NetWorker Storage Node (servidor de armazenamento) e recuperar os dados durante um processo de recuperação.

O *NetWorker Storage Node* possui a função de organizar os dados provenientes dos clientes de backup e gravá-los no dispositivo de backup. Os nós de armazenamento enviam os metadados sobre os dados gravados no dispositivo de backup ao servidor de backup.

O *NetWorker Server* (servidor de backup) direciona e apoia os clientes de backup e as operações de recuperação de dados. O servidor de backup recebe os metadados de backup dos clientes de backup e dos nós de armazenamento, armazenando-as no catálogo de backup.

O NetWorker possui funcionalidades que garantem alta desempenho e escalabilidade. Ele fornece uma plataforma comum com amplo suporte de opções para proteção de dados, backup em disco, replicação contínua, e deduplicação em ambientes físicos e virtuais (DUBOIS E AMATRUDA, 2010).

Os Storage Nodes funcionam como servidores secundários de backup, compartilhando a carga de trabalho, é uma boa prática configurá-los diretamente em uma SAN, aliviando o fluxo pela LAN.

O *Backup to disk,* permite o backup diretamente em disco, aumentando o desempenho, além de recursos de clonagem de disco para fita com maior rapidez. *Backup to tape*, suporte para variados tipos de fitas, unidades e bibliotecas, utiliza o recurso OTF (*Open Tape Format*) que é uma estrutura que permite dados heterogêneos residir na mesma fita, garantindo a portabilidade entre diversas plataformas de fitas. *Snapshot backup,* suporte para tecnologia de replicação baseada em array e em softwares de criação e uso de snapshots.

O *Block-based backup*, atráves de um módulo adicionável, o *SnapImage*, é possível a realização de backup e recuperação de dados a nível de bloco. *NDMP-based backup*, suporte ao protocolo NDMP (Network Data Management Protocol). *Deduplication com EMC Avamar*, utilizado junto com o Avamar, realiza a deduplicação na origem dos dados, diminuindo a carga do tráfego pela LAN.

A *Interface Networker Management Console* (Fig. 1) possui cinco (5) botões: E*vents, Enterprise, Libraries, Reports* e *Setup* (EMC CORPORATION, 2007b): Estes botões apresenta novas janelas na *Interface Networker Management Console* .

A janela *Events* fornece informações sobre os eventos gerenciados. Através da janela *Enterprise* é possível selecionar um servidor NetWorker para gerenciar, e monitorar o servidor e seus clientes de backup. A janela *Enterprise* permite abrir a janela *Administration* para acesso a um servidor NetWorker.

A janela *Libraries* apresenta um resumo informativo e gerencia as bibliotecas para todos os servidores *Networker*. A janela *Administration* também pode ser aberta a partir desta localização. A janela *Reports* permite a configuração e visualização de relatórios. Na janela *Setup* existem funções de controle administrativo: gerenciamento de usuários e de licenças.

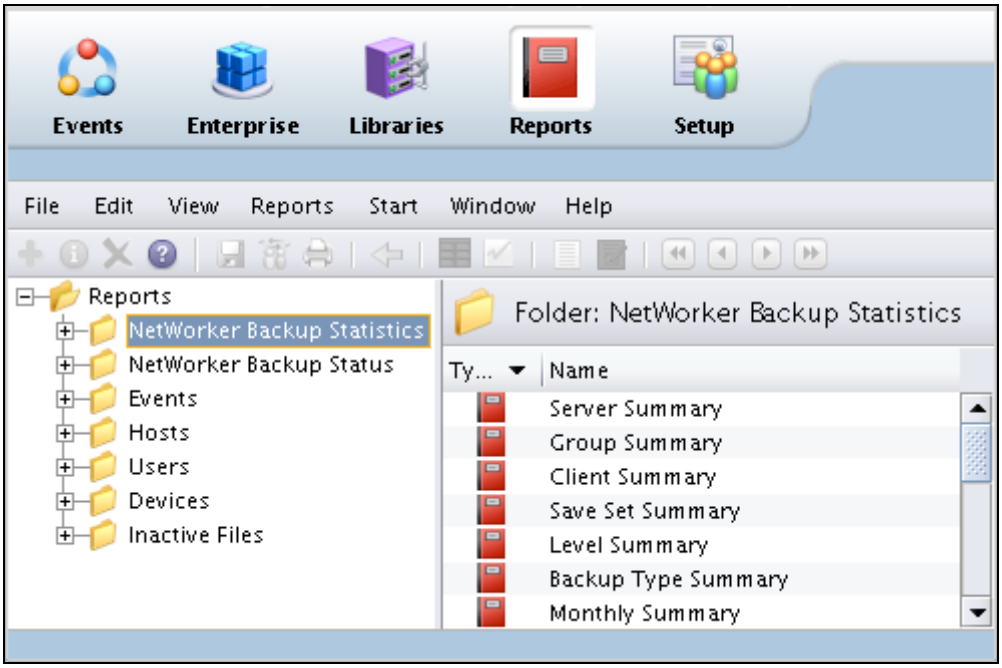

Figura 1 - Interface *Networker Management Console*

A Interface *Networker Administration* (Fig. 2) é acionada pelo botão Administration da Interface *Networker Management Console.* Seus quatro (4) botões permitem acesso as janelas *Monitoring, Configuration*, *Devices e Media*.

As diversas atividades relacionadas ao servidor NetWorker são monitoradas através da janela *Monitoring*, como o progresso de um backup programado e visualização de alertas. O servidor NetWorker e seus recursos são gerenciados através da janela *Configuration*, como clientes, agendamentos de backup e políticas.

A janela *Devices* permite a inclusão, configuração e operação de um ou mais dispositivos, bibliotecas e silos no servidor NetWorker. A janela *Media* possibilita o gerenciamento das atividades e recursos relacionados aos volumes de backup.

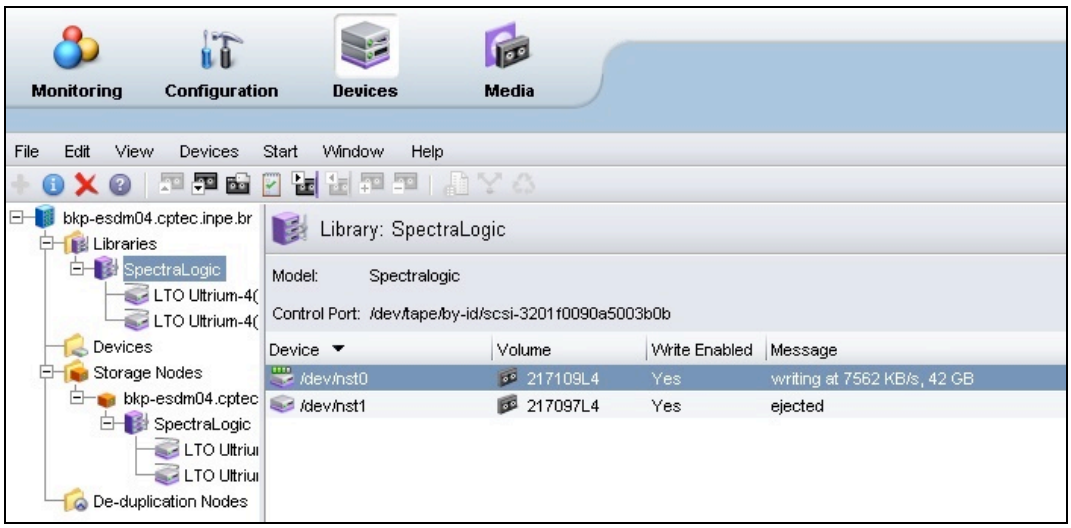

Figura 2 - Interface *Networker Administration*

A Fig. 3 exibe os grupos de *backup* atualmente definidos e o *status* de sua última execução na interface *EMC Networker Administration.* Os estados sucesso (verde), falha (vermelho) e backup abortado inesperadamente são apresentados na primeira coluna. O tempo de duração de todos os dados de um determinado grupo de backup encontra-se na coluna *duration*.

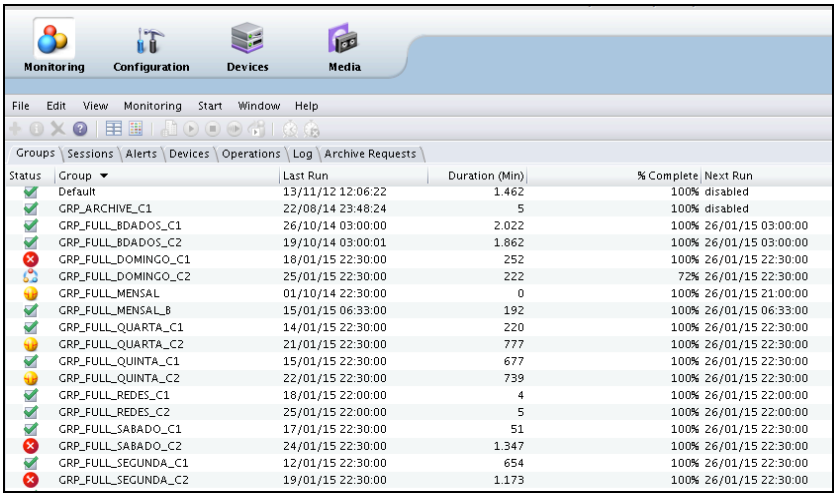

Figura 3 - Interface do EMC Networker Administration

4. Cenário atual do backup

A infraestrutura de backup (Fig. 4) analisada consiste dos servidores que terão seus dados salvaguardados *(backup clients)*, interligados através de uma rede local (LAN) ao servidor de backup *(backup server)*/servidor de armazenamento *(storage node)*. O dispositivo de armazenamento de backup (*backup storage device*) interconecta-se através de uma rede de armazenamento (SAN) ao servidor de backup/servidor de armazenamento.

As licenças servidor e cliente do software EMC NetWorker, versão 7.4 SP5, encontram-se instalados nesses servidores, que está homologado pela EMC para os sistemas operacionais *Microsoft Windows*, *GNU/Linux, Unix* e *Mac OS*.

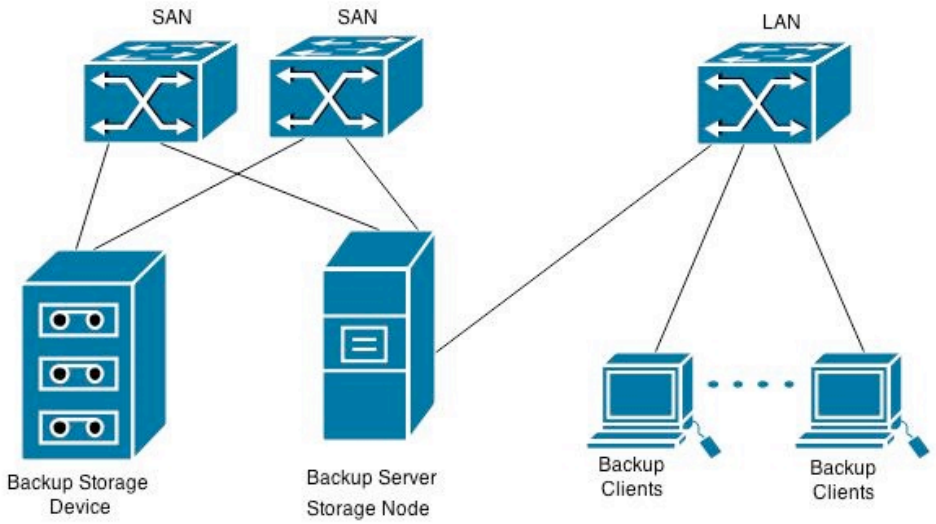

Figura 4 - Infraestrutura de backup

As funções de servidor de armazenamento e de backup estão disponibilizadas em um único servidor, com as seguintes características: dois processadores *AMD Opteron™ Quad-Core* (oito cores), 15 GB de memória, 127 GB de disco rígido (HD) e Sistema Operacional (S.O.) *Suse Enterprise Server 10 SP2*. Sua função é definir as políticas globais de retenção, inserir e retirar áreas de backup, horários de execução, e gerenciar todos os clientes instalados. Diversos dados estatísticos são disponibilizados, podem ser apresentados por grupos, cliente específico ou períodos estabelecidos.

Cada cliente de backup possui um agente do *NetWorker* instalado. Este agente se comunica diretamente com o servidor de armazenamento, na operação de backup envia os dados de origem e no *recovery* os dados são recuperados da biblioteca de fita.

[redacted]

Duas unidades (drives) da biblioteca de fitas *(tape library) Spectra® T-Finity's (Spectra Logic)* estão disponíveis para backup. Nesta biblioteca estão reservadas 2000 fitas LTO-4 com capacidade de armazenamento individual de 800 GB, sem compressão, totalizando 1,6 PB de espaço máximo disponível para o backup corporativo. As fitas podem ser sobrescritas, conforme a necessidade e o tempo estipulado de retenção dos dados.

Os backups da rede interna e rede externa (DMZ) utilizam a estratégia D2T, sendo configurados 29 áreas para backup *(save sets)* que estão distribuídos por 23 servidores clientes. A volumetria dos dados para backup é de aproximadamente 3.35 TB. Nenhuma opção de deduplicação encontra-se ativa, sendo que a volumetria original da área é a mesma transferida para as fitas.

As políticas utilizadas foram *backup full (semanalmente)* e *incremental* (diariamente). O *backup full* inicia às 20h30 (GMT) e o incremental por volta das 1h (GMT). Os grupos de backup são divididos em dois conjuntos (C1 e C2) e cada um deles estão organizados por um *pool* de 10 fitas.

Os conjuntos diferem no dia de sua inicialização. O grupo de *backup full C1* está agendado para iniciar em toda segunda-feira da primeira e terceira semana do mês. O grupo C2 está configurado para começar na segunda e quarta semana do mês. Desta forma, ambos os conjuntos são complementares. Um determinado número de clientes é atribuído a cada grupo de backup.

O tempo de retenção da maior parte dos dados é de quinze dias. Algumas exceções incluem a área home dos usuários do supercomputador, com tempo de retenção de seis meses., O sistema de backup libera a fita ao término do tempo de retenção, podendo a mesma ser reutilizada.

A restauração dos dados é executada a partir do cliente através do comando *recover*. O agente instalado no cliente se conecta ao servidor do *EMC Networker* que recupera os arquivos selecionados das fitas. Esses dados podem ser transferidos para o local original ou outro definido pelo operador do sistema de backup.

O tempo de restauração dos dados está diretamente relacionado a quantidade de dados salvo, a divisão dos dados nas fitas e a taxa de transferência da LAN. Em caso de desastre com a máquina cliente, também é possível a realização da restauração dos dados a partir de um outro cliente.

5. ANÁLISE

A implementação de backup analisada não considerou a quantidade de dados a serem salvos e não houve uma análise posterior do RPO e RTO. Apenas preocupou-se em salvar as informações importantes para a rotina operacional.

Todos os dados foram tratados como não estruturados, inclusive os armazenados em banco de dados. O backup dos dados é realizado em diretórios pré-determinados, copiando os arquivos fielmente de sua origem para fita da biblioteca de fitas, sem sofrerem qualquer tipo de tratamento ou classificação.

A política de backup completa foi parcialmente descrita. Como consequência, é o os usuários desconhecem as informações salvas e seus RPOs e RTOs.

O software *EMC Networker* oferece informações sobre o processo de backup, permitindo acompanhar a taxa de crescimento dos dados de backup e da janela de backup. Definiu-se que a frequência dos relatórios seja mensal, referentes a duração total do backup no mês com os números de sucesso e falhas, e a evolução da quantidade de dados para cada dia da semana (GB).

O gráfico da Figura 5 mostra a distribuição semanal da quantidade de dados em GB (volumetria) copiadas para a fita na operação de *full backup* (ao longo da semana em grupos diários).

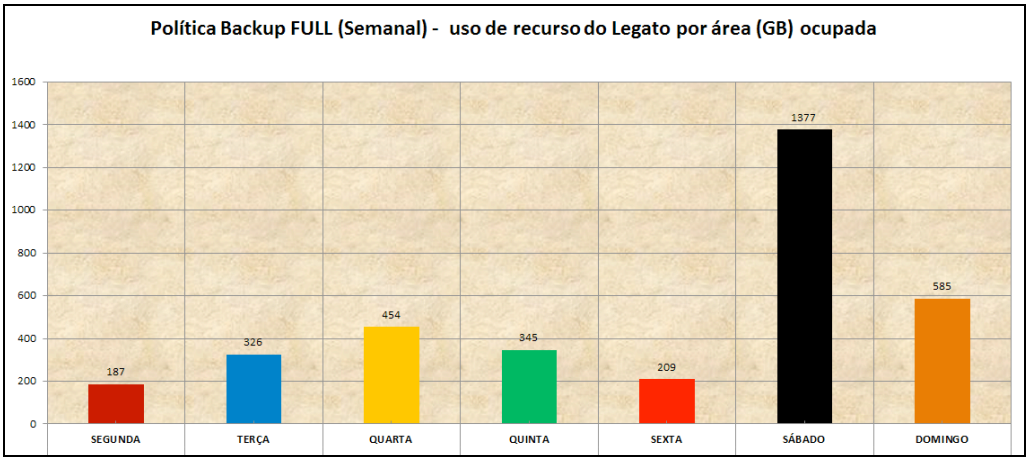

Figura 5 – Tamanho dos *backups full* (média de setembro de 2014)

A Figura 6 apresenta o gráfico com a distribuição semanal do tempo de execução (em horas) do backup. Observa-se um alto RPO, tem um impacto significante no restabelecimento dos dados.

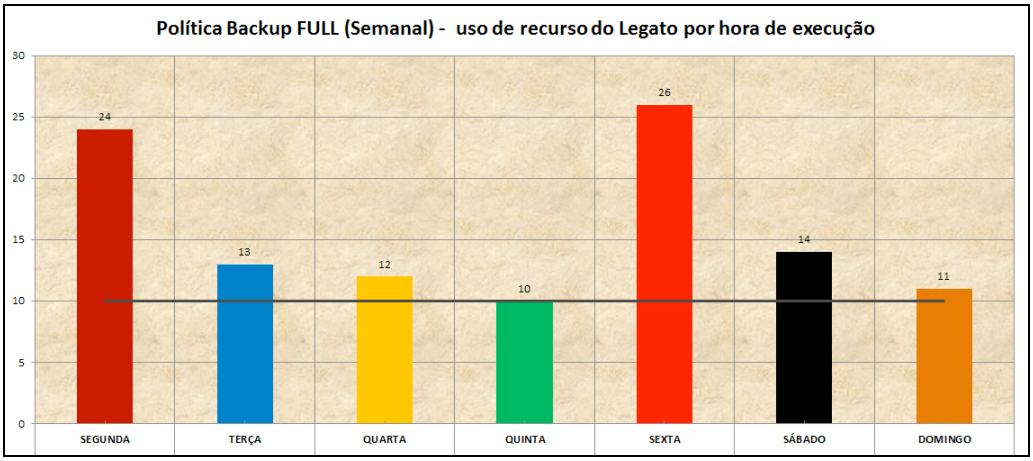

Figura 6 - Tempos de execução dos *backup full* (média de setembro de 2014)

A janela de backup varia de dez horas a 26 horas, com efeitos na consistência dos dados e no restabelecimento dos negócios. Na segunda e na sexta-feira, os tempos de backup de 24 e 26 horas afetam outros grupos de backup, pois concorrem com os jobs agendados.

Comparando os dois gráficos, nota-se claramente que não existe relação entre volume dos dados de uma área e o tempo de conclusão do backup. No grupo de sexta-feira temos um total de 209 GB e no de segunda 187 GB (os menores tamanhos da semana), entretanto são os dias com o pior desempenho de backup.

O baixa desempenho do backup pode ser atribuído ao tráfego de dados elevado na LAN, discos lentos, S.O. desatualizado e arquivos muito fragmentados. Segundo Somasundaram e Shrivastava (2011), a lentidão do backup e da recuperação pode ser atribuído ao grande número de arquivos pequenos em um sistema de arquivos, uma vez que um grande número de entradas no sistema torna lenta a pesquisa em todo o sistema de arquivos.

A estratégia D2T não é recomendada para backups operacionais, com dados sendo gravados frequentemente e com tempo de retenção relativamente curto, um vez que a prioridade são valores baixos de RPO e de RTO. O uso da estratégia D2T é recomendada para

armazenamento externos onde se deseja arquivar arquivos por períodos longos, necessitando de ambientes controlados para evitar que as fitas se danifiquem (SOMASUNDARAM E SHRIVASTAVA, 2011).

6. CONCLUSÕES

A velocidade e volume de dados capturados e gerados pelos laboratórios que executam aplicações científicas aumenta dia a dia. Salvaguardar esses dados, garantindo sua recuperação em caso de danos é parte das atividades das equipes de TI.

O objetivo deste trabalho é analisar o projeto de backup de um centro meteorológico, uma vez que um projeto de backup incorreto e/ou desatualizado pode impactar nos negócios de uma organização.

As saídas da execuções diárias dos modelos numéricos de previsão de meteorológica e ambiental são uma das origens da geração de informações meteorológicas. Esse dados modelos são produzidos quatro vezes ao dia.

Um alto RPO prejudica os resultados acumulados destas previsões de tempo futuras e um alto RTO compromete o restabelecimento de algum serviços críticos em caso de falha, estes serviços essenciais incluem os alertas para a defesa civil, centro de desastres naturais, órgãos do governo, produtos pagos, atualização do website do centro e impactos administrativos como serviços de e-mail e documentações regulamentadoras.

A estratégia D2T utilizada causa altos valores de RPO e RTO, pois a movimentação sequencial da fita e o uso de várias fitas para uma mesma área torna demasiadamente lento o processo de backup e recuperação. Outra desvantagem é uma janela de backup longa, que afeta a consistência de dados já armazenados e compromete o restabelecimento da última informação gerada.

Uma política de backup completa e amplamente debatida com os demais departamentos estratégicos, torna-se um instrumento seguro e padroniza as ações da equipe técnica de TI.

6. REFERÊNCIAS

CHEVANCE, R. J. Server architectures: Multiprocessors, clusters, parallel systems, web servers, storage solutions. Digital Press, 2004.

CHEVERNAK, A.; VELLANKI, V.; KURMAS, Z. Protecting file systems: A survey of backup techniques. In Joint NASA and IEEE Mass Storage Conference, 1998.

DHARMA R.; SAKE, S.; MANUEL, M. Backup and Recovery in a SAN. EMC^2 Techbooks, 2013.

DUBOIS, L.; AMATRUDA, R. Backup and Recovery: Accelerating Efficiency and Driving Down IT Costs Using Data Deduplication, *EMC Corporation,* 2010.

EMC Corporation, EMC NetWorker Release 7.4 Service Pack 1 Multiplatform Version Installation Guide, Dec, 2007a

EMC Corporation, EMC NetWorker Release 7.4 Service Pack 1 Multiplatform Version Administration Guide, Dec, 2007b

FURTADO, V. Tecnologia e Gestão da Informação na Gestão Pública.Rio de Janeiro: Garamond, 2002.

ISMAIL, B. I.; MYDIN, M.; NIZAM, M.; KHALID, M. F. Architecture of scalable backup service for private cloud. In Open Systems (ICOS), 2013 IEEE Conference on (pp. 174-179), 2013.

HEY, T.; TANSLEY, S.; TOLLE, K. The Fourth Paradigm: Data-Intensive Scientific Discovery, Redmond, WA: Microsoft Research, 2009.

ISO/IEC 27002:2005 Information Technology-Security Techniques-Code of Practice for Information Security Management, 2005

JUNIOR, M. F. Uso da técnica de deduplicação para armazenamento de dados biológicos em storage (Doctoral dissertation, Universidade de Brasília), 2011.

MAGRYŚ, M.; POGODA, M.; SUŁKOWSKI, G.; TWARDY, M.; WINIARCZYK, P. Snapshot backup system in distributed computing environments-problems, solution, results. Bio-Algorithms and Med-Systems, 7, 2011.

MASON, B. J. The role of meteorology in the national economy. Weather, 21(11), 382-393, 1966.

MCDOWALL, R. D. Computer (In) security–2: computer system backup and recovery. The Quality Assurance Journal, 5(3), 149-155, 2011.

NAIK, D. C. Backup and Restore - Technologies for Windows, 2003. Disponível em: http://www.informit.com/articles/article.aspx?p=99985, acessado em: 10 fev. 2015.

PING, Y.; BO, K.; JINPING, L.; MENGXIA, L. Remote disaster recovery system architecture based on database replication technology. In Computer and Communication Technologies in Agriculture Engineering (CCTAE), 2010 International Conference On (Vol. 1, pp. 254-257), 2010.

SOMASUNDARAM, G.; SHRIVASTAVA, A. Armazenamento e Gerenciamento de Informações: como armazenar, gerenciar e proteger informações digitais. Porto Alegre. Ed. Bookman, 2011.

VERAS, M. Datacenter: componente central da infraestrutura de TI. Rio de Janeiro. Ed. Brasport, 2010.