

Dissertação de Mestrado

Estudo e Implementação de Códigos Corretores de Erros no Sistema de Arquivos Distribuído do Hadoop

Celina d' Ávila Samogin
Instituto de Computação — UNICAMP

12 de novembro de 2012

Orientadora: Profa. Dra. Islene Calciolari Garcia

1 Introdução

2 Introdução

3 Referências Bibliográficas

Agenda

1 Introdução

2 Introdução

3 Referências Bibliográficas

Motivação

- Este trabalho é uma contribuição para *software* livre em sistemas distribuídos.
- Armazenamento de arquivos é um componente essencial na computação de alto desempenho.
- Códigos Corretores de Erro (*Erasure codes*) introduzem redundância e tem sido utilizados em sistemas para alcançar confiabilidade e redução do custo de armazenamento.

- Alguns sistemas que utilizam códigos corretores de erros:
 - ▶ *NASA's Deep Space Network* no envio e na recepção de sinais e dados de telemetria (*downlinks*) vindos de veículos espaciais (*very distant spacecrafts*) e para enviar telecomandos (*uplinks*) para veículos espaciais [?, ?, ?, ?];
 - ▶ *Delay and Disruption Tolerant Networks*, redes de sensores e redes *peer-to-peer* [?, ?, ?, ?, ?, ?, ?];
 - ▶ armazenamento de grande volume de dados [?, ?, ?, ?, ?, ?, ?], como também o sistema de arquivos distribuído do Hadoop (HDFS) [?].

Motivação

- O HDFS, por padrão, implementa alta disponibilidade dos dados via replicação simples dos blocos de dados. Esta abordagem acarreta um alto custo de armazenamento para garantir que os dados estarão sempre disponíveis.
- Esforços iniciais nessa linha foram feitos utilizando técnicas de *Redundant Array of Independent Drives* (RAID) [?, ?] e mais recentemente do algoritmo Reed-Solomon (RS) [?, ?, ?].

Motivação

- RAID
- Reed-Solomon

Objetivos deste trabalho

- avaliação de desempenho, ganhos, e custos de diferentes estratégias de códigos corretores de erro;
- implementação de novos algoritmos (e.g., Tornado codes) e extensão da interface atual para aceitá-los;
- integração do código atual com o HDFS.

Agenda

1 Introdução

2 Introdução

3 Referências Bibliográficas

Motivação

- Software Livre, Open-source
- Sistemas Distribuídos

Agenda

1 Introdução

2 Introdução

3 Referências Bibliográficas

Referências Bibliográficas I