1
Introdução

1.1
Motivação

O crescimento e a expansão da *Internet* aumentaram o número de acessos aos portais e aplicações. Para atender à crescente demanda, as grandes empresas de *Internet* têm hoje uma vasta infra-estrutura composta de centenas de servidores.

Estes servidores geram uma grande quantidade de informação conforme os usuários da *Internet* acessam as páginas e as aplicações que são por eles servidas. Toda esta informação gerada está representada por registros em arquivos de texto, conhecidos como arquivos de log. Estes arquivos são importantes tanto para atividades técnicas de detecção e resolução de problemas, quanto para atividades de negócio que desejam conhecer o acesso aos produtos e o comportamento dos usuários.

Além do grande volume que estes arquivos hoje representam, os arquivos de log geralmente estão distribuídos pelo parque de servidores. Isto torna a atividade de análise e correlação de todo esse grande e disperso volume de informação um desafio nas empresas de *Internet*, como é o caso da Globo.com.

Aliado a isto, a análise e correlação de logs deve produzir resultados em tempo real. Não é possível esperar mais do que poucos minutos antes de detectar e, idealmente, disparar uma ação para resolver um problema técnico. Da mesma forma, as equipes de negócio precisam cada vez mais de informação sobre os acessos as páginas e comportamento dos usuários e esta informação deve estar disponível em tempo real para que ações que influenciem a demanda possam ser tomadas.

Nesse sentido, existe hoje uma necessidade grande de se estudar e propor técnicas e sistemas que sejam capazes de analisar e correlacionar à informação produzida em forma de arquivos de log.

1.2 Proposta da dissertação

Esta dissertação se propõe a enxergar cada registro de log como um *evento de log*, e a seqüência destes eventos como *fluxos de eventos de logs*. Estes fluxos precisam, então, ser analisados e correlacionados em tempo real para produzir informação relevante tanto para o negócio, quanto para a gerência de tecnologia. Além da análise e correlação dos fluxos de eventos de log, existem outras funções envolvidas com o gerenciamento de logs, como o armazenamento, a pesquisa e a visualização, entre outras, que não podem ser esquecidas. Contudo, o foco desta pesquisa será sobre as atividades de correlação de eventos de log.

Do ponto de vista de sistemas, um sistema que se proponha a gerenciar eventos de log precisa ser capaz de lidar com diferentes formatos, com o fato de estes arquivos estarem distribuídos, com o fato de estes eventos acontecerem em grande volume e, além disso, este sistema não pode impactar no desempenho dos sistemas que estão produzindo os logs.

Nesse sentido, a proposta desta tese é estudar os logs e sistemas que se propõem a gerenciá-los, estudar técnicas de correlação de eventos que produzam resultados em tempo real, estudar arquiteturas de sistemas adequadas para gerenciar eventos distribuídos, para propor uma arquitetura de sistema capaz de gerenciar eventos de log em tempo real e, finalmente, desenvolver um protótipo capaz de realizar uma prova de conceito baseada na realidade de uma empresa de *Internet*, a Globo.com.

1.3 Organização da dissertação

Esta dissertação está organizada da seguinte forma. O Capítulo 2 traz algumas definições básicas sobre log, explora a idéia de se enxergar o log como um fluxo de eventos, apresenta diferentes padrões de arquivos de logs e apresenta o que são conhecidos como *Sistemas de Gerenciamento de Logs* e as atividades relacionadas com este tipo de sistema.

O Capítulo 3 explora o conceito de eventos e métodos e sistemas capazes de realizar o processamento e a correlação de eventos. Entre os diversos métodos foram analisados mais profundamente os baseados em regras e os baseados em consultas, representados por sistemas conhecidos como *Data*

Stream Management Systems (DSMS), ou *Sistemas de Gerenciamento de Fluxos de Dados*.

O Capítulo 4 especifica os requisitos de um sistema para gerenciamento de eventos de log, apresenta os conceitos relacionados ao padrão de arquitetura orientada a eventos e propõe uma arquitetura orientada a eventos para o gerenciamento de eventos de log.

O Capítulo 5 apresenta o protótipo desenvolvido e as ferramentas utilizadas. Mostra também a prova de conceito realizada e os testes realizados para garantir a qualidade funcional do protótipo desenvolvido. Ao final, a Conclusão resume os resultados desta dissertação e apresenta sugestões de trabalhos futuros sobre este tema.