

Universidade Federal de Ouro Preto - UFOP
Instituto de Ciências Exatas e Biológicas - ICEB
Departamento de Computação - DECOM

ADICIONANDO ESCALABILIDADE AO *FRAMEWORK*
DE RECOMENDAÇÃO IDEALIZE

Aluno: Alex Amorim Dutra
Matricula: 07.1.4149

Orientador: Álvaro Rodrigues Pereira Júnior
Co-Orientador: Felipe Santiago Martins Coimbra de Melo

Ouro Preto
2 de outubro de 2010

Universidade Federal de Ouro Preto - UFOP
Instituto de Ciências Exatas e Biológicas - ICEB
Departamento de Computação - DECOM

ADICIONANDO ESCALABILIDADE AO *FRAMEWORK* DE RECOMENDAÇÃO IDEALIZE

Proposta de monografia apresentada ao curso de Bacharelado em Ciência da Computação, Universidade Federal de Ouro Preto, como requisito parcial para a conclusão da disciplina Monografia I (BCC390).

Aluno: Alex Amorim Dutra
Matrícula: 07.1.4149

Orientador: Álvaro Rodrigues Pereira Júnior
Co-Orientador: Felipe Santiago Martins Coimbra de Melo

Ouro Preto
2 de outubro de 2010

Resumo

Resumo

Palavras-chave: Escalabilidade. *Hadoop.* *Framework.* Recomendação. Idealize. *Framework* de Recomendação.

Desde a antiguidade o homem utiliza meios para fazer recomendações a outras pessoas. Sistemas de recomendação na web têm a cada dia deixado de ser uma novidade e tem se tornado uma necessidade para os usuários da web, devido ao volume de informações disponíveis. Para conclusão da disciplina Monografia I (BCC390), estarei tratando do *Framework* de Recomendação Idealize em especial na parte de escalabilidade utilizando o *framework Hadoop*, pelo fato de ser *open-source* e ser utilizado para computação escalável e distribuída [2]. O *Hadoop* utiliza do paradigma de programação *Map Reduce* para realizar a computação distribuída em *clusters* contendo centenas ou até milhares de máquinas.

Sumário

1	Introdução	1
2	Justificativa	2
3	Objetivos	3
3.1	Objetivo geral	3
3.2	Objetivos específicos	3
4	Metodologia	5
5	Cronograma de atividades	6

Lista de Figuras

1	Arquitetura final	3
2	Pseudo-Escala	5

Lista de Tabelas

1	Cronograma de Atividades.	6
---	-----------------------------------	---

1 Introdução

Com a crescente explosão de informações principalmente na web, temos ao alcance informações que caracterizamos relevantes em diversas áreas. Algumas vezes quando estamos procurando sobre determinado assunto, produto, ou qualquer outro item acabamos não encontrando o que desejamos, devido ao grande conteúdo que consideramos irrelevantes serem encontrados. Sistemas de recomendação têm a finalidade de levar ao usuário o que realmente é relevante, por ser algo que o usuário procurava ou um conteúdo novo que o interessa. Neste trabalho estarei tratando da escalabilidade do *Framework* de Recomendação Idealize e desta forma adaptando alguns métodos de recomendação para utilizarem desta arquitetura escalável. Escalabilidade tem o objetivo de "Distribuir recursos computacionais ao longo do tempo de forma que as tarefas do sistema preservem seus requisitos temporais." [4]

As aplicações de recomendações que utilizam o *framework* hoje, são aplicações que fazem recomendação baseada em conteúdo, filtragem colaborativa e dados de uso. Recomendações baseadas em conteúdo tem a origem na recuperação da informação e recomendam um item para um usuário com base em uma descrição do item e a descrição dos interesses no perfil do usuário [6]. Recomendações por filtragem colaborativa tem sua origem na mineração de dados e é o processo de filtragem ou avaliação dos itens através de múltiplos usuários [3].

2 Justificativa

Como citado, sistemas de recomendação tem a finalidade de levar o que realmente interessa aos usuários, evitando assim um gasto de tempo maior com a procura de conteúdo relevante.

Hoje grandes empresas da web como Facebook, Yahoo, Google, Twitter e Amazon armazenam dados da ordem de Petabytes [1]. Assim como nestas empresas a movimentação e processamento de informações são altas, várias aplicações de recomendação também envolvem grandes quantidades de dados e um grande número de usuários. A escalabilidade tem a função principal de distribuir os componentes e serviços de forma a aumentar o desempenho, diminuindo assim o tempo de resposta dos pedidos de recomendações.

Sistemas escaláveis são mais estáveis e flexíveis para o atendimento de grandes quantidades de usuários e informações, ou seja, aumentam consideravelmente o poder de processamento destes dados.

3 Objetivos

3.1 Objetivo geral

- O objetivo ao final do curso de Monografia I (BCC390) e Mografia II (BCC391) é estar com o *Framework* Idealize, permitindo assim as aplicações de recomendações utilizarem a arquitetura escalável. Podendo os desenvolvedores utilizarem um *cluster* com quantas máquinas forem necessárias para suas aplicações de recomendação, como apresentado na figura 1. Um objetivo a cumprir também será a criação de uma aplicação, sendo ela de recomendação por filtragem colaborativa ou recomendação baseada em conteúdo, que utilize desta arquitetura escalável implantada no *framework*.

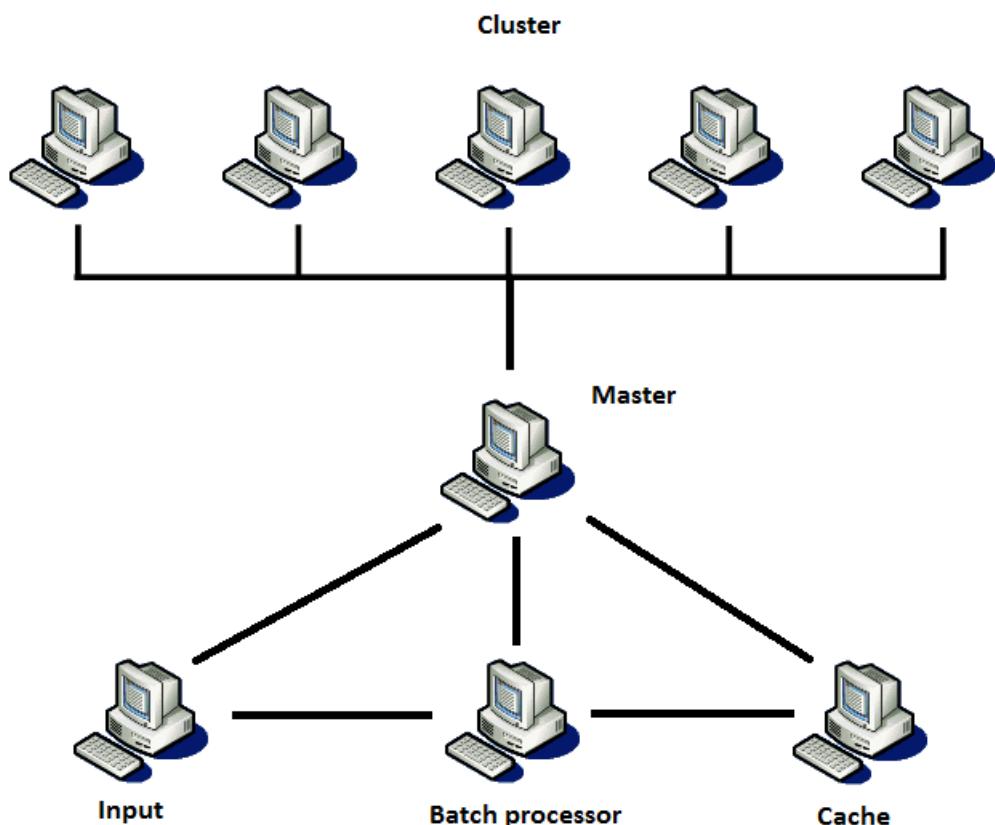


Figura 1: Arquitetura final

3.2 Objetivos específicos

- Instanciar máquinas utilizando uma pseudo-escala, para recomendações baseada em conteúdo.

O primeiro objetivo é fazer a utilização de uma pseudo-escala onde se tem três máquinas. Uma máquina fazendo o processamento em lote, uma outra máquina

permite a inserção, remoção ou atualização de itens e uma terceira máquina armazena dados em cache, podendo ser estes dados listas de recomendações processadas, *dataModel*, e outros dados que o desenvolvedor de aplicativos que utilizem o *framework* julgar necessário. Nesta fase estarei utilizando RMI (Remote Method Invocation) é uma das abordagens da tecnologia Java para prover as funcionalidades de uma plataforma de objetos distribuídos [5].

- Implementar uma arquitetura escalável ao *Framework* de recomendação Idealize. Esta tarefa uma parte estará sendo realizada tanto na Monografia I (BCC390) quanto na Monografia II (BCC391).

Nesta fase pretendo estar realizando estudos sobre sistemas escaláveis e com isto estar desenvolvendo a arquitetura deste tipo para o *Framework*. O objetivo principal nesta fase é estar utilizando o *Hadoop* e montar um *cluster* de máquinas, onde as máquinas possam comunicar com a máquina *master* e esta distribuir o processamento ao *cluster* de máquinas.

- Desenvolver uma aplicação que utilize a arquitetura escalável.

Por fim tenho o objetivo de estar desenvolvendo uma aplicação que utilize uma determinada abordagem de recomendação, podendo ser de filtragem colaborativa ou baseada em conteúdo. Esta aplicação servirá de modelo para outros desenvolvedores que desejarem utilizar a arquitetura escalável.

4 Metodologia

Esta metodologia será seguida para as disciplinas Monografia I (BCC390) e Monografia II (BCC391). Este trabalho é de caráter exploratório, deseja-se melhorar o tempo de processamento de recomendações. O trabalho será dividido em fases, onde a primeira consistirá de um estudo sistematizado feito com base em publicações livros, artigos, revistas. Em principal análise da arquitetura desenvolvida no Idealize *Framework* de Recomendação utilizando filtragem colaborativa.

A segunda fase será a implementação de uma arquitetura disposta de três máquinas para recomendação baseada em conteúdo, sendo estas três máquinas como descrito na figura 2.

A terceira fase será a realização de testes envolvendo ferramentas próprias e análise dos resultados encontrados. Esta análise poderá ser feita comparando resultados encontrados na literatura.

A quarta fase será o estudo aprofundado sobre escalabilidade e *Hadoop*, feito com base em aplicações, publicações, livros, artigos. Em seguida será a implementação de uma arquitetura que utilize o *Hadoop* para tornar o sistema escalável e distribuído.

Em seguida faremos novos testes e análise dos resultados envolvendo a escalabilidade, analisando e comparando com os dados obtidos anteriormente.

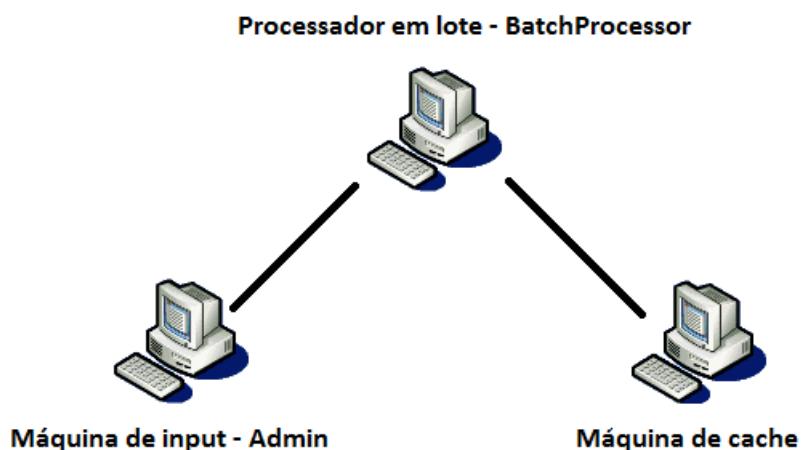


Figura 2: Pseudo-Escala

5 Cronograma de atividades

Na Tabela 1, segue o cronograma das atividades que estarei seguindo para ao fim do curso obter os resultados esperados.

Atividades	Ago	Set	Out	Nov	Dez
Estudo de escalabilidade	X				
Estudo e início de implementações	X	X			
Implementações de Escalabilidade			X	X	
Implementação Redigir a Monografia				X	X
Apresentação do Trabalho					X

Tabela 1: Cronograma de Atividades.

Referências

- [1] <http://escalabilidade.com/2010/05/18/>, Maio 2010.
- [2] Apache. Hadoop. Disponível em <http://hadoop.apache.org/>, setembro 2010.
- [3] Jon Herlocker Shilad Sen J B Schafer, Dan Frankowski. Collaborative filtering recommender systems. 4321:291–324, 2007.
- [4] Raimundo José Macêdo. Escalabilidade. Disponível em www.macedo.ufba.br, agosto 2009.
- [5] Oracle. <http://download.oracle.com/javase/tutorial/rmi/>, Setembro 2010.
- [6] Michael J. Pazzani and Daniel Billsus. Content-based recommendation systems. pages 325–341, 2007.