Key Features

- A complete turnkey solution bundled with supporting software stack
- Supports multiple file systems and data sources including HDFS, GPFS and any file system with HDFS plug-in
- Self-monitoring and troubleshooting capabilities
- Support for up to 300 job trackers across a shared pool of resources
- Application and job level predictability via sophisticated job scheduler to guarantee SLA requirements
- Support for homogeneous and heterogeneous network environments
- Future proof solution equipped with 10G Ethernet out of the box
- Support for a wide variety of operating systems and for cloud deployment - Hypervisor support coming out soon

Simplicity to the Core

Hadoop is an ensemble where each piece including Sqoop, Flume, Pig and Hive etc. are integral pieces of the puzzle. Being tuned to your needs, PHAT-Data is simple to deploy with a click of a button. The entire provisioning process is automated and customized based on your requirements.

Performance Tuning Capability

Big Data solutions using Hadoop consist of an ensemble of tools which require precise tuning and configuration for optimum efficiency and maximum performance. For example, a Hadoop based database utilizes HBase - a real-time database which can use compression to minimize network overhead – which requires specific tuning to extract the full potential. AMAX understands your application and tunes each turnkey cluster so that guesswork is taken out of the equation.

Open Architecture

As a part of our commitment to an open architecture design philosophy, AMAX PHAT-Data does not include any proprietary components. The entire solution offers 100% compatibility with Java based MapReduce applications and support heterogeneous environments and open standards right out of the box.

High Availability

The Hadoop NameNode can become single point of failure if the cluster is not configured for high availability. The AMAX PHAT-Data includes 2 name nodes configured with DRBD/Heartbeat or RSF-1 in master-slave configuration with a dedicated heartbeat network to address this issue. The auxiliary NameNode automatically takes over in case of a failure. Our solution promises no downtime and absence of a single point of failure. The MapReduce engine offers automatic job recovery without the need to restart in case of JobTracker/TaskTracker failure.

Ready for Production Environment

Each and every component selected in the PHAT-Data cluster solution passes through AMAX's stringent 5 stage quality test which includes individual component tests as well as a complete system integration test. Our ISO 9001 certified manufacturing facility and 3 decades of engineering expertise promises utmost care at every level irrespective of the volume.
System Configuration

Standard configuration includes:

- 2x Master Nodes (NameNode/JobTracker)
- 18x Slave Nodes (DataNode/TaskTracker)
- 1x 10GbE Switch
- 1x 1GbE Switch

Master Node:

- 2x Intel® Xeon® 5600 or Xeon® E5 processor series
- 24GB ECC Registered DDR3 RAM (Up to 192GB)
- 6x 3.5" 7.2K/15K 2TB RPM SATA/SAS drives (Up to 12 drives)
- 2x Intel 82575 Gigabit Ethernet
- 1x Emulex OneConnect OCe 11102-NX Dual port 10GbE Converged Network Adapter
- 1x Intel RMM3/4 with complete IPMI 2.0 support

Slave Node:

- 2x Intel® Xeon® 5600 or Xeon® E5 processor series
- 48GB ECC Registered DDR3 RAM (Up to 192GB)
- 6x 3.5" 7.2K/15K RPM 2TB SATA/SAS drives (Up to 12 drives)
- 2x Intel 82575 Gigabit Ethernet
- 1x Emulex OneConnect OCe 11102-NX Dual port 10GbE network adapter
- 1x Intel RMM3/4 with complete IPMI 2.0 support

Network Fabric

- 1x 10GbE 24/48 10GbE Layer 2/3 switch with cut-through mode support
- 1x 100M/1GbE switch dedicated for management

Software Components

- Apache Hadoop – Open source software framework for scalable, reliable distributed computing.
  - Hadoop Common
  - Hadoop Distributed File System (HDFS)
  - Hadoop MapReduce
- Apache HBase – Database for real-time, random access to data stored using Hadoop.
- Apache Hive – Data warehouse using Hadoop for data summarization, ad-hoc queries and analysis.
- Apache Pig – High level language for expressing data analysis.
- Apache Sqoop – SQL to RDBMS integration tool.
- Apache Flume – Distributed, reliable service to aggregate log data.
- Apache Whirr – A set of libraries to run cloud services.
- Apache Zookeeper – Coordination service for all components of Hadoop.
- Oozie – Workflow engine to manage Hadoop jobs.
- Hue – Framework to create interactive web applications.

Supported Operating Systems

- Red Hat Enterprise Linux 5 and 6
- Suse Enterprise Linux Server 10 and 11
- CentOS 5 and 6
- Ubuntu Server 10, 11
- Microsoft Windows Server 2008 R2