

MapReduce

- ✓ Fornece um mecanismo que permite abstrair a complexidade envolvida no paralelismo das aplicações [1]
- ✓ Baseia-se em linguagens funcionais de alto nível como LISP.
- ✓ **Proposta concebida para uso em grandes clusters, com recursos (processadores, memória, armazenamento e rede) homogêneos.**
- ✓ O *Hadoop* é a implementação *open source* mais popular do *MapReduce*, proposta pela *Apache Software Foundation* e usada por corporações como Yahoo, Amazon e FaceBook
- ✓ A adaptação do modelo MapReduce em ambientes Heterogêneos ainda é um campo a ser explorado.

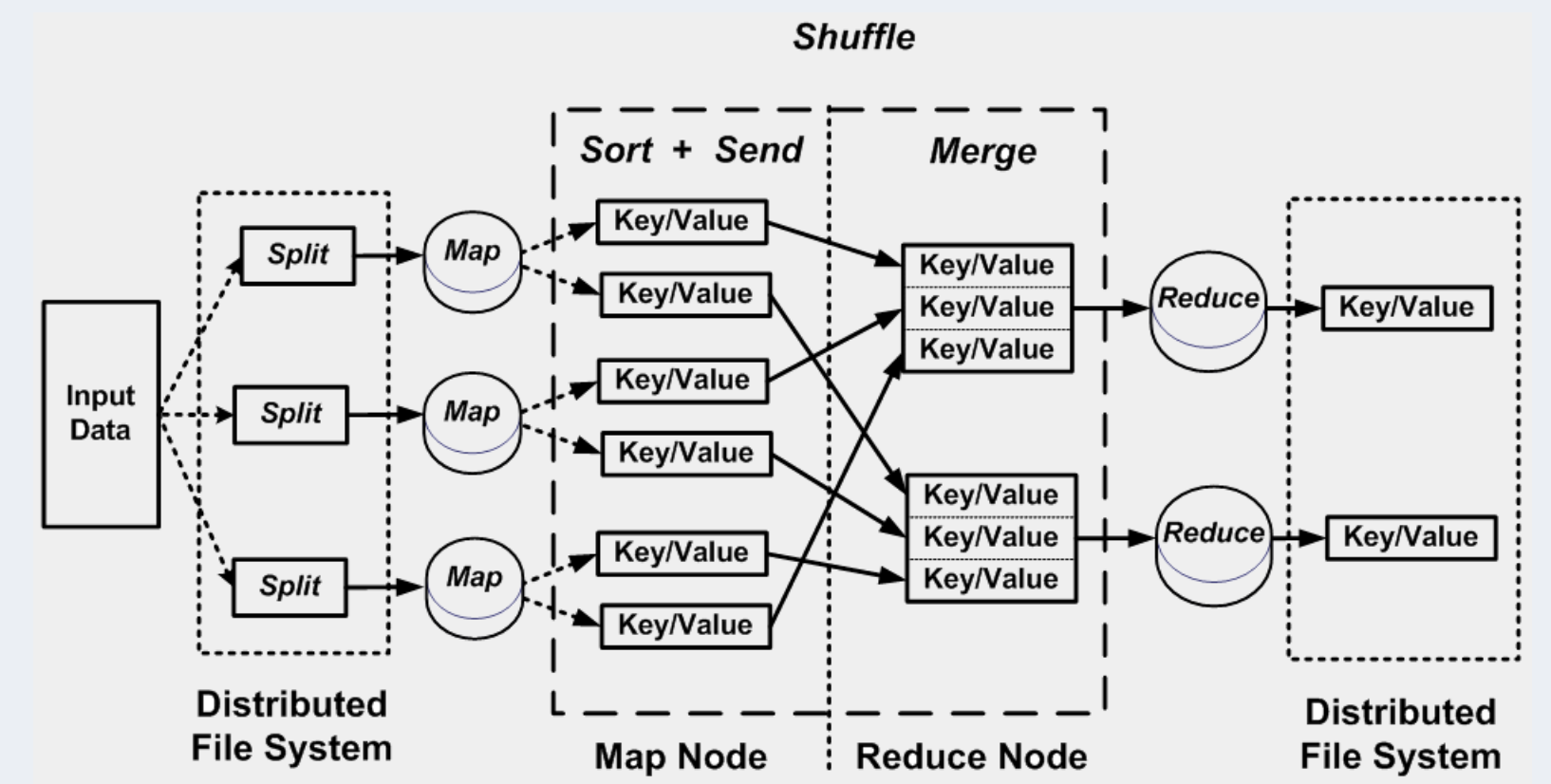


Figura: Estrutura do Modelo de Programação

O Simulador MRSG

- ✓ Simulador desenvolvido pelo grupo de pesquisa, apresentado em [2]
- ✓ Alternativa para testes em ambientes reais, onde é necessário uma grande quantidade de recursos.
- ✓ Fornece uma série de facilidades para implementação de algoritmos teóricos, mudanças na configuração do sistema sendo testado, além disso, não é necessário grandes recursos físicos.

Objetivo do Trabalho: Analisar o comportamento do *Hadoop* em ambientes Heterogêneos e testar, com diferentes parametrizações, os algoritmos desenvolvidos pelo grupo de pesquisa para adaptação do modelo MapReduce para tais ambientes.

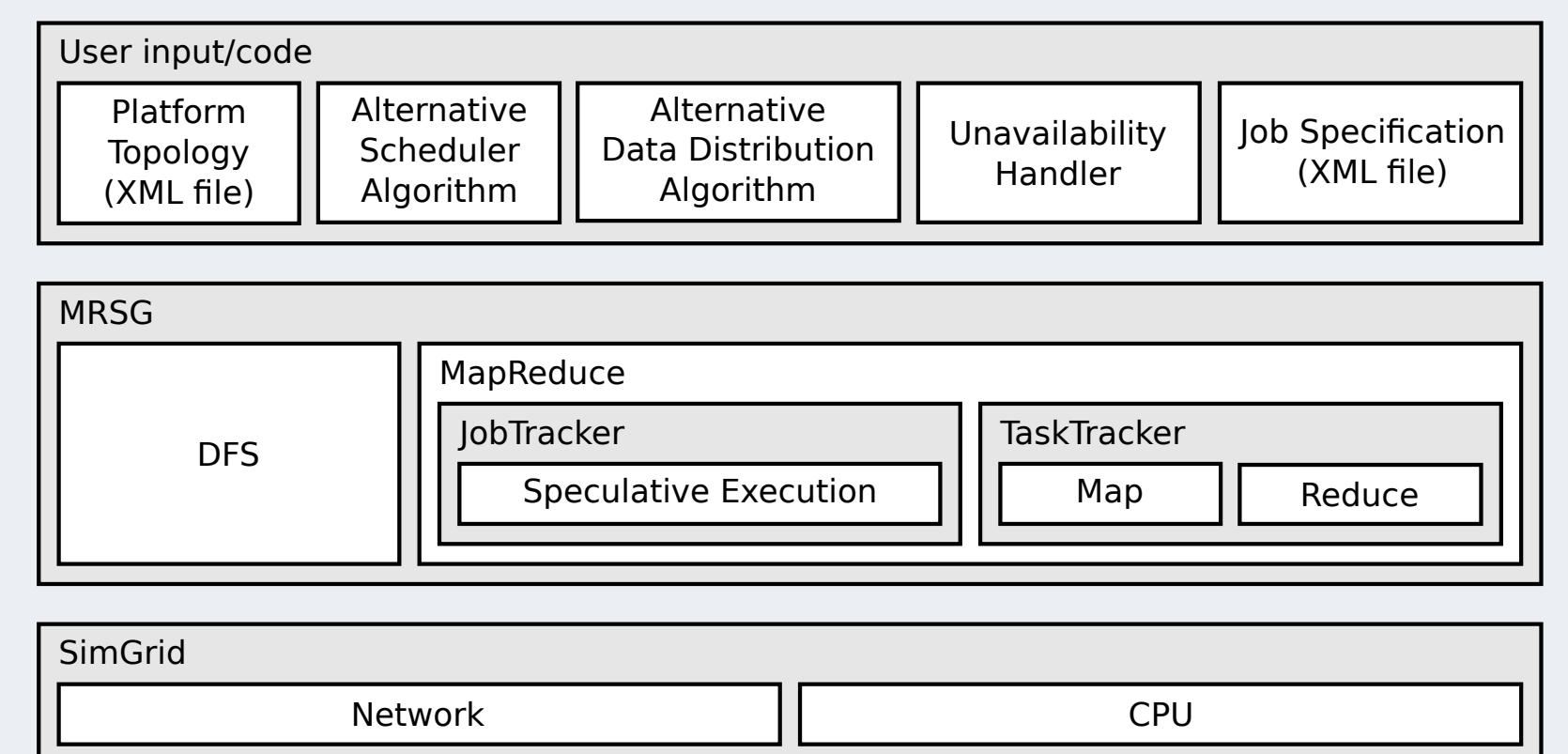


Figura: Arquitetura do simulador MRSG

Atividades

- ✓ Experimentos 2^k , utilizando o simulador MRSG, com algoritmos modificados e a comparação de seus resultados com os algoritmos originais.
- ✓ Instalação e configuração do novo Cluster do Grupo de Pesquisa.
- ✓ Testes em ambientes reais, com o Hadoop, para um trabalho de validação do simulador submetido ao ParCo.
- ✓ Testes em ambientes Simulados, com o simulador MRSG, também para o trabalho de validação.
- ✓ Publicação de um artigo no WSCAD-WIC 2011 [3].

Sobre o artigo Publicado no WSCAD-WIC

- ✓ Objetivo: Analisar, em ambiente simulado, o comportamento do MapReduce em um ambiente com heterogeneidade de rede, em comparação com um ambiente com rede homogênea.
- ✓ As Máquinas tinham poder computacional homogêneo.
- ✓ Testes com diferentes numeros de máquinas, com diferentes tamanhos de entrada, com duas aplicações.
- ✓ A configuração da rede Homogênea foi setada com links fixos de 100 Mb/s.
- ✓ A configuração da rede Heterogênea foi setada de duas formas, visando o melhor e o pior caso, de dados obtidos por uma pesquisa do Instituto Brasileiro de Opinião Pública e Estatística (IBOPE).
- ✓ Os Resultados mostram que o desempenho do modelo sem adaptações para um ambiente com heterogeneidade de rede é extremamente inferior, comparado com o desempenho do modelo em ambientes com links fixos.

Algumas Referências

- [1] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, January 2008.
- [2] Wagner Kolberg and Julio Anjos. MRSG: A MapReduce Simulator for Desktop Grids. Technical report, Universidade Federal do Rio Grande do Sul, April 2011.
- [3] Alexandre Miyazaki, Pedro de Botelho Marcos, and Cláudio R. Geyer. Análise do desempenho do mapreduce com rede heterogênea. *WSCAD - Workshop de Iniciação Científica*, October 2011.

Agradecimentos

Este trabalho foi parcialmente financiado pelo Edital FAPERGS/CNPq n. 008/2009 Projeto de Pesquisa: **“GREEN-GRID: Computação de Alto Desempenho Sustentável”**.

Este trabalho foi feito com colaboração do Subgrupo de pesquisa em MapReduce, pertencente ao Grupo de Processamento Paralelo e Distribuído.