

System Management

Enterprise-Grade Reliability, Availability and Scalability with Cluster Economics

PHEMI Central uses well-established and industry-leading big data technologies to reliably store the curated digital assets at scale. PHEMI Central leverages this base “operating system” capability to build powerful features that index, protect, and transform data for business use.

Beyond Hadoop (Or Standards Plus+)

PHEMI draws on the extensive Hadoop ecosystem to leverage innovative features, integrating and leveraging key capabilities.

- **Hadoop Distributed File System (HDFS)** provides linear scale and reliable data storage across large cluster of low-cost commodity servers.
- **Accumulo** is a distributed database on top of the HDFS distributed file system. Developed by the NSA, Accumulo provides high-performance storage and retrieval with fine-grained privacy access controls.
- **Ambari** is an open framework to provision, manage and monitor Apache Hadoop clusters.
- **Hive** delivers interactive and batch SQL query capabilities into PHEMI Central in order to interoperate with analytics tools and pre-existing applications.
- **YARN** provides resource management and distributed computing for the PHEMI Central system.

However, PHEMI Central adds significantly more functionality. In contrast to conventional Hadoop-based systems, PHEMI Central is a fully integrated enterprise-grade system. Users don't need to worry about digging deep into the world of Hadoop, MapReduce, YARN, Pig, HIVE, Sqoop, HBase, Zookeeper, Accumulo, and so on. At the same time, at no time can users, applications or external systems bypass the PHEMI policy enforcement engine to access data directly.

graphic or illustration
TBD

Commodity Hardware

PHEMI Central eliminates the cost and performance bottlenecks associated with expensive Storage Area Network (SAN) or Network Attached Storage (NAS) architectures. The system uses low cost, commodity hardware components and direct attached disk drives to significantly lower the cost of ownership compared to traditional enterprise data warehouse systems.

Scalability and Performance

PHEMI Central can easily aggregate structured and unstructured data, scaling storage and compute linearly from terabytes to petabytes with each additional hard drive and node.

Reliability and Availability

All data stored in PHEMI Central is replicated three times across the system to ensure high availability and resiliency in the event of a hardware failure. Direct attached hard drives can be hot swapped without impacting performance or data

availability while larger or faster drives and nodes can be absorbed into the system and load balanced automatically.

LDAP integration

PHEMI Central can operate with your existing LDAP or Active Directory identity management systems or via internally managed PHEMI Central user accounts.

Maintenance and Support

PHEMI Central provides clear visibility into overall system health, diagnostics, troubleshooting, capacity and digital assets under management so that IT administrators can quickly configure and test all nodes. Additionally, with Apache Ambari, system management capabilities can be integrated with IT infrastructure monitoring tools such as Microsoft System Center, Teradata Viewpoint, Nagios and Ganglia.

SCREENSHOT TO GO WITH “SYSTEM MGMT”?

DO WE WANT TO INCLUDE THIS?

ANY COPY TO GO WITH?

