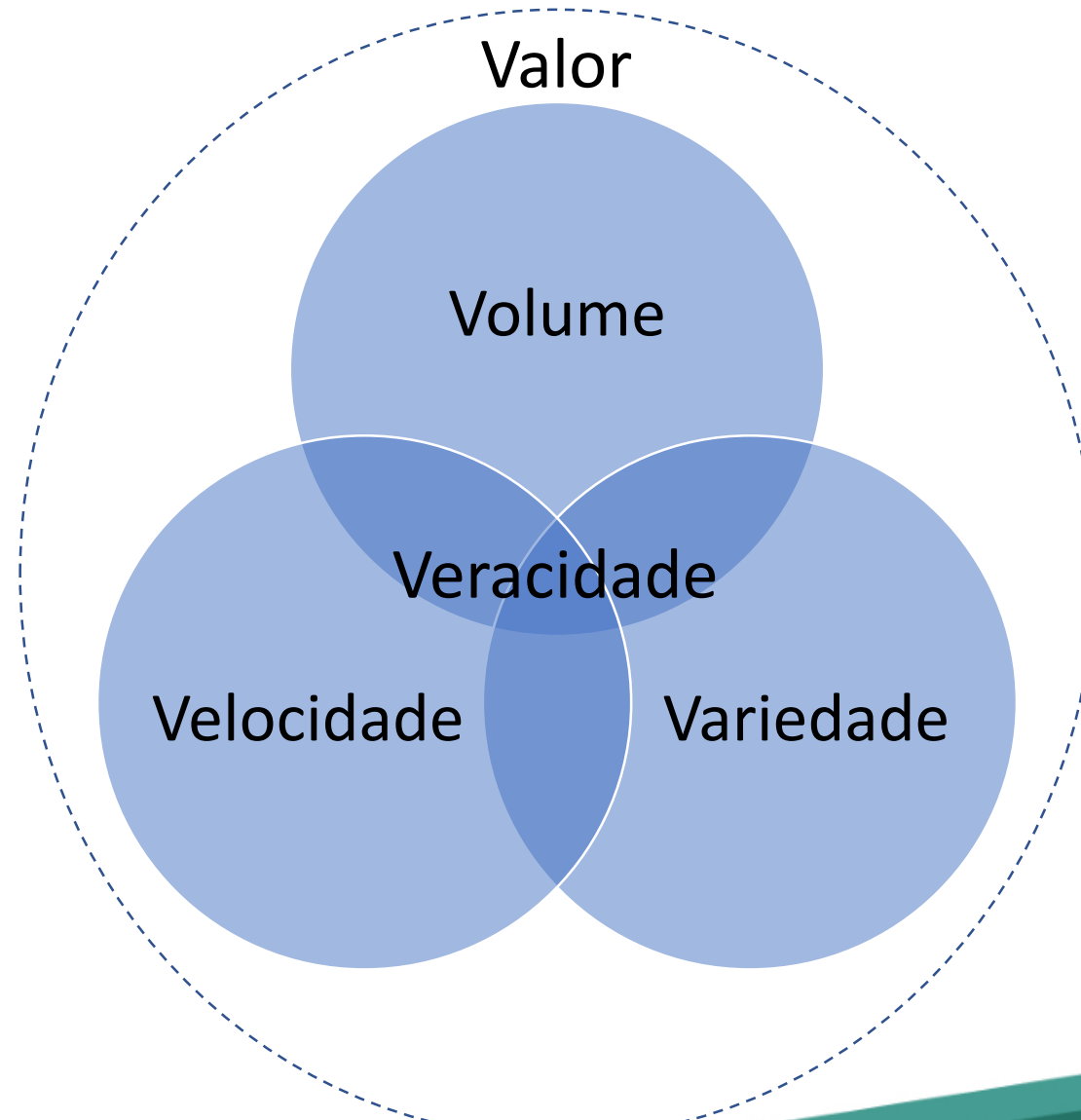


BIG DATA



Big Data



Infraestrutura para Big Data



Computação de Alta Performance

- ✓ Computação de alto desempenho ou HPC (do inglês High-performance computing) é o uso de supercomputadores ou clusters de vários computadores em tarefas que requerem grandes recursos de computação
- ✓ Aplicação em simulações numéricas muito complicadas



Exemplos de HPC

- HPC
- MPP
- SAN
- Mainframe
- Appliance Databases
- RAID

Computação na Nuvem

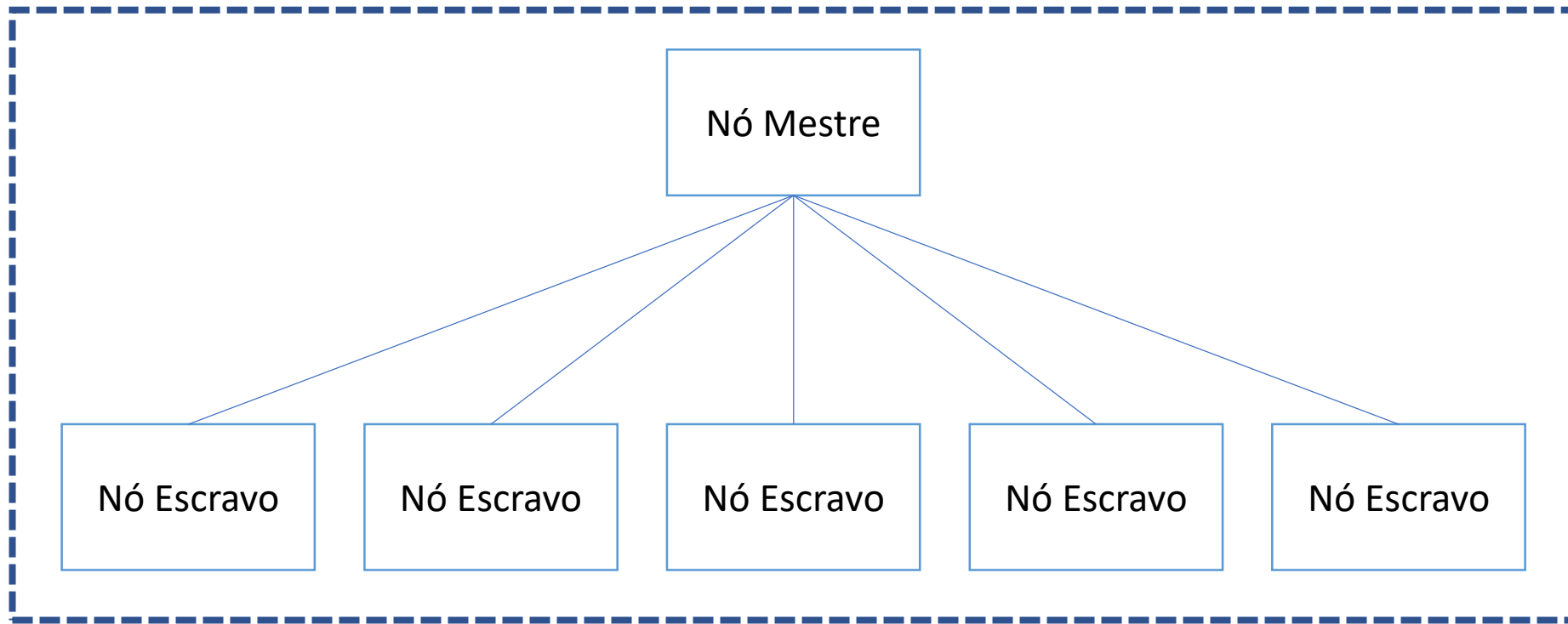
- ✓ O conceito refere-se à utilização da memória e da capacidade de armazenamento e processamento de computadores e servidores compartilhados e interligados por meio da Internet



Principais Tipos de Computação na Nuvem

- ✓ **IaaS** - Infrastructure as a Service
- ✓ **PaaS** - Plataform as a Service
- ✓ **DaaS** - Development as a Service
- ✓ **SaaS** - Software as a Service
- ✓ **DBaaS** - Data Base as a Service

Computação Distribuída - Cluster



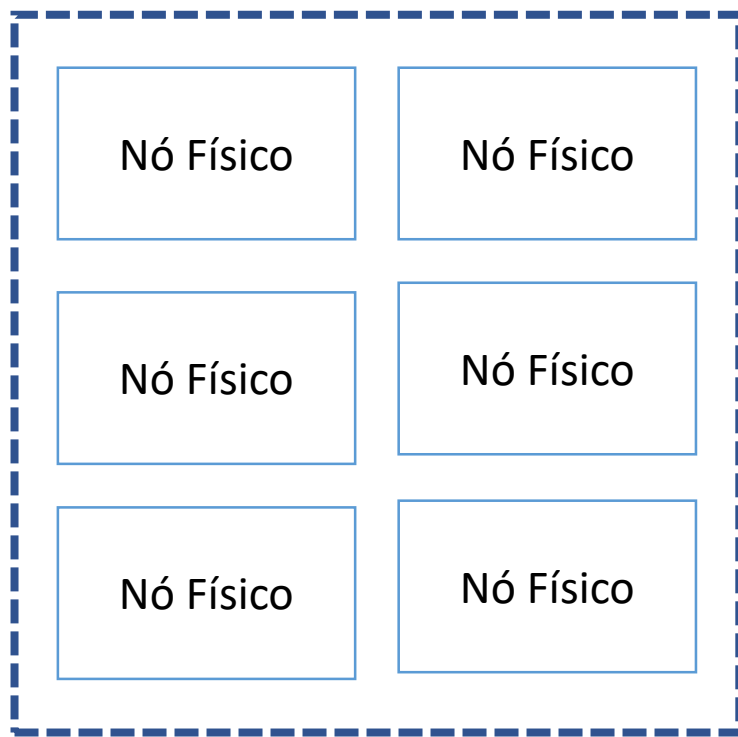
Computação Massiva Paralela

- ✓ É o uso de uma grande números de processadores para executar um conjunto coordenado de tarefas computacionais simultaneamente

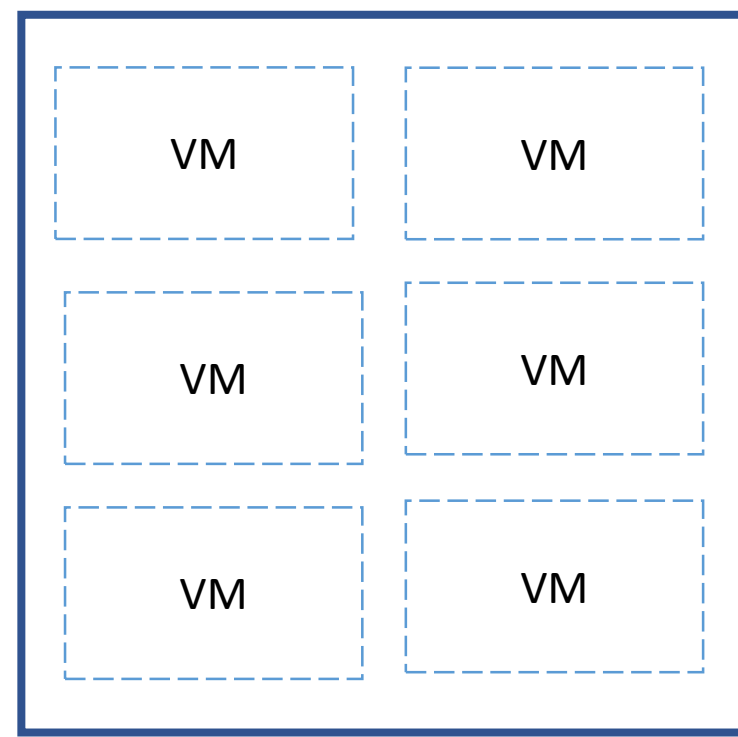


Computação Distribuída x Virtualização

Servidor Virtual



Servidor Físico



Evolução do armazenamento de dados

Capacidade dos discos x velocidade de acesso

1990 - 1.370 MB Drive, taxa de transferência 4.4 MB/s - 5 minutos

2010 - 1TB, taxa de transferência 100 MB/s - 2,5h

100 x 1TB em paralelo – 2 minutos -> Solução usada no Hadoop

Princípios do Hadoop

Scale-Out, não Scale-Up

Trazer o processamento para os dados, não o contrário

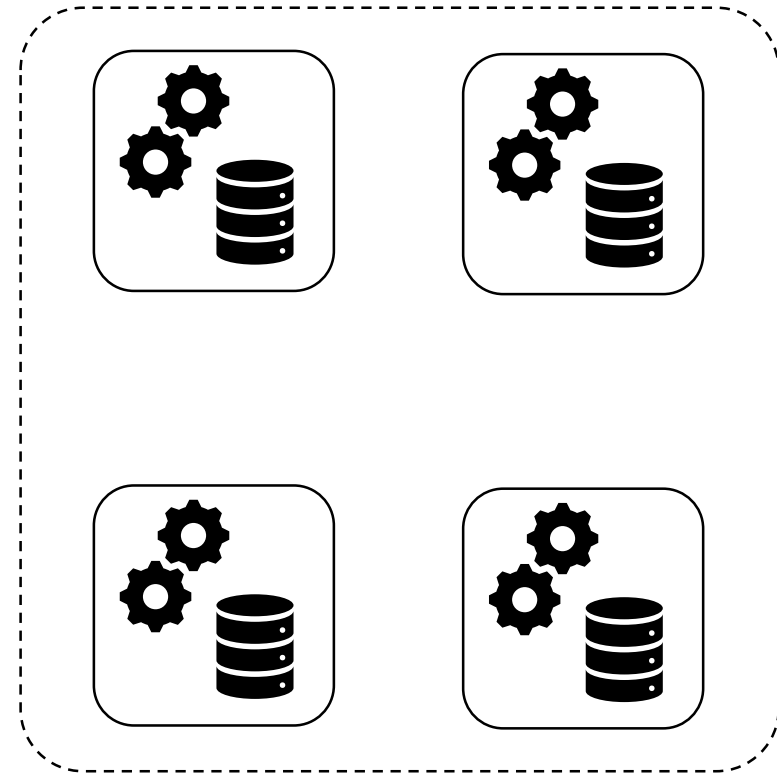
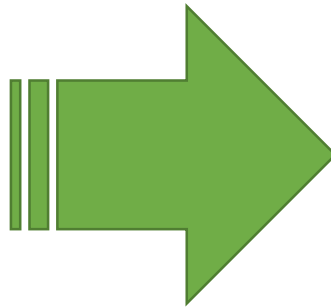
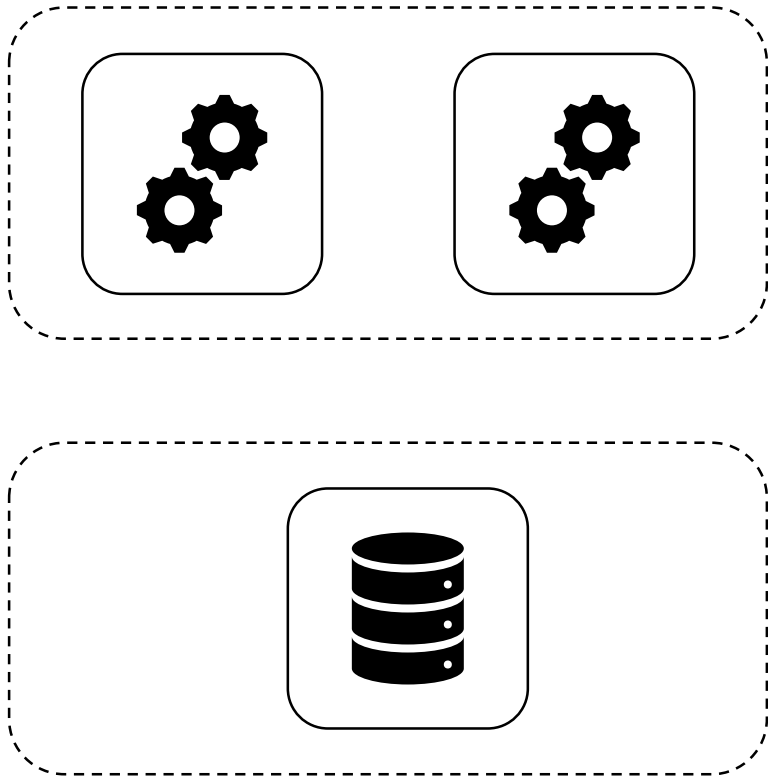
Falhas acontecem, então lide com elas

Abstrair a complexidade de aplicações concorrentes e distribuídas

Scale-Out, não Scale-Up

- Scale-Up (upgrade / escalabilidade vertical)
- É mais difícil adicionar mais recursos para um servidor já existente (CPU,RAM)
- Novas unidades devem ser adquiridas se os recursos não poderem ser adicionados
- Lei de Moore não acompanha a taxa de crescimento do Big Data
- Scale-Out (escalabilidade horizontal)
- Adicionar mais servidores para um sistema distribuído
- Transparência, abstração facilitada por camada de software

Trazer o processamento para os dados, não o contrário



Em sistema distribuídos, falhas são comuns

- Hadoop lida com as falhas
- Os dados são replicados
- Os trabalhos são retomados

Hardware Commodity

Hardware commodity não significa desktops e servidores muito obsoletos

Appliance vs DIY

