

PHEMI Central Big Data Warehouse

Innovative Features — Powerful, Integrated Solution

COLLECT



PHEMI Central can collect virtually any kind of data—structured (such as database records), semi-structured (such as Microsoft Excel or genomic files) or unstructured (such as images, machine-collected data, or documents). Data can be ingested and aggregated from multiple disparate sources. During collection, PHEMI Central tags each raw data item with user-defined metadata, then stores the tagged raw data in PHEMI's fast, powerful data store. The data store is key-value-based and schemaless, so data can be ingested without complex or brittle schema-mapping exercises.

CURATE



PHEMI Central's powerful data store automatically indexes and catalogs all ingested data, so that even the raw data becomes a digital asset. Next user-specifiable data processing functions cleanse, parse, and structure the data—transforming the raw data into analytics-ready digital assets. Transformations include sophisticated cataloging and indexing, graphs, and tags for quickly finding specific digital assets within petabytes of data, plus privacy and sensitivity tags per individual cell of data. Audit, version control, and rollback features deliver full governance of findable, manageable digital assets ready for queries by users and consumption by applications and analytics tools.

CONSUME



PHEMI Central allows you to build datasets on demand without having to manage multiple data marts or complex MapReduce or YARN processes. Whether the dataset is consumed by a user or exported to an application or analytics tool, PHEMI Central strictly enforces your organization's privacy and security policies to ensure appropriate access to data.

Privacy, Security, and Governance



PHEMI Central manages crucial aspects of privacy and security such as data encryption, audit logs, and data validation. Privacy and security are designed in from the ground up to be able to accurately reflect your organization's governance policies.

Data Management



PHEMI Central's data management features allow you to manage your organization's data according to the privacy and access policies that govern digital assets being managed within PHEMI Central.

System Management



PHEMI Central runs on commercial servers and commodity disk drives, driving down hardware costs and allowing the system to scale to petabytes without SAN/NAS costs or performance bottlenecks. Automatic replication and load balancing means data is always available and performance is optimized across system nodes.