



<b>Evento</b>	Salão UFRGS 2018: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS - FINOVA
<b>Ano</b>	2018
<b>Local</b>	Campus do Vale - UFRGS
<b>Título</b>	SMART-SENT: Desenvolvimento de uma Plataforma para a Análise de Big Data em Aplicações de Cidades Inteligentes
<b>Autor</b>	GABRIEL JOB ANTUNES GRABHER
<b>Orientador</b>	CLAUDIO FERNANDO RESIN GEYER

## **TÍTULO DO PROJETO: SMART-SENT: Desenvolvimento de uma Plataforma para a Análise de Big Data em Aplicações de Cidades Inteligentes.**

Aluno: Gabriel Job Antunes Grabher

Orientador: Cláudio Fernando Resin Geyer

### **RESUMO DAS ATIVIDADES**

---

#### 1. Introdução:

A instrumentação da sociedade, a utilização de redes sociais, mídias digitais e aplicações para dispositivos móveis estão crescendo constantemente. Esse cenário impulsionou o que se conhece atualmente como cidades inteligentes. Desse modo, os dados oriundos dos vários dispositivos estão interconectados por diversas fontes, entre os próprios usuários - em seus dispositivos, em centros de dados, em equipamentos de rede e assim por diante.

A plataforma SMART-SENT foi proposta para suprir os desafios no processamento intensivo de dados em áreas onde é de fundamental importância lidar com a coleta dos dados, armazenamento em larga escala, análise e o processamento em tempo real, para assim obter informações que potencializam a melhora de produtos e serviços para a população.

Para a análise de seu comportamento, serão utilizados simuladores e dados de entrada reais públicos de ambientes urbanos que irão validar o funcionamento da infraestrutura. Entre os simuladores utilizados, estão o MRSRG e o MRA, que simulam o modelo de programação paralelo e distribuído MapReduce em ambientes de computação homogêneos e heterogêneos respectivamente, junto ao BigHybrid que é uma composição dos dois.

Durante as atividades realizadas, foram estudadas as características e os comportamentos do processamento em larga escala de dados através dos simuladores mencionados em sistemas homogêneos e heterogêneos.

#### 2. Atividades realizadas:

**Atividade 1:** Estudo dos conceitos relacionados ao modelo MapReduce e ao toolkit BigHybrid.

Período: Out/2017 a Nov/2017

Descrição: Leitura de artigos referentes ao estado da arte sobre benchmarks para aplicações Big Data e sobre o modelo MapReduce, que é utilizado para o processamento de dados em larga escala. Dentre tais artigos, foram realizadas leituras das publicações de autoria do grupo relacionadas à simulação para Big Data, nomeadamente sobre o BigHybrid, o MRA++ e o MRSRG.

**Atividade 2:** Estudar e compreender as APIs S4U e XBT do Simgrid.

Período: Nov/2017 a Dez/2017

Descrição: Os simuladores MRSRG, MRA e BigHybrid foram desenvolvidos utilizando o SimGrid, uma ferramenta para a simulação de aplicações distribuídas. Entre suas APIs oferecidas, usa-se a S4U, para definir a abstração

do ambiente distribuído e suas técnicas de interação, mais a XBT, que permite uma manipulação mais simples de estruturas de dados e memória em geral. Na última versão do Simgrid, a API MSG foi substituída pela S4U, porém os simuladores mencionados ainda utilizavam versões mais antigas.

**Atividade 3:** Atualização do MRSG para a versão mais recente do Simgrid.

Período: Dez/2017 a Jul/2018

Descrição: Nesta atividade foi feita uma conversão completa do simulador MRSG para a última versão do Simgrid (versão 3.20). Para a validação do funcionamento do simulador foi comparado o resultado da execução de um caso de teste presente no artigo[5], cujos dados foram obtidos na Grid5000.

**Atividade 4:** Atualização do MRA para a versão mais recente do Simgrid.

Período: Jul/2018 a Ago/2018

Descrição: Realizaram-se mudanças no código para o simulador MRA executar na última versão do Simgrid (versão 3.20), utilizando de maneira híbrida funções da nova API S4U e da antiga MSG.

### 3. Objetivos atingidos:

- Revisão sobre o estado da arte do modelo MapReduce, simuladores e benchmarks;
- Estudo e compreensão da nova API S4U do Simgrid;
- Migração completa do simulador de MapReduce em ambientes homogêneos (MRSG) para uma versão atual e funcional no SimGrid 3.20;
- Migração parcial do simulador de MapReduce em ambientes heterogêneos (MRA) para uma versão atual e funcional no SimGrid 3.20.

### 4. Resultados obtidos:

- Atualização de versão do simulador de aplicações MapReduce em ambientes homogêneos (MRSG);
- Validação da última versão do simulador MRSG comparando com os resultados obtidos de execuções nas versões anteriores;
- Atualização de versão do simulador de aplicações MapReduce em ambientes heterogêneos (MRA).

### 5. Conclusão:

Os resultados obtidos durante este estágio tornaram possível o avanço na análise do comportamento para uma infraestrutura de sensoriamento urbano, assim como também facilitarão o desenvolvimento de seus módulos futuramente.

Para dar continuação a esse avanço, tem-se como objetivo migrar os simuladores MRSG, MRA e BigHybrid completamente para o uso da API S4U, e elaborar um relatório técnico sobre o funcionamento dos mesmos.

## Referências

- [1] DEAN, Jeffrey; GHEMAWAT, Sanjay. **MapReduce: Simplified Data Processing on Large Clusters**. Symposium on Operating Systems Design & Implementation, v. 6, dez., 2004.
- [2] DEAN, Jeffrey; GHEMAWAT, Sanjay. **MapReduce – a flexible data processing tool**. Communications of the ACM, v. 53, n.1, p. 72–77, jan., 2010.
- [3] KOLBERG, Wagner; MARCOS, Pedro de Botelho; ANJOS, Julio Cesar Santos; MIYAZAKI, Alexandre Kenta Salgueiro; GEYER, Cláudio Fernando Resin; ARANTES, Luciana Bezerra. **MRSB – A MapReduce simulator over SimGrid**. Parallel, v. 39, n. 4-5, p. 233–244, abr./maio, 2013.
- [4] ANJOS, Julio Cesar Santos; CARRERA, Ivan; KOLBERG, Wagner; TIBOLA, André Luis; ARANTES, Luciana Bezerra; GEYER, Cláudio Fernando Resin. **MRA++: Scheduling and Data Placement on MapReduce for Heterogeneous Environments**. Future Generation Computer Systems, v. 42, p. 22–35, jan., 2015.
- [5] ANJOS, Julio Cesar Santos; FEDAK, Gilles; GEYER, Cláudio Fernando Resin. **BIGHybrid: A Simulator for MapReduce Applications in Hybrid Distributed Infrastructures Validated with the Grid5000 Experimental Platform**. Concurrency and Computation: Practice and Experience, v. 28, n. 8, p. 2416–2439, jun., 2016.
- [6] CHINESE ACADEMY OF SCIENCES. Institute of Computing Technology. **Handbook of BigDataBench (version 3.1) – A Big Data Benchmark Suite**. Report.