

# Armazenamento Distribuído

## Aula 2

Prof. Mario A. R. Dantas

Departamento de Informática e Estatística (INE)  
Universidade Federal de Santa Catarina (UFSC)

Apresentado por Eduardo Camilo Inacio  
Programa de Pós-Graduação em Ciência da Computação (PPGCC)  
[eduardo.camilo@posgrad.ufsc.br](mailto:eduardo.camilo@posgrad.ufsc.br)

INE5625 - Sistemas Distribuídos  
INE5418 - Computação Distribuída

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Sistemas de arquivos
  - **Abstração** para utilização de dispositivos de armazenamento
- Arquivos, metadados e estrutura de diretórios
- Conjunto de operações para arquivos, metadados e diretórios
  - Criação, remoção, escrita, leitura, etc.
- Estrutura em camadas
  - Interface com sistema de arquivos, virtual file system, sistemas locais e remotos

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Sistemas de arquivos distribuídos
  - Dados em servidor(es) remoto(s)
  - Compartilhamento de arquivos entre processos
- Arquiteturas de SAD
  - Cliente-servidor
  - Baseado em cluster
  - Simétrico

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## ■ Network File System (NFS)

- Arquitetura básica
- Comunicação cliente/servidor via RPC
- Espaço de nomes e montagem
- Caching no lado cliente
- Prós e contras

# Sistemas de Arquivos Paralelos

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

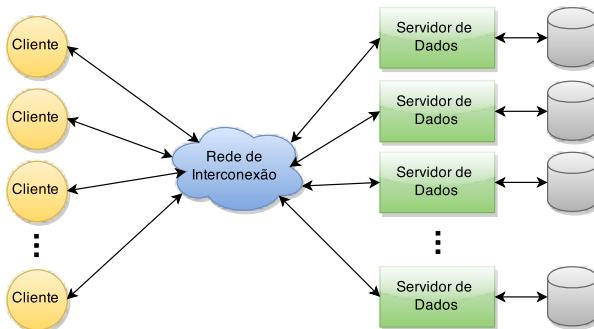
### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## ■ Sistemas de Arquivos Paralelos (SAP)

- Conceitos básicos
- Arquitetura
- Particionamento e Distribuição
- Operação



# Conceitos Básicos

## Na Aula Anterior

### SAP

Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais

- Arquitetura baseada em **cluster**
- Projetado para **paralelismo**
  - Acesso concorrente de **muitos** processos
- Projetado para **alto desempenho**
  - Operação sobre redes de interconexão de **alta velocidade**
  - Otimizações de E/S para **máximo desempenho**

# Conceitos Básicos

## Na Aula Anterior

### SAP

#### Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais

- **Agregação de recursos**
  - Escalabilidade em **espaço de armazenamento**
  - Otimização da **utilização** de recursos
- **Particionamento e distribuição** de arquivos
  - Prover **maior desempenho**

## Na Aula Anterior

### SAP

Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações Finais

# Qual a diferença entre Sistema de Arquivos Paralelo (SAP) e Sistema de Arquivos Distribuído (SAD)???





## Na Aula Anterior

### SAP

#### Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais

- Classificação de TANENBAUM; STEEN, 2006
  - **Arquitetura**
- Sistema de Arquivos Distribuídos
  - Cliente-servidor
  - **Baseado em cluster**
    - Sistema de Arquivos Paralelo (SAP)**
  - Simétrico

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Nomenclatura **não uniforme** para sistemas de arquivos com arquitetura baseada em cluster, porém **diferente dos SAPs**:
  - Distributed file systems (mais comum)
  - Internet services file systems
  - Cluster file systems
- Adotaremos o nome Sistema de Arquivos Distribuídos (SAD) para esse grupo

## Na Aula Anterior

### SAP

#### Conceitos Básicos

Arquitetura

Particionamento e

Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais

## ■ Simetria

- SADs usualmente executam em arquiteturas onde a aplicação e os dados de entrada (arquivos) são **co-aloçados**  
Ex.: HDFS, GoogleFS
- Nos SAPs, armazenamento e computação estão, em geral, **fisicamente separados**  
Ex.: OrangeFS/PVFS2, Lustre

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## ■ Carga de trabalho

- SAD: em geral, aplicações distribuídas **fracamente acopladas**

Ex.: Data-intensive, big data

- SAP: acessos **coordenados**, com requisitos de taxas de transmissão massivas

Ex.: aplicações científicas e de engenharia

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Tolerância a falta (falta → erro → falha)
  - SADs costumam assumir responsabilidades de tolerância a falta
  - SAPs usualmente "confiam" na arquitetura de armazenamento

# Arquitetura

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

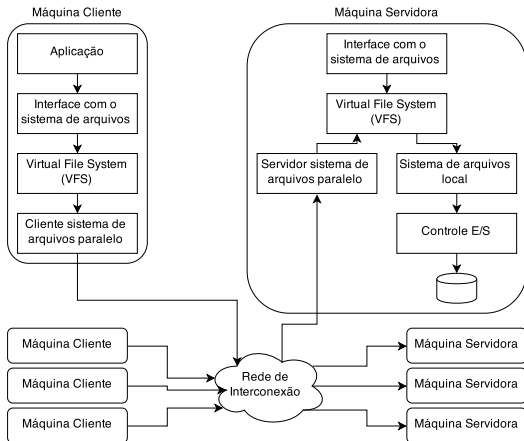
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



## Arquitetura básica de um sistema de arquivos paralelo

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

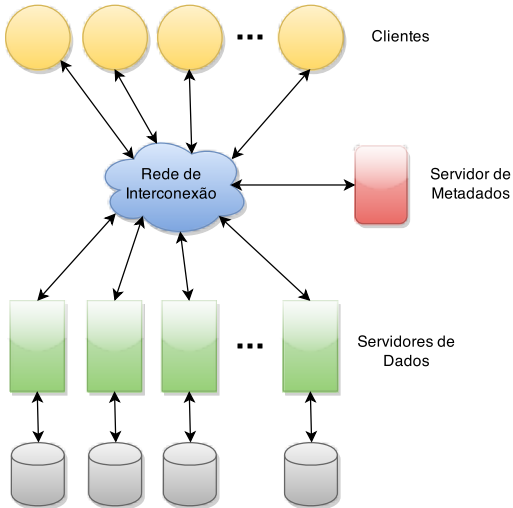
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



Componentes de um sistema de arquivos paralelo

## Na Aula Anterior

### SAP

- Conceitos Básicos
- Arquitetura**
- Particionamento e Distribuição
- Operação

### OrangeFS/PVFS2

- Visão Geral
- Arquitetura

### HDFS

- Hadoop MapReduce
- Visão Geral
- Arquitetura
- HDFS x OrangeFS

### Considerações Finais

## Servidor de dados

Responsável pela **persistência do conteúdo** dos arquivos. Sua função é interagir com o(s) dispositivo(s) de armazenamento controlado(s) por ele para armazenar ou recuperar arquivos.



## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## Servidor de metadados

Responsável por manter atualizadas **informações referentes aos arquivos** (metadados). Isso inclui não apenas a manutenção de dados como nome e permissões, mas também a **estrutura de diretórios** como um todo.

- Centralizado: Ceph, Lustre
- Distribuído: OrangeFS/PVFS2, GPFS

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## Cliente

Possibilita a **interação com o sistema** de arquivos paralelo. Esse componente está usualmente localizado nos **nodos de computação** e é utilizado pelas aplicações para realização de operações sobre o sistema de arquivos.

- Interface (API) nativa
- POSIX, MPI-IO, e outras

# Particionamento e Distribuição

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

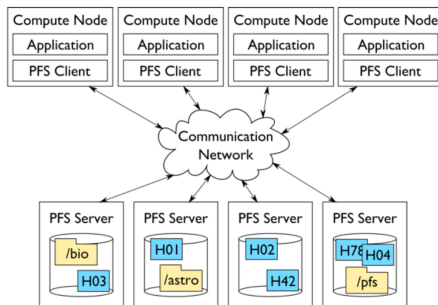
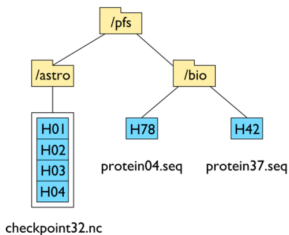
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



Armazenamento de uma estrutura de diretórios em um SAP

# Particionamento e Distribuição

File striping

## Na Aula Anterior

### SAP

Conceitos Básicos

Arquitetura

Particionamento e Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

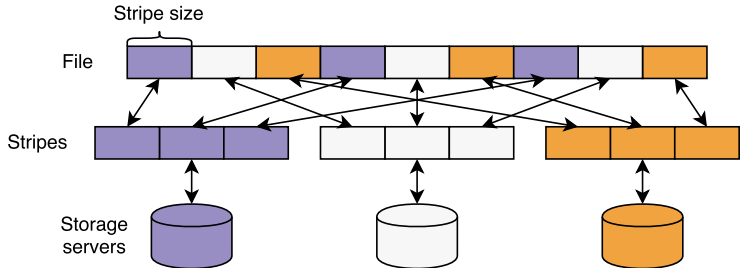
Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais



Processo de divisão de arquivo (file striping)

### Na Aula Anterior

#### SAP

- Conceitos Básicos
- Arquitetura
- Particionamento e Distribuição
- Operação

#### OrangeFS/PVFS2

- Visão Geral
- Arquitetura

#### HDFS

- Hadoop MapReduce
- Visão Geral
- Arquitetura
- HDFS x OrangeFS

#### Considerações Finais

## File striping

**Dividir** o arquivo em segmentos consecutivos, ou **faixas** (do inglês, stripes), e **distribuir** essas faixas entre os servidores de dados. Os tamanhos das faixas são em geral configuráveis no sistema de arquivos paralelo, podendo ser otimizadas para a carga de trabalho em questão.

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Permite que o armazenamento e a recuperação do conteúdo de um arquivo sejam realizados de forma **paralela**
- Apresenta **degradação de desempenho** para arquivos pequenos
  - Custos de processamento, comunicação, etc.

# Particionamento e Distribuição

Layout de distribuição

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

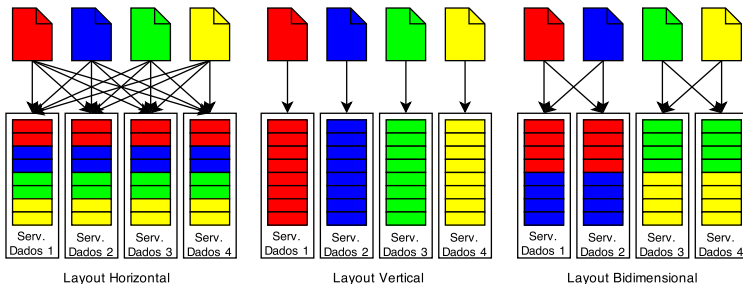
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



## Layouts de distribuição de arquivos em SAPs

# Particionamento e Distribuição

Layout de distribuição

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- **Horizontal** (ou round robin): faixas distribuídas entre todos os servidores de dados (máximo de paralelismo)
- **Vertical**: todas as faixas de um arquivo em um mesmo servidor de dados (sem paralelismo)
- **Bidirecional**: híbrido dos anteriores; faixas são distribuídas entre um subconjunto de servidores de dados



## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Uma aplicação criando e escrevendo 1 MB de dados em um arquivo armazenado em um SAP com 8 servidores de dados, um servidor de metadados e tamanho de faixa (stripe size) igual a 64 KB
  - 1 Aplicação solicita (ao cliente) a criação do arquivo
  - 2 Cliente envia solicitação para o servidor de metadados
  - 3 Servidor de metadados cria registro para o arquivo
  - 4 Aplicação faz uma requisição de escrita (ao cliente) com 1 MB de dados
  - 5 Cliente divide a requisição em 16 faixas de 64 KB
  - 6 Cliente envia 2 faixas para cada um dos 8 servidores de dados

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Uma aplicação faz a leitura de 1 MB de dados de um arquivo armazenado em um SAP com 8 servidores de dados, um servidor de metadados e tamanho de faixa (stripe size) igual a 64 KB
  - 1 Aplicação solicita (ao cliente) a localização do arquivo
  - 2 Cliente envia a solicitação para o servidor de metadados
  - 3 Servidor de metadados informa localização das faixas do arquivo
  - 4 Aplicação faz uma requisição de leitura (ao cliente) de 1 MB de dados
  - 5 Cliente solicita aos 8 servidores de dados as faixas do arquivo
  - 6 Cliente "reconstrói" arquivo e retorna a aplicação

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- OrangeFS/PVFS2
  - Visão geral
  - Arquitetura



## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Sistema de arquivos paralelo com arquitetura baseada em cluster
- Extensão do Parallel Virtual File System v2 (PVFS2)
  - Desenvolvido pelo Argonne National Laboratory (ANL) e pela Clemson University
- Atualmente suportado pela Omnibond
- Open source

# Visão geral

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

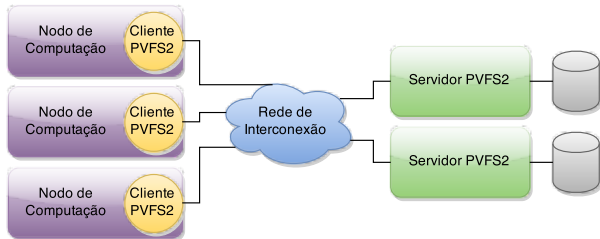
Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Múltiplos servidores colaborando para prover um sistema de arquivos com um único espaço de nomes
- Possibilidade de acessos concorrentes aos objetos do sistema de arquivos
- Comunicação confiável: TCP/IP



# Visão geral

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

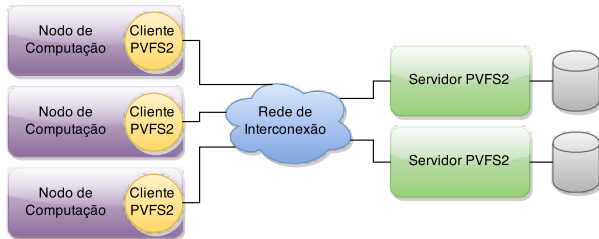
Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Distribuição do conteúdo de um arquivo entre múltiplos servidores
- Servidor PVFS2 pode ser configurado como servidor de dados e/ou de metadados
- Utiliza o sistema de arquivos local do servidor de dados



### Na Aula Anterior

#### SAP

Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

#### OrangeFS/PVFS2

Visão Geral

Arquitetura

#### HDFS

Hadoop MapReduce

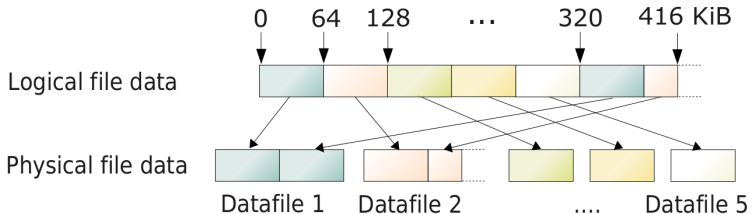
Visão Geral

Arquitetura

HDFS x OrangeFS

#### Considerações Finais

- Layouts de distribuição customizáveis
- Stripe size configurável (default 64 KB)
- Cada faixa (datafile) é armazenada em apenas um servidor de dados



## Na Aula Anterior

### SAP

Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais

- Não possui mecanismos de tolerância a falta a nível de software
  - Assume hardware confiável
- Alta disponibilidade a nível de hardware requer armazenamento compartilhado
  - Ex.: Storage Area Network (SAN)
- Múltiplos servidores podem acessar o mesmo dispositivo de armazenamento



# Arquitetura

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

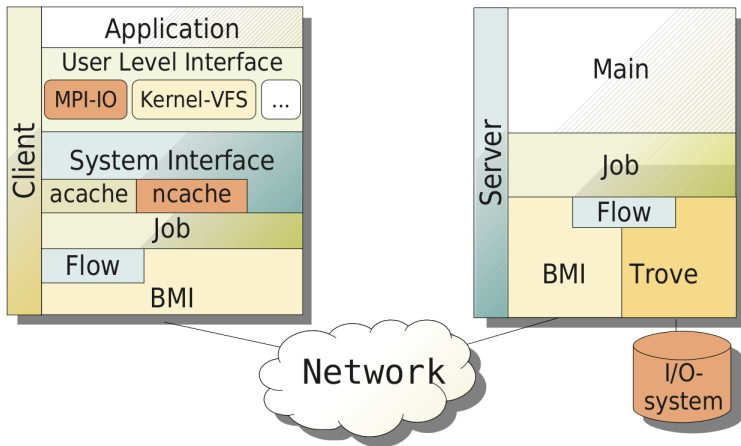
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



Arquitetura do OrangeFS/PVFS2

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

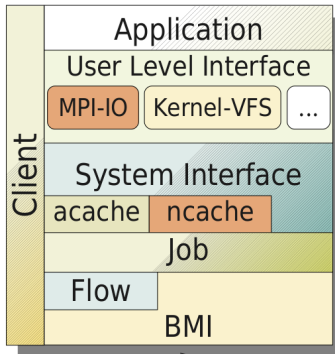
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



## ■ Interface de nível de usuário

- Suporte POSIX via módulo de kernel para integração com VFS
- Módulo ROMIO para MPICH2

## ■ Interface de sistema

- API para manipulação de objetos do sistema
- Caches para hierarquia de diretórios e metadados

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

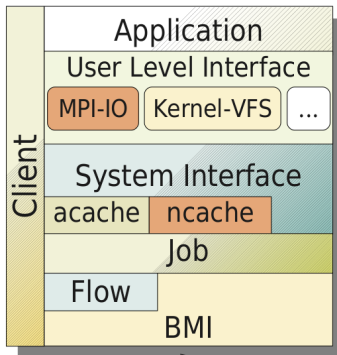
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



### ■ Job

- Camada de abstração e controle de camadas inferiores

### ■ Flow

- Confiabilidade de transferência de dados
- Políticas de fluxo

### ■ BMI

- Interface de rede: TCP, Myrinet, IB

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

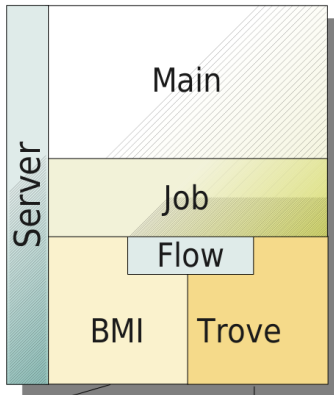
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



- Processo principal (Main)
  - Atendimento de requisição
  - Inicialização de máquinas de estados
- Job, Flow, BMI
  - Mesma função do cliente
- Trove
  - Camada de persistência
  - Metadados: Berkeley DB
  - Dados: sistema de arquivos local

# Hadoop File System (HDFS)

## Na Aula Anterior

### SAP

Conceitos Básicos

Arquitetura

Particionamento e  
Distribuição

Operação

### OrangeFS/PVFS2

Visão Geral

Arquitetura

### HDFS

Hadoop MapReduce

Visão Geral

Arquitetura

HDFS x OrangeFS

### Considerações

### Finais

## ■ Hadoop File System (HDFS)

- Hadoop MapReduce
- Visão Geral
- Arquitetura
- HDFS x OrangeFS



# Hadoop MapReduce

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

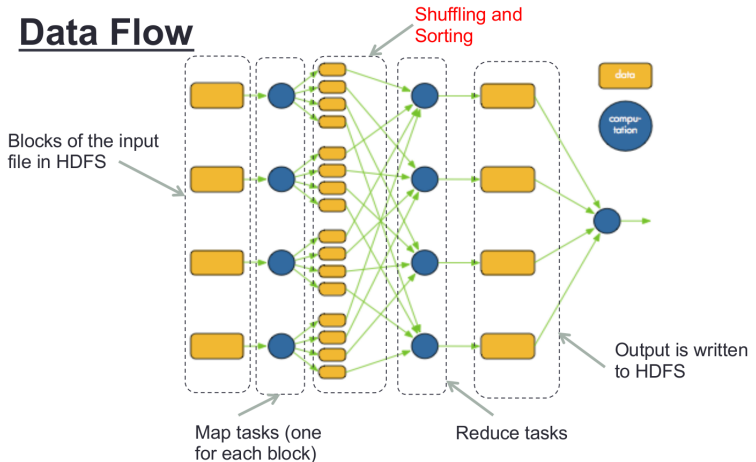
Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## Data Flow



# Hadoop MapReduce

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Implementação open-source do Google MapReduce
- Framework Java
- Localidade de dados
  - Computação "vai" até os dados
- Chave-valor
  - Processamento da entrada
  - Mapeamento dos *reduces*

# Hadoop MapReduce

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

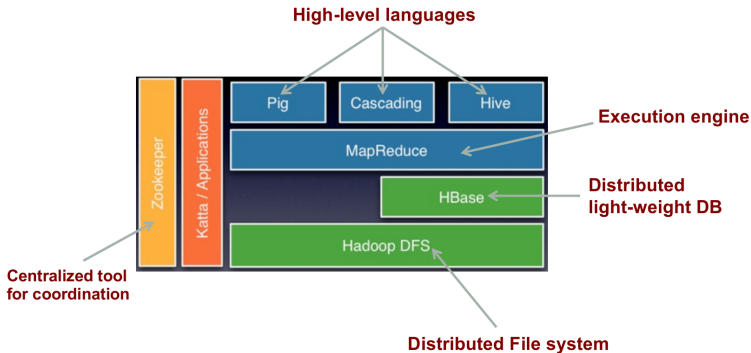
### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais



MapReduce e HDFS são suficientes!



## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Implementação open-source do Google File System (GFS)
- Escrito em Java
- Otimizado para o MapReduce
- Armazena os dados em múltiplos servidores
- Executa sobre hardware comum

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
**Visão Geral**  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Tolerância a falta
  - Replicação de dados entre nodos e *racks*
  - Quantidade de réplicas configurável
  - Arquitetura *master/worker*

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

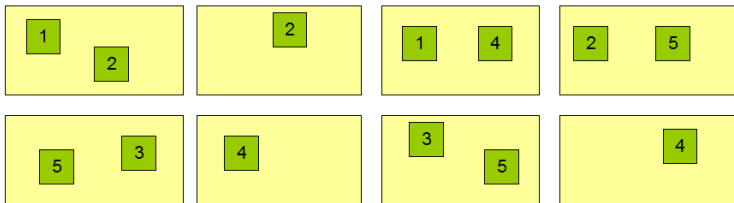
Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## Block Replication

Namenode (Filename, numReplicas, block-ids, ...)  
/users/sameerp/data/part-0, r:2, {1,3}, ...  
/users/sameerp/data/part-1, r:3, {2,4,5}, ...

## Datanodes



## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

#### Block Replication

Namenode (Filename, numReplicas, block-ids, ...)  
/users/sameerp/data/part-0, r:2, {1,3}, ...  
/users/sameerp/data/part-1, r:3, {2,4,5}, ...

#### Datanodes



- Namenode
  - Manutenção dos metadados

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## Block Replication

Namenode (Filename, numReplicas, block-ids, ...)  
/users/sameerp/data/part-0, r:2, {1,3}, ...  
/users/sameerp/data/part-1, r:3, {2,4,5}, ...

## Datanodes



## ■ Datanode

- Armazenam blocos de tamanho fixo dos arquivos (default = 64 MB)
- Blocos são replicados para tolerância a falta e rápido acesso (default = 3)
- Enviam *heartbeats* periodicamente para o namenode

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

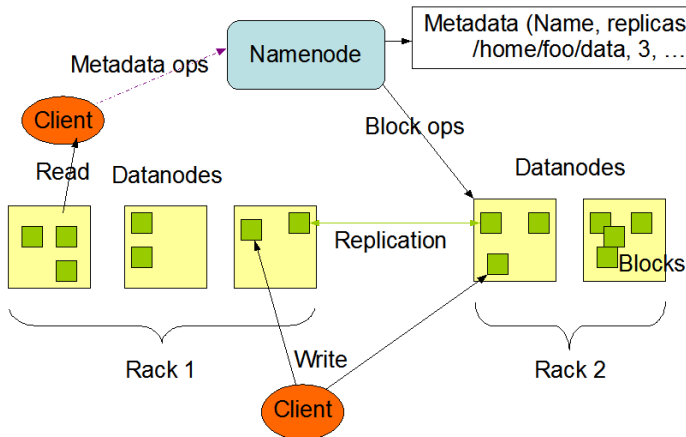
Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## HDFS Architecture



## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Datanodes podem ser organizados em racks
  - Políticas orientadas pelo rack
- Política de alocação de dados
  - **1a. cópia:** nodo que escreve o dado
  - **2a. cópia:** nodo no mesmo rack
  - **3a. cópia:** nodo em um rack diferente
  - Balancimento de carga, acesso rápido e tolerância a falta
- Podem haver mais Namenodes
  - Apenas um ativo

# HDFS x OrangeFS

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

## Algumas diferenças entre o HDFS e o OrangeFS

Característica	HDFS	OrangeFS
Arquitetura	Localidade de dados	Computação e armazenamento separados
Tolerância a faltas	Replicação de blocos	Hardware "confiável"
Paralelização de acesso	Replicação de blocos	Particionamento e distribuição
Interface de comunicação	POSIX (parcial)	API nativa, POSIX, MPI-IO
Metadados	Único Namenode	Servidores distribuídos



# Considerações Finais

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- Sistemas de arquivos paralelos
- Principais características:
  - Cluster de armazenamento
  - Particionamento e distribuição de arquivos
  - Paralelização de operações
- Diferenças entre SAP e SAD
  - Simetria
  - Carga de trabalho
  - Tolerância a falta

# Considerações Finais

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

## Considerações Finais

- Arquitetura e principais componentes
  - Cliente
  - Servidor de metadados
  - Servidor de dados
- Processo de file striping
- Layouts de distribuição

# Considerações Finais

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

### Considerações Finais

- OrangeFS/PVFS2
- Sistema de arquivos paralelo open source
- Altamente configurável
  - Stripe size, layout de distribuição, servidores de dados e de metadados
- Suporte para POSIX e comunicação TCP/IP
- Utiliza o sistema de arquivos local no servidor de dados

# Considerações Finais

## Na Aula Anterior

### SAP

Conceitos Básicos  
Arquitetura  
Particionamento e  
Distribuição  
Operação

### OrangeFS/PVFS2

Visão Geral  
Arquitetura

### HDFS

Hadoop MapReduce  
Visão Geral  
Arquitetura  
HDFS x OrangeFS

## Considerações Finais

### ■ Hadoop File System (HDFS)

- Hadoop MapReduce
- Visão Geral
- Arquitetura
- HDFS x OrangeFS