UNIVERSIDADE DO VALE DO ITAJAÍ
CENTRO DE CIÊNCIAS TECNOLÓGICAS DA TERRA E DO MAR
CURSO DE CIÊNCIA DA COMPUTAÇÃO

IMPLANTAÇÃO DE UMA SOLUÇÃO DE ALTA DISPONIBILIDADE PARA O SISTEMA EMISSOR DE NOTAS FISCAIS DE SERVIÇOS DA PREFEITURA MUNICIPAL DE ITAJAÍ

por

Diogo Roedel

Itajaí (SC), dezembro de 2012

UNIVERSIDADE DO VALE DO ITAJAÍ
CENTRO DE CIÊNCIAS TECNOLÓGICAS DA TERRA E DO MAR
CURSO DE CIÊNCIA DA COMPUTAÇÃO

IMPLANTAÇÃO DE UMA SOLUÇÃO DE ALTA DISPONIBILIDADE PARA O SISTEMA EMISSOR DE NOTAS FISCAIS DE SERVIÇOS DA PREFEITURA MUNICIPAL DE ITAJAÍ

Área de Redes de Computadores

por

Diogo Roedel

Relatório apresentado à Banca Examinadora do Trabalho Técnico-científico de Conclusão do Curso de Ciência da Computação para análise e aprovação.
Orientador: Fabrício Bortoluzzi, M.Sc.

Itajaí (SC), dezembro de 2012

RESUMO

ROEDEL, Diogo. **Implantação de uma solução de alta disponibilidade para o Sistema Emissor de Notas Fiscais de Serviços da Prefeitura Municipal de Itajaí.** Itajaí, 2012. 90 folhas. Trabalho Técnico-científico de Conclusão de Curso (Graduação em Ciência da Computação) – Centro de Ciências Tecnológicas da Terra e do Mar, Universidade do Vale do Itajaí, Itajaí, 2012.

Nos dias atuais, empresas estabelecidas no município de Itajaí funcionam ininterruptamente, não havendo um horário regular definido para operações de faturamento ou expedição, desta forma o sistema emissor de notas fiscais de serviço da Prefeitura de Itajaí não pode ter sua operacionalidade comprometida, seja por uma hora, ou até por alguns minutos, dependendo do horário. Um pequeno período fora de operação pode representar grandes prejuízos. Uma estrutura redundante se faz necessária, a fim de diminuir os riscos e elevar a disponibilidade do sistema. Na tentativa de prover uma solução adequada de disponibilidade de dados para suas organizações, muitos gestores enfrentam sérios problemas para identificar qual a real necessidade de sua empresa. Um projeto mal dimensionado pode custar excessivamente caro ou não atingir aos níveis de disponibilidade pretendidos. Uma mínima diferença no percentual de disponibilidade pretendida pode elevar consideravelmente os custos do projeto, em alguns casos inviabilizando sua implantação. Este projeto descreve os principais componentes do ambiente computacional da Prefeitura de Itajaí tornando-o altamente disponível, citando as tecnologias envolvidas e recomendações para a implantação de uma solução de sucesso. A implantação do projeto consistiu em detectar todos os pontos únicos de falha, formulando propostas par eliminá-los, através da redundância de recursos, gerenciamento das falhas, replicação dos dados entre dois servidores com sistema operacional Linux. Caso ocorrer alguma grande falha ou desastre que inviabilize o servidor primário, um servidor secundário assumirá suas tarefas assumindo a disponibilidade da aplicação sem a necessidade de intervenção humana. O desenvolvimento do projeto baseou-se em pesquisas na Internet, manuais de fabricantes de soluções, livros e artigos específicos sobre o assunto.

Palavras-chave: Alta Disponibilidade. Redundância. Recuperação de Desastres.

ABSTRACT

Nowadays, companies located in the city of Itajai operate continuously, without a regular schedule set for billing operations or dispatch, so the system of issuing invoices to service the City of Itajai can not have his operation compromised, either by an hour, or even for a few minutes, depending on the time. A short period out of operation may represent large losses. A redundant structure is necessary in order to reduce risks and increase system availability. In an attempt to provide an adequate solution to data availability for their organizations, many managers face serious problems to identify what the real needs of your business. A poorly sized project may cost too expensive or not achieving the desired levels of availability. A minimal difference in the percentage of availability desired can considerably raise project costs, in some cases impeding its implementation. This project describes the main components of the computing environment of the City of Itajai making it highly available, citing the technologies involved and recommendations for implementing a successful solution. The implementation of the project was to detect all single points of failure, formulating proposals pair eliminate them through redundancy features, management failures, replication of data between two servers with Linux operating system. Should be some great disaster or failure that could prevent the primary server, a secondary server assumes its duties by assuming the availability of the application without the need for human intervention. The development project was based on Internet searches, manufacturers solutions manuals, books and articles on the subject specific.

Keywords*: High availability. Redundancy. Disaster Recovery.*

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE EQUAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

AS	Automated System
ARP	Address Resolution Protocol
BGP	Border Gateway Protocol
CARP	Common Address Redundancy Protocol
GE	General Electric
HA	High Availability
HVAC	Heating, Ventilation, and Air Conditioning
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISS	Imposto Sobre Serviço
IGP	Interior Gateway Protocol
ISO	International Organization for Standardization
ISP	Internet Service Provider
LAN	Local Area Network
MAC	Media Access Control
MTBF	Mean Time Between Failures
MTTF	Mean Time to Fail
MTTR	Mean Time to Repair
NFSE	Nota Fiscal de Serviço Eletrônica
OSI	Open Systems Interconnection
OSPF	Open Shortest-Path First
PDU	Power Distribution Unit
PIB	Produto Interno Bruto
PMI	Prefeitura Municipal de Itajaí
PUF	Ponto Único de Falha
RIP	Routing Information Protocol
TCP	Transmission Control Protocol
TI	Tecnologia da Informação
TTC	Trabalho Técnico-científico de Conclusão de Curso
UNIVALE	Universidade Vale do Rio Doce
UNIVALI	Universidade do Vale do Itajaí
UPS	Uninterrupted Power Supply

SUMÁRIO

1 INTRODUÇÃO

A Nota Fiscal de Serviço Eletrônica (NFSe), foi instituída no município de Itajaí no final do ano de 2011 e teve início de sua utilização na primeira semana do ano de 2012. Trata-se da emissão de documentos em formato digital, processados em sistemas de informação, sob responsabilidade da administração municipal. Deve se emitida na prestação de serviços substituindo aos documentos impressos antes tradicionalmente utilizados em blocos. (PMI, 2012).

Ao final do mês de junho de 2012, totalizavam duas mil quinhentas e nove empresas que faturaram seus serviços através da emissão de NFSe em Itajaí. Entre essas encontram-se o Porto Municipal de Itajaí, APM Terminais e uma variedade de empresas no segmento de transportes e logística, que movimentam a economia de Itajaí, fazendo dela a primeira em comércio exterior e o segundo maior PIB (Produto Interno Bruto) do Estado de Santa Catarina. (SPG, 201?)

No primeiro semestre de 2012, foram emitidas aproximadamente duzentas e setenta e duas mil notas fiscais por meio do sistema emissor, representando uma arrecadação de R$ 35.130.412,66 (trinta e cinco milhões, cento e 130 mil, quatrocentos e doze reais e centavos) em impostos sobre serviços de qualquer natureza (ISS). Em média simples, o valor arrecadado divididos por seis meses, por trinta dias e por vinte e quatro horas, resultam em R$ 8.132 (oito mil cento e trinta e dois reais) de prejuízos por hora de indisponibilidade do sistema (PMI, 2012).

Justamente pela importância, alta utilização e o crescimento das emissões de NFSe, o sistema emissor deve estar preparado para processar inúmeras informações simultâneas com alta performance, baixo tempo de resposta e principalmente alta disponibilidade (HA - *high availability*).

Em tecnologia da informação, alta disponibilidade pode ser definida como um padrão de projeto que assegura manter um sistema de informação disponível ao longo do tempo, podendo utilizar mecanismos de detecção, recuperação e mascaramento de falhas, visando manter o funcionamento dos serviços durante o máximo de tempo possível, sem interrupções ou falhas, inclusive quando efetuando manutenções programadas. Também chamada de disponibilidade continuada, alta disponibilidade portanto, descreve uma variedade de ações e requisitos técnicos, a partir do hardware e infraestrutura, objetivando minimizar o tempo de

indisponibilidade dos sistemas, podendo ser mensurada relativa ao “100% operacional” ou “livre de falhas”. (AHLUWALIA; JAIN, 2006).

Nesse contexto, disponibilidade refere-se à capacidade de um usuário de determinado sistema acessar, incluir ou modificar os dados existentes, assegurando a integridade de quaisquer alterações realizadas em qualquer intervalo de tempo. Caso, por qualquer que seja o motivo, um usuário que não tenha acesso a todo ou parte fundamental desse sistema, é dito então que ele está indisponível, sendo o tempo total de indisponibilidade conhecido pelo termo *downtime*.

Em uma visão geral, o projeto de Alta Disponibilidade pode ser dividido em dois grupos, tolerância a falhas e gerencia de falhas (*Fault Tolerance* e *Fault Management*) (AHLUWALIA; JAIN, 2006).

a) Tolerância a falhas: Pode-se fazer um sistema ou parte dele tolerante a falhas, a partir da redundância dos recursos. Assume-se que para cada recurso haverá um ou mais idênticos exercendo a mesma função, mantendo o funcionamento do sistema em estado continuo e sem perdas ou atrasos.
b) Gestão de falhas: Ao detectar uma falha, o sistema procura por recursos disponíveis que possam atender a demanda daquele inoperante, repassando a carga de trabalho a um ou distribuindo entre mais recursos.

A proposta deste trabalho foi buscar a melhor combinação de recursos de alta disponibilidade para tornar o serviço de emissão de notas fiscais o mais disponível possível dentro de parâmetros a serem formalizados neste Trabalho Técnico-científico de Conclusão de Curso.

Em medição de disponibilidade realizada com base nos valores apresentados pelo virtualizador hospedeiro sistema emissor de notas fiscais, gerenciador UPS e concessionária de enlace externo, para o período do primeiro semestre de 2012, constatou-se a disponibilidade próxima a 99,85%. Isto é equivalente a 6 horas e 27 minutos que o sistema esteve indisponível aos seus usuários. Ocasionados por faltas no enlace externo, ausência de energia elétrica por período superior à autonomia do UPS e tempos de serviços para restauração.

A expectativa da Coordenação Tecnológica da Prefeitura Municipal de Itajaí é estabelecida sob a premissa básica de que o referido sistema não deveria parar por um tempo superior a 1 hora por mês. Por convenção estabeleceu-se, portanto, o objetivo de alcançar a medida de disponibilidade de 99,90%, que representa um tempo aceitável de paradas com

duração de 43 minutos em um período de 30 dias, com possibilidades de manutenção e intervenção humana. Fazendo deste projeto economicamente viável de implementação.

Um mínimo aumento de 99,90% para 99,99% de disponibilidade pretendida, conforme valores apresentados mais adiante na subseção 2.4.1 , diminuiria o período máximo aceitável de indisponibilidade para 4 minutos e 31 segundos. Não sendo tempo suficiente para restaurações de *backups*, desta forma todos os recursos devem ser duplicados, redundantes e todas as tarefas automatizadas, elevando consideravelmente o valor do projeto e desta forma inviabilizando-o.

Este trabalho justifica-se como Trabalho Técnico-Científico em Ciência da Computação porque exigiu do acadêmico a aplicação direta dos conhecimentos sobre *clusters*, redundância e balanceamento de carga obtidos na disciplina de Sistemas Distribuídos, enlaces, redes e aplicações nas duas disciplinas de Redes de Computadores e paralelismo e programação concorrente da disciplina de Sistemas Operacionais.

1.1 PROBLEMATIZAÇÃO

1.1.1 Formulação do Problema

Anteriormente não existia um plano de disponibilidade para os sistemas, que utilizavam equipamentos robustos e confiáveis, mas de forma recorrente ocorriam paradas não esperadas, algumas com tempo de restauração de minutos e outras levanto até horas, em pleno período crítico onde o sistema é mais exigido por seus usuários. As falhas e períodos de indisponibilidade são apresentados na Tabela 1 a seguir.

Tabela 1 - Disponibilidade e *downtime* apresentados nos seis primeiros meses de 2012

Período	**Falha**	**Tempo para restauração**	**Erro apresentado**	**Disponibilidade mês**	***downtime* mês**
Janeiro	Interrupção energia elétrica.	20 min.	Nenhuma, o suplemento de energia ininterrupto atendeu como fonte alternativa.	99,95%	20 min.
	Falha no enlace externo	20 min.	Indisponibilidade total do sistema aos usuários.		

Fevereiro	Falha no enlace externo	15 min.	Indisponibilidade total aos usuários.	99,96%	15 min.
Março	Interrupção energia elétrica.	5 horas	Indisponibilidade total por 2 horas e 40 min. O suplemente de energia ininterrupto atendeu por 2 horas e 20 min.	99,00%	340 min.
	Falha no enlace externo	2 horas e 20 min.	Indisponibilidade total aos usuários		
	Falha no comutador de núcleo	3 horas e 20 min.	Indisponibilidade total		
Abril	Falha no enlace externo	8 min.	Indisponibilidade total aos usuários	99,98%	8 min.
Maio	Nenhuma	-		100%	-
Junho	Falha no enlace externo	4 min.	Indisponibilidade total aos usuários	99,99%	4 min.
Total					387 min.

Somados todos os tempos de indisponibilidade do sistema (*downtimes*) totalizam aproximadamente 6 horas e 27 minutos. Ocasionados por falhas no enlace externo, ausência de energia elétrica por período superior à autonomia do UPS e tempo para reparação.

Baseando-se no tempo total de indisponibilidade, é possível calcular a disponibilidade global próxima aos 99,85%. Conforme possível observar na Tabela 1, os valores de disponibilidade apresentam variação, em função da atual estrutura não possuir recursos suficientes para garantir um tempo máximo aceitável para paradas.

Buscando por soluções comerciais, foi proposta a criação de um novo *datacenter backup*, replicando todos os recursos disponíveis da antiga estrutura, entre servidores, comutadores de rede, roteadores, UPSs, softwares de gerenciamento e sistema de armazenamento em rede, com valores superiores a um milhão de reais.

Na tentativa de prover uma solução adequada de disponibilidade de 99,90%, foi considerado o uso de tecnologias gratuitas, como exemplo o sistema operacional FreeBSD e protocolo CARP para implementar um roteador redundante, reconfigurando os recursos já disponíveis na Prefeitura de Itajaí, desta forma com a menor oneração possível.

1.1.2 Solução Proposta

Este trabalho propôs alterações no ambiente tecnológico, especificamente na disponibilidade do sistema emissor de notas fiscais da Prefeitura de Itajaí, baseando-se nos conceitos das tecnologias envolvidas e soluções semelhantes.

Esta solução visa propor um conjunto de técnicas, que aplicadas possam garantir uma disponibilidade global mínima de 99,90%, dando condições para estabelecer um acordo de nível de serviço a ser mantido.

Este trabalho tem por objetivo manter o Sistema emissor de notas fiscais da Prefeitura de Itajaí, em funcionamento contínuo, com paradas não superiores aos 43 minutos, sejam elas programadas ou não.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo geral deste TTC é empregar um conjunto de técnicas de alta disponibilidade que eleve a disponibilidade global do serviço de emissão de notas fiscais de serviços da Prefeitura Municipal de Itajaí.

1.2.2 Objetivos Específicos

O objetivo geral deste trabalho será alcançado através do cumprimento de quatro objetivos específicos:

1. Garantir uma disponibilidade global de no mínimo 99,90% para o sistema emissor de notas fiscais de serviço.

2. Pesquisar a combinação de recursos de alta disponibilidade mais adequada para atender às necessidades do serviço de emissão de notas fiscais da Prefeitura Municipal de Itajaí.

3. Implantar a combinação mais adequada conforme estabelecido no item anterior.

4. Aferir a disponibilidade obtida e sua aderência com a métrica definida no objetivo geral.

1.3 METODOLOGIA

Para atender as necessidades do projeto proposto, as atividades foram divididas e realizadas da seguinte forma:

- Definição do Tema: nesta etapa foram definidos assuntos relacionados ao tema, como tecnologias a serem estudadas, como o tema se aplicaria e como poderia de tornar um trabalho técnico científico;

- Definição de Introdução, Escopo e Objetivos: após definido o tema, foi necessário elaborar uma Introdução ao trabalho, que foi de ajuda também para melhor exemplificar a ideia central do trabalho e definir o escopo ao qual este projeto se propõe. Pesquisas quanto a modelos de Alta Disponibilidade e tecnologias relacionadas foram realizadas a fim de definir e chegar a um consenso do que seria o escopo do projeto;

- Estudos: levantamento bibliográfico foi realizado nessa etapa com o propósito de verificar o material necessário para a escrita da Fundamentação Teórica e de base para construção do projeto. Foram utilizados recursos como livros, dissertações e artigos. Após o levantamento do material necessário, deu-se inicio ao processo de leitura do mesmo, realizando anotações durante e escrevendo a Fundamentação Teórica em paralelo;

- Desenvolvimento: após a conclusão dos estudos, deu-se inicio ao capítulo referente ao desenvolvimento do projeto, levando em conta os objetivos definidos no inicio das atividades e as soluções já levantadas. Durante a escrita do Projeto, ainda foram realizadas leituras sobre tecnologias envolvidas e projetos similares.

1.4 ESTRUTURA DO TRABALHO

Este documento está estruturado em quatro capítulos. Em seguida, é feita uma breve descrição sobre eles.

O Capítulo 1, Introdução, apresentou uma visão geral do sistema emissor de notas fiscais de serviços da Prefeitura Municipal de Itajaí, a motivação para a escolha do tema. Foram definidos os objetivos e detalhado a problematização ao qual este trabalho se refere, além da metodologia aplicada ao projeto.

No Capítulo 2, Fundamentação Teórica, é apresentada uma revisão bibliográfica sobre temas relevantes ao projeto proposto, como objetos de aprendizagem, infraestrutura local, infraestrutura computacional, soluções acessíveis e um geral sobre alta disponibilidade.

O Capítulo 3 apresenta o Desenvolvimento detalhado da solução proposta, especificando-o textualmente e por meio de diagramas. Este capítulo também descreve os pontos únicos de falha no cenário antes existente, apontando as alterações necessárias para que fosse possível firmar um compromisso com a disponibilidade pretendida, e descreve o desenvolvimento apresentando detalhadamente dos principais prontos.

Concluindo, no Capítulo 4, apresentam-se as Considerações Finais, onde são apresentados os resultados finais e possíveis contribuições a trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, será revisada a fundamentação teórica, apresentando o conceito e definição das tecnologias abordadas ao longo deste trabalho, contemplando alta disponibilidade, técnicas para alcançar alta disponibilidade, pontos críticos de falha, principais causas de falhas e necessidades para se alcançar alta disponibilidade.

2.1 INFRAESTRUTURA LOCAL

A infraestrutura local pode ser compreendida como todos os recursos presentes em um ambiente físico, capaz de prover condições de funcionamento aos componentes computacionais, entre eles, energia elétrica, climatização e segurança, que serão mais bem detalhados nas subseções seguintes.

Segundo Marcus e Stern (2003), os acessos aos ambientes computacionais precisam ser controlados com objetivo de evitar que pessoas não autorizadas tenham acesso aos equipamentos e possam efetuar sabotagens e atos terroristas.

2.1.1 Energia elétrica

Em um ambiente computacional, o combustível que alimenta os equipamentos é a energia elétrica. Portanto, a falha mais visível é a falta de energia elétrica. Falhas de energia podem ocorrer no fornecimento da concessionária por interrupções devido a manutenções ou falhas nas redes de transmissões, ou por problemas internos ocasionados por falhas nos dispositivos elétricos, entre eles: disjuntores e condutores elétricos (Weygant, 2002).

As concessionárias de energia elétrica devem manter seus clientes informados referente ao compromisso de maior tempo possível para interrupções nos abastecimentos. Informação esta que deve estar presente no contrato de fornecimento da concessionária (Jayaswal, 2006).

A produção de energia difere de país para país e de estado para estado. Alguns países e estados produzem energia excedente que vendem a seus vizinhos, enquanto outros têm de comprar energia para atender suas necessidades. Cada equipe de manutenção de infraestrutura local deve estar familiarizada com a estrutura elétrica e preparado para reagir rapidamente às informações recebidas da concessionária de energia elétrica. Mesmo a concessionária de energia mais confiável não pode garantir o fornecimento em 24×7×365 (Jayaswal, 2006).

Para reduzir ou inibir falhas no fornecimento de energia elétrica é recomendado que o ambiente tecnológico possua uma infraestrutura de energia que consiste em:

- Um UPS (*Uninterruptible Power Supply*), suplemento de energia ininterrupto, também chamado *nobreak*, contendo módulos baterias, capaz de atender às necessidades em curto período de tempo;
- Um gerador de energia capaz de fornecer energia alternativa por longos períodos de tempo; e
- Um sistema de distribuição que forneça energia para os servidores individuais.

2.1.1.1 Suplemento de Energia Ininterrupto

Para uma elevada disponibilidade em um ambiente computacional, é necessária uma fonte contínua e confiável de energia. Se a energia da rede elétrica falhar, o sistema UPS deve ser capaz de fornecer energia para todos os equipamentos, condicionadores de ar, luzes, leitores de acesso, e assim por diante por algum período de tempo, ou pelo menos, até que a fonte de energia seja transferida para outras alternativas, tais como geradores de *backup*. Esse período de tempo pode ser de 15 minutos para algumas horas. A bateria de um UPS deve ser capaz de fornecer a energia para períodos com o consumo em picos, que é maior do que o consumo em execução normal. Um dispositivo com consumo normal de 1.000 watts pode exigir 1.500 watts em períodos de pico (Jayaswal, 2006).

Um UPS tem geralmente baterias que mantêm o fornecimento de energia, enquanto o gerador dá a partida. Esse processo geralmente leva entre 15 segundos e 1 minuto. Um cuidado com as baterias é que elas devem ser substituídas a cada cinco anos ou menos. O *nobreak* deve ser confiável, redundante, e capaz de fornecer energia suficiente. Não sendo apenas necessário para cobrir a falta de energia, mas também para realizar a manutenção no sistema primário de energia elétrica (Jayaswal, 2006).

2.1.1.2 Gerador de energia

Geradores de energia elétrica devem ser instalados se as interrupções de energia forem frequentes ou por períodos superiores a 15 minutos. O custo par aquisição de um gerador é elevado e uma única interrupção de 20 minutos a cada ano pode não justificar o retorno de seu investimento (ROI). No entanto, é comum que *datacenters* com compromissos de níveis

altíssimos de disponibilidade e serviços de missão crítica justifiquem a necessidade de geradores de energia (Jayaswal, 2006).

Existem vários requisitos que devem ser atendidos além da aquisição de um gerador, como exemplo ter um contrato com o fornecedor de combustível que irá verificar regularmente e encher os tanques, se necessário. De nada servirá um gerador com o tanque de combustível quase vazio quando necessário utiliza-o (Jayaswal, 2006).

Um gerador é constituído por:

- Um motor para produzir energia; e
- Uma bobina para converter a energia em eletricidade.

Os geradores mais populares utilizam motores diesel com quatro e até seis cilindros. Grandes fabricantes de motores diesel, incluindo Caterpillar, Tata, Stuart e Stevenson. Outra alternativa é a utilização de motores de turbinas. Semelhantes aos utilizados nos aviões. Turbinas são fabricadas pela General Electric (GE) e Rolls Royce. Estes motores geralmente são abastecidos por gás natural ou diesel e produzem muito mais energia que os motores diesel de tamanho semelhante. No entanto, os motores de turbina são caros e projetado para serem utilizados continuamente. Eles são, portanto, raramente utilizados para energia de emergência (Jayaswal, 2006).

2.1.1.3 Unidades de Distribuição de Energia

Uma PDU (*Power Distribution Unit*), unidade de distribuição de energia, é uma maneira de combinar os disjuntores, cabos e tomadas em um lugar central no *datacenter*, de onde é fácil alimentar vários racks e equipamentos. A PDU aumenta a facilidade de manutenção e upgrades, mas existem algumas desvantagens. Elas devem ser projetadas para as normas elétricas de cada localidade (Jayaswal, 2006).

Segundo Jayaswal (2006), PDUs devem ser projetadas para oferecer confiabilidade e segurança. Todos os comutadores e dispositivos no painel frontal devem ser rebaixados. Disjuntores e interruptores devem ser protegidos ou cobertos para evitar que sejam acidentalmente ligado ou desligado. A PDU deve ter embutidos supressores de surtos e filtros de entrada para proteger contra a entrada de surtos e ruídos. Deve ser possível trabalhar em qualquer disjuntor único sem derrubar o resto do PDU.

2.1.2 Climatização

Os equipamentos de um *datacenter* ou ambiente computacional produzem uma grande quantidade de calor. Um *datacenter* típico consome cerca de 4.000 a 40.000 kW de potência de energia. A maior parte desse consumo é data por seus equipamentos (tais como servidores, armazenamento em rede, e assim por diante) e o sistema de aquecimento, ventilação e ar condicionado (HVAC). Sistemas utilizados para manter o *datacenter* fresco e seco. Temperaturas e umidade relativa fora de um intervalo operacional podem levar a falhas de dispositivos e comportamento não confiável (Jayaswal, 2006).

Segundo Marcus e Stern (2003), independente do tamanho do *datacenter*, é aconselhável que se tenha duas fontes independentes de refrigeração ou no mínimo um equipamento de *backup* que possa entrar em operação e amenizar o problema, caso ocorra a falha de um equipamento que esteja em operação.

A Tabela 2 mostra a temperatura e a umidade relativa do ar ideal e aceitável dentro do centro de um ambiente computacional. Embora os limites aceitáveis para os equipamentos são amplos, o *datacenter* deve ser mantido perto do nível ótimo para aumentar a vida útil e confiabilidade dos equipamentos (Jayaswal, 2006).

Tabela 2 - Temperatura e umidade relativa ideal e aceitável para um *datacenter*.

Fatores ambientais	**Temperatura**	**Umidade Relativa**
Melhor intervalo	21ºC a 23ºC(70ºF a 74ºF)	45-50%
Intervalo aceitável	10ºC a 32ºC(70ºF a 74ºF)	25-75%

Fonte: Adaptado de Jayaswal (2006).

2.2 INFRAESTRUTURA COMPUTACIONAL

A infraestrutura de um sistema computacional pode ser definida como toda a pilha de recursos, incluindo softwares e hardware que provem serviços para aplicação, middleware, ou outras categorias, mas não estão integrados. Como todos os outros componentes, para alcançar alta disponibilidade em infraestrutura computacional, são necessários redundância e robustez (Schmidt, 2006).

Os componentes atuais mais comumente citados em infraestrutura são Rede e Virtualização.

2.2.1 Rede

Apesar de não ser a maior causa de falha nos sistemas, conforme apresentado na Figura 7, a rede é o componente mais importante e exigido para um modelo de alta disponibilidade. A capacidade de comunicação entre computadores é obrigatória para muitas aplicações, restando poucas aplicações de classe empresarial que dispensam o uso de rede (Schmidt, 2006).

Algumas aplicações são explicitamente projetadas para suportar a perda de comunicação em rede por algum tempo. Mas isso é possível porque seu modo de operação é local, e os usuários não precisam acessar um servidor através de uma rede. Mas as aplicações cliente/servidor ou multicamadas são como padrões para quase todas as outras áreas de aplicação (Schmidt, 2006).

Muitos protocolos de comunicação são utilizados em rede. A Organização Internacional para Padronização, ISO (*International Organization for Standardization*), desenvolveu seu próprio padrão, o modelo de referência OSI (*Open Systems Interconnection*). Esse modelo é muito valioso para discussões de problemas de rede, uma vez que fornece abstrações e possui terminologia bem definida. O Quadro 1 apresenta as sete camadas do modelo de referencia OSI (Schmidt, 2006).

Camada	**Nome**	**Foco**
7	Aplicação	Interface de rede para aplicação
6	Apresentação	Gerenciar a representação de dados, codificação de valores, a compressão de dados e encriptação
5	Sessão	Gerenciar a troca de dados entre duas aplicações: estabelecimento, terminação e pontos de verificação de uma conexão
4	Transporte	Transferência transparente de dados entre aplicações, controla a confiabilidade de uma ligação
3	Rede	Transferir dados entre dois sistemas que podem não estar na mesma rede, suporte a redes conectadas, use endereços lógicos com um esquema de endereçamento hierárquico, encontrar o caminho entre dois computadores em redes diferentes
2	Enlace	Transferir dados diretamente entre dois sistemas que estão na rede, use endereçamento físico com um esquema de endereçamento plano
1	Físico	Propriedades elétricas e físicas de dispositivos: layout de

		pinos, voltagens, terminação, especificações de cabos, a representação de sinais em diferentes mídias

Quadro 1 - Modelo de referencia OSI
Fonte: Adaptado de Schmidt (2006)

Em discussões sobre rede, alguns termos precisam ser conceituados claramente, são estes:

1. Rede de área local (LAN – *Local Area network*): Uma rede de computadores que abrange uma pequena área geográfica e que normalmente não envolvem as linhas de comunicação alugadas (externas). Normalmente, a rede que cobre um pequeno campus ou um site é chamada de LAN. LANs geralmente têm a largura de banda alta e muito baixa latência para conexões de rede.

2. LAN Virtual (VLAN – *Virtual local area network*): A virtualização, uma rede que é logicamente independente e que foi construída em cima de uma LAN física. São utilizadas para a segmentação (para criar domínios de transmissão Ethernet) e para ser capaz de compartilhar o equipamento ao longo de várias estruturas independentes. Mas isso não é relevante para o tema de alta disponibilidade.

3. Rede de área metropolitana (MAN – *Metropolitan area network*): Esta é uma rede que cobre um grande campus ou uma cidade. É composto de vários sites LANs que estão ligadas por ligações de alta velocidade. Dentro de uma MAN, é comum ampla largura de banda e latência compatível. Até 100 Mbits/s e 100 km de comprimento do cabo, podemos mesmo esperar latência baixa, apenas com velocidades de 1 Gbit/s ou superior e distâncias de 20 km ou mais é que vamos chegar a latência média.

4. Rede de área ampla (WAN – *Wide area network*): uma rede que cobre uma ampla região geográfica, maior do que uma cidade. É composta pela conexão de muitas LANs ou MANs por meio de roteadores, muitas vezes ligadas por linhas alugadas. Embora hoje pode-se encontrar WAN com largura de banda elevada, mas ainda de grandeza menor do que as LANs. A latência em WAN muitas vezes é bastante ruim e os projetos quem envolvem o uso de LAN WAN devem levar isso em conta;

5. SAN são modelos de armazenamento de dados em rede, que tem como a principal finalidade, proporcionar uma infraestrutura lógica e física para a transferência dos dados entre aplicações de sistema e os dispositivos de armazenamento. A arquitetura SAN é uma infraestrutura de rede dedicada ao compartilhamento de dispositivos de armazenamento para servidores de aplicação, proporcionando flexibilidade, alta disponibilidade e escalabilidade para os sistemas corporativos, bem como o armazenamento de todas as informações dos servidores de aplicação em um único ponto de armazenamento de dados (FONSECA, 2008).

6. Ativos de rede: São os componentes que fazem a rede: switches, roteadores, firewall e etc.

2.2.1.1 Dispositivos de rede

Dispositivos de rede, também chamados ativos de rede, são sistemas que consistem de hardware e softwares, que são utilizados para construir redes de comunicação de dados. Para conectar os recursos computacionais e desfrutar dos benefícios da computação em rede, é necessário utilizar ativos de rede. Pode-se dividir elementos de rede de acordo com o número de camada OSI em que operam, como ilustrado na Figura 1. Roteadores IP operam com roteamento de Camada 3, enquanto os comutadores operam a comutação da Camada 2. Modems, hubs e multiplexadores operam na Camada 1. (MCCROSKY; MINOLI; INIEWSKI, 2008)

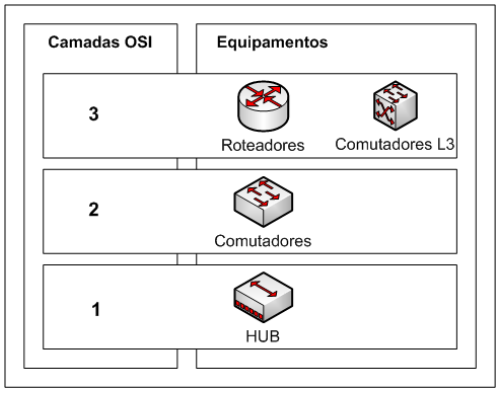

Figura 1 - Dispositivos de rede classificados por nível da camada OSI.

Fonte: Adaptado de Mccrosky, Minoli e Iniewski (2008)

Eles formam uma topologia em estrela, que os computador são ligados aos switches, e switches estão interligados uns aos outros. Portanto, a disponibilidade de switches e

roteadores é importantes para a disponibilidade de toda a rede, são típicos pontos únicos de falha de uma rede (Schmidt, 2006).

- Comutador de rede, também conhecido por *switch*, como o nome indica, alterna quadros entre vários dispositivos. Ele atribui a largura de banda dedicada para ser designado para cada dispositivo na rede conectado. Comutadores dividem grandes redes em segmentos menores, diminuindo um número de usuários que compartilham os mesmos recursos. Os switches na maioria das vezes atuam na camada 2, que gerenciam a comunicação de sistemas conectados com diferentes endereços MAC, não sendo visíveis no nível IP e não têm capacidades de roteamento. Há também switches de camada 3 ou multicamadas; de uma visão abstrata funcional, eles são uma forma especializada de roteadores (SCHMIDT, 2006; MCCROSKY; MINOLI; INIEWSKI, 2008).

- Roteadores de pacotes de dados diretos, seguindo as regras IP, a partir de uma rede para outra. Para realizar a tarefa de examinar o conteúdo de pacotes de dados que fluem através deles. Os roteadores devem determinar o caminho mais eficiente através da rede usando complexos algoritmos de roteamento. Depois de encontrar o caminho mais eficiente, roteadores mudam quadros entre vários portos. Neste sentido, roteadores executam a mesma função de comutadores. Os roteadores são dispositivos muito mais complexos, porém, como eles devem lidar com grandes tabelas de roteamento e encontrar os endereços apropriados de roteamento. Comutadores, por outro lado, apenas alternam quadros conhecendo apenas os dispositivos próximos. Em termos de funções de processamento de pacotes, roteadores executam muito mais operações em comparação com comutadores.

2.2.1.2 Protocolos

Mccrosky, Minoli e Iniewski (2008), em seu livro apresentam detalhadamente os conceitos e fundamentos para uma rede computacional de alta disponibilidade. A partir desta obra, a seguir será apresentado uma compilação dos principais protolocos utilizados em redes computacionais altamente disponíveis.

- Border Gateway Protocol (BGP): é o protocolo de roteamento usado para trocar informações de roteamento entre sistemas *gateways* em um sistema autônomo (AS) de rede, criado para uso nos roteadores principais da Internet. Um AS é uma rede ou grupo de redes sob uma administração comum e com políticas comuns de roteamento. BGP é usado para a troca de informação de roteamento entre sistemas gateway (cada um com seu próprio roteador) em uma rede de ASs. BGP é usado para trocar informações de roteamento para a Internet e é o protocolo usado entre os prestadores de serviços de Internet (ISPs), isto é, BGP é muitas vezes o protocolo usado entre *hosts* de *gateways* na Internet. Quando um roteador se conecta à rede pela primeira vez, os roteadores BGP trocam suas tabelas de rotas completas. De maneira similar, quando a tabela de rotas muda, roteadores enviam a parte da tabela que mudou. Roteadores BGP não enviam regularmente atualizações de roteamento planejadas e as atualizações de rotas informam somente a trajetória ótima para uma rede. Empresas e instituições geralmente utilizam uma IGP, como OSPF para troca de informações de roteamento dentro de suas redes.

- Interior Gateway Protocol (IGP): um protocolo de roteamento para a troca de informação entre gateways (hosts com roteadores) dentro de uma rede autônoma.

- Link Aggregation: uma característica da Camada 2 que permite que as ligações de rede múltiplas para ser combinado formando um canal de alta velocidade única.

- Open Shortest Path First (OSPF): uma característica da Camada 3 que suporta o cálculo de uma árvore do caminho mais curto e mantém uma tabela de roteamento para reduzir a quantidade de saltos (e latência) que leva para chegar ao destino.

- Routing Information Protocol (RIP): um recurso da Camada 3 que apoia a determinação de uma rota com base na contagem menor custo entre origem e destino.

2.2.1.3 Segmentação de LAN

A rede de dados da camada de enlace ou Camada 2 é um segmento de rede LAN onde a comunicação não é roteada. Todos os sistemas em uma rede desse tipo podem alcançar uns aos outros por endereçamento físico. Cada interface de rede que está conectada a essa rede tem um endereço de rede física chamado de endereço MAC. Pacotes Ethernet contém pacotes IP ou partes deles. Tecnicamente, uma rede de dados da camada de enlace também é chamado de domínio de broadcast Ethernet (MCCROSKY; MINOLI; INIEWSKI, 2008).

A necessidade de redundância dentro de um segmento de LAN é melhor ilustrada na Figura 2, que exemplifica um cluster *failover* com dois nós lógicos de rede L1 e L2, cada um com um servidor físico F1 e F2. O cliente C, em um terceiro segmento de rede, precisa se comunicar com os servidores através de um Roteador R.

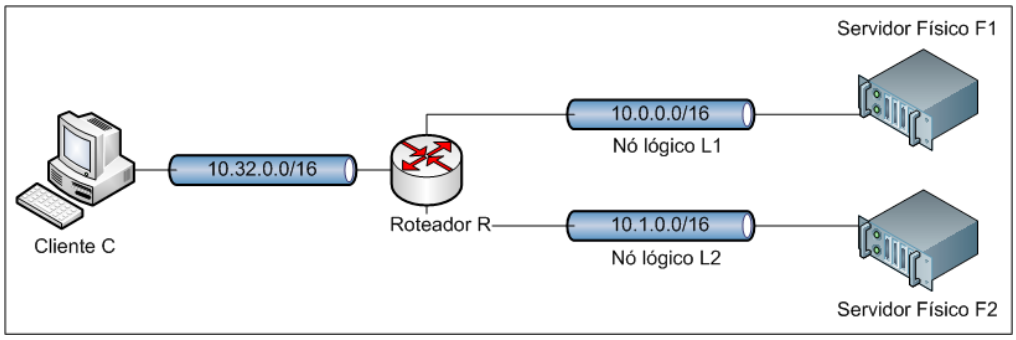

Figura 2 - Cluster *failover* em dois segmentos de LAN.
Fonte: Adaptado de Mccrosky, Minoli e Iniewski (2008)

No caso de uma falha no cluster, um nó lógico mudaria para o outro nó físico. Supondo que L2 mude para P1. Em seguida, P1 tem duas interfaces de rede lógica. Mas sem roteamento dinâmico, o roteador nunca saberia que L2 é subitamente ativada em uma rede à qual ele não pertence. L2 tem um endereço IP de 10.1.0.1, o roteador R nunca vai pedir para o endereço MAC do L2 na rede 10.0.0.0/16 e, portanto, não será capaz de comunicar com L2. O *failover* todo não é utilizável, como C não pode chegar a L2 (MCCROSKY; MINOLI; INIEWSKI, 2008).

Uma vez que não se utiliza o roteamento dinâmico, ambos os nós físicos devem estar no mesmo segmento de LAN, conforme ilustrado na Figura 3. Ela reforça que um segmento de LAN deve ser redundante.

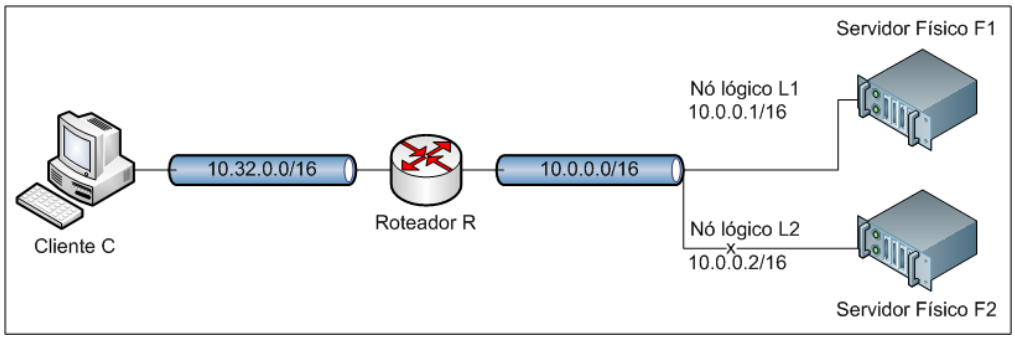

Figura 3 - Cluster *failover* em um único segmento de LAN

Fonte: Adaptado de Mccrosky, Minoli e Iniewski (2008)

Para um segmento de LAN redundante, são necessárias três característica:

1. Todas as conexões redundantes;
2. Os comutadores de rede redundantes; e
3. As configurações de redundância devem ser feitas nos comutadores e sistemas operacionais.

É possível ver na Figura 4, o exemplo de um cluster *failover* utilizando LAN redundante de camada 2. Possuindo dois switches com conexões redundantes.

Ambos os servidores são conectados a cada um dos switches, mas apenas uma das conexões é ativa, isso é chamado de configuração *multipath*.

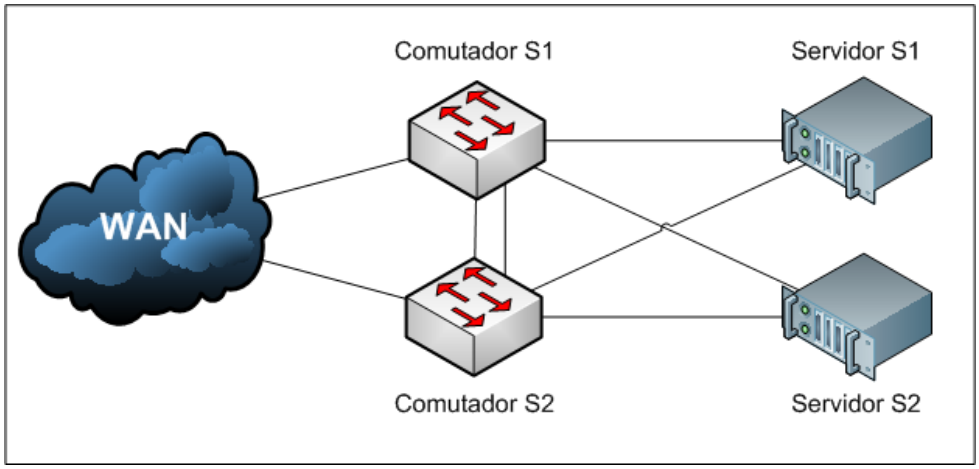

Figura 4 - LAN redundante de camada 2

Fonte: Adaptado de Mccrosky, Minoli e Iniewski (2008)

2.2.2 Virtualização

O conceito de virtualização, embora pareça algo recente e novo, suas origens remetem ao início da história dos computadores, no final dos anos de 1950 e início de 1960. Desenvolvida pela IBM, as máquinas virtuais originalmente concebidas para centralizar os sistemas de computador utilizados no ambiente VM/370. Naquele sistema, cada máquina virtual simulava uma réplica física da máquina real e os usuários tinham a ilusão de que o sistema está disponível para seu uso exclusivo (Laureano, 2006).

Esta tecnologia permite que vários sistemas operacionais e aplicações de diversas plataformas possam, de forma simultânea ser executados na mesma máquina física. Tornando o equipamento mais eficiente ao diminuir a ociosidade de recursos. A reutilização do mesmo equipamento trará economia de espaço físico, tempo, dinheiro e simplifica a estrutura de suporte de TI.

Segundo a VMware (2012), empresa desenvolvedora da solução de virtualização de servidores existentes na Prefeitura de Itajaí, muitas empresas ainda executam um único aplicativo por servidor e, muitos desses aplicativos utilizam apenas de 5% a 15% de toda a capacidade da CPU de seus servidores.

Uma máquina virtual é um ambiente criado por um monitor de máquina virtual (VMM – *Virtual Machine Monitor*), também denominado *hypervisor*. O VMM é um componente de software que hospeda as máquinas virtuais, responsável pela virtualização e controle dos recursos compartilhados pelas máquinas virtuais, tais como, processadores, dispositivos de entrada e saída, memória, armazenagem. Também é função do VMM escalonar qual máquina virtual vai executar a cada momento, semelhante ao escalonador de processos do Sistema Operacional (Laureano, 2006).

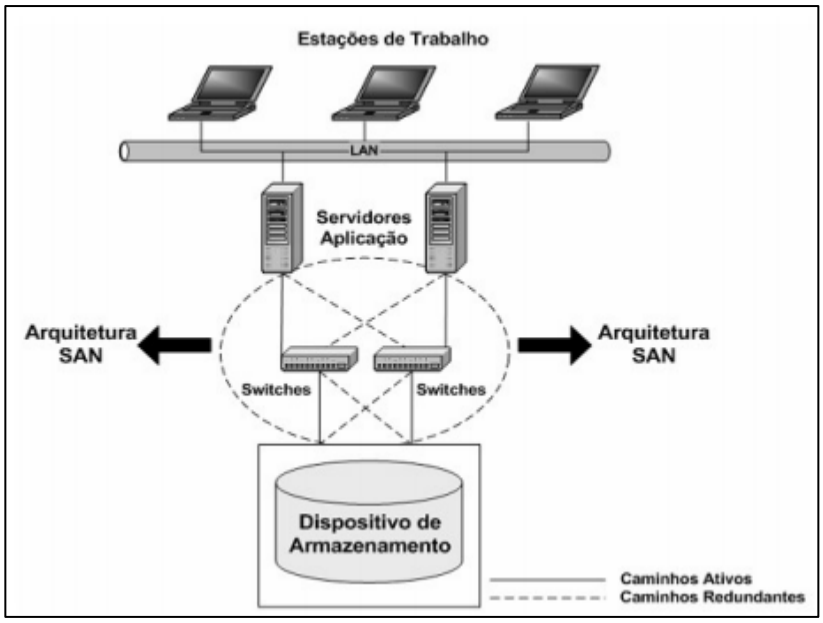

Figura 5 - Armazenamento de dados em modelo SAN
Fonte: Adaptado de LAUREANO (2006)

O VMM é executado no modo de supervisor, no entanto as máquinas virtuais são executadas em modo de usuário. Como as máquinas virtuais são executadas em modo de usuário, quando estas tentam executar uma instrução privilegiada, é gerada uma interrupção e o VMM se encarrega de emular a execução desta instrução.

O software de máquina virtual cria um ambiente através do monitor de máquina virtual, que é o computador com o seu próprio sistema operacional dentro de outro sistema operacional hospedeiro (Laureano, 2006).

Com base nas máquinas virtuais, onde os arquivos são executados em servidores de virtualização, intenciona-se disponibilizá-los em uma *storage*, tornando-os assim, disponíveis para todos os servidores de virtualização que estiverem conectados a ela, permitindo a intervenção dos servidores envolvidos nesse processo quando um deles falhar. Para que os servidores possam utilizar esse recurso de armazenamento, eles devem possuir um dispositivo chamado SAN (*Storage Area Network*), para utilização de seu armazenamento de dados.

2.3 SOLUÇÕES ACESSÍVEIS PARA ATIVOS REDUNDANTES.

Ao procurar por soluções de alta disponibilidade que proporcionam o desenvolvimento de um sistema robusto, estável e redundante através do uso de ferramentas acessíveis e livres de aquisição de licenças, encontrou-se a monografia de Botelho (2006), que será usado como

referencia para a fundamentação do sistema operacional FreeBSD e demais tecnologias, essas idênticas as utilizas pela Prefeitura de Itajaí em suas produções.

2.3.1 FreeBSD

Desenvolvido pela Universidade de Berkeley, é um sistema operacional *open source*, denominado Software Livre1, descendente dos sistemas Unix. Seu primeiro lançamento oficial foi o FreeBSD 1.0 em dezembro de 1993, coordenado por Jordan Hubbard, Nate Williams e Rod Grimes. Publicado com licença BSD, estabelece que os créditos dos autores iniciais devem ser mantidos e o profissional não tem a obrigação de disponibilizar o código fonte. Seu foco principal é a performance, facilidade no uso e estabilidade, principalmente nos serviços de rede e pilha de protocolos TCP/IP. Possui ampla compatibilidade com plataformas de mercado, entre estas Intel x86, DEC Alpha, Sparc, PowerPC e PC-98. Assim como para as arquiteturas baseadas em processadores de 64 bits.

Devido ao sucesso em sua performance da pilha TCP/IP, o FreeBSD é uma excelente opção para a implementações de *firewalls*, nome dado ao dispositivo que tem por função encaminhar e/ou filtrar o trafego entre redes distintas, impedindo a transmissão de dados nocivos ou não autorizados de uma rede a outra, atuando na camada 4, de transporte do modelo referencia OSI.

Nos primórdios do projeto FreeBSD, a árvore de desenvolvimento do sistema foi dividida em dois ramos. Um ramo foi chamado -STABLE e o outro -CURRENT. O FreeBSD-STABLE é direcionado para Provedores de Serviços de Internet e para outros empreendimentos comerciais que não pretendem conviver com mudanças bruscas ou testar novas características experimentais do sistema. Ele recebe apenas código que tenha sido totalmente testado, correções de problemas e outras pequenas inovações incrementais (FREEBSD, 2012).

Também existindo na forma de *appliance*, combinação de software e hardware, sua instalação depende do tamanho da rede, da complexidade das regras que autorizam o fluxo de entrada e saída de informações e do grau de segurança desejado. Entretanto, existem 2 tipos de filtragem de pacotes no nível da camada de rede:

1. Stateless ou estático: cada pacote é analisado de forma independente, sem nenhuma associação com possíveis pacotes já processados. Esta opção é a mais simples, porem, consome mais recursos dos dispositivos;

2. Stateful ou dinâmico: são filtragens mais refinadas e que oferecem um desempenho visivelmente melhor. Nesta filtragem cada pacote é analisado e associado (ou não) a uma conexão já existente. Este processo permite que os pacotes associados as conexões estabelecidas passem automaticamente, diminuindo o overhead de analise e ação sobre cada pacote.

Entre as possibilidades de subsistemas de filtro de pacotes, o FreeBSD possui uma excelente integração entre PF (*Packet Filter*) e o protocolo CARP (*Common Address Redundancy Protocol*). Estes dois que serão detalhados nas subseções seguintes.

2.3.1.1 Packet Filter

Packet Filter, ou simplesmente PF, foi originalmente desenvolvido por Daniel Hartmeier. Foi inicialmente implementado pelo sistema operacional OpenBSD para realizar filtragem de pacotes e traduções de rede NAT (*Network Address Translator*) em TCP/IP. Possuindo também a capacidade de realizar normalização e condicionamento do trafego TCP/IP, controlar banda e priorização de pacotes.

Utilizando a interface de rede chamada PFSync, *Packet Filter State Table Loggin Interface*, permite sincronizar suas conexões estabelecidas através do PF entre os *Firewalls*. Monitorado através do tcpdump, podem ser observadas as alterações na tabela de estados em tempo real. A interface PFSync pode enviar mensagens de alterações de estado para outros nós rodando PF, possibilitando unir as alterações em suas próprias tabelas de estado, desta forma trabalhando de forma sincronizada.

Para solucionar o problema de repetições dos endereços IP é utilizado o CARP, que será detalhado na subseção seguinte.

2.3.1.2 CARP

CARP, *Common Address Redundancy Protocol*, como seu próprio nome sugere, é um protocolo de redundância de endereço comum, permitindo que múltiplos *hosts* no mesmo segmento de rede compartilhem um endereço IP. Este grupo de hosts é referido como um

grupo de redundância. O mesmo é atribuído a um endereço IP que é compartilhado entre os membros do grupo. Dentro do grupo, um *host* é designado o *master* , sendo este o principal e o restante como *slaves*, backups ou contingencia. O *host master* responde a qualquer trafego ou requisições ARP direcionadas para o IP compartilhado. Cada *host* pode pertencer a mais de um grupo de redundância por vez.

Usualmente, o CARP é utilizado para criar um grupo de *firewalls* redundantes. O IP Virtual, que é atribuído ao grupo, configurado nas maquinas clientes como o gateway padrão. Caso o Firewall sofra uma falha, ou seja desligado, o IP se moverá para um dos Firewalls *backup* e o serviço continuará sem ser afetado.

2.3.1.3 Considerações sobre a seção

Esta seção revisou os conceitos de componentes utilizados para a implementação de um modelo computacional de alta disponibilidade. Foram apresentado em detalhes a infraestrutura local necessária para acomodar os equipamentos computacionais, que também chamados de infraestrutura computacional.

Na próxima seção, será revisado o conceito de Alta Disponibilidade, métricas e necessidades para alcançá-la.

2.4 ALTA DISPONIBILIDADE

Alta disponibilidade é um termo que todos parecem saber o que é, mas é difícil encontrar uma definição ampla e precisa. De forma geral, é a capacidade de um sistema proteger-se ou recuperar-se de pequenas falhas em um curto espaço de tempo por meios amplamente automatizados. Não importando se as falhas que venham a causar interrupções de serviços ocorram no próprio sistema, ou no meio ambiente ou são resultado de erro humano. Um sistema altamente disponível tem por opção abortar sessões atuais, ou seja, o usuário vai ser notificado, mas espera-se que sua recuperação ocorra num curto espaço de tempo, restaurando os serviços e disponibilizando-os novamente aos usuários (Schmidt, 2006).

Disponibilidade ou “disponível” é definido como a quantidade de tempo que um sistema é capaz de oferecer serviços aos seus usuários, sem interrupções ou falhas. Podendo ser expressada por números inteiros, medidos em horas, minutos ou segundos, ou então em percentual relativo ao 100% funcional. Qualquer interrupção de um serviço, seja esta planejada ou resultante de erro, é reconhecida como falha (*outage*) e o tempo de duração da

falha que mantém o serviço indisponível, de indisponibilidade (*downtime*) (MARCUS; STERN, 2003).

Se o serviço requer um tempo de duração de falhas o mais próximo possível à zero, então o requisito desse serviço é ser um sistema tolerante a falhas (*fault tolerance*). Sistemas de alta disponibilidade são projetados para suportar um tempo reduzido de duração da falha, combinando investimentos e custos de manutenção menores. Técnicas de alta disponibilidade serão conceituadas mais a frente.

Schmidt (2006) define disponibilidade como medida de quanto tempo que um serviço ou componente de um sistema fica disponível para uso. A disponibilidade pode ser medida pela divisão entre tempo de funcionamento por tempo total, que é a soma dos tempos de funcionamento e parada, podendo ser expressado por:

$$Disponibilidade = \frac{Tempo\ funcionando}{Tempo\ total} \qquad \text{Equação 1}$$

Disponibilidade (*Availability*), denotado por A(t), é a fração média de tempo ao longo do intervalo [0, t] que o sistema está em funcionamento. Esta medida é adequada para aplicações em que uma possível parada possa ser tolerada, desde que a recuperação ocorra num curto espaço de tempo. Expressada por:

$$Disponibilidade = \lim_{t \to \infty} A(t) \qquad \text{Equação 2}$$

Podendo ser interpretada como a probabilidade de que o sistema estará disponível em algum período do tempo, e só é significativa para sistemas que incluem componentes de reparação de falhas.

A disponibilidade está relacionada à taxa de ocorrências de falhas nos componentes de um sistema. O tempo médio até a falha, denotado por MTTF (*Mean Time To Fail*), e o tempo médio entre falhas, MTBF (*Mean Time Between Failures*). O primeiro é o tempo médio do sistema operando até ocorrer uma falha, ao passo que o segundo é o tempo médio entre duas falhas consecutivas. A diferença entre os dois é devido ao tempo necessário para reparar o sistema após a primeira falha, chamado por tempo médio de reparação e denotado por MTTR (*Mean Time To Repair*). Podendo ser expressado por:

$$MTBF = MTTF + MTTR$$ Equação 3

A Figura 6 exemplifica ao longo do tempo t o tempo médio até a falha, o tempo médio entre falhas e o tempo médio para reparação.

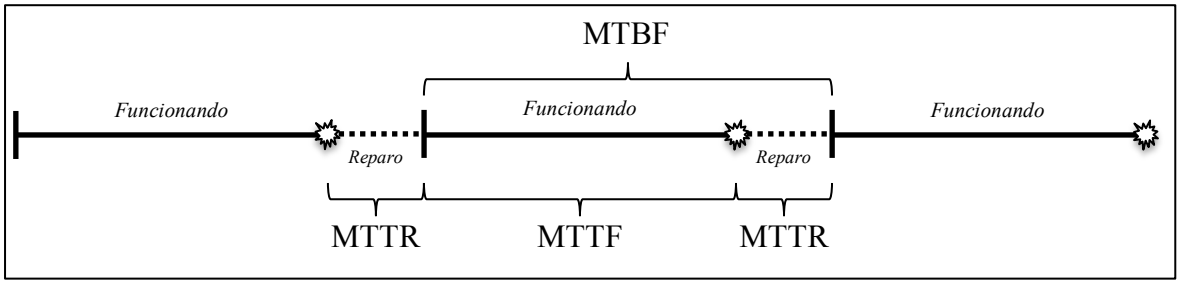

Figura 6 - Alternância de períodos de funcionamento e reparos.
Fonte: Adaptado de Weber (2002)

A disponibilidade a longo prazo pode ser calculada a partir MTTF, MTBF, MTTR e como apresentada a seguir:

$$Disponibilidade = \frac{MTTF}{MTBF} = \frac{MTTF}{MTTF + MTTR}$$ Equação 4

Um sistema tem um MTBF de apenas 1 hora e, consequentemente, uma baixa confiabilidade, no entanto, sua disponibilidade é elevada: A = 3599/3600 = 0,99972 (KOREN; KRISHNA, 2007).

Em práticas atuais, os objetivos de disponibilidades são expressados por números de noves, tipicamente em intervalo de 3NOVES a 5NOVES (99,9% ~ 99,999%). Aplicações de missão crítica, como exemplos em telecomunicação devem estar próximos ou superiores aos 5NOVES (AHLUWALIA; JAIN, 2006).

Schmidt (2006) apresenta três fatores que devem ser considerados em sistemas de Alta Disponibilidade. São estes:

1) **Categorização de falhas**. Esta é uma condição que dirá se possui domínio do problema e solução. É preciso conhecer as potenciais falhas para os serviços e as menores paradas aceitáveis para esses.

2) **Categorização do sistema**: São definidos quais os tempos máximos para interrupção do sistema. Somente quando esses tempos são baixos aplica-se Alta Disponibilidade. Quando um sistema pode ficar uma semana inteira fora, este não utiliza Alta Disponibilidade.

3) **Recuperação ou proteção automatizada**: Tecnologias e soluções abordadas também exercem influência em relação a necessidade de alta disponibilidade. A mesma exigência pode ser resolvida por dois serviços distintos de duas maneiras diferentes: uma precisa de alta disponibilidade, outra não.

A partir das falhas apresentadas na Tabela 1, pode-se calcular os valores para MTTF, MTTR, MTBF e então obter a disponibilidade dos recursos presentes no sistema emissor de notas ficais de serviço. Conforme apresentados a seguir na Tabela 3.

Tabela 3 - Falhas e tempos médios até falhar e reparar

Falha	MTTF	MTTR	Disponibilidade
Interrupção energia elétrica	130.880 min.	160 min.	99,88%
Falha no enlace externo	52,376,6 min.	39,4 min.	99,92%
Falha no comutador de núcleo	261.880 min.	200 min.	99,92%

2.4.1 Avaliando a disponibilidade

Uma forma comum de qualificar a importância de um sistema é caracteriza-lo através do percentual de disponibilidade. Esta qualificação pode aparecer em muitos contratos ou ofertas de *outsourcing,* ou seja, uma concessionária de internet pode prometer disponibilidade de 99,90% em seu material de marketing (Schmidt, 2006).

Esses números são mais frequentemente listados no Acordo de Nível de Serviço (SLA – *Service Level Agreement*). Além do percentual de disponibilidade acordado, é importante constar em números absolutos o tempo máximo de parada aceitado, em n minutos por mês ou x horas anos, para não gerar dúvidas ou duplo entendimento (Schmidt, 2006).

Observando a Tabela 4, para muitos sistemas, 99% de disponibilidade pode ser o suficiente. Se o sistema puder ficar indisponível em média 1 hora e meia por semana, priorizando períodos de menor utilização, como exemplo, Domingo entre as 3 e 4h e meia da madrugada.

Tabela 4 - Disponibilidades e seus *downtimes* em minutos (m).

Percentual	Dia	Semana	Mês	Mês (horas)
99,00	14,400	100,800	438,000	7,300
99,10	12,960	90,720	394,200	6,570
99,20	11,520	80,640	350,400	5,840
99,30	10,080	70,560	306,600	5,110
99,40	8,640	60,480	262,800	4,380
99,50	7,200	50,400	219,000	3,650
99,60	5,760	40,320	175,200	2,920
99,70	4,320	30,240	131,400	2,190
99,80	2,880	20,160	87,600	1,460
99,90	1,440	10,080	43,800	0,730
99,99	0,144	1,008	4,380	0,073
99,999	0,014	0,101	0,438	0,007
99,9999	0,001	0,010	0,044	0,001

Fonte: Adaptado de Schmidt (2006).

Os valores foram apresentados em minutos para dia, semana e mês. Este último (mês) também apresentado em horas.

2.4.2 Falha, erro e defeito

Para melhor entendimento de falha, erro e defeito, sugere-se a interessante definição de Weber, conforme apresentado a seguir.

> Um defeito (*failure*) é definido como um desvio da especificação. Defeitos não podem ser tolerados, mas deve ser evitado que o sistema apresente defeito. Define-se que um sistema está em estado errôneo, ou em erro, se o processamento posterior a partir desse estado pode levar a um defeito. Finalmente define-se falha ou falta (*fault*) como a causa física ou algorítmica do erro. Falhas são inevitáveis. Componentes físicos envelhecem e sofrem com interferências externas, sejam ambientais ou humanas. O software, e também os projetos de software e hardware, são vitimas de sua alta complexidade e da fragilidade humana em trabalhar com grande volume de detalhes ou com deficiências de especificação. Defeitos são evitáveis usando técnicas de tolerância a falhas (Weber, 2002).

Alguns autores nacionais traduzem as palavras inglesas *failure* como falha e *fault* como falta. Para ser coerente com essa ultima tradução a area deveria se chamar tolerância a faltas, pois *failures* não podem ser toleradas. Este trabalho utiliza a palavra falha para mencionar qualquer comportamento fora do seu estado normal, sejam eles falhas ou faltas.

2.4.3 Classificando Falhas

Marcus e Stern (2003) apresentam uma interessante visão das principais falhas ou paradas não planejadas, que apontam o Software como o principal responsável, subintende-se desde o sistema operacional até a aplicação.

Uma explicação para os valores apresentados na Figura 7, que apresentam o software acima do hardware como principal responsável por paradas não planejadas, seria em função dos diversos problemas que podem ocorrer na pilha de softwares, entre eles vírus, configurações e erros diversos. Contrastando com o aumentando na confiabilidade do hardware, que pode utilizar recursos redundantes (MARCUS; STERN, 2003).

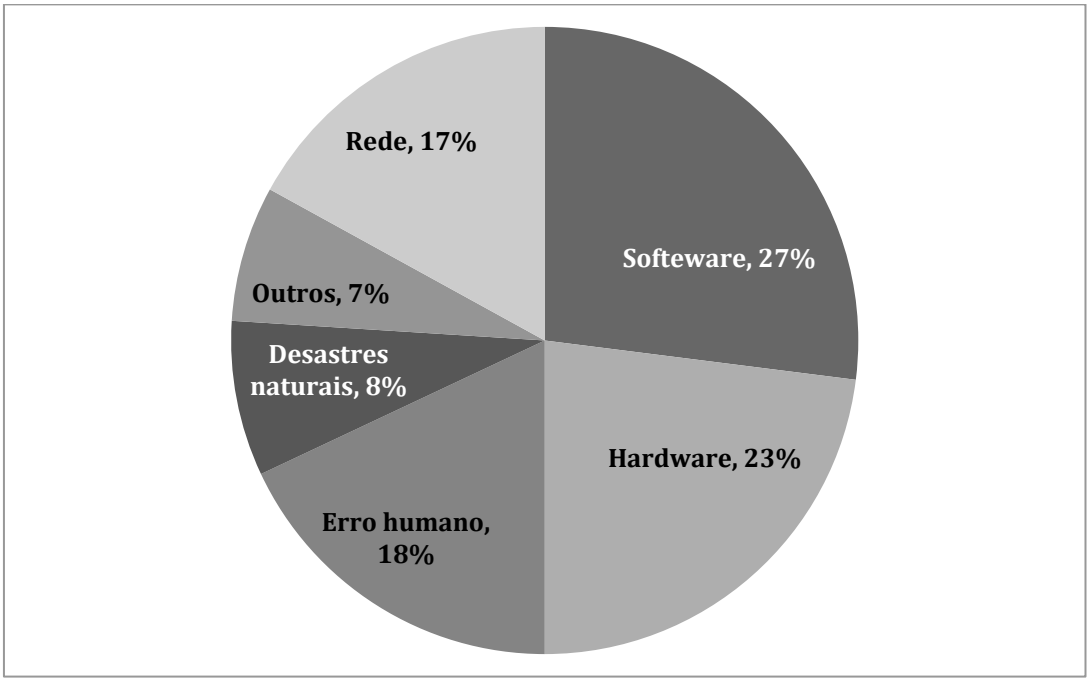

Figura 7 - Principais causas de falhas / paradas não planejadas
Fonte: Adaptado de Marcus e Stern (2003).

Falhas de hardware podem ser classificadas de acordo com vários aspectos. Quanto à sua duração, falhas de hardware podem ser classificadas em permanente, transitório ou intermitente.

- Falha permanente é aquele que assume um estado final e irreversível. Como exemplo de uma falha permanente uma lâmpada queimada.

- Falha transiente é aquela que ocasiona o mau funcionamento de um componente por algum tempo, depois desaparece reestabelecendo o estado

normal. Como exemplo, uma interferência de ruído aleatório durante uma conversa telefônica.

- Falha intermitente nunca se vai embora completamente; oscila entre inativo e ativo. Um exemplo para uma falha intermitente é um fio solto com mau contato. (KOREN; KRISHNA, 2007)

2.4.4 Custo da Alta Disponibilidade

Conforme Weygant (2002), o custo de um sistema altamente disponível depende do grau de disponibilidade desejada. O valor do sistema de alta disponibilidade computacional desejado está diretamente relacionado ao custo de sua parada. Quanto maior o custo de uma parada, mais fácil de justificar o investimento em sistemas. Quanto mais o nível de disponibilidade aproximar-se de 100%, maior e mais rapidamente ele crescerá. O custo de um sistema de 99,95% de disponibilidade é muito maior do que 99,5% de disponibilidade. O custo de 99,5% é muito maior do que o custo de um sistema com disponibilidade de 99% e assim por diante.

2.4.5 Ponto Único de Falha

Mesmo um sistema *stand-alone*, que disponibiliza serviços apenas para LAN e possuindo alta confiabilidade, poderá possuir inúmeros pontos únicos de falha (PUF), também referenciados por SPOF (*single point of failure*). Um ponto único de falha pode estar presentes de várias formas, pode ser um componente de hardware ou softwares que quando falhar causará a indisponibilidade de um serviço de forma parcial ou total. Normalmente estão associados a componentes que não utilizam mecanismos de tolerância a falhas, ou não possuem recursos redundantes, desta forma tornando-se este um ponto único de falha. Com diferentes graus de risco, cada um destes pontos únicos de falha podem colaborar com a queda de um sistema computacional. Identificá-los e programar soluções de continuidade do recurso é basicamente a função de um sistema de alta disponibilidade (Weygant, 2002).

Considerando a instalação de um sistema típico cliente/servidor, os clientes terão suas aplicações rodando em suas estações conectadas sobre uma rede ao servidor de aplicação que está executando alguma atividade em sua CPU. O servidor lê e escreve dados de seus clientes em arquivos de seu rígido. O sistema operacional manipula as conexões com os clientes, a

transferência de dados, a alocação de memoria e outras funções que proporcionam o funcionamento do sistema.

No Quadro 2 estão relacionadas algumas situações de componentes associados com uma possível falha e qual solução que poderia minimizar sua ocorrência.

Componente	**Falha verificada**	**Forma eliminar pontos únicos de falha**
Única CPU	O serviço ficará indisponível até que a CPU seja substituída.	Prover backup de CPU para a aplicação. Por exemplo, criar um cluster de sistemas.
Única interface de rede	Conectividade do cliente é perdida.	Instalar interfaces de placas de rede redundantes.
Único disco rígido	Serviço é perdido até que o disco seja trocado.	Utilizar mecanismos de espelhamento de discos
Único conjunto de dados	Dado é perdido.	Utilizar mecanismos de espelhamentos de discos, backup ou sistema de armazenamento em rede.
Ponto único de alimentação elétrica	Serviço é interrompido até o restabelecimento da alimentação elétrica	Utilizar mais de uma fonte de alimentação com UPS ou gerador de energia.
Única controladora de discos rígidos	Serviço é interrompido até a substituição da placa defeituosa.	Utilizar mais de uma controladora para os discos rígidos.
Programas aplicativos	Serviço é interrompido até o restabelecimento do aplicativo.	Prover reinicialização do programa aplicativo.
Sistema Operacional	Serviço é interrompido até o sistema operacional reiniciar.	Prover capacidade de *failover* no nó afetado.
Comportamento humano	Serviço é interrompido até que o erro humano seja corrigido.	Automatizar a maior quantidade possível de operações.

Quadro 2 - Exemplos de possíveis pontos únicos de falha
Fonte: Adaptado de Weygant (2002).

Conforme conceituado anteriormente, sistemas de alta disponibilidade em sua necessidade mais básica, a identificação dos pontos únicos de falha e programar soluções de continuidade através de mecanismos tolerantes a falhas ou gerenciamento delas, mecanismos estes que serão detalhados nas seções seguintes.

2.4.6 Tolerância a falhas

Um sistema ao exigir os mais altos índices de disponibilidade, mecanismos de recuperação e gerenciamento de falhas não serão suficientes. Para estes casos, o sistema deverá ser construído utilizando técnicas de tolerância a falha. Essas técnicas garantem o bom funcionamento do serviço aos seus usuários, mesmo que venham a ocorrer falhas em seus

componentes, baseando-se todas elas em redundância. A tolerância a falhas não dispensa as técnicas prevenção e remoção das falhas, que estão presentes em gerenciamento de falhas. Sistemas construídos com componentes frágeis e técnicas inadequadas podem não ser confiáveis por simplesmente aplicar tolerância a falhas (Weber, 2002).

Tolerância a falhas de hardware é a área mais madura na visão geral de computação tolerante a falhas. Muitas técnicas de tolerância a falhas em hardware têm sido desenvolvidas e utilizadas na prática em aplicações críticas. Por muito tempo, o principal obstáculo para utilização de tolerância a falhas tem sido o custo do hardware extra necessário. Com a redução do custo do hardware ao passar do tempo, esse impedimento não é mais uma desvantagem significativa, ocasionando o aumento no uso de hardwares tolerantes a falha. No entanto, outras restrições podem inviabilizar o uso desses recursos, como o consumo de energia elétrica (KOREN; KRISHNA, 2007).

2.4.6.1 Redundância e Replicação;

Redundância é a habilidade de um sistema continuar operante caso ocorrer uma falha em algum componente, utilizando técnicas de replicação gerenciada.

Sua precaução base é proporcionar um componente ou sistema de backup. Este pode ser uma parte duplicada, ou um sistema alternativo, ou uma localização alternativa. Pode ser apenas um componente de backup duplicado ou pode haver vários componentes. Todas as repetições de componentes têm um objetivo em comum: evitar pontos únicos de falha (Schmidt, 2006).

Todas as técnicas de tolerância a falhas envolvem alguma forma de redundância, estando tão intimamente relacionado que, na indústria nacional, o termo usado para designar um sistema tolerante a falhas é sistema redundante (Weber, 2002).

As variadas formas de redundância, de hardware, de software, temporal e de informação, tem um impacto no sistema, seja no custo, no desempenho, na área física, ou no consumo de recursos externos. Portanto, apesar de ser a principal solução para tolerância a falhas, o uso de redundância em qualquer projeto deve ser bem ponderada.

Redundância tanto pode servir para detecção de falhas como para mascaramento de falhas. O grau de redundância em cada caso é diferente. Sendo necessários mais componentes para mascarar falhas do que para detectar falhas. Por exemplo, para detectar falhas em um

microprocessador, muitas vezes é usado outro microprocessador idêntico, sincronizado ao primeiro, além de um comparador de sinais na saída de ambos (duplicação e comparação). Qualquer diferença na comparação indica que o par de microprocessadores está em desacordo, e que portanto um dos dois está danificado (ou sofreu uma falha temporária). Entretanto está falha não pode ser mascarada. O resultado da comparação não indica quais as saídas são as corretas (Weber, 2002).

2.4.7 Tipo de redundância

A aplicação de redundância para técnicas de tolerância a falhas podem aparecer de várias formas, entre elas:

- redundância de informação;
- redundância temporal;
- redundância espacial.

Para cada dessas técnicas antes relacionadas, serão detalhadas a seguir nas subseções seguintes.

2.4.7.1 Redundância de informação

Em redundância de informações, juntamente com os dados serão transmitidos sinais ou bits extras, sem qualquer tipo de informação, apenas utilizados para detecção de erros ou mascaramento de falhas. O mascaramento pode ser feito utilizando redundância de informação juntamente com códigos de correção de erros, como ECC (*error correction code*), utilizados com muita frequência em memórias ou transferências entre memórias e processadores (Weber, 2002).

2.4.7.2 Redundância temporal

A redundância temporal consiste em repetir a operação após a detecção de uma falha, evitando o custo de hardware adicional, aumentando o tempo necessário para realizar determinada operação. É usada em sistemas onde o tempo não é crítico, ou o processador trabalha com ociosidade. Essa estratégia não é adequada para falhas permanentes, porque os resultados repetidos serão sempre iguais. Em falhas transitórias, ao repetir uma operação, resultados diferentes são uma forte indicação de falha transitória (Weber, 2002).

2.4.7.3 Redundância espacial

Técnicas de redundância de hardware, podem ser implementadas de formas diferentes. Entre os principais tipos de implementação destacam-se: redundância de hardware passiva e dinâmica, detalhadas a seguir.

a) Redundância de hardware passiva. Na redundância de hardware passiva os elementos redundantes são usados para mascarar falhas. Todos os elementos executam a mesma tarefa e o resultado é determinado por votação. Quando a existência de três ou mais elementos realizando a mesma tarefa, na existência de resultados diferentes, aquele que se repetiu na maioria dos elementos é assumido como verdadeiro (Weber, 2002).

b) Redundância dinâmica. Na redundância dinâmica ou ativa, a tolerância a falhas é realizada por utilizar técnicas de detecção, localização e recuperação. A redundância empregada neste caso não suporta mascaramento. Este tipo de redundância é utilizada em aplicações que suportam permanecer em estado de erro em um curto espaço de tempo, tempo necessário para detecção do erro e recuperação para um estado livre de falhas. Desta forma, fazendo desta técnica mais acessível, por seu menor custo em relação as necessidades para mascaramento de falhas (Weber, 2002).

2.4.8 Recuperação de desastres

"Recuperação de desastres é a habilidade de continuar com os serviços, no caso de grandes interrupções, muitas vezes com capacidades ou desempenho reduzidos. Soluções para Recuperação de desastres geralmente envolvem atividades manuais." (Schmidt, 2006)

A recuperação de desastres é assumida quando todo o sistema torna-se inoperante, seja pela parada de um ponto único de falha, ou por falha de diversos pequenos componentes. Tratando o caso quando as operações não pode ser reiniciada no mesmo sistema ou no mesmo local. Em vez disso, um sistema de contingencia ou de backup é ativado e as operações continuam a partir daí. Este sistema de backup pode estar no mesmo local que o sistema primário, mas normalmente está localizada em outro lugar. Uma vez que esta é a reação a uma grande falha, que se espera que aconteça raramente, recuperação de desastres frequentemente restaura apenas recursos restritos. (Schmidt, 2006)

2.5 TRABALHOS RELACIONADOS

Serão apresentados nesta seção, trabalhos de outros autores que implementam os conceitos revisados no Capítulo 2, definindo-os como casos de sucesso. Com o objetivo de consolidar os conceitos e soluções adotadas para a proposta deste trabalho, foram analisados outros trabalhos científicos em que seus temas principais travam o assunto Alta Disponibilidade.

Ao final desta seção é apresentada uma tabela comparado os trabalhos relacionais à este presente, destacando suas diferenças e principais contribuições.

2.5.1 Alta disponibilidade em firewall utilizando pfsync e carp sobre freebsd

O trabalho de Botelho (2006) apresenta uma solução para ambientes que exigem alta disponibilidade dos recursos computacionais através da redundância do hardware e software – PFSync e CARP sobre o sistema operacional FreeBSD.

Nesta monografia é apresentada a técnica de montagem e a implementação de uma solução de Alta Disponibilidade para o conjunto de firewall e roteador interno, na Universidade Vale do Rio Doce – UNIVALE, e demonstrando o aumento da confiabilidade e disponibilidade dos serviços prestados a toda comunidade acadêmica daquela instituição.

Esta solução tem como finalidade garantir que os serviços informatizados da instituição estejam disponíveis o máximo de tempo possível. Auxiliando na definição do conceito e aplicação do protocolo CARP em ambiente simular ao existente na Prefeitura de Itajaí.

Após a implantação da solução de Alta Disponibilidade através da utilização simultânea dos protocolos CARP e PFSync houve o aumento da disponibilidade dos serviços oferecidos pela Universidade Vale do Rio Doce – UNIVALE, atendendo satisfatoriamente em tempo integral a todos os seus usuários.

Através desta pesquisa comprovou-se a facilidade de implementação desta tecnologia, que permite assegurar o funcionamento constante, através da redundância de hardware e software, do sistema de firewall - equipamento essencial para o funcionamento de uma rede de dados.

Contribuições

O cenário base para a implementação deste trabalho é muito similar ao antes existente na Prefeitura Municipal de Itajaí, no que diz respeito às tecnologias existentes, complexidade e necessidade de prover uma solução adequada de alta disponibilidade, que não onere um alto custo de implementação. Portanto, o trabalho foi de grande valor para a consolidação do conceito e servindo como material de apoio e referencia para a implantação da solução CARP + PFSync no ambiente computacional da Prefeitura de Itajaí.

2.5.2 Estudo dos Recursos de Alta Disponibilidade e Implementação de um Modelo de Pequeno Porte.

O trabalho de Schneider (2006) apresenta uma tentativa de prover uma solução adequada de disponibilidade de dados para a organização. Expõe os problemas enfrentados em identificar qual a real necessidade das empresas.

O estudo em questão, visa identificar os principais componentes para prover a redundância de servidores de uma forma genérica. Citando as tecnologias envolvidas, as funcionalidades de storage, abordando o impacto negativo no negocio gerado pela indisponibilidade dos serviços de Informática e descreve as principais arquiteturas de discos.

Diante do trabalho de pesquisa e implementação de algumas soluções propostas, foi gerado um quadro comparativo (Tabela 5) onde o autor procurou atribuir notas para cada solução. Os critérios avaliados tendem variar de uma organização para outra, bem como as notas atribuídas. O objetivo deste quadro foi de definir um ponto de partida objetivando avaliar mais de uma solução de disponibilidade aplicada a um ambiente de pequeno porte.

Tabela 5 - Análise comparativa entre soluções de alta disponibilidade para servidores.

Critério	**Modelo tradicional**	**Solução 1 Heartbeat + rsync**	**Solução 2 Programa + rsync**	**Solução 3 Heartbeat + DRBD**
	Nota	**Nota**	**Nota**	**Nota**
Tempo de indisponibilidade total	0	3	2	3
Intervalo da cópia dos dados	ND	2	2	3
Intervalo de verificação de disponibilidade	ND	3	1	3
Integridade dos dados	1	1	1	3

Dificuldade de implementação	3	2	1	2
Custo da solução	3	2	2	2
Redundância da aplicação	1	3	3	3
Dificuldade de manutenção	3	2	1	2
Especialização da mão de obra	3	2	1	2
Redundância do hardware	1	3	3	3
Pontuação total	**15**	**23**	**17**	**26**

O autor atribui notas de 0 a 3, onde 0 é considerada o pior caso e 3 o melhor. Desta forma a análise comparativa entre as soluções testadas, aponta a terceira (Heartbeat + DRDB) com a melhor. Entretanto, não aplicável na Prefeitura de Itajaí por já existir outra solução de clusterização.

Contribuições

A principal contribuição do trabalho em questão, foi consolidar a possibilidade de replicar os servidores, especificamente o sistema de arquivos através do rsync de forma assíncrona. Inicialmente foi considerada a possibilidade de clusterizar o servidor de produção com o novo de contingencia, utilizando o Heartbeat, que fora devidamente analisado e comparado com outras soluções pelo autor.

Porém, na fase de implementação da solução para replicação dos servidores. Em contato com o desenvolvedor e responsável pelo sistema emissor, houve a recomendação para a clusterização através dos recursos já existentes na camada de *middle-ware* do servidor de aplicação Java Jboss. Recomendação essa, que foi acatada e assim implementada.

2.5.3 Alta Disponibilidade em ambientes Virtualizados

O artigo de Silveira (2009) apresenta a pesquisa realizada sobre virtualização de servidores, com ênfase em alta disponibilidade. Baseia-se na utilização de ferramentas de virtualização sem custo de licença da VMware, objetivando a alta disponibilidade de servidores.

Com os altíssimos valores das soluções de virtualização em conjunto com alta disponibilidade, foi proposto então o desenvolvimento desse trabalho, focando em prover um cenário sem custos de licenças de softwares com hardwares de baixo custo. Permitindo assim

que, empresas de pequeno e médio porte possam ter uma solução utilizando sistemas virtualizados com alta disponibilidade.

O autor conclui que a solução apresentada não pode ser comparada tecnologicamente com as atuais soluções pagas do mercado, limitando-se a sua capacidade tecnológica.

Contribuições

A virtualização de servidores é uma tecnologia já utilizada pela Prefeitura Municipal de Itajaí, possuindo módulos de gerenciamento *entreprise*. Portanto, o artigo serviu como base para consolidar as necessidades em manter a tecnologia proprietária da VMWare, já existe, onde esta possui recursos que não conseguem ser substituídos por solução sem custos com licenciamento.

2.5.4 Comparativo entre soluções similares e o presente trabalho

De forma sintetizada e para rápida análise, o Quadro 3 apresenta os trabalhos relacionados e seus objetivos, também suas contribuições e diferenças em relação a este trabalho.

Trabalho / Autor	Objetivo	Contribuições	Diferenças
2.5.1 Alta disponibilidade em firewall utilizando pfsync e carp sobre freebsd (BOTELHO, 2006).	Implantar uma solução de alta disponibilidade em firewall que atenda as necessidades da UNIVALE.	Consolidação em adotar o protocolo de redundância CARP e o módulo PFsync para sincronização de estados das interfaces.	O trabalho relacionado propõe e implementa uma solução para firewall, solução essa utilizada como referencia para uma das diversas técnicas implementadas neste trabalho.
2.5.2 Estudo dos Recursos de Alta Disponibilidade e Implementação de um Modelo de Pequeno Porte (SCHNEIDER, 2006).	Estudo comparativo entre soluções para clusterização de servidores visando alta disponibilidade.	Consolidação do uso do rsync para sincronização assíncrona entre servidores de produção e contingencia.	O trabalho relacionado visa uma solução simples para clusterização. Este trabalho por sua vez busca por um conjunto de técnicas que atendam a

			disponibilidade pretendida.
2.5.3 Alta Disponibilidade em ambientes Virtualizados (SILVEIRA, 2009).	Proposta por uma solução de alta disponibilidade em ambientes virtualizados não onerando com licenciamento.	Serviu como base para consolidar as necessidades em manter a tecnologia proprietária da VMWare,	O artigo visa uma solução apenas para o ambiente virtualizado. Este trabalho propõe um conjunto de técnicas e soluções que atendam toda a infraestrutura.

Quadro 3 - Comparativo sintetizados dos trabalhos relacionados

Por não ser o objetivo deste trabalho, não foram realizadas pesquisas exaustivas por trabalhos similares que estudem uma proposta de solução para alta disponibilidade em ambientes da administração pública. Estas pesquisas não retornaram algum trabalho similar. Desta forma, foram pesquisados por trabalhos que consolidaram os conceitos e soluções adotadas.

3 DESENVOLVIMENTO

Este capítulo relata o desenvolvimento da solução proposta por este TTC. Um estudo da cenário existente no ambiente computacional do Centro Tecnológico da Prefeitura Municipal de Itajaí, resultado no aumento da disponibilidade do Sistema Emissor de Notas Fiscais de Serviço da Prefeitura de Itajaí. Neste capitulo também é mencionado o problema encontrado com o protocolo CARP e suas alterações necessárias. Foram levados em consideração os objetivos previamente definidos, definição de escopo e são apresentados gráficos da disponibilidade global para melhor entendimento do produto final. Ao final deste capítulo, são descritos os testes realizados e os resultados obtidos.

3.1 CENÁRIO ATUAL

A metodologia empregada para descrever e definir a contribuição deste estudo consiste na apresentação detalhada da situação existente da pilha de hardware e software empregados para o Sistema Emissor de Notas Fiscais de Serviço e o cenário proposto objetivando 99,90% de disponibilidade.

Com o raciocínio de julgamento para a definição do que deveria ser alterado/melhorado, usando-se a investigação da existência de pontos únicos de falha.

Os objetos que fazem parte do comparativo são todos aqueles que de fato sofrem alterações de suas características. A Figura 8 descreve a situação existente no início deste projeto.

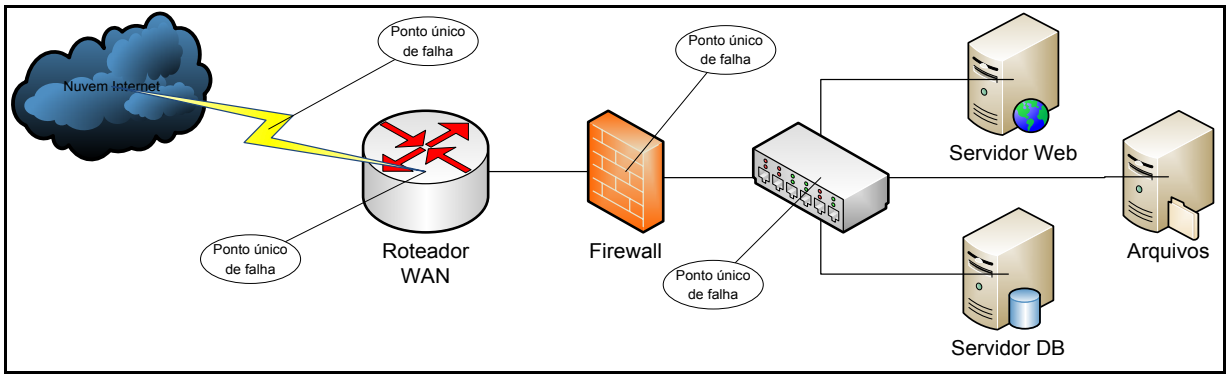

Figura 8 - Abstração do cenário existente e seus pontos únicos de falha.

3.1.1 Rede

Conforme descrito no Capitulo 2, quatro itens são considerados críticos para a implantação de um sistema de Alta Disponibilidade. Nessa subseção, são descritos estes itens e como estavam configurados no ambiente da Prefeitura Municipal de Itajaí.

3.1.1.1 Enlace Externo

Serviço contratado junto à operadora de serviços de telecomunicação, o link de dados externo possui velocidade de 30Mbps. O meio físico de transporte desta é através de fibra óptica, que parte do município de Balneário Camboriú por dois caminhos distintos e independentes, porém sua última milha ocorre por uma única conexão. Ela chega pela Av. José Eugênio Muller e entra na Prefeitura de Itajaí por galeria subterrânea, conforme apresentado na Figura 9.

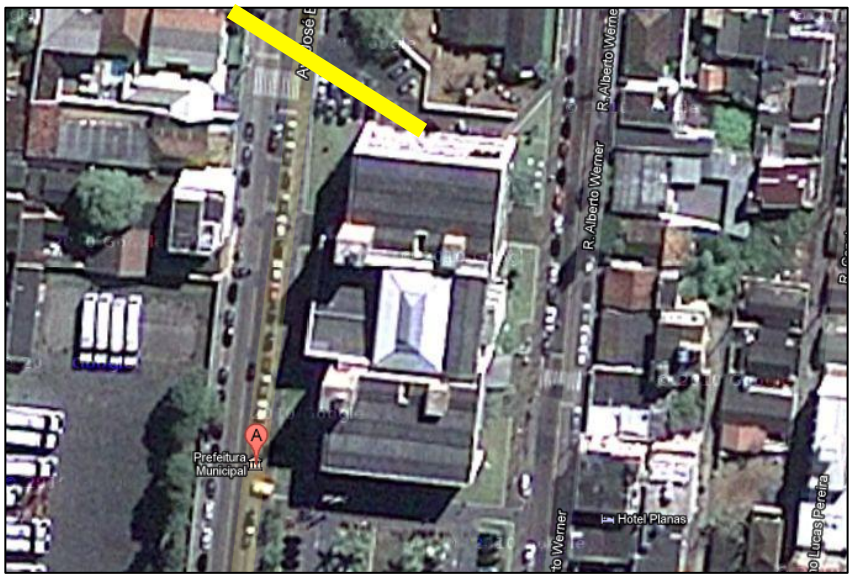

Figura 9 - Entrada do link de fibra óptica.
Fonte: Adaptado de Google (2012).

Ao chegar na sala dos servidores, a fibra óptica é conectada ao modem óptico Overtek, na porta 01, transformando a mídia óptica em pares metálicos. Estes por sua vez são conectados ao Roteador Cisco 1805 na sua única porta WAN. A conexão entre o roteador e o firewall é realizada por um único cabo, padrão *patch cord ethernet* categoria 5e, conectado em uma interface de rede PCI do servidor que atende como firewall.

Na Figura 10 pode-se visualizar de forma mais detalhada como se dá a configuração da camada de Enlace Externa no cenário existente.

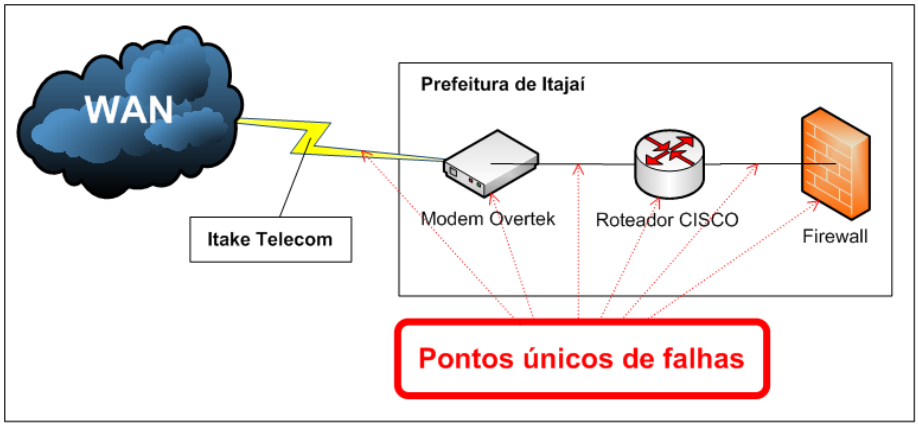

Figura 10 - Enlace externo e seus pontos únicos de falha.

3.1.1.2 Firewall

O *firewall* por seu conceito básico, atendente como filtro de rede, aplicando regras para entrada e saída em seus diferentes segmentos de rede. Entretanto, a solução de *firewall* adotada e implementada na Prefeitura Municipal de Itajaí, também atende como *Gateway* e desta forma assumindo o papel de roteador interno da rede. Diante do entendimento de sua definição e funcionalidades adotadas, na continuidade deste trabalho será referenciado apenas como *firewall*.

O firewall era configurado utilizando um servidor HP Proliant DL365, este tolerante a falhas em uma plataforma ultra densa com alimentação redundante, ventiladores redundantes e capacidade de RAID incorporado. Entre suas principais especificações encontram-se dois processadores AMD Opteron™ Dual-Core, 2.4GHz por núcleo, 4GB DDR2 ECC 667 MHz instaladas em módulos 1GB cada e ocupando 04 dos 08 bancos de memória disponíveis na placa mãe. A Figura 11 apresenta um diagrama dos componentes internos do servidor utilizado como firewall.

Eram utilizadas três interfaces de rede, duas delas on-board com chip Broadcom NetExtreme II e uma Intel 82574L Gigabit Ethernet Controller PCIe, que conectam três segmentos de rede: local, DMZ e válida, respectivamente, através cabos de padrão ethernet cat5e nas portas 21 e 22 do comutador de núcleo e roteador, conforme apresentado no item 3.1.1.1.

Dois discos SAS de 2,5”10k rpm de 72GB, configurados em RAID 1 por sua controladora Smart Array P400, são apresentados como um único volume ao sistema operacional, FreeBSD.

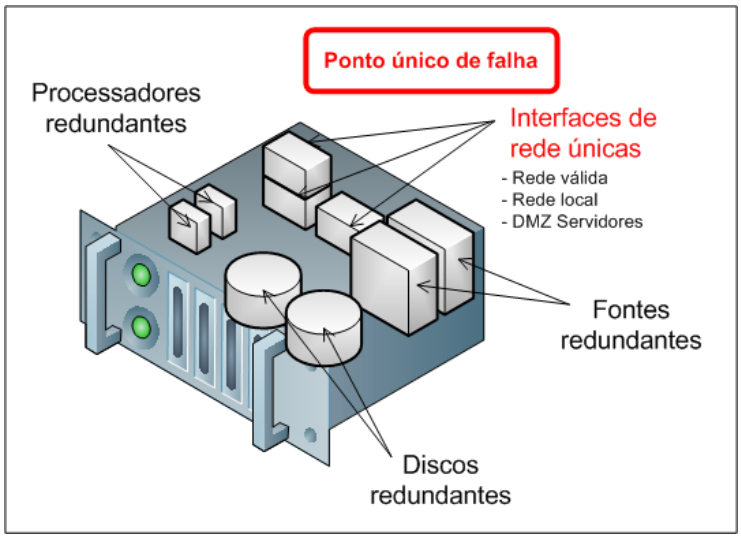

Figura 11 - Componentes internos do servidor firewall

FreeBSD na versão 8.2 em 64bits, instalado com o *kernel* genérico e posteriormente recompilado para atender os recursos presentes no servidor em modo estável.

O subsistema de filtro de pacotes utilizado, PF (*Packet Filter*), portado do OpenBSD, proporciona as regras de controle nos tráfegos de entrada e saída em cada segmento de rede.

3.1.1.3 Enlace Interno

Todos os segmentos de rede estão conectados ao comutador de núcleo, também denominado switch core. O produto utilizado é o D'LINK DES-3226L, possuindo como principais características 24 portas Gigabit ethernet e 04 Combo SFP 1Gbps.

A separação dos segmentos ocorre por VLANs, aplicando *tags* de identificação, direcionando o trafego aos destinos apropriados, de forma isolada e segura.

3.1.2 Servidores

Servidores do tipo lâmina são acomodados no gabinete HP c7000 BladeSystem, que oferece infraestrutura compartilhada de energia, refrigeração e I/O para suportar servidores modular. O gabinete possui 10U de altura e capacidade para até 16 servidores e/ou lâminas de armazenamento, mais modulo redundante de rede e interconexão de armazenamento. Fornecido junto com o sistema de gerenciamento HP Insight Control, que proporciona plena capacidade de gestão adequada da linha ProLiant.

Sua energização ocorre por quatro fontes redundantes conectadas em pares nas duas PDUs disponíveis no rack.

3.1.2.1 Lâminas

As seis lâminas de processamento, ou servidores, instaladas, estão conectadas nas seis primeiras baias do gabinete c7000. Todas de mesma configuração, possuem como principais características dois processadores Intel Xeon Six-Core de 2.80GHz por núcleo, 48GB de memória DDR3 ECC. Suas conexões com o gabinete c7000 são realizadas por barramentos metálicos.

3.1.2.2 Virtual Connect

Utilizado o HP Virtual Connect, a forma mais convergente e flexível para conectar as lâminas de servidores virtualizados a quaisquer sistemas de armazenamento em rede, agregando as conexões LAN e SAN. As conexões realizadas com o comutador de núcleo e sistema de armazenamento em rede, ocorrem por conexões ethernet cat5e e fibre channel respectivamente. No total são quatro módulos Virtual Connect, redundantes e configurados em modo ativo/ativo. (Butow, Dicke, & Joyal, 2011)

3.1.2.3 Servidor de gerenciamento

Para gerenciamento é utilizado um servidor HP Proliant ML110 G6, possuindo como principais características processador Intel Xeon Quad-Core, 4GB de memória DDR2 ECC 667Mhz e controladora SCSI conectada a biblioteca de fitas de backup que será melhor a seguir, no item 3.1.6.1.

3.1.3 Sistema de armazenamento em rede

O sistema de armazenamento em rede EMC CLARiiON CX3 modelo 40, possui duas controladoras redundantes em modo ativo/ativo, com 4GB de memória cache e processadores Intel Xeon. 45 discos SAS 3,5” 15k rpm com 300GB e interface de conexão FC com velocidade de 4Gbps proporcionam capacidade de armazenamento útil de 8TB, configuradas em grupos RAID 5. Todos os recursos presentes são redundantes e tolerantes a falha, entre eles fontes de alimentação com baterias, unidades de disco, gavetas de discos, cabos, ventiladores e comutadores FC.

3.1.3.1 Organização

Para melhor organização, os discos são agrupados e configurados em RAID 5. Cada agrupamento de discos é particionado logicamente, cada partição recebe um número de identificação, denominado LUN. As LUNs são apresentadas aos servidores, conforme políticas de leitura/escrita definidas para cada associação.

A LUN23 atende a maquina virtual, instalada com o sistema operacional Linux Ubuntu Server em sua versão 10.04 de 64bits, que disponibiliza o servidor de aplicação JBOSS para a aplicação do sistema emissor de notas fiscais.

3.1.4 Virtualizador de servidores

O sistema de virtualização de servidores utilizado é o VMware ESX, na versão 5.0. Do tipo X, conforme apresentada no Capítulo 2, instalado em cada lâmina HP Proliant, possui um ambiente gerencial centralizado, denominado vCenter e instalado no servidor de gerenciamento, conforme apresentado no item 3.1.2.3.

Todos os servidores de lamina, executando o ambiente virtualizador, são apresentado a um cluster denominado PMI.

Os segmentos de rede, são apresentados através de comutadores virtualizados, que comunicam-se com o comutador de núcleo, através das *tags* de identificação.

3.1.4.1 Fault Tolerance

O ambiente virtualizador possui configurado o recurso de FT (*Fault Tolerance*), fornecendo disponibilidade contínua para aplicações em caso de falhas da lâmina, criando uma instância de sombra em tempo real de uma máquina virtual, mantendo-as em sincronia. Em caso de falha de hardware, o FT elimina até mesmo a menor chance de perda de dados ou interrupção.

3.1.4.2 Distributed Resource Scheduler - DRS

Configurado no cluster PMI, permite ao *datacenter* virtual fazer balanceamento de carga das maquinas virtuais para adequar-se a mudanças na demanda de cada aplicação/maquina virtual.

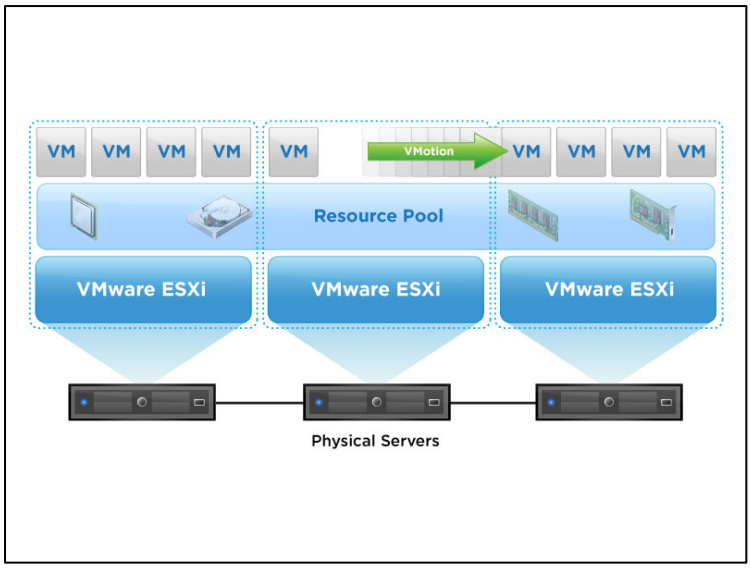

Figura 12 - Funcionamento do VMware Distributed Resource Scheduler – DRS

Fonte: VMware (2012)

3.1.5 Infraestrutura local

Todos os servidores e recursos computacionais, disponíveis que atendem ao serviço emissor de notas fiscais e demais, estão acomodados no paço municipal, mais precisamente no segundo andar do lado norte.

O espaço reservado para acomodar os equipamentos possui 12 metros quadrados de área. Suas paredes são de divisórias, assim como sua porta.

Embora apresente fragilidade de segurança física, a salda dos servidores está localizada dentro das dependências do CTIMA. Local de acesso restrito e protegido por uma porta externa.

3.1.5.1 Disposição do ambiente

Dentro da sala dos servidores, possuem três racks, acomodando os equipamentos de rede, servidores e sistema de armazenamento em rede.

O primeiro rack, com 40U de altura, serve exclusivamente aos equipamentos de rede. Possuindo duas PDUs, conectadas ao *nobreak*.

O segundo rack, com 40U de altura, serve aos servidores, possuindo duas PDUs conectadas ao *nobreak*.

3.1.5.2 Fornecimento de energia

Possuindo apenas uma única concessionária de energia elétrica em Itajaí e em Santa Catarina, o principal fornecimento de energia elétrica parte CELESC. Atendendo por apenas uma única entrada subterrânea.

Como fonte alternativa de energia, possui um *nobreak* de 10kva, modelo APC SURT10000XLI, com três módulos de baterias que garantem autonomia de 2horas e meia. Todas as PDUs dos racks estão conectadas ao mesmo *nobreak*

3.1.5.3 Climatização

A climatização do ambiente ocorre apenas por injeção ar refrigerado, por três aparelhos condicionadores de ar independentes, sem gerenciamento inteligente, porém ligados simultaneamente, mantendo a temperatura de 18º centigrados mesmo quando falhar um. O monitoramento da temperatura é feito por um termômetro fixado em uma da paredes, sem qualquer sistema automatizado. Equipamentos estes revisados periodicamente por empresa especializada e contratada por processo licitatório.

3.1.6 Backup e recuperação

Utiliza-se o sistema gerenciador de backups HP Data Protector, agentes são instalados nos servidores proporcionando cópias quentes, ou seja, mesmo em produção é possível copiar os dados de cada servidor.

3.1.6.1 Biblioteca de fitas

A biblioteca de fitas modelo HP LTO 4, conectada ao servidor de gerenciamento (item 3.1.2.3) grava os dados em conjunto de 8 fitas, com capacidade total próxima aos 13 Terabytes.

3.1.6.2 Cópia de segurança

Por meio de scripts e agendamentos via HP Data Protector, o conjunto de dados a ser copiado é definido por seleções granulares de arquivos ou até mesmo a cópia total de uma máquina virtual.

3.1.6.3 Restauração

Caso necessário restaurar alguma máquina virtual ou arquivos diversos, o sistema gerenciador de backup cataloga todas as versões disponíveis para o conjunto de dados de interesse. Para a restauração é necessário posicionar na biblioteca de fitas, o conjunto indicado. O tempo de restauração, dependendo do volume, pode demorar de minutos até mesmo horas de trabalho.

3.1.6.4 Política

Diariamente são copiados para fitas os arquivos ou blocos que sofreram alterações. Semanalmente são copiados um determinado conjunto de dados. Mensalmente é copiado um conjunto de máquinas virtuais e por fim, semestralmente são copiadas todas as máquinas virtuais.

3.2 PONTOS ÚNICOS DE FALHA ENCONTRADOS

A seguir é apresentado o Quadro 4, relacionando os pontos únicos de falha encontrados e as formas para eliminá-los.

Componente	**Falha verificada**	**Forma eliminar pontos únicos de falha**
Acesso à internet	Se ficar indisponível, todo o serviço ficará inacessível aos clientes externos até que seja reestabelecido o fornecimento.	Instalação de um link redundante, criando sistema autônomo e utilizando o protocolo BGB através de duas operadoras.
Comutador de núcleo.	Em caso de falha, a segmentação da rede entre servidores e internet ficará indisponível.	Instalar um par de switches empilhados e redundantes.
Conexões de rede	Conexões únicas, por um único cabo, em caso de rompimento, disponibilizará a comunicação entre dispositivos.	Instalar um par de cabos para cada conexão.
Firewall	Servidor único com interfaces únicas para cada segmento, caso falhar, interromperá todas as transferências de dados.	Instalar um novo equipamento, compatível com o atual. Utilizar o sistema operacional FreeBSD + Protocolo CARP, replicando todo o firewall e proporcionando redundância

Servidor físico	Embora o servidor físico seja inteiramente redundante, caso ficar indisponível por um desastre maior, toda a aplicação ficará indisponível até disponibilização de um novo servidor e a restauração dos backups.	Instalar um novo servidor em ambiente físico distante, replicando todo o conteúdo do servidor de produção, capaz de atender prontamente o direcionamento do trafego quando necessário.
Energia elétrica	Caso ficar indisponível, todos os equipamentos ficarão inoperantes e acessíveis até que o serviço seja reestabelecido.	Instalar *nobreak*s redundantes, com autonomia de no mínimo 3 horas em baterias. Toda a instalação elétrica local deve ser redundante, permitindo a conexão dos equipamentos em dois *nobreak*s distintos.

Quadro 4 - Pontos únicos de falha encontrados no cenário atual

Cada ponto único de falha identificado e apresentado no Quadro 4 possui uma proposta de alteração no subcapítulo 3.3 e respectiva alteração realizada no subcapítulo 3.4.

3.3 ALTERAÇÕES NECESSÁRIAS

Buscando eliminar os pontos únicos de falha encontrados no cenário existente, apresentado no item 3.2, nesta seção estão descritas as soluções propostas para cada ponto crítico. O cenário planejado possui a necessidade de diversas alterações. Todas elas estão descritas e detalhadas nas subseções seguintes.

Recurso	**Alterações necessárias**	**Valor estimado**
Enlace externo	Instalar novo link redundante.	R$ 10.000,00/mês
Firewall	Instalar novo firewall redundante e replicando o atual.	R$ 2.800,00
Enlace interno	Instalar dois novos comutadores de núcleo redundantes.	R$ 87.000,00
Servidores	Instalar novo servidor replicando o sistema emissor de forma assíncrona.	R$ 2.800,00
Sistema de armazenamento em rede (*storage*)	Nenhuma alteração necessária.	-
Virtualizador	Nenhuma alteração necessária	-

Fornecimento de energia elétrica	Buscar um acordo de nível de serviço com a concessionária.	-
Fornecimento de energia elétrica	Instalar novos *nobreaks* com autonomia suficiente para cobrir o SLA.	R$ 50.000,00
Disposição do ambiente	Nenhuma alteração necessária	-
Climatização	Nenhuma alteração necessária	-
Backup e recuperação	Nenhuma alteração necessária	-

Quadro 5 - Alterações necessárias no cenário atual

Cada item que compõe os Quadros 4 e 5 é discutido individualmente nas próximas seções.

3.3.1 Rede

3.3.1.1 Enlace externo

Instalar um novo roteador com roteamento dinâmico BGP na Prefeitura e construir uma segunda abordagem por outro caminho físico totalmente separado. Uma sessão BGP para Operadora A e outra para Operadora B. O registro do sistema autônomo proporcionará o uso de uma faixa de endereçamento IP independente de operadora, desta forma pode-se garantir o funcionamento do serviço mesmo se ocorrer alguma falha parcial ou em todos os recursos.

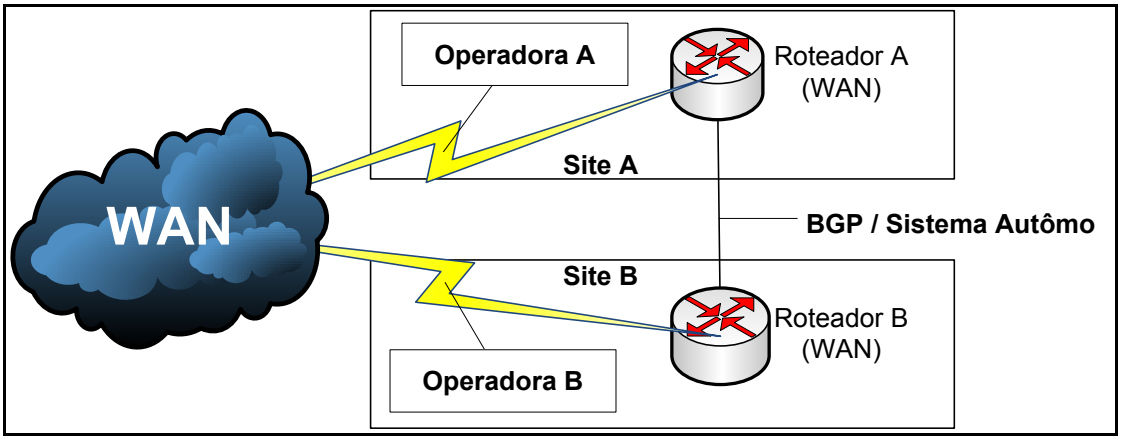

Figura 13 - Enlace externo redundante.

3.3.1.2 Firewall

Instalar um novo servidor para firewall, possuindo hardware similar e com o mesmo número de interfaces de rede. Configurar o protocolo CARP para replicar todas as funcionalidades do firewall que executam no FreeBSD.

3.3.1.3 Enlace Interno

Instalar dois novos comutadores de núcleo, configurados para utilizar o protocolo Link Aggregation, instalados em locais separados, denominados Site A e Site B. A interligação se realizará por cabos de fibra óptica operando na velocidade de 10 Gigabits. Todo o controle e configuração das VLANs serão replicados e atenderão com um único dispositivo de rede.

3.3.2 Servidores

Instalar um novo servidor físico no ambiente remoto, executando o sistema operacional Linux + rsync mantendo uma cópia assíncrona do sistema de arquivos contendo os arquivos de remessas das notas fiscais transmitidas.

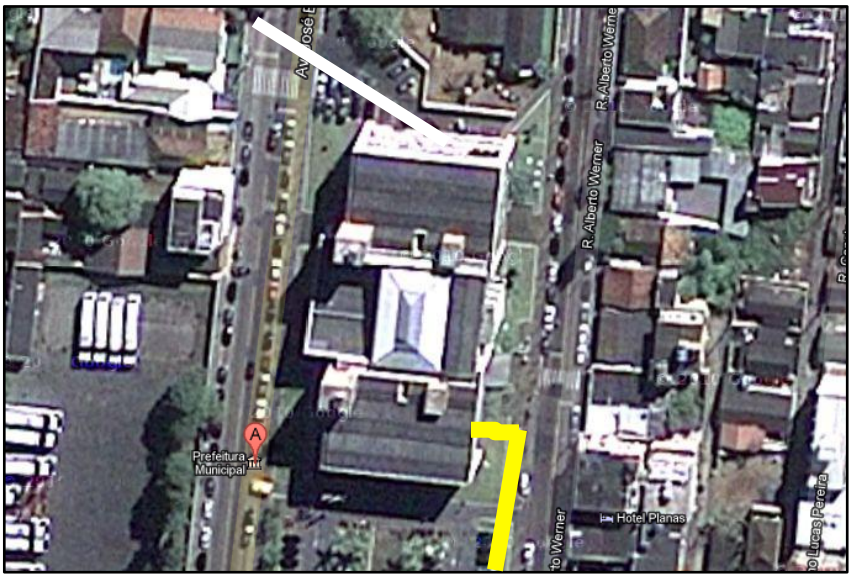

Figura 14 - Segunda abordagem de enlace externo por caminhos físicos separados.

Fonte: Adaptado de Google (2012)

3.3.3 Infraestrutura local

Disponibilizar um novo ambiente, capaz de acomodar os equipamentos de contingencia, entre eles roteador, comutador, servidor e *nobreak*. O espaço deve possuir pelo

menos 2 metros quadrados de área, acesso restrito e climatização próxima aos 20º centígrados.

3.3.3.1 Fornecimento de energia

Instalar um novo *nobreak* junto à infraestrutura atual, permitindo conectar pelo menos uma PDU de cada rack em *nobreak*s diferentes. Os *nobreak*s deverão possuir módulos de bateria com autonomia mínima de 3 horas de funcionamento quando ausente de rede elétrica.

Sugere-se a instalação de um gerador de energia elétrica, atendendo paradas superiores a autonomia das baterias.

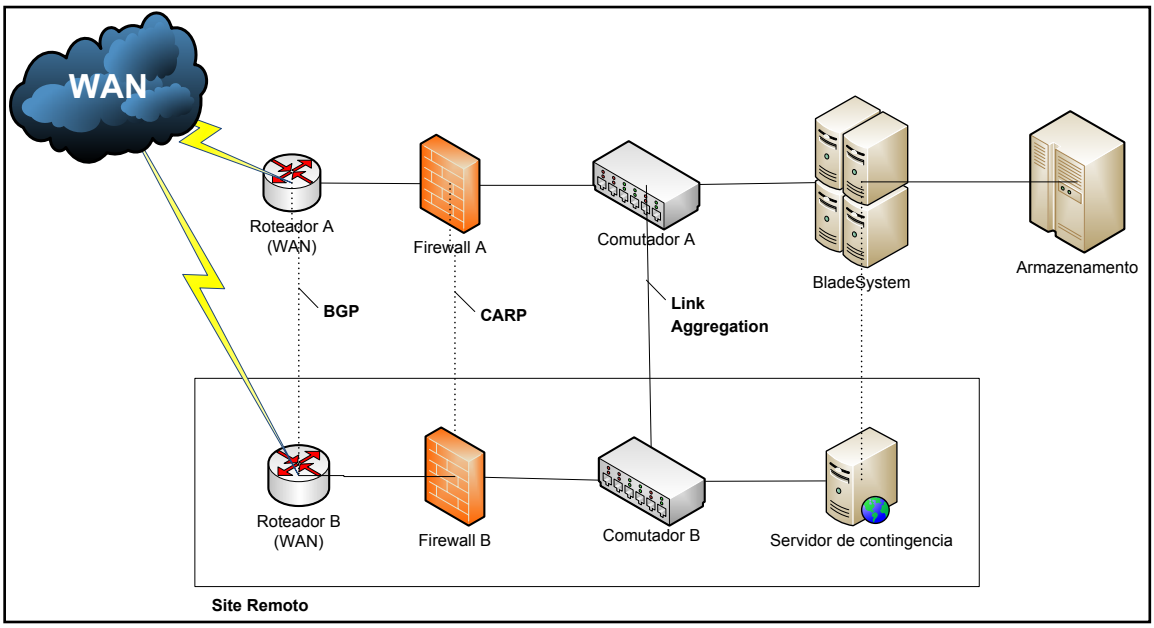

Figura 15 - Diagrama do cenário planejado

3.4 ALTERAÇÕES REALIZADAS

Esta seção descreve as alterações realizadas para cada um dos pontos únicos de falha identificados no item 3.3, resultando em um novo cenário doravante chamado Cenário Atual.

A seguir, no Quadro 6, são apresentadas, de forma sintetizada, as alterações e os investimentos realizados.

Recurso	Alterações realizadas	Valor investido
Enlace externo	Instalado novo enlace redundante.	R$ 1.800,00/mês
Firewall / roteador	Instalados dois novos firewalls redundantes.	R$ 0,00

Firewall / roteador	Configuração das conexões de rede	R$ 0,00
Enlace interno	Instalados dois novos comutadores de núcleo redundantes.	R$ 74.000,00
Servidores	Instalado novo servidor, replicando o de produção.	R$ 0,00
Sistema de armazenamento em rede (*storage*)	Nenhuma alteração necessária.	-
Virtualizador	Nenhuma alteração necessária	-
Fornecimento de energia elétrica	Não foi possível firmar o acordo de nível de serviço conforme esperado.	-
Fornecimento de energia elétrica	Não foi possível adquirir nova solução redundante para UPS.	-
Disposição do ambiente	Nenhuma alteração necessária	-
Climatização	Nenhuma alteração necessária	-
Backup e recuperação	Nenhuma alteração necessária	-

Quadro 6 - Alterações realizadas e respectivos investimentos.

Os itens apresentados no Quadro 6 são os mesmos que antes apresentados no Quadro 5, entretanto, antes referenciados como alterações necessárias passam a ser listadas como alterações realizadas e respectivos desembolsos.

3.4.1 Rede

3.4.1.1 Enlace externo

A concessionária de enlace externo já existente entregou um segundo enlace redundante. Este possui origem e caminho físico distinto até seu destino final, chegando ao Centro Tecnológico da Prefeitura Municipal de Itajaí pela rua Alberto Werner, conforme ilustrado na Figura 16.

Figura 16 – Redundância do enlace externo pode caminhos diferentes.
Fonte: Adaptado de Google (2012)

Depois da alteração, hoje há dois enlaces com origem em diferentes meios físicos que oferecem a mesma conectividade IP, isto foi possível devido à ativação do protocolo BGO nesta contratação de tal modo que um sistema autônomo (AS) foi operacionalizado conforme planejado.

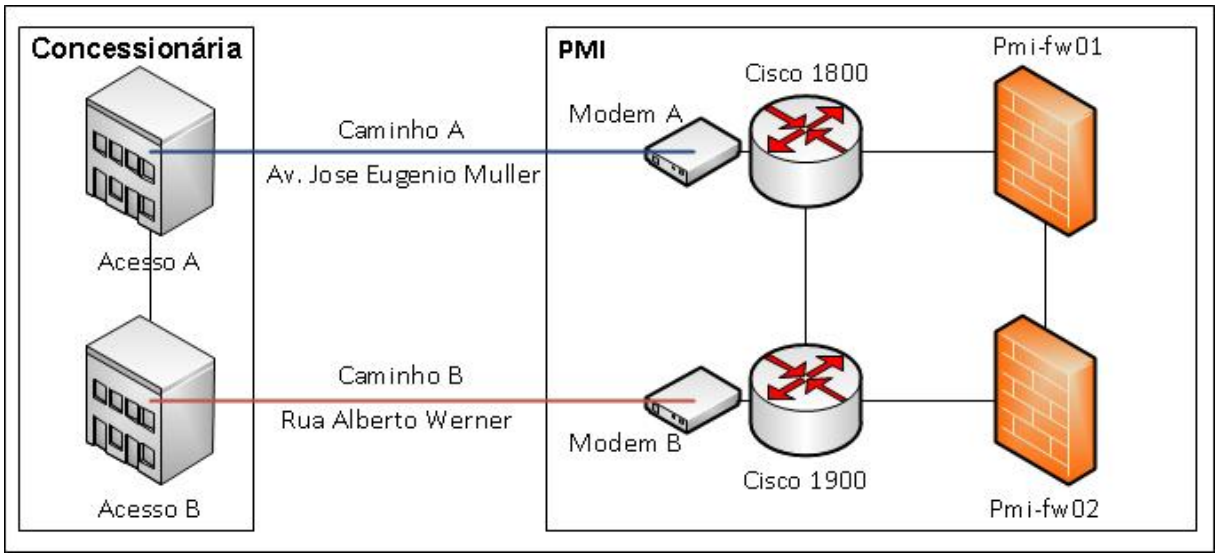

Figura 17 - Novo enlace externo redundante.

A Figura 17 ilustra o novo enlace externo, agora redundante. Cada um dos enlaces é conectado a um modem óptico, Overtek E8110T. Os modens por sua vez conectam-se a roteadores Cisco 1800/1900 series. Portanto, cada enlace possui um conjunto de modem óptico e roteador conectando-os aos novos Firewalls. As conexões entre os roteadores da concessionária e os firewalls ocorrem por cabos padrão ethernet 5e.

3.4.1.2 Firewall

A instalação de uma nova solução para *firewall* redundante, foi realizada reutilizando dois computadores servidores, idênticos, de modelo Sun Fire™ V20z Server, já existentes e que estavam sem uso. Desta forma, não onerando o município com a aquisição de novos equipamentos.

Os servidores possuem entre suas principais características, processador AMD Opteron 200 Series de 2.4Ghz e 1MB de Cache L2, memória DDR ECC com 4GB RAM, dois discos *hot-swap* Ultra320 SCSI de 73GB e 10K RPM, implementando Raid1.

Após a instalação do sistema operacional FreeBSD 9.0 64 bits, em ambos os servidores, o *kernel* foi recompilado, transformando-o em *stable,* garantindo uma maior confiabilidade e implementando os módulos necessários para uso do protocolo CARP e pfsync, conforme apresentado no Quadro 7.

```
...
device          carp
device          pflog
device          pfsync
device          pf
...
```

Quadro 7 – Arquivo de configurações do *kernel*: /sys/amd64/conf/PMI

O quadro anterior apresentou parte do arquivo de configuração do *kernel*, acrescido da primeira linha, que juntamente com as demais de mesma categoria, habilitam o protocolo CARP e o módulo de sincronização pfsync ao sistema operacional.

Problemas enfrentados

Visando otimizar a performance da rede, juntamente com a implementação do protocolo de redundância CARP, foram habilitados recursos de balanceamento de carga (*load balance*) que utilizam um algoritmo de escalonamento *Round-Robin*, distribuindo as conexões e carga de trabalho entre os dois servidores. Entretanto, os sistemas legados de uso da Prefeitura Municipal em seus setores internos, implementam um modelo de conexão com o banco de dados, fazendo que esta seja estabelecida ao iniciar e mantendo-a persistente ao longo de seu tempo de execução.

Essas conexões estabelecidas eram interrompidas quando o roteamento do firewall distribuía a carga de trabalho, ocasionando erros e encerrando o sistema de forma inesperada.

Solução empregada

Foi adotado o modelo de *failover* para a sincronização dos servidores que atuam como firewall e roteadores internos. Desta forma, o pmi-fw01 foi definido como principal ou *master*, e o pmi-fw02 definido como secundário ou *slave*. O servidor *master* ficou responsável por receber todo o tráfego de rede e intermediar as conexões entre segmentos. Quando ocorrer uma falta em algum dos segmentos, o servidor *slave* automaticamente recebe o estado de *master* naquele segmento e passa a responder pelo endereço IP compartilhado. Desta forma resultando em uma perda mínima de pacotes de rede.

Os novos firewalls foram configurados utilizando *hostnames* e seus endereços de IP privados, conforme apresentado no Quadro 8.

Hostname	**Preferencia**	**IP privado (LAN)**	**IP privado (DMZ)**	**IP privado (WAN)**
pmi-fw01	Principal	10.1.1.253/16	10.10.1.253/16	187.44.99.73/29
pmi-fw02	Secundário	10.1.1.252/16	10.10.1.252/16	187.44.99.72/29

Quadro 8 - Configurações novos *firewalls*

Os novos servidores utilizam pmi-fw01 e pmi-fw02 respectivamente como seus nomes de rede ou *hostnames*. Cada um deles possui um endereço IP único e exclusivo para cada segmento de rede, denominado IP privado. A denominação privado não faz referencia ao IP não válido, de uso somente em rede local. Os endereço IP do segmento WAN possui acesso a partir de qualquer dispositivo conectado à internet, ou seja, trata-se de um IP válido. O protocolo de compartilhamento CARP, possibilita à dois ou mais dispositivos compartilharem o mesmo endereço IP, assim denominado IP público ou compartilhado.

Anteriormente, cada segmento era conectado a sua única interface física de rede. Desta forma a vazão do segmento estava limitado a vazão da interface.

Na solução implementada, os segmentos foram separados por redes virtuais (VLAN) e as interfaces dos firewalls foram agrupadas com a implementação do protocolo *link-aggregation*. Desta forma, recebendo o trafego de todos os segmentos e distinguidos por suas respectivas identificações (*tags*).

O novo modelo proporciona redundância de interfaces e se necessário uma maior vazão em algum dos seus segmentos, conforme ilustrado na Figura 18.

Anteriormente, as interfaces eram físicas e independente uma das outras. Caso falhar uma desses ou suas conexões, o segmento falhado ficaria indisponível. No modelo atual, as interfaces dos segmentos locais são agrupadas. Desta forma proporcionando redundância entre interfaces e agregação da largura de banda e vazão.

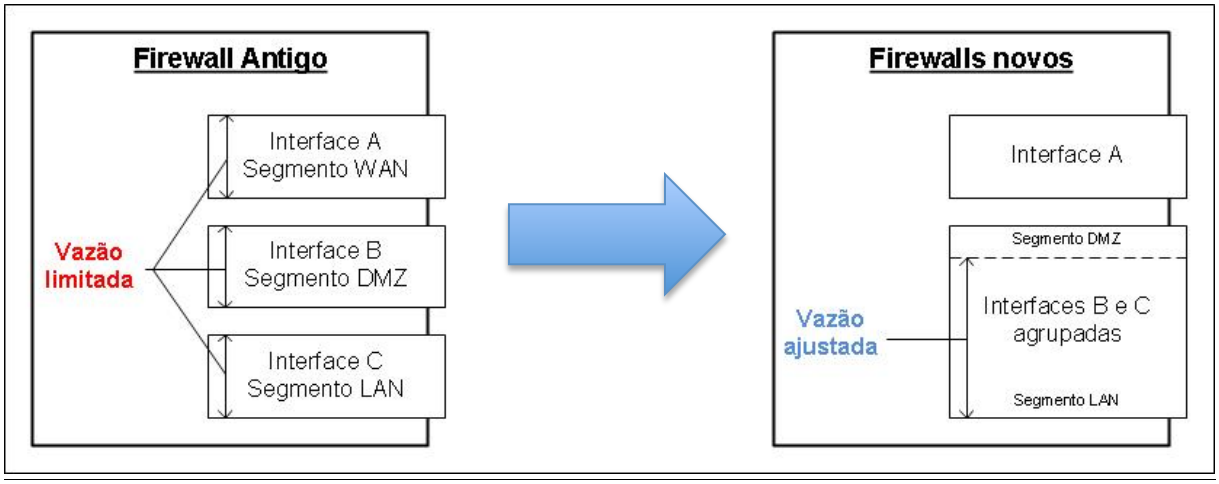

Figura 18 – Interfaces dos novos firewalls

Com a ausência de segmentação física, as interfaces locais passam a receber todo tráfego de rede por meio das mesmas conexões físicas, entretanto, mantendo a distinção lógica dos segmentos. Atendendo as necessidades, foram criadas interfaces virtuais capazes de identificar e separar de maneira segura o tráfego de cada segmento de rede. O Quadro 9 exibe os segmentos pertinentes ao projeto e seus tipos de interface.

Segmento	**Tipo de interface**	**Identificação**
DMZ	virtual	vlan6
LAN	virtual	vlan7
WAN	física	re0

Quadro 9 - Segmentos de rede e identificações

Apenas os segmentos locais foram agrupados e transformados em interfaces virtuais, mas não por isso que o segmento de rede WAN ficará menos seguro. Este último assume a redundância entre os firewalls, que possuem uma interface cada de uso exclusivo.

Agregação de interfaces no FreeBSD

A configuração do sistema operacional FreeBSB, teve por seu primeiro passo o agrupamento das interfaces físicas que são utilizadas para o segmentos locais, para que estas possam responder por uma única. O Quadro 10 apresenta os parâmetros necessários em /etc/rc.conf para ativar a agregação. Devendo estes, serem configurados em ambos os servidores que atendem por firewall.

```
ifconfig_bge0="up"
ifconfig_bge1="up"
ifconfig_lagg0="up"
ifconfig_lagg0="laggproto lacp laggport bge0 laggport bge1"
```

Quadro 10 – Configuração de *link-aggregation* em ambos os servidores.

As interfaces são inicializadas utilizando os parâmetros ifconfig_bge0=”up” e ifconfig_bge1=”up”. Após inicializar as interfaces, uma nova interface agregada é criada e inicializada através do parâmetro ifconfig_lagg0=”up”.

Em seguida, deve-se informar qual o protocolo e quais portas físicas devem ser utilizadas para a agregação. Para isso, usa-se o parâmetro ifconfig_lagg0="laggproto lacp laggport bge0 laggport bge1".

Configuração de interfaces virtuais

A configuração das interfaces virtuais acontecem no arquivo /etc/rc.conf, devendo estas serem realizadas em ambos os servidores que atendem por firewall.

O Quadro 11 apresenta as configurações utilizadas no pmi-fw01 para configurar suas interfaces virtuais e separação lógica de seus segmentos.

```
#DMZ
ifconfig_vlan6="inet 10.10.1.253 netmask 255.255.0.0 vlan 6 vlandev lagg0"

#LAN
ifconfig_vlan7="inet 10.1.1.253 netmask 255.255.0.0 vlan 7 vlandev lagg0"
```

Quadro 11 – Configuração das interfaces virtuais em pmi-fw01.

Os parâmetros ifconfig_vlan<id> recebem como argumento o endereço IP por qual a interface virtual responde, sua máscara de rede, qual o ID da rede virtual (ou tag da VLAN) e por fim o dispositivo de rede. Para os segmentos locais, utiliza-se a mesma interface de rede, a lagg0.

O Quadro 12 apresenta as configurações utilizadas no pmi-fw02 para configurar suas interfaces virtuais e separação lógica de seus segmentos.

```
#DMZ
ifconfig_vlan6="inet 10.10.1.252 netmask 255.255.0.0 vlan 6 vlandev lagg0"

#LAN
ifconfig_vlan7="inet 10.1.1.252 netmask 255.255.0.0 vlan 7 vlandev lagg0"
```

Quadro 12 – Configuração das interfaces virtuais em pmi-fw02.

Os parâmetros apresentados são os mesmo utilizados anteriormente para o pmi-fw01, porém os endereços IP foram alterados de acordo com os valores previamente definidos para uso exclusivo.

Configuração do protocolo CARP

A configuração do protocolo de redundância CARP, ocorreu em ambos os servidores de forma idêntica, com exceção de seus *hostnames* e VHIDs (Identificação do host virtual). Após criar as interfaces virtuais, o protocolo CARP foi configurado para compartilhar os endereços IP públicos que serão acessados pelos demais dispositivos de rede, sejam estes nos segmentos internos ou através da internet. O Quadro 13 mostra as configurações realizadas no pmi-fw01 em seu /etc/rc.conf.

```
 #WAN
 ifconfig_carp9="vhid 10 advbase 20 advskew 0 pass Ctim123 187.44.99.66/28"

 #LAN
 ifconfig_carp11="vhid 12 advbase 20 advskew 0 pass Ctim123 10.1.1.254/16"

 #DMZ
 ifconfig_carp100="vhid   101   advbase   20   advskew   0   pass   Ctim123
10.10.1.254/16"
```

Quadro 13 – Configurações de CARP em pmi-fw01.

Os parâmetros ifconfig_carp<id> recebem os argumentos vhid, advbase, advskew, pass e endereço IP compartilhado. Seus valores são detalhados a seguinte:

- vhid: Número inteiro utilizado para identificar o agrupamento.
- advbase: Tempo (em segundos) que o host envia um sinal por broadcast para informar seu estado em relação aos demais.
- advskew: Utilizado para definir a prioridade entre os servidores. O menor terá preferencia sobre os demais.

- pass: Senha utilizada para a sincronização dos hosts. Devendo ser idêntica entre os hosts de mesmo grupo.

O firewall secundário pmi-fw02 foi configurado para o compartilhamento de endereço IP por protocolo CARP conforme apresentado no Quadro 14.

```
#WAN
ifconfig_carp9="vhid 10 advbase 20 advskew 1 pass Ctim123 187.44.99.66/28"

#LAN
ifconfig_carp11="vhid 12 advbase 20 advskew 1 pass Ctim123 10.1.1.254/16"



#DMZ
ifconfig_carp100="vhid 101 advbase 20 advskew 1 pass Ctim123
10.10.1.254/16"
```

Quadro 14 – Configurações de CARP em pmi-fw02.

Os parâmetros são próximos aos configurados anteriormente em pmi-fw01, alterando sua prioridade em relação ao anterior.

A segmentação da rede através de redes virtuais, somente foi possível de implementação por existir comutadores compatíveis e capazes de repassar o trafego de diferentes segmentos para portas específicas. Esses por sua vez estão posicionados no enlace interno e detalhado na seção seguinte.

3.4.1.3 Enlace Interno

Foram instalados dois novos comutadores de núcleo, marca Extreme Summit x460. Os novos dispositivos L3, entre suas principais características destacam-se a densidade de portas de alta performance, tecnologia de empilhamento e alimentado pelo sistema operacional ExtremeXOS®. Suportando a capacidade empilhamento por portas específicas, com tecnologia proprietária para empilhamento de comutadores com velocidade de 80 Gbps. Além das portas de empilhamento o x460 possui portas de 10 Gigabit Ethernet e SFP+ para conectividade com *hosts* e servidores.

Alta disponibilidade é algo que está fortemente ligado a essa linha de produtos, através de recursos presentes entre *link aggregation*, núcleo de implantação de *backbone*, conexões para energia elétrica para ambientes AC (corrente alternada) e DC (corrente contínua). O sistema operacional ExtremeXOS oferece alta disponibilidade e simplicidade com um sistema operacional em toda a rede.

Os recursos presentes nos novos comutadores de núcleo, proporcionam a interconexão entre dispositivos, fazendo que estes apareçam para a rede como um único.

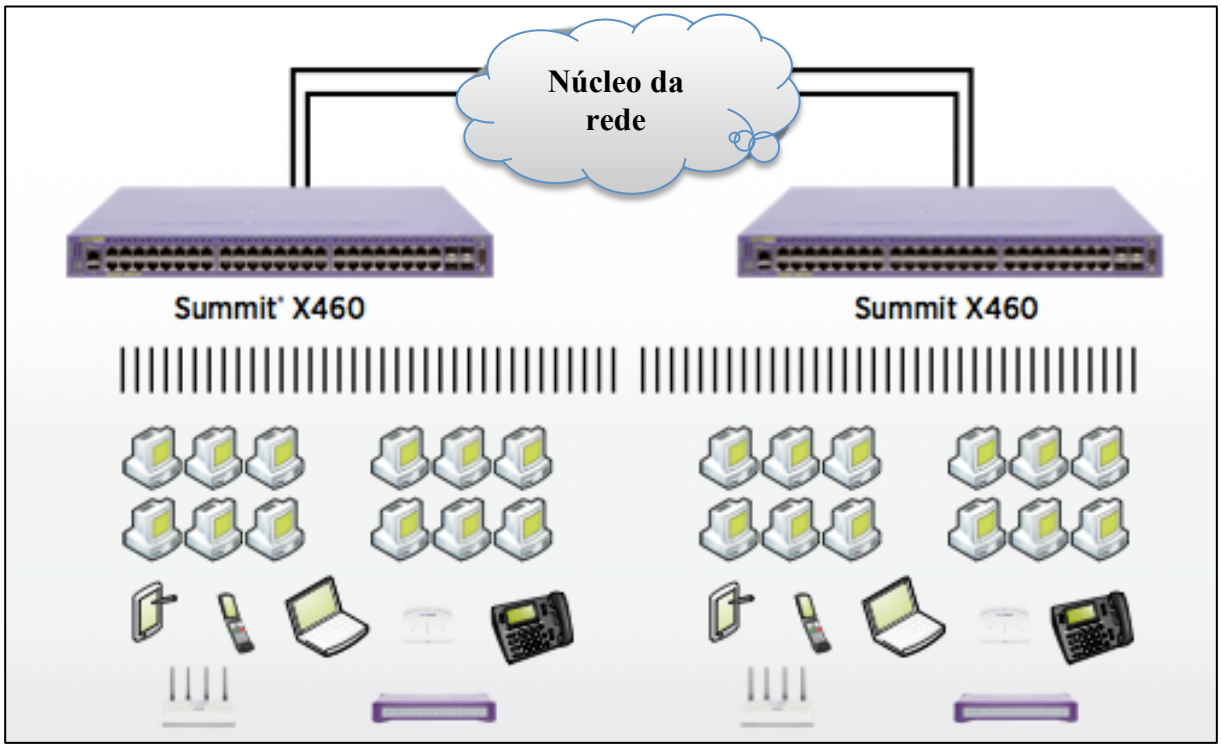

Figura 19 - Interligação dos comutadores de núcleo
Fonte: Adaptado de Extreme (2012).

A interligação entre os comutadores foi realizada através de suas portas para uso específico, que utilizam tecnologia proprietária e de fácil gerência através de seu sistema operacional, conforme ilustrado na Figura 20. Desta forma sem exigir grandes esforços de implementação e configuração.

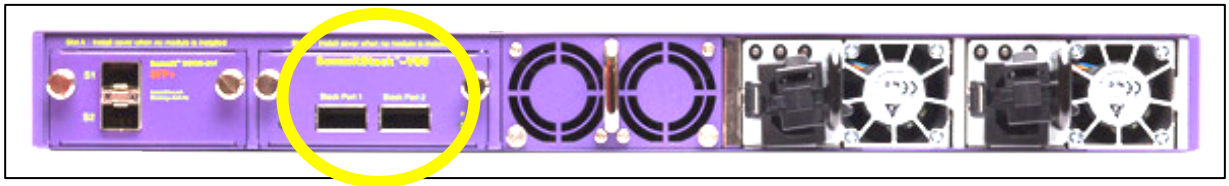

Figura 20 - Visão traseira dos comutadores de núcleo destacando portas para empilhamento.
Fonte: Adaptado de Extreme (2012).

A configuração dos segmentos de rede apresentados ao Firewall e também aos servidores, de forma agregada e separados por redes virtuais, é detalhada nos quadros seguintes.

```
#DMZ
create vlan PMI-DMZ description `VLAN-DMZ` tag 6

#LAN
create vlan PMI-LAN description `VLAN-LAN` tag 7
```

Quadro 15 – Criando vlan nos comutadores de núcleo.

Acessando o console de configuração, os comandos anteriores foram utilizados para criar os segmentos DMZ e LAN. Os argumentos vlan, description e tag são detalhados a seguite:

- vlan: Uma espécie de chave primária utilizado para identificar o segmento de rede.
- description: Descrição detalhada do segmento de rede.
- tag: Valor inteiro utilizado para marcar os pacotes de rede, informando a qual rede virtual este pertence.

Após configurar as redes virtuais nos comutadores de núcleo, foi necessário adicionar as portas que devem receber o tráfego de redes virtuais. A configuração das porstas realizadas no console de configuração é apresentada no Quadro 16.

```
#DMZ
Configure vlan PMI-DMZ add ports 1:27, 1,28, 2:27, 2:28 tagged

#LAN
Configure vlan PMI-LAN add ports 1:27, 1,28, 2:27, 2:28 tagged
```

Quadro 16 – Configurando portas para tráfego de redes virtuais.

Os parâmetros utilizados para configuração das portas, incluem vlan, add ports e tagged. A utilização de cada é detalhada a seguir:

- vlan: A identificação da rede virtual que será configurada.
- add ports: Identifica as portas que devem receber o trafego das redes virtuais. A identificação 1:27 refere-se ao comutador 1 e porta 27, 2:27 por sua vez refere-se a porta 27 do comutador 2.
- tagged: Determina que as portas informadas receberam apenas pacos marcados. Os que não atenderem a exigência devem ser descartados.

Apresentados à rede de forma transparente, como um único comutador de núcleo, os computadores são conectados aos Servidores *Blade* por 4 cabos SFP+ de 10 Gbps, vide

Figura 21. Partindo dois cabos de cada comutador e conectados a cada um dos HP VirtualConnect. Desta forma proporcionando redundância de 4/1 e *link-aggregation* de até 40 Gbps.

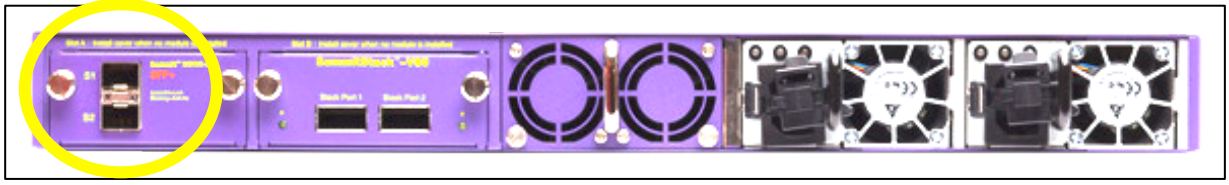

Figura 21 - Visão traseira dos comutadores de núcleo destacando portas 10 Gbps SFP+.
Fonte: Adaptado de Extreme (2012).

3.4.2 Servidores

O novo servidor backup, com replicação assíncrona foi instalado em uma máquina virtual com características e configuração idênticas ao servidor de produção. Em modelo *failover*, caso ocorrer uma indisponibilidade do servidor de produção, os acessos são direcionados ao servidor backup. Sua implementação foi realizadq utilizando o sistema Operacional Linux CentOS, em plataforma de 64bits.

Assumindo como pré-requisito para a instalação dos servidores em *failover*, o sistema operacional Linux, configurações de rede, máquina virtual Java e servidor de aplicações Java JBoss. Todos devidamente instalados e configurados.

Configurações servidor de produção

No servidor principal, foram realizados ajustes no arquivo de configuração do servidor de aplicações Java JBoss, em jbm-configuration.xml, informando a existência e parâmetros de seu servidor de backup. Informação essa realizada através da especificação de um elemento `backup-connector-ref`. Este elemento referencia um conector, atribuindo-o informações de como conectar-se ao servidor de backup. O Quadro 17 apresenta um trecho do jbm-configuration.xml no servidor de produção, que foi configurado com o servidor principal.

```
<backup-connector-ref connector-name="backup-connector"/>
<!-- Connectors -->
<connectors>
   ...
<!-- Este conector especifica como conectar-se ao servidor de backup -->
   <connector name="backup-connector">
     <factory-class>
org.jboss.messaging.integration.transports.netty.NettyConnectorFactory
     </factory-class>
     <param key="jbm.remoting.netty.port" value="5445" type="Integer"/>
```

```
    </connector>
</connectors>
```

Quadro 17 – Configurações JBoss no servidor principal

Os parâmetros apresentados no Quadro 17, configuram o método de conexão com o servidor backup e qual porta deve realizar a conexão.

No servidor backup, o arquivo de configuração jbm-configuration.xml deve conter o elemento backup que por sua vez deve estar definido como *true*, conforme apresentado no Quadro 18.

```
<backup>true</backup>
```

Quadro 18 – Configuração JBoss no servidor backup

O elemento presente no Quadro 18 apenas informa ao servidor JBoss que estará trabalhando como um backup de outro servidor.

Por fim, após realizar as configurações de *failover* nos servidores de aplicação do Sistema Emissor de Notas Fiscais de Serviço, o sistema de arquivos que armazena os arquivos de lote transferidos, precisam ser sincronizados. Para sincronização dos sistemas de arquivos utiliza-se o rsync.

```
# Nome da sincronizaçãoo
[nfse]
# caminho de destino
path = /lotes
# Hosts que podem copiar a informação
hosts allow = 10.10.1.35
hosts deny = *
list = true
uid = root
gid = root
read only = false
```

Quadro 19 – Configuração rsync no servidor backup em /etc/rsyncd.conf

As configurações presentes no servidor backup e apresentadas no Quadro 19, ajustam o rsync à permitir sincronizar o conteúdo originado pelo *host* com endereço IP 10.10.1.35 e destino /lotes.

No servidor de produção foi agendado um tarefa, executando a sincronização de tempos em tempos, conforme apresentado no Quadro 22.

```
# Adicionando agendamento para sincronização dos dados
[root@nfse ~]#
crontab –e

# executar de hora em hora
00 * * * * rsync -avz /lotes/ 10.10.1.36::nfse
```

Quadro 20 – Sincronização no servidor de produção

O agendamento antes exibido, executa a toda hora e com privilégio de *root* o comando "`rsync -avz /lotes/ 10.10.1.36::nfse`".

3.5 PROBLEMAS E DIFICULDADES

A solução de alta disponibilidade proposta foi possível de implementação e desenvolvimento em sua totalidade, em função da ocorrência de eventos não previstos anteriormente e que serão descritos nessa sessão.

3.5.1 Fornecimento de energia

Após identificar pontos únicos de falha no fornecimento de energia elétrica, em virtude da existência de um único sistema para suplemento de energia ininterrupto (UPS). Foi proposto como solução a instalação de novos UPS redundantes, capazes do prover autonomia para períodos superiores a 3 horas de indisponibilidade. Desta forma cada PDU existente nos raques deveriam ser conectadas a UPS distintos.

Entretanto, em função deste ano possuir processo eleitoral, amparado no artigo 42 da Lei 9.504/97, ficou impossibilitado a abertura de processo licitatório para aquisição de uma solução UPS redundante. Conforme apresentado na integra a seguir.

> "Art. 42. É vedado ao titular de Poder ou órgão referido no art. 20, nos últimos dois quadrimestres do seu mandato, contrair obrigação de despesa que não possa ser cumprida integralmente dentro dele, ou que tenha parcelas a serem pagas no exercício seguinte sem que haja suficiente disponibilidade de caixa para este efeito."

Contudo, conforme lei, por não existir caixa ou previsão orçamentária para o exercício 2012. Diferentemente dos comutadores de núcleo, não foi possível adquirir os UPS necessários para a execução total deste projeto.

A concessionária de energia elétrica (CELESC) negou firmar um SLA, comprometendo-se manter um período máximo para a restauração dos serviços em caso de

falhas. Negação esta justificada pela necessidade da Prefeitura Municipal de Itajaí investir em obras de melhorias na região, como condição para que o acordo pudesse ser firmado.

3.6 TESTES

Nesta seção é apresentada a sequencia de testes realizadas na solução implantada, aferindo a disponibilidade ao Sistema Emissor de Notas Fiscais de Serviço.

Foram realizados uma sequência total de 25 testes, distribuídos estes entre enlace externo, firewall, enlace interno e UPS. Todos os recursos modificados com a implantação da solução proposta foram testados e seus resultados obtidos foram confrontados com os resultados esperados, desta forma assumiu-se sucesso ou fracasso.

Visando monitorar a disponibilidade global resultante da solução proposta por este documento, foi utilizada a ferramenta externa host-tracker.com. Seu resultado é apresentado no final deste capítulo.

Todas as sequencias de testes relatadas nas subseções seguintes, foram realizadas no dia 30 de outubro. Porém, antes e durante a implantação dos novos recursos, foram realizados testes de validação que não são apresentados nas sequencias seguintes.

3.6.1 Enlaces externo

No cenário antes existente, as falhas no enlace externo foram as mais frequentes e recorrentes, conforme apresentadas na Tabela 1. Portanto, os testes da nova solução redundante são os que produzem maior expectativa com seus resultados. Afinal o enlace externo é o responsável pelo acesso e entrega das funcionalidades aos usuários externos.

Para a execução dos testes com os enlaces redundantes, foram posicionados dois computadores, que monitoraram a funcionalidade dos enlaces e disponibilidade do sistema emissor. O primeiro, conectado à internet por meio de telefonia móvel 3G e independente de qualquer recurso da Prefeitura de Itajaí. Já o segundo, conectado à rede local.

Ambos os computadores acessaram o sistema emissor de notas fiscais de serviço, mantendo-o aberto enquanto disparando pings por ICMP ao endereço IP 187.44.99.66, resolvido pelo *hostname* nfse.itajai.sc.gov.br.

Teste	Resultado esperado	Resultado apresentado	Situação
Desconectar conexão de fibra óptica do enlace A	Respostas dos pings sem perda de pacotes. Continuidade dos acessos ao sistema emissor.	Respostas dos pings sem perda de pacotes. Os acessos mantiveram-se em perfeito funcionamento, sem perdas ou falhas.	Sucesso.
Desconectar conexão de fibra óptica do enlace B	Respostas dos pings sem perda de pacotes. Continuidade dos acessos ao sistema emissor.	Respostas dos pings sem perda de pacotes. Os acessos mantiveram-se em perfeito funcionamento, sem perdas ou falhas.	Sucesso.
Desligar roteador Cisco 1800 series, que atende ao enlace A.	Respostas dos pings sem perda de pacotes. Continuidade dos acessos ao sistema emissor.	Respostas dos pings sem perda de pacotes. Os acessos mantiveram-se em perfeito funcionamento, sem perdas ou falhas.	Sucesso.
Desligar roteador Cisco 1900 series, que atende ao enlace B.	Respostas dos pings sem perda de pacotes. Continuidade dos acessos ao sistema emissor.	Respostas dos pings sem perda de pacotes. Os acessos mantiveram-se em perfeito funcionamento, sem perdas ou falhas.	Sucesso.

Quadro 21 - Testes com enlaces externo redundantes.

O novo enlace externo alcançou o sucesso em todos os testes realizados. Estes, desconectando fisicamente seus cabos de dados e posteriormente energia elétrica.

3.6.2 Firewall

Para execução dos testes dos novos firewalls, foram disparados pings por meio de um dispositivo externo ao host nfse.itajai.sc.gov.br, que atende ao endereço IP 187.44.99.66 e também por meio de um dispositivo posicionado no segmento de rede local ao host nfse.itajai.local, que atende ao endereço IP 10.10.1.35.

Teste	Resultado esperado	Resultado apresentado	Situação
Desconectar conexão de	Respostas dos pings sem	Respostas dos pings sem	Sucesso.

rede 01 do pmi-fw01	perda de pacotes.	perda de pacotes.	
Desconectar conexão de rede 02 do pmi-fw01	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso.
Desconectar conexão de rede 03 do pmi-fw01	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso.
Desconectar conexões de rede 01 e 02 do pmi-fw01	Respostas dos pings sem perda de pacotes.	Resposta dos pings com perda mínima de pacotes.	Sucesso.
Desconectar conexões de rede 02 e 03 do pmi-fw02	Respostas dos pings sem perda de pacotes.	Resposta dos pings com perda mínima de pacotes	Sucesso.
Desconectar conexão de rede 01 do pmi-fw02	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso.
Desconectar conexão de rede 02 do pmi-fw02	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso.
Desconectar conexão de rede 03 do pmi-fw02	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso.

Quadro 22 - Testes das conexões redundantes dos novos firewalls / roteadores.

Os testes apresentados no quadro anterior, foram realizados simulando eventuais problemas físicos que possam vir a ocorrer em suas conexões. Para isso, foram fisicamente desconectados e desta forma alcançando o sucesso.

Os testes de redundância por compartilhamento de endereços IP através da implementação do protocolo CARP foram realizados conforme Quadro 23, disparando pings icmp aos endereços IP em cada segmento de rede, entre eles local 10.1.1.254, DMZ 10.10.1.254 e WAN 187.44.99.66.

Teste	**Resultado esperado**	**Resultado apresentado**	**Situação**
Desativar vlan6 no pmi-fw01.	Respostas dos pings sem perda de pacotes. O pmi-fw02 deve chavear o estado para master e responder as solicitações.	Respostas dos pings com perda mínima, 01 pacote no IP 10.1.1.254. O pmi-fw02 assumiu com o papel de *master*.	Sucesso.
Desativar vlan7 no pmi-fw01.	Respostas dos pings sem perda de pacotes. O pmi-fw02 deve chavear o estado para master e	Respostas dos pings com perda mínima, 01 pacote no IP 10.10.1.254. O pmi-fw02 assumiu com o	Sucesso.

	responder as solicitações.	papel de *master*.	
Desativar re0 no pmi-fw01.	Respostas dos pings sem perda de pacotes. O pmi-fw02 deve chavear o estado para master e responder as solicitações.	Respostas dos pings com perda mínima, 01 pacote no IP 187.44.99.66	Sucesso.
Desativar vlan6 no pmi-fw02.	Respostas dos pings sem perda de pacotes. O pmi-fw01 deve retornar ao estado de máster e responder as solicitações.	Respostas dos pings com perda mínima, 01 pacote no IP 10.1.1.254	Sucesso.
Desativar vlan7 no pmi-fw02.	Respostas dos pings sem perda de pacotes. O pmi-fw01 deve retornar ao estado de máster e responder as solicitações.	Respostas dos pings com perda mínima, 01 pacote no IP 10.10.1.254	Sucesso.
Desativar re0 no pmi-fw02.	Respostas dos pings sem perda de pacotes. O pmi-fw01 deve retornar ao estado de máster e responder as solicitações.	Respostas dos pings com perda mínima, 01 pacote no IP 187.44.99.66	Sucesso.
Desconectar o cabo de energia do pmi-fw01 ocasionado uma parada inesperada.	Respostas dos pings sem perda de pacotes. O pmi-fw02 deve chavear o estado para master todos os seguimentos.	Respostas dos pings com perda mínima, 01 pacote em todos os seguimentos de rede.	Sucesso.
Desconectar o cabo de energia do pmi-fw02 ocasionado uma parada inesperada.	Respostas dos pings sem perda de pacotes. O pmi-fw01 deve chavear o estado retornando o estado master em todos os seguimentos.	Ao religar o pmi-01, desligado no teste anterior, automaticamente retornou o estado de máster em todos os seguimentos de rede com perda mínima, 01 pacote em todos os seguimentos. Deligando o pmi-fw02 todos os pacotes foram respondidos em perdas.	Sucesso.

Quadro 23 - Testes dos endereços IP compartilhados por protocolo CARP

Os testes realizados e apresentados no quadro anterior, buscaram aferir o funcionamento do protocolo de redundância CARP. Desconectando as conexões físicas e

desativar as interfaces virtuais através de comandos ifconfig <interface> down e posteriormente religa-las através de comandos ifconfig <interface> up.

3.6.3 Enlace interno

Os testes do enlace interno foram realizados e acompanhados disparando pings icmp aos endereços IP em cada segmento de rede, entre eles local 10.1.1.254, DMZ 10.10.1.254 e WAN 187.44.99.66.

Teste	Resultado esperado	Resultado apresentado	Situação
Desconectar cabos de energia elétrica do comutador de núcleo 01.	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso
Desconectar cabo de energia elétrica do comutador de núcleo 02.	Respostas dos pings sem perda de pacotes.	Respostas dos pings sem perda de pacotes.	Sucesso.

Quadro 24 - Testes do enlace interno

Os testes nos comutadores de núcleo, que representam o enlace interno, foram realizados representando sua falha máxima. Desta forma simulados por desconectar seus cabos de energia elétrica.

3.6.4 Servidores

Para execução dos testes dos servidores foram disparados pings por meio de um dispositivo externo ao host nfse.itajai.sc.gov.br, que atende ao endereço IP 187.44.99.66 e também por meio de um dispositivo posicionado no segmento de rede local ao host nfse.itajai.local, que atende ao endereço IP 10.10.1.35.

Após a realização de cada etapa de teste, um acesso ao sistema emissor deve ser realizado validando a disponibilidade do sistema. Um nota fiscal de teste deve ser emita e após religar o servidor testado a sincronização dos dados é verificada.

Teste	Resultado esperado	Resultado apresentado	Situação
Desligar a máquina virtual que atende ao servidor de produção do	Respostas dos pings sem perda de pacotes. Acesso ao sistema com	Resposta com perda mínima dos pings. A conexão existente foi	Sucesso.

sistema emissor.	sucesso. Sucesso ao emitir nova nota fiscal.	interrompida e necessário novamente conectar ao sistema.	
Religar máquina virtual que atende ao servidor de produção.	Respostas dos pings sem perda de pacotes. Acesso ao sistema com sucesso. Sucesso ao emitir nova nota fiscal. Sucesso na replicação dos dados.	Resposta com perda mínima dos pings. A conexão existente foi interrompida e necessário novamente conectar ao sistema.	Sucesso.

Quadro 25 - Testes dos servidores

Os testes realizados nos servidores, simularam uma parada geral do servidor de produção, desligando-os de forma inesperada. Antes de desligar o servidor, foi aberto uma sessão e esta veio a travar logo após desligar. Porém, logo ao fechar o navegador da internet e abrir o site nfse.itajai.sc.gov.br a aplicação já estava disponível para acessos. Desta forma classificou-se como sucesso.

3.6.5 Fornecimento de energia elétrica

O teste do UPS foi realizado desconectando-o da rede elétrica e monitorando o funcionamento dos sistemas a partir dos segmentos locais e via internet.

Teste	**Resultado esperado**	**Resultado apresentado**	**Situação**
Desligar o UPS da rede de energia elétrica	Manter fonte secundária de energia por período superior a 3 horas corridas.	O UPS foi reconectado após 20 min. de teste. Mantendo todos os sistemas em funcionamento.	Sucesso.

Quadro 26 – Teste do UPS

Os testes no UPS foram realizados por apenas 20 minutos, não podendo verificar o tempo máximo provido por módulos de baterias. Enquanto realizavam-se os testes com o UPS, foram acompanhados os valores apresentados pelo monitoramento em seu ambiente gerencial. Valores esses que indicam o consumo atual (em *watts*), consumo médio dos últimos minutos, carga dos módulos de baterias e sua projeção de autonomia (em minutos). A autonomia apresentada baseia-se nos valores médios de consumo e carga atual dos módulos de baterias.

Embora a autonomia apresentada para oferta de energia alternativa por módulos de baterias era próxima à 3 horas, os testes foram interrompidos aos 20 minutos para que não prejudicassem as baterias. Também não correndo o risco de uma parada total do sistema de forma inesperada, que desta forma prejudicaria também os demais sistemas que estavam em produção e não são relatados por este trabalho.

3.6.6 Host-Tracker

O resultado aferido e acompanhado pelo sistema externo host-tracker.com, conforme Figura 22, apresentou a disponibilidade de 99,90% para o Sistema Emissor de Notas Fiscais de Serviço. O monitoramento teve início a partir do período de testes, totalizando dois meses.

O Host-Traker foi configurado para realizar verificações periódicas, através de diferentes origens geograficamente distribuídas. Realizando conexões na porta padrão (TCP/80) da aplicação web.

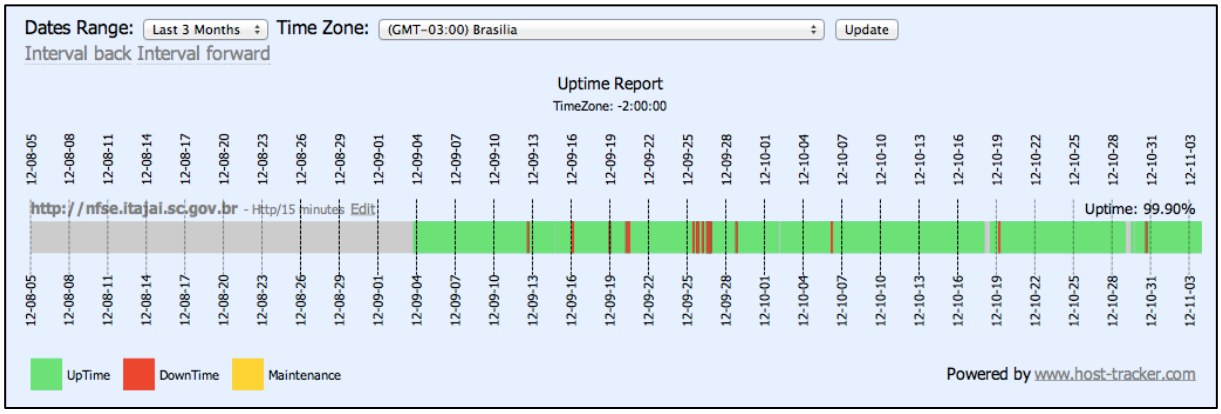

Figura 22 - Disponibilidade aferida pelo host-tracker.com
Fonte: HostTracker (2012).

O gráfico exibido na Figura 22, mostra o monitoramento dos últimos três meses, porém existindo dados somente a partir do dia 04/09/2012, quando após implantação e validação dos novos recursos, efetivamente a ferramenta passou a monitorar a disponibilidade do sistema.

A linha predominantemente verde, indica o período que o sistema esteve disponível para seus usuários, com pequenas interrupções indicadas em vermelho. O Host-Tracker quando perceber a indisponibilidade do sistema faz uma nova verificação após 5 minutos. Portanto, apresenta um MTTR mínimo de 5 min.

Os períodos de indisponibilidade da aplicação, por manutenção de terceiros, estão contidos no gráfico apresentado pelo Host-Tracker na Figura 22, no período entre os dias 25 e 28 de setembro de 2012. Podendo ser observados por recorrentes *downtimes*, concentrados e marcados em vermelho.

Ressalta-se o fato que o presente trabalho propõe e implementa uma solução para alta disponibilidade, atendendo toda a infraestrutura relacionada ao sistema emissor de notas fiscais presente na Prefeitura Municipal de Itajaí. Entretanto, a aplicação, compreendida pelo *software* que implementado e mantido por empresa terceira também pode apresentar períodos de indisponibilidade, quando em atualizações ou manutenções.

Para a disponibilidade da infraestrutura, considerando os valores apresentados pelo virtualizador, UPS e enlace externo, conforme apresentado na Tabela 6, percebe-se uma diferença em relação a apresentada pelo HostTracker por não considerar alguma indisponibilidade da aplicação.

Tabela 6 - Disponibilidade e *downtime* da infraestrutura nos últimos três meses de 2012.

Período	**Falha**	**Tempo para restauração**	**Erro apresentado**	**Disponibilidade mês**	***downtime* mês**
Outubro	Falha no enlace externo	3 min e 28 seg.	Indisponibilidade total aos usuários	99,99%	3 min e 23 seg.
Novembro	Enlace interno	3 min.	Por manutenção o conjunto de firewalls foi reiniciado	99,99%	4 min.
	Falha enlace externo	1 min	Perdendo pacotes, foi necessário reiniciar o roteador		
Dezembro	Falha enlace externo	-	-	100%	-
Total				99,993%	7 min e 23 seg.

No mês de novembro, uma perda de pacotes era percebida. Situação essa que forçou uma manutenção nos firewalls e posteriormente suas reinicializações. Entretanto, sem solução o mesmo procedimento foi realizado nos roteadores do enlace externo, que por sua vez reestabeleceu a normalidade.

4 CONCLUSÃO

Este trabalho apresentou uma proposta de uma solução de alta disponibilidade para ser implantando na Prefeitura de Itajaí, atendendo especificamente o sistema emissor de notas fiscais de serviço. A forma anterior de disponibilidade dos serviços não permitia um compromisso com o tempo máximo de indisponibilidade do sistema emissor de notas ficais de serviços.

A partir desse objetivo, surgiu a necessidade de conhecer diferentes metodologias e, ao mesmo tempo, entender o processo de implementação de servidores e recursos redundantes. Esse conhecimento foi alcançado ao término da Fundamentação Teórica, através da apresentação dos principais conceitos referentes ao assunto e do estudo de alguns ambientes similares em ambientes de médio e grande porte. Com base nessas informações, foi proposta e desenvolvida uma solução de disponibilidade para o cenário antes existente.

Neste projeto foram tratados a redundância de servidores e recursos, mas para um ambiente de alta disponibilidade deve-se procurar em primeiro lugar eliminar os demais pontos únicos de falha citados na fundamentação teórica.

Houveram problemas com a aquisição de novos equipamentos, mas embora tenha sido um contratempo, a solução desenvolvida garantiu a disponibilidade pretendida, mesmo assumindo a permanência de um ponto único de falha no UPS.

Embora atendendo ao geral objetivo deste trabalho, recomenda-se a aquisição de uma nova solução redundante para UPS. Sendo a energia elétrica o recurso fundamental para o funcionamento de sistemas eletrônicos.

Soluções similares às realizadas neste trabalho foram orçadas para a viabilidade comercial por valores não inferiores a um milhão de reais, conforme proposta Anexo I. Este projeto alcançou o nível pretendido de alta disponibilidade por meio da aplicação de tecnologias cujo conhecimento foi adquirido ao longo dos estudos e proporcionou um inestimável aprendizado para o autor, além da evidente economia de recursos para o ambiente contemplado.

REFERENCIAS

AHLUWALIA, Kanwardeep Singh; JAIN, Atul. **High Availability Design Patterns**. In: PATTERN LANGUAGES OF PROGRAMS (PLOP) CONFERENCE, 2006, Portland. Anais… New York: ACM, 2006.

BOTELHO, Marcos A. Faria. **Alta Disponibilidade em firewall utilizando Pfsync e CARP sobre freebsd.** 2006. 53 f. Monografia (Pós-graduação em Administração em Redes Linux) – Faculdade de Ciência da Computação, Universidade Federal de Lavras, Lavras, 2006.

BUTOW, Eric; DICKE, Bill; JOYAL, John. **HP virtual connect**: for Dummies. Indianapolis: Wiley Publishing, 2011.

EXTREME: **Extreme Networks.** 2012. Disponível em: < http://www.extremenetworks.com >. Acesso em: 13 set. 2012.

GOOGLE: **Google maps.** 2012. Disponível em: < https://maps.google.com >. Acesso em: 04 jun. 2012.

HOSTTRAKER: **HostTracker.** 2012. Disponível em: < http://www.host-tracker.com >. Acesso em: 03 set. 2012.

JAYASWAL, Kailash. **Administering data centers:** Servers, Storage, and Voice over IP. Indianapolis: Wiley Publishing, 2006.

KOREN, Israel; KRISHNA, C. Mani. **Fault tolerant systems.** San Francisco: Elsevier, 2006.

LAUREANO, Marcos. **Máquinas virtuais e emuladores:** conceitos, técnicas e aplicações. São Paulo: Novatec, 2006.

MARCUS, Evan; STERN, Hal. **Blueprints for high availability**. 2.ed. Indianapolis: Wiley Publishing, 2003.

MCCROSKY, Carl; MINOLI, Daniel; INIEWSKI, Krzysztof. **Network infrastructure and architecture:** designing high-availability networks. New Jersey: John Wiley & Sons, 2008.

MESINER, David et al. **Power routing:** dynamic power provisioning in the data center. In: ARCHITECTURAL SUPPORT FOR PROGRAMMING LANGUAGES AND OPERATING SYSTEMS, 15., 2010, Pittsburgh. Anais… New York: ACM, 2010.

SILVEIRA, Leonardo Salgado. **Alta Disponibilidade em ambientes virtualizados.** 2009. 35 f. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação), Universidade Luterana, Guaíba, 2009.

PMI. **Prefeitura Municipal de Itajaí.** 2012. Disponível em: < http://www.itajai.sc.gov.br >. Acesso em: 06 jul. 2012.

PMI. **Itajaí presta contas.** 2012. Disponível em: < http://prestacontas.itajai.sc.gov.br >. Acesso em: 06 jul. 2012.

SCHNEIDER, Edson Neri. **Estudo dos recursos de alta disponibilidade e implementação de um modelo de pequeno porte.** 2006. 92 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação), Universidade do Vale do Itajaí, Itajaí, 2006.

SCHMIDT, Klaus. **High availability and disaster recovery**: concepts, design, implementation. Berlin: Springer, 2006.

SPG. **Secretaria de Estado do Planejamento**. 201?. Disponível em: < http://www.spg.sc.gov.br/dados_munic.php >. Acesso em: 03 mar. 2012.

VMWARE. **Distributed resource scheduling, distributed power management of server resources.** 2012. Disponível em: < http://www.vmware.com/products/drs/overview.html >. Acesso em: 01 jun. 2012.

VMWARE: **VMWare Fault Tolerance (FT), High Availability for Virtual Machines & Applications.** 2012. Disponível em: < http://www.vmware.com/products/fault-tolerance/overview.html >. Acesso em: 01 jun. 2012.

WEBER, Taisy Silva. **Um roteiro para exploração dos conceitos básicos de tolerância a falhas**. 2002. 62 f. Monografia (Pós-graduação em Redes e Sistemas Distribuídos) - Instituto de Informática. Universidade Federal do Rio Grande do Sul. Porto Alegre, 2002.

WEYGANT, Peter S. **Clusters for high availability:** a primer of HP solutions. 2.ed. New Jersey: Prentice Hall, 2002.

ANEXO I

Proposta Comercial

Preço Solução:

MODELO	R$	QTDE	TOTAL
Storage 3Par F400	R$ 645.800,00	1	R$ 645.800,00
C7000 c/ Flex Fabric	R$ 95.800,00	1	R$ 95.800,00
BL460G7	R$ 25.710,00	8	R$ 205.680,00
Switch SAN 8/24	R$ 24.595,00	2	R$ 49.190,00
MSL4048	R$ 59.210,00	1	R$ 59.210,00
Data Protector	R$ 31.560,00	1	R$ 31.560,00
VMware Ent. Plus 1P	R$ 9.170,00	16	R$ 146.720,00
Rack 10642	R$ 11.050,00	1	R$ 11.050,00
Switch A7506	R$ 158.450,00	1	R$ 158.450,00
Nobreak APC 10 kVA	R$ 10.700,00	1	R$ 10.700,00
TOTAL			**R$ 1.414.160,00**

1. **Frete:** C.I.F.

2. **Prazo de Entrega:** 40 a 60 dias.

3. **Prazo de Pagamento:** 30 Dias após o Faturamento(*)

   (*) Sujeito a aprovação de crédito

4. **Validade desta Proposta:** 15 Dias.