

PHEMI Central Big Data Warehouse

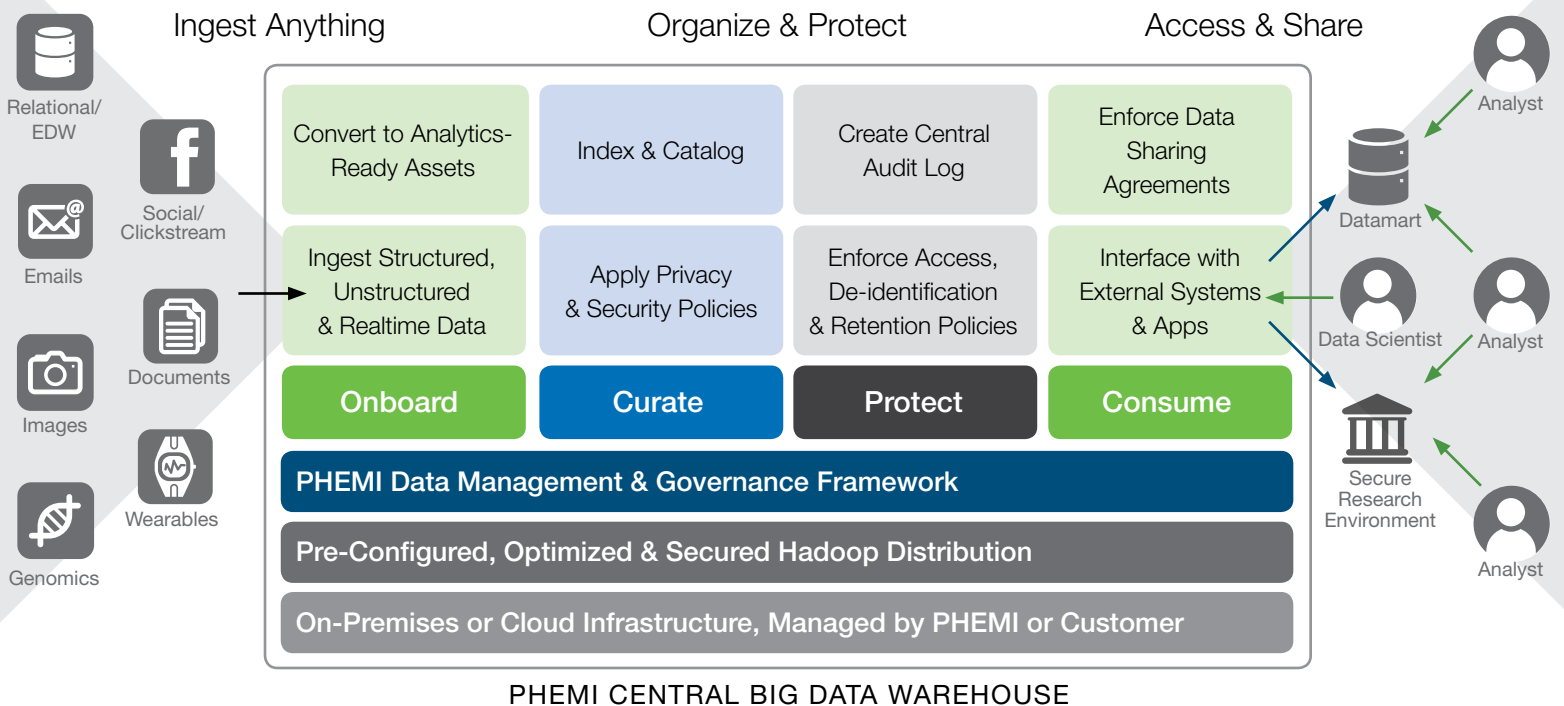
PHEMI Central™ is a Hadoop-based big data warehouse for structured, unstructured, and real-time data with built-in data governance, privacy, and security. PHEMI Central consolidates data from diverse sources, annotates it with relevant metadata, organizes it, and makes it available for further processing, BI, and advanced analytics.

Increase Analytics ROI, Decrease Risk & TCO

Over half of all big data implementations fail to achieve expected return on investment due to lack of data governance, integration complexity, and shortage of skills. PHEMI helps improve your time-to-value and analytics ROI by complementing Hadoop with features to simplify data ingestion, curation, governance, and accessibility. A PHI-ready cluster can be deployed on premise in less than two weeks and you can be delivering insights in less than a month. With the Data Science Tool Kit and Integration APIs, PHEMI Central is ready to be integrated into your enterprise data pipeline and identity management infrastructure.

Improve Governance, Privacy, and Accessibility

PHEMI Central's flexible and easy-to-use data governance framework helps resolve the tensions between data accessibility and privacy protection which have typically restricted the value enterprises can extract from their data. PHEMI makes Hadoop data easily discoverable and consumable, while ensuring that privacy, access controls, and life cycle management policies are enforced and audited at an individual data element level. Data analytic workloads run in-situ on the PHEMI Central cluster, without transferring large volumes of data to external systems, thus protecting privacy and delivering faster performance. PHEMI Central can also be used for publishing policy-compliant data sets suitable for data sharing or downstream analytics pipelines. The system automatically enforces data retention policies and data sharing agreements. Robust version control and rollback capabilities built on the notion of immutable data mean that data is never lost, corrupted, or overwritten.



PHEMI CENTRAL BIG DATA WAREHOUSE

Reduce big data complexity.

PHEMI Central is simple to setup and operate without previous Hadoop expertise, thanks to the capabilities formerly only available in traditional data warehouses. From data ingestion and curation, to governance and analytics, PHEMI is simple to use. The system enforces segregation of data administration, curation, and policy management duties. Attribute-Based Access Control (ABAC) features and a powerful privacy framework simplify policy creation, administration, and compliance auditing. The Data Processing Framework provides a fast and flexible parallel ingestion pipeline to transform data from complex multi-structured data sources into annotated, indexed data assets that are ready for analysis, reporting, and collaboration.

Data scientists can use the preconfigured Spark-based data science environment with a simple interface to R or SAS, while business analysts can use their favorite BI tools such as Tableau, Qlik, or PowerBI for working with published data sets.

Flexible Deployment

PHEMI Central is compatible with Cloudera and Hortonworks Hadoop distributions and available as on-premise software, a cloud service (managed or self-hosted), or preinstalled on an Oracle Big Data Appliance, depending on your IT preferences.

All our deployment methods align with appropriate privacy and security requirements, including Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) Act, as well as Canadian federal and provincial legislation.

Ready for Enterprise Integration

PHEMI Central is designed to integrate with your existing Active Directory/LDAP infrastructure and meets enterprise security, availability, privacy, and access control requirements. PHEMI features data encryption at rest and in transit, Kerberized services, and a central audit log of all data access events regardless of the interface used to access data.

PHEMI can ingest data from HDFS and Kafka-compatible streaming interfaces, HTTP(s) file uploads, REST API, or via ODBC/JDBC-compatible drivers. Export destinations include Hive, REST API, JDBC-compatible destinations, or CSV files.

PHEMI Central's Unique Capabilities

Data Processing Framework

Distributed Data Processing Functions (DPFs) are computing modules that run on a cluster to parse, recognize, extract, cleanse, normalize, encrypt, de-identify, or otherwise transform incoming data into analytics-ready data assets. All transformations are nondestructive and create a new version of the transformed data element, preserving complete transformation history.

Metadata Framework

PHEMI Central stores raw and processed data assets, along with extensible metadata in a high-performance distributed key-value store. The system annotates each data element with a number of metadata attributes, such as provenance, version, data classification, access control, privacy, consent, retention, data sharing, value description, ranges, and business context. PHEMI Central automatically indexes and catalogs all data for sub-second lookups, regardless of data volume.

Privacy and Data Governance

PHEMI Central's attribute-based Data Governance operates on an individual data element level. The system enforces compliance of access controls, de-identification rules, data retention, and data sharing policies based on the metadata attributes.

Ease your entry into big data and grow your solution at your own pace.

KEY BENEFITS

- Integrate data from silos
- Curate and catalog data, making it available for self-serve analytics
- Enforce privacy, security, and data sharing policies
- Reduce Hadoop implementation and integration time
- Use your preferred analytics tools, proprietary or third party
- Enable collaboration across internal and external stakeholders
- Eliminate performance bottlenecks
- Free up scarce IT resources
- Adapt to quickly-changing requirements
- Manage data retention, versioning, and compliance reporting
- Simplify HIPAA compliance

SPECIFICATION

On-Premise Deployment	Cloud Deployment
-----------------------	------------------

- | | |
|--|---|
| <p>4 Data Nodes (20 TB)</p> <ul style="list-style-type: none"> • 32 x 2.2GHz Core • 256 GB RAM • 64 TB Direct Attached Storage <p>3 Master Nodes</p> <ul style="list-style-type: none"> • 12 x 2.2 GHz Core • 192 GB RAM • 3 TB Direct Attached Storage <p>1 Management Node:</p> <ul style="list-style-type: none"> • 4 x 2.2 GHz Core • 64 GB RAM • 300 GB RAID1 Storage • 1 TB Direct Attached Storage <p>10 Gigabit Ethernet Network</p> | <ul style="list-style-type: none"> • Available as a managed service running on Amazon, Microsoft, Oracle, or ClearData HIPAA Compliant clouds. • Cloud service grows from 1 TB storage capacity. <p>All our deployments align with appropriate privacy and security requirements, including Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) Act as well as Canadian federal and provincial legislation.</p> |
|--|---|



Data Ingest Protocols	Data Export Protocols
-----------------------	-----------------------

- | | |
|--|---|
| <ul style="list-style-type: none"> • SFTP File Transfer • HTTP/HTTPS Upload • REST API • ODBC/JDBC | <ul style="list-style-type: none"> • Excel/CSV Download • REST API • ODBC/JDBC • Hive/Spark |
|--|---|

Analytics & Data Visualization	Data Processing Functions
--------------------------------	---------------------------

- | | |
|--|--|
| <p>Supports leading analytics tools, including:</p> <ul style="list-style-type: none"> • Tableau • Qlik • Power BI • R • SAS • SPSS | <ul style="list-style-type: none"> • Excel/CSV Reader • Genomics Reader (VCF/gVCF) • JSON Reader • XML/HL7 Reader • Custom DPFs |
|--|--|

For more information visit www.phemi.com/next-step

