

Execução de aplicações saco-de-tarefas com o Iguassu

EQUIPE

Francisco Vilar Brasileiro
Universidade Federal de Campina Grande (UFCG)

Thiago Emmanuel Pereira
Universidade Federal de Campina Grande (UFCG)

SITE

<http://cloudlab-brasil.rnp.br/iguassu>

PARCEIROS

Rede Nacional de Ensino e Pesquisa (RNP)

CONTATO

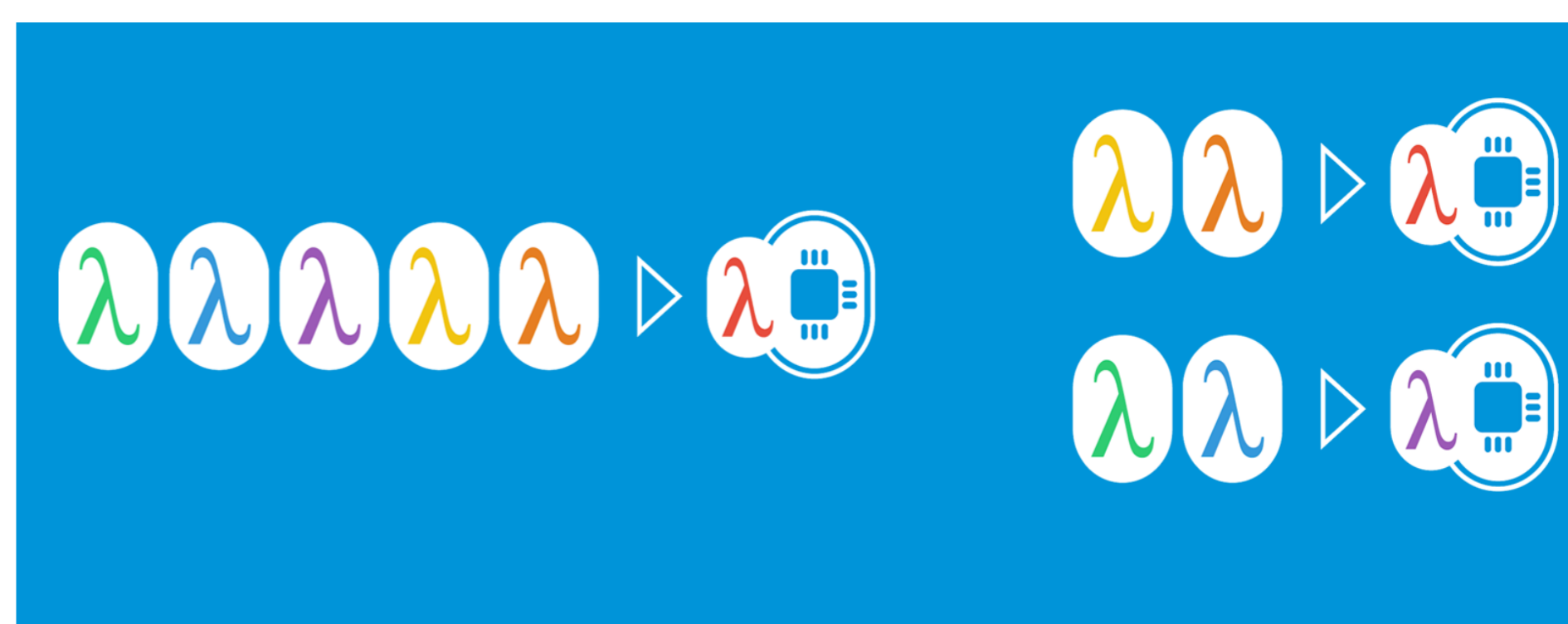
fubica@computacao.ufcg.edu.br
temmanuel@computacao.ufcg.edu.br



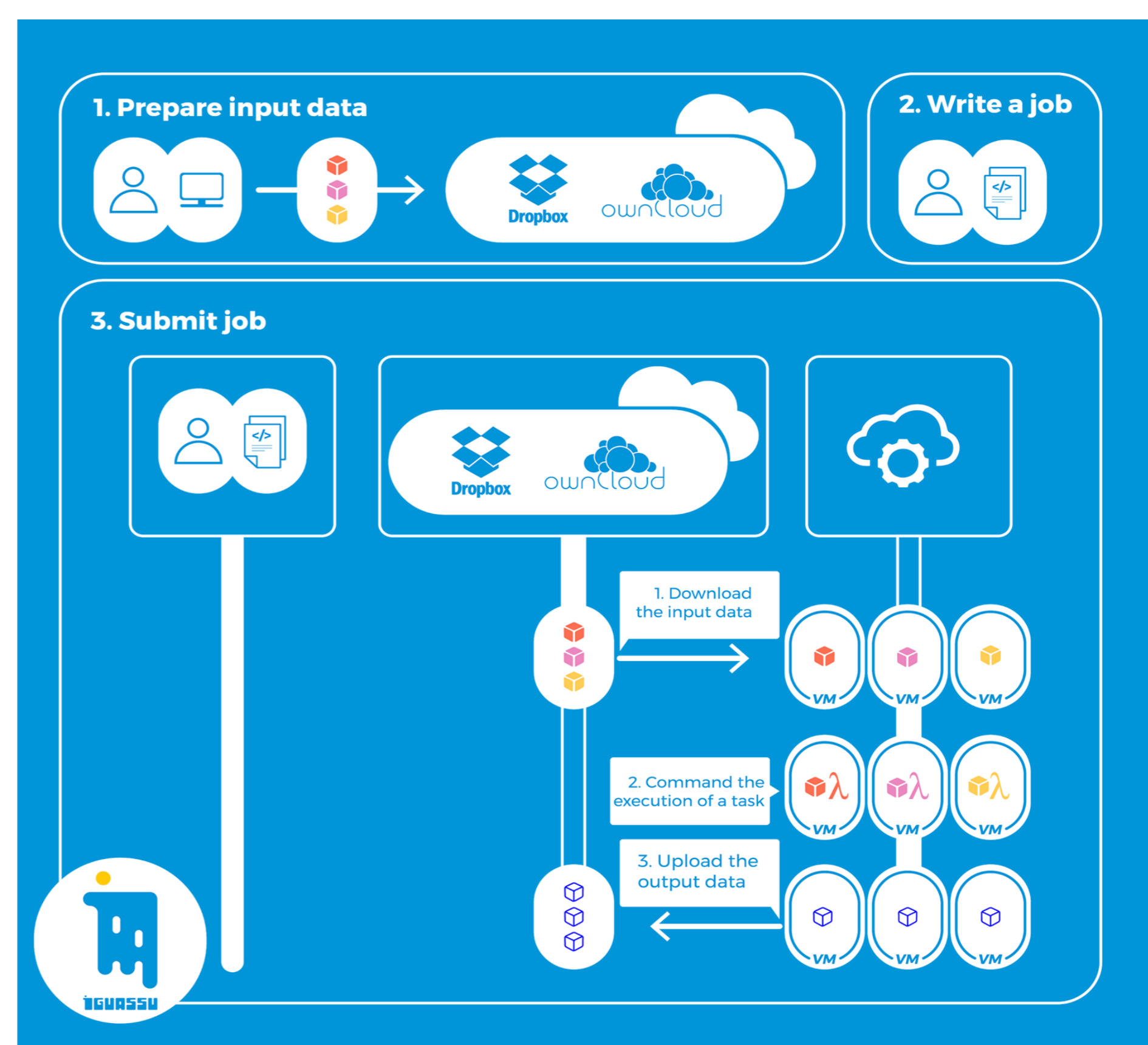
DESCRIÇÃO

O Iguassu é um serviço de alta-velocidade para execução de aplicações saco-de-tarefas. Muitas aplicações de e-ciência seguem o modelo saco-de-tarefas. É o caso, por exemplo, de simulações Monte-Carlo, renderização de imagens e algoritmos de busca em bases de dados genômicas. A execução dessas aplicações em nuvens computacionais não é fácil; é necessário alocar as máquinas virtuais para a execução das tarefas, monitorar os recursos alocados (para, por exemplo, tratar falhas), entre outras preocupações. O Iguassu automatiza todos esses aspectos, de modo que os seus usuários possam se dedicar somente em escrever suas aplicações e analisar os resultados obtidos. Nós apresentamos a arquitetura que permite ao Iguassu implementar essas funcionalidades, incluindo a integração com o serviço de armazenamento Edudrive (<https://edudrive.rnp.br/>), operado pela RNP, e o middleware de integração para orquestradores de nuvem Fogbow (<http://www.fogbowcloud.org/>) que atua como um broker para várias soluções disponíveis no mercado (OpenStack, CloudStack, Opennebula, Azure e Amazon EC2).

Aplicações intensivas em computação, proporcionalmente, executam muito mais instruções aritméticas do que instruções que movimentam dados. Em muitos casos, essas aplicações são compostas por conjuntos de tarefas similares (p.ex o mesmo programa executado com diferentes parâmetros). Esse é o caso de aplicações de *protein folding* e simulações de modo geral.

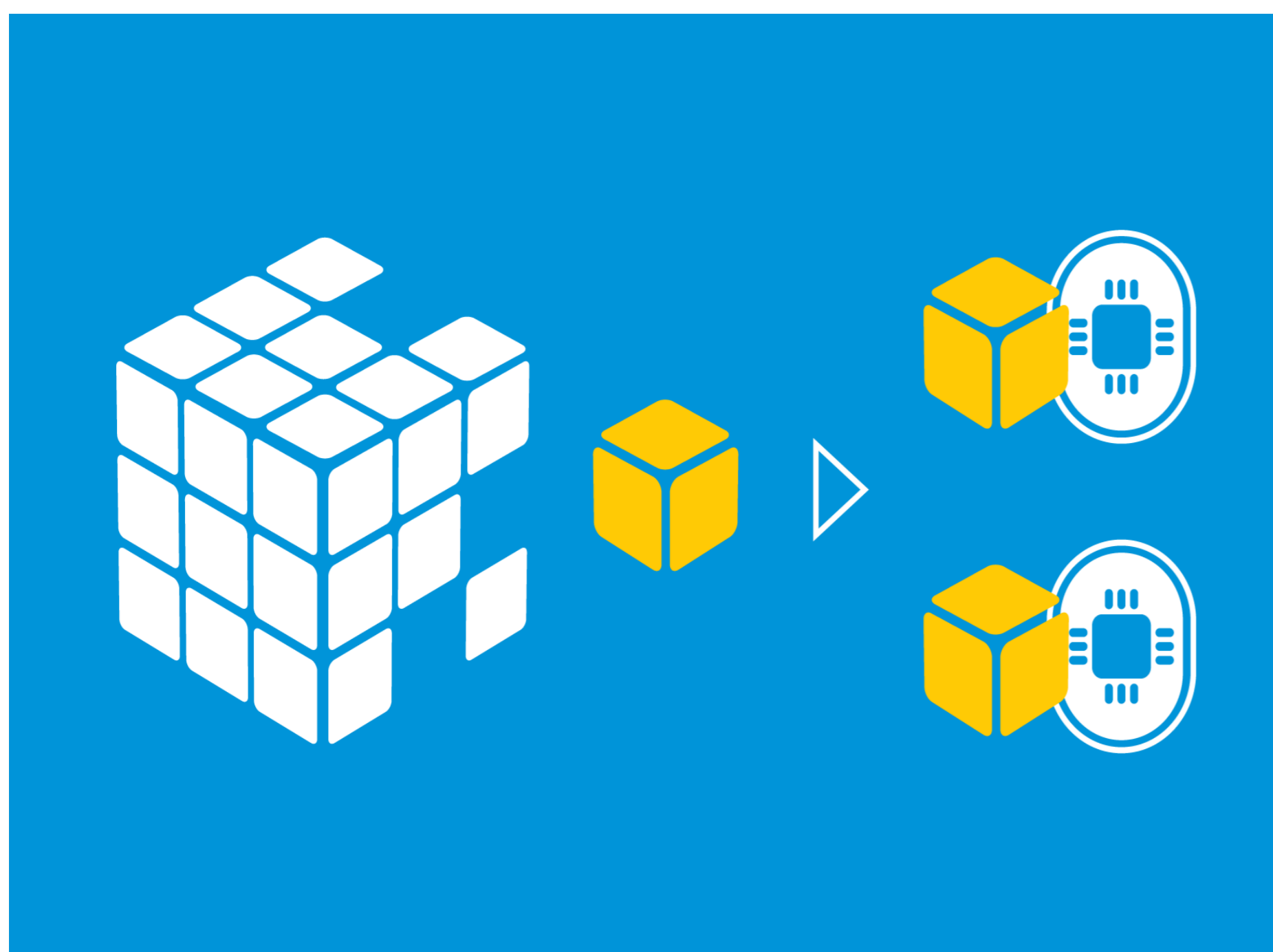


Arquitetura Iguassu



Aplicações saco-de-tarefas

Aplicações intensivas em dados dedicam boa parte de seu tempo de execução manipulando dados de entrada e de saída. Uma otimização importante quando o volume de dados consiste em dividir a coleção de dados em pedaços menores, enviadas para unidades de processamento independentes, e processadas em paralelo.



1. Prepare input data



2. Write a job



3. Submit job

