



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO  
CENTRO DE INFORMÁTICA  
2014.1

---

BALANCEAMENTO DE *CLUSTERS* SEMÂNTICOS POR MEIO DA  
IDENTIFICAÇÃO DE *HOT-PEERS*

---

**PROPOSTA DE TRABALHO DE GRADUAÇÃO**

**Aluno:** João Paulo Aguilera Borges (jpab@cin.ufpe.br)

**Orientador:** Ana Carolina Salgado (acs@cin.ufpe.br)

Recife, Maio de 2014

# 1. Resumo

*Peer Data Management Systems* (PDMSs) são considerados uma extensão natural de sistemas de integração de dados. Com o diferencial na representação e localização dos dados através da utilização de semântica, o projeto *SPEED* (*Semantic PEER-to-Peer Data Management System*) foi proposto. Por se tratar de um sistema distribuído baseado em clusters semânticos apresenta problemas singulares no balanceamento de carga. O objetivo geral desse Trabalho de Graduação é a especificação e implementação de um mecanismo para solucionar esse problema por meio da identificação de *peers* sobrecarregados (*hot-peers*). O trabalho aborda ainda a identificação de métricas qualitativas para o balanceamento dos *clusters*. Os algoritmos serão implementados e avaliados como parte integrante do *SPEED*.

## 2. Contextualização

A informação sempre foi e será a base da interação humana. Ao lado da explosão do conhecimento científico e tecnológico, a informática trouxe um aumento considerável da oferta de informação e das possibilidades de sua disseminação. Nesse contexto, o gerenciamento de dados acessados a partir de fontes de dados distribuídas, e a integração dos mesmos para responder às consultas efetuadas, apresenta um tremendo desafio. Os sistemas de gerenciamento de dados Peer-to-Peer (P2P) ou PDMS (*Peer Data Management System*) representam uma evolução dos tradicionais sistemas de integração de dados, pois provêem autonomia, representatividade de consultas, eficiência, qualidade de serviço, tolerância a falhas e segurança.

O *SPEED* é um PDMS que utiliza as seguintes topologias de rede *peer-to-peer* em sua arquitetura: *DHT* (*Distributed Hash Tables*) e *super-peer*. Parte de sua arquitetura é formada pelos *peers* semânticos interligados em uma rede *DHT*. A outra parte, a *super-peer*, é composta por *peers* de integração e *peers* de dados. Esses *peers* de integração e de dados estão agrupados em *clusters* semânticos, que compõem comunidades semânticas.

Em qualquer sistema distribuído, o balanceamento justo de recursos é um dos principais fatores a serem atingidos. Infelizmente, não é possível atingir uma distribuição equitativa de recursos naturalmente, já que, normalmente, há uma diferença de capacidade entre cada *peer* e cada requisição e, para tal, mecanismos de balanceamento de carga são necessários.

Em sistemas distribuídos organizados por *cluster* o correto balanceamento de carga requer a determinação da carga de cada nó (*peer*) do *cluster* e a migração de nós sobrecarregados (*hot-peers*).

### 3. Objetivos

O objetivo geral desse Trabalho de Graduação é a especificação e implementação de um mecanismo de balanceamento de carga por meio da identificação de *peers* sobrecarregados (*hot-peers*). O trabalho aborda ainda a identificação de métricas de balanceamento que serão utilizadas para classificar a qualidade dos *clusters*.

No processo da construção do Trabalho de Graduação, os algoritmos para o balanceamento serão implementados utilizando a Linguagem Java e serão incluídos à solução SPEED já existente. Esse projeto comportará ainda uma fase de testes e avaliação de desempenho para averiguar a qualidade da solução proposta.

### 4. Cronograma

Nesta seção, é apresentado o cronograma de atividades previsto para o desenvolvimento desse Trabalho de Graduação (Tabela 1).

Atividade	Maio			Junho				Julho			Agosto		
Levantamento do estado da arte e definição do escopo	X	X	X										
Implementação dos algoritmos		X	X	X	X	X							
Testes e experimentos			X	X	X	X	X	X					
Análise dos resultados						X	X	X	X				
Elaboração do relatório			X	X	X	X	X	X	X	X	X	X	
Preparação e defesa										X	X	X	

Tabela 1: Cronograma de atividades

### 5. Possíveis Avaliadores

Os possíveis avaliadores para o resultado a ser obtido ao final de todas as etapas da proposta descrita neste documento são:

- Bernadette Farias Lóscio
- Fernando da Fonseca de Souza

## 6. Assinaturas

---

João Paulo Aguilera Borges  
**Orientando**

---

Ana Carolina Salgado  
**Orientador**

Recife, Maio de 2014