



# Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

## Data Lake – Design, Projeto e Integração

### Recomendações de Hardware Para o Cluster Hadoop



Vejamos quais são as recomendações de hardware para um cluster Hadoop.

### Recomendações de hardware para os nós slaves:

Configuração	Descrição
Storage	Em um ambiente de intensivo de I/O, recomenda-se 12 discos SATA 7200 RPM de 2 TB cada um, para balanceamento entre custo e performance. RAID não é recomendado em máquinas com serviços slaves do Hadoop.
Memória	Nodes slaves requerem normalmente entre 24 e 48 GB de memória RAM. Memória não utilizada será consumida por outras aplicações Hadoop.
Processador	Processadores com clock médio e menos de 2 sockets são recomendados.
Rede	Cluster de tamanho considerável, tipicamente requer links de 1 GB para todos os nodes em um rack com 20 nodes.

### Recomendações de hardware para os nós master:

Configuração	Descrição
Storage	Deve-se utilizar 2 servidores: um para o NameNode Principal e outro para o Secundário. O Master deve ter pelo menos 4 volumes de storage redundantes, seja local ou em rede.
Memória	64 GB de RAM suportam aproximadamente 100 milhões de arquivos
Processador	16 ou 24 CPU's para suportar o tráfego de mensagens

**Estas são apenas recomendações e que podem variar de acordo com os fatores para o planejamento do cluster: objetivo, serviços e layout.**

**Para o nosso Data Lake, vamos simular um ambiente virtual usando VirtualBox!**