Why Virtualize Hadoop Nodes?

Boost Hardware utilization
Bare metal Hadoop deployments average 10-20% CPU utilization, a major waste of hardware resources and datacenter space. Virtualizing Hadoop allows for better hardware utilization and flexibility.

Dynamically Grow and Shrink Capacity
Expand and contract MapReduce data crunching capacity with the addition and removal of nodes based on load. Avoid the perils of physical server inflexibility.

Allow DevOps & IT Ops to live harmony
Big Data scientists demand performance, reliability, and a flexible scale model. IT Ops relies on virtualization to tame server sprawl, increase utilization, encapsulate workloads, manage capacity growth, and mitigate disruptive hardware downtime. By virtualizing Hadoop, Data Scientists and IT Ops mutually achieve all objectives while preserving autonomy and independence for their respective charters.

Isolate jobs – Make Hadoop and Enterprise Apps play nice
Buggy MapReduce jobs can quickly saturate hardware resources, creating havoc for remaining jobs in the queue. Virtualizing Hadoop clusters encapsulates and isolates MapReduce jobs from other important sorting runs.

Batch Scheduling & Stacked workloads
Allow all workloads and applications to co-exist, e.g. Hadoop, virtual desktops and servers. Schedule MapReduce job runs during off-peak hours to take advantage of idle night time and weekend hours that would otherwise go to waste.

New Hadoop Economics
Bare metal implementations are expensive and difficult to predict size and scale. Downtime and underutilized CPU consequences of physical servers can jeopardize project viability. Virtualizing Hadoop reduces complexity and ensures success for sophisticated projects with a scale-out grow as you go model – a perfect fit for Big Data projects.
**Nutanix & Hadoop = Enterprise-Grade Big Data**

**Business Continuity and Data Protection:**
As Hadoop clusters gradually become mission critical, enterprise-grade data management features are essential ingredients for successful big data projects. Nutanix has built-in thin provisioning, snapshots, clones, compression, and array-level replication that complements Hadoop projects that mature from PoC to production.

**High Availability:**
Distributed file system name nodes are single points of failure. Nutanix has built-in network RAID and high-availability features to secure all pieces of Hadoop data, including the file system name node.

**Change Management:**
Traditional deployment models make it very difficult to run multiple MapReduce jobs side by side. Nutanix allows you to create parallel MapReduce branches that can execute simultaneously on the same production data using Nutanix Snapshots and FastClone capabilities.

**Flash SSDs for fast NoSQL:**
Business reports created from roll-up summaries posted to NoSQL databases like HBase are typically memory and I/O constrained. Nutanix direct-attached Flash and local SATA SSD storage help accelerate data sorting with its innovative heat-optimized tiering technology. Data is migrated up and down tiers transparently to assist with I/O-thirsty workloads.

**Intuitive Cluster Management:**
Nutanix employs an Apple-like approach to managing large clusters, including a modern dashboard that displays a single pane of glass for servers and storage. This open-source Bonjour mechanism auto-discovers and configures new nodes as they enter and leave the cluster.

**Combine Hadoop with your Data Sources**
Programmers  Data Scientists  Power Users  Business Users

**Scale with Your Business, One Node at a Time**

Visit [www.nutanix.com](http://www.nutanix.com) for more information. Follow us @nutanix Email learnmore@nutanix.com

**Blazing fast performance**
Start with 2,000 MB/s of sequential throughput in a compact 2U 4-node cluster. A Terasort benchmark yields 250 MB/s in the same 2U cluster.

Linear Scale-out Performance
Throughput scales linearly by stacking more blocks. 2x better throughput compared to closest bare metal equivalents.

**Nutanix Complete Cluster at a Glance**
Nutanix SAN-Free architecture:
Delivers best of both worlds: local bus speed performance to virtualized Hadoop workloads with all the benefits of shared storage.

Time-sliced clusters:
Like Amazon EC2, Nutanix clusters coupled with VMware vCloud Director, can run Hadoop and other virtualized workloads on the same shared physical hardware.

High-density Hadoop:
Nutanix uses a hyperscale server architecture in which 8 sockets of Intel fit in a single 2U spread over 4 motherboards. Coupled with data archiving and compression, Nutanix can shrink Hadoop hardware footprints by up to 4x and reduce completion times by 5x. Large clusters set up in hours, not months.

Software Defined Data Tiering and Compression:
Place data on the best performing Flash tier for most frequently accessed data. Compress cold data on conventional high-capacity spinning disks. Maximize the performance/cost advantages of every storage medium (PCIe Flash, SSDs, HDDs).

Just-in-time turnkey infrastructure:
Capacity augmentation reduces unnecessary up-front planning and capacity waste, enabling a more predictable and manageable sizing approach. Now you can dynamically grow or shrink with a single click with zero downtime to end-users or application availability.