PHEMI

# PHEMI Central™ Big Data Warehouse

## Product Description

# PHEMI Central Big Data Warehouse —Fully Integrated, Fully Adaptable

Designed to adapt to constantly evolving business demands, PHEMI Central unlocks data silos and transforms structured and unstructured data into analytics-ready digital assets for users to gain better insights, faster. With increased agility in data collection and increased power in data inventory and curation, organizations can conceive new business applications and rapidly build new solutions to support business objectives.

PHEMI

# Table of Contents



TO BE UPDATED & ACTUAL PAGE #S TBD

# PHEMI Central Goes Far Beyond the “Data Lake”

## Introducing a powerful, new way to manage your organization’s data — collect, curate and consume data at speed and scale

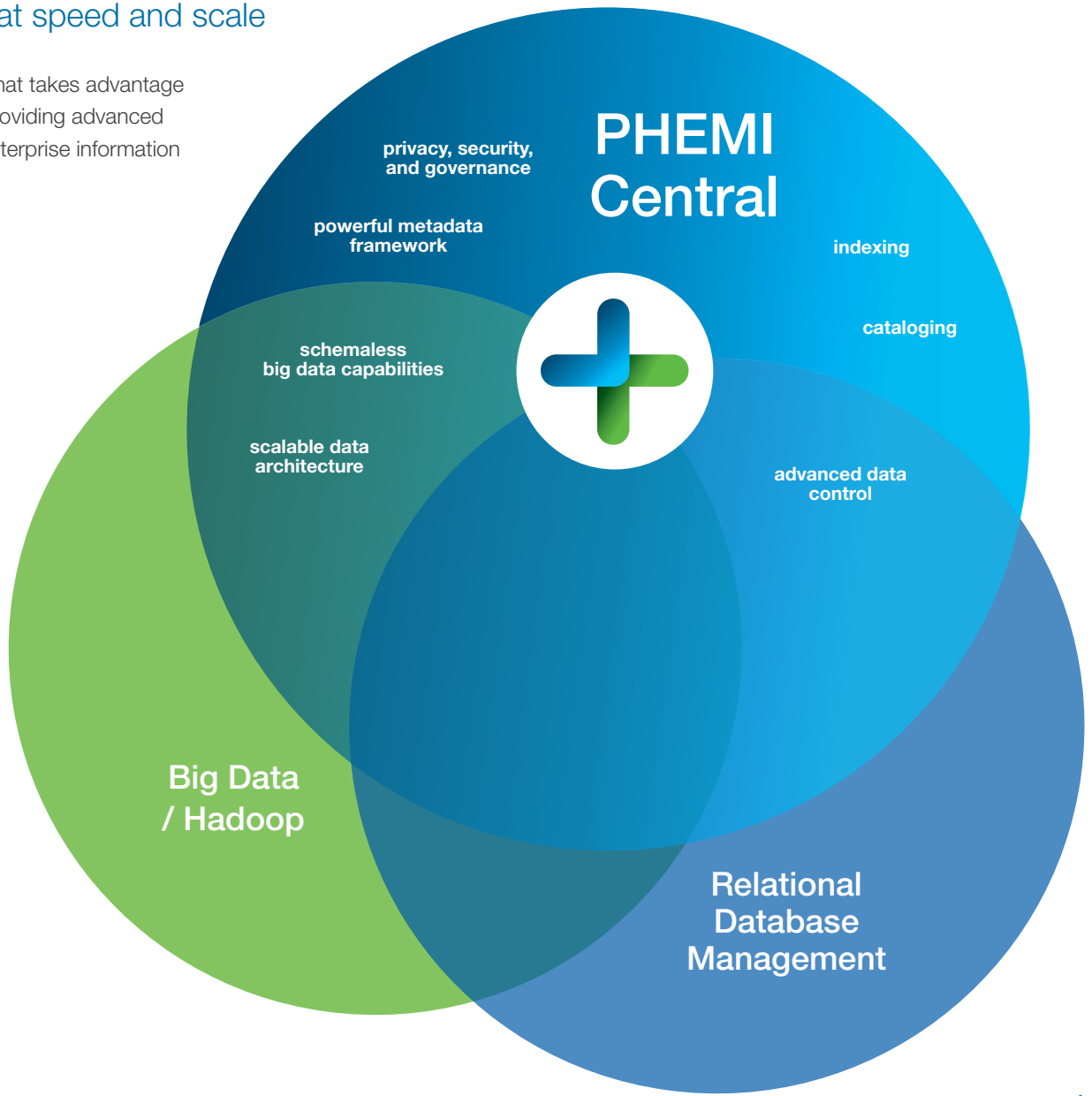
PHEMI Central is a fully integrated big data warehouse that takes advantage of the power, scalability and flexibility of Hadoop while providing advanced privacy, security, and governance along with powerful enterprise information management capabilities — all built right in.

### Beyond the “Data Lake”

Many big data approaches today just dump data in its native format into a repository, without any oversight or governance. The result is often an incomprehensible “data lake,” with data that’s almost impossible to find and retrieve. PHEMI Central’s powerful metadata framework automatically indexes and catalogs all of your digital assets, so that you can find them quickly and easily—making sense of the data lake even as it adds privacy, security, and governance to your data to ensure rightful access at speed and scale.

### PHEMI Central Has it All

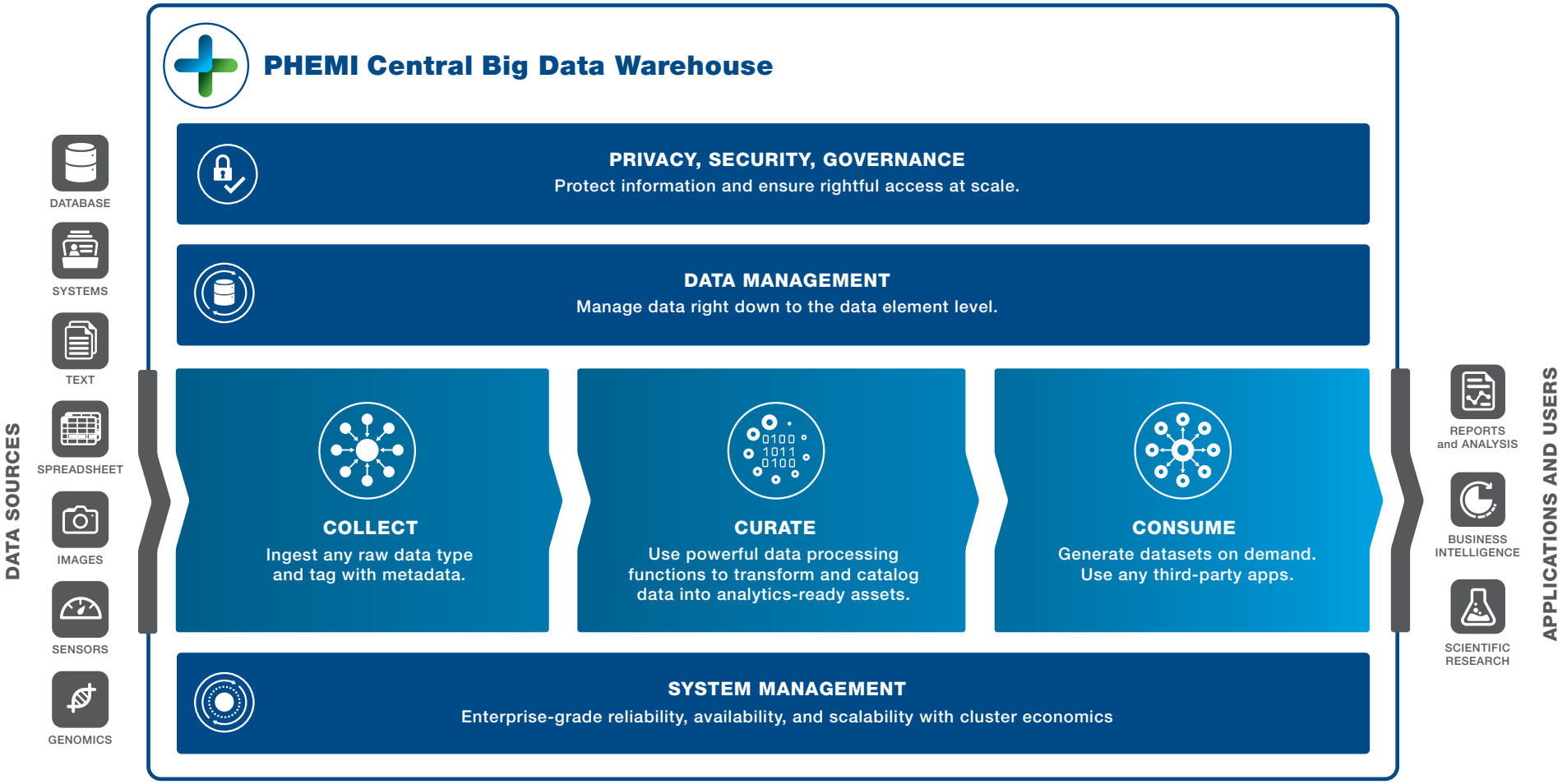
Other big data platforms may use Hadoop’s distributed file system capabilities. But PHEMI Central goes far beyond “plain old Hadoop,” providing a fully integrated solution that combines the Hadoop ecosystem with industry-pioneering approaches to privacy, security, and governance, along with sophisticated indexing, cataloging, and data management features. PHEMI takes the best of big data and traditional enterprise information management—then outstrips them both to offer a robust yet flexible solution for collecting, curating, and consuming sensitive data at scale.



# Introducing PHEMI Central Big Data Warehouse

## Collect, Curate, and Consume Your Data with Privacy, Security, and Governance

PHEMI Central is a new class of data warehouse that uses big data technologies to allow your organization to collect, curate, and consume of any volume and variety of data, while meeting, at scale, your organization's standards for data management, privacy, and governance.



# PHEMI Central Features Overview

## 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.

# PHEMI Central Big Data Warehouse —Collect, Curate, and Consume Data with Privacy, Security, and Governance

For the first time, organizations that need to protect and govern the use of their information can take advantage of big data technology to access, catalog, and analyze their digital assets at scale.

PHEMI

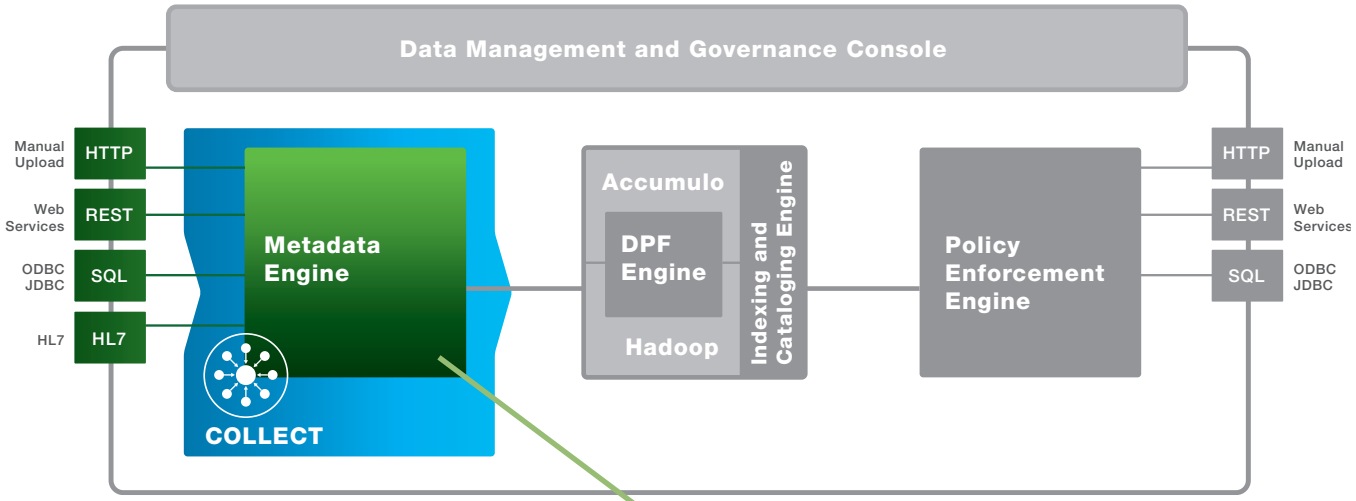
# Collecting Data

## Ingest and Manage Any Type of Data

### Data Sources

PHEMI Central can ingest virtually any kind of data, so organizations can consolidate their data silos. Data sources can include any data type from small kilobyte messages to large terabyte files:

- **Database Records** — Data extracted from information systems, databases, etc.
- **Structured Non-Relational Data** — Spreadsheets, GIS datasets, genomics, machine data, XML, JSON, HL7, PHEMI Clinical, PHEMI Forms Server, etc.
- **Semi-structured** — ECGs, tabular documents, etc.
- **Unstructured files & datasets** — Images, consult letters, reports, emails, customer feedback, social media, etc.




### Data Import

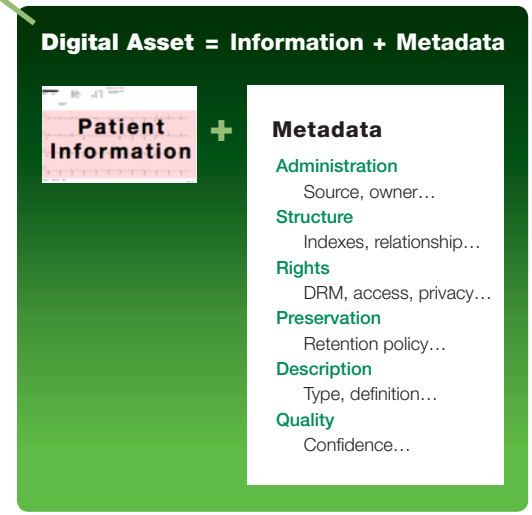
PHEMI Central can import data in a variety of ways:

- **Streaming** — Machine-to-machine data sources, such as telemetry and hospital bedside monitors, can stream data to PHEMI Central by means of the PHEMI REST interface.
- **Push** — Data sources can publish to PHEMI Central using either JDBC or the PHEMI REST interface.
- **Pull** — Custom connectors based on PHEMI's REST can be developed to allow PHEMI Central to fetch source data from sources.
- **Manual** — Files can be manually uploaded to PHEMI Central from a standard browser window.
- **Store by Reference and Action**

PHEMI Central can reference remote data or a remote dataset through a URL, stored procedure, SQL query or REST API. Store by Reference and Store by Action operations are ideal for collaboration projects between organizations or when accessing third-party datasets where data sharing agreements restrict data from being replicated.

### Metadata Tagging on Ingest

Immediately on ingest, PHEMI Central catalogs and describes raw data as part of cross-system data management. Describing and indexing data at the data element level but at petabyte scale provides users with unprecedented flexibility for accessing and using digital assets.

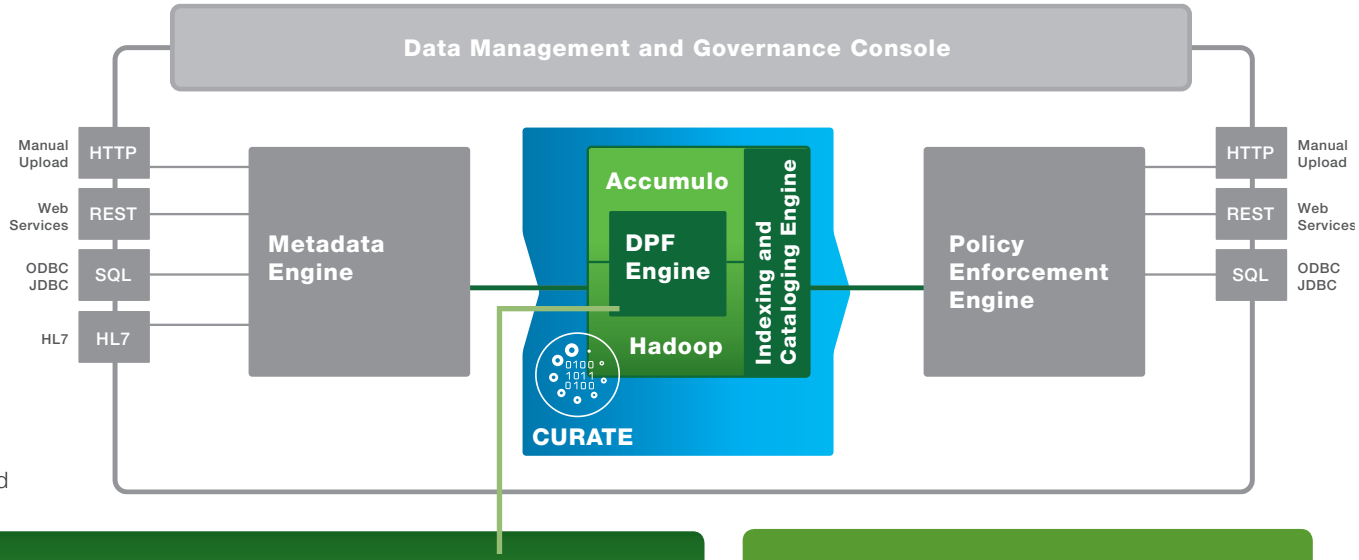


# Curating Data

## Convert Raw Data into Analytics-Ready Digital Assets

### Data Processing Functions

A data processing function (DPF) is an executable piece of code that transforms the original raw data (for example, a log message or medical report) into analytics-ready, governance-compliant digital assets specifically targeted at your organization's needs (such as a temperature reading or blood glucose measurement). The DPF is uploaded as a code archive into PHEMI Central using the Data Management and Governance Console. The code is executed by the PHEMI Central DPF Engine, which uses it to direct curation of the data.




The DPF supplies the instructions for parsing the raw data, extracting key content and performing data cleansing, ontology matching, enhanced indexing and cataloging, and structuring data according to the organization's needs. Standard PHEMI DPFs are included to index and describe structured data, such as database records or strongly typed XML/JSON, for consumption of the data through a REST or SQL interface. User-defined DPFs can also be developed for advanced needs, such as analysing semi-structured data or performing natural language processing on free text. Or, DPFs can catalog data and standardize it into ontologies such as SNOMED or LOINC, making it easier for data analysts to find the right information in the right format. DPFs can also analyze streams of machine data to find patterns and exceptions, calculating aggregates and converting the telemetry into an analytics-ready state for trending and predictive analysis. As the organization's needs evolve and as knowledge advances, DPFs can be updated and re-executed, to leverage the value of your historical data in new ways.

**The PHEMI standard DPF library includes:**

| Excel Reader | VCF Reader |
|---|---|
| Ingested Microsoft Excel spreadsheets and comma-separated value (CSV) files are converted into fine-grained analytics-ready digital assets, with each cell governed by the parent file's data sharing agreement. | Ingested genomic Variant Call Format (VCF) files are converted into a series of analytics-ready variants, with each variant governed by the parent file's data sharing agreement. |

**The powerful concept of DPFs is unique to PHEMI.** DPFs enable data scientists and programmers to write rich, customized transform functions in common programming languages (including Python, Java, and C++) using standard development tools. No specialized expertise in MapReduce or YARN is required. For parsing unstructured documents such as scans or X-rays, the DPF can include specialized parsing functions such as OCR or image parsing. Your DPF can be written by PHEMI, by your organization's in-house programmers, or by third- party developers.

As the organization evolves and as knowledge advances, DPFs can be updated and re-executed, to leverage the value of your historical data in new ways.

# Curating Data continued

### Indexing and Cataloging

Indexing and cataloging functions occur automatically in PHEMI Central's Indexing and Cataloging Engine, making it easier and faster to find and consume data. User-defined DPFs enable deeper and more sophisticated indexing and cataloguing, and second-order indexes and graph relationships allow data analysts to quickly find and build datasets across petabytes of heterogeneous digital assets. Linking datasets with common keys makes it faster and easier to build meaningful datasets across many sources. These powerful indexing features mean that data can be accessed in milliseconds, without having to wait for MapReduce or YARN jobs to complete.

### Schemaless Storage

PHEMI Central's data store is schemaless: both raw and curated data items are stored in a binary format that is unaffected by the source and destination schema. This approach means that organizations can quickly aggregate new data sources without costly redefinition of old schemas. Schemaless storage also permits the organization to extend uses or imagine new uses for data as knowledge advances and needs evolve, without concern for migrating rigid predefined schemas. Instead, PHEMI Central uses a flexible, powerful, distributed key-value store and sophisticated metadata tagging to manage, describe, and govern the data it stores. Curated digital assets derived from the raw data are linked to the original raw data, but PHEMI Central's SQL and REST interfaces abstract away from internal linkages and structures, so users and applications can focus on data use rather than data janitorial work.

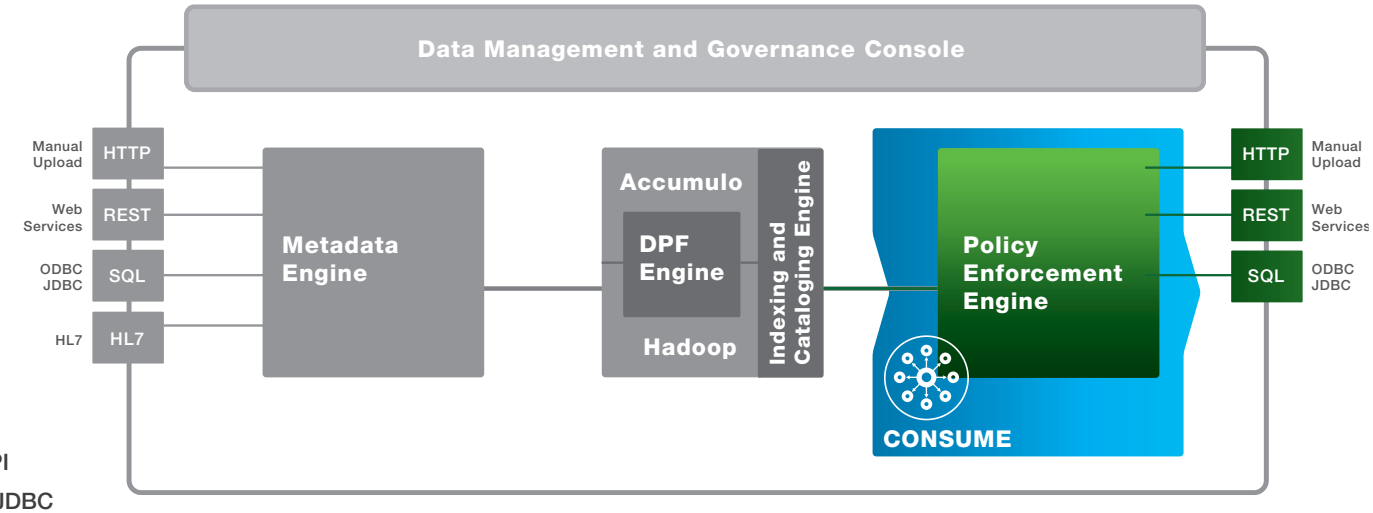
Powerful indexing features mean that data can be accessed in milliseconds without having to wait for MapReduce or YARN jobs to complete.

Schemaless storage permits the organization to extend uses or imagine new uses for data as knowledge advances and needs evolve, without concern for migrating rigid predefined schemas

# Consuming Data

## Build Innovative New Applications while Leveraging Your Existing Analytics Tools and Software

PHEMI Central integrates with your existing infrastructure, applications, and analytics tools to let you immediately take advantage of the big data warehouse. Multiple users can concurrently interact with the system, accessing datasets via SQL, data exports, and using custom applications, breaking down the costly data silos spread throughout your organization.




### Dataset Access

Datasets can be accessed in a variety of ways:

- **Export to a Microsoft Excel spreadsheet or Comma Separated Value (CSV) file**
- **Custom application via the PHEMI REST API**
- **Applications and analytics tools via ODBC/JDBC connectors**
- **SAP HANA via the SAP Smart Data Access connector**

### On Demand Datasets

Datasets are instantiated when information is needed, without having to predefine or navigate complex relational database schema. Data views are logical mappings that define a subset of the cataloged digital assets in the system across a specified collection of data. They can be created, altered, and discarded, and, since they are virtual constructs, they eliminate any need for data marts.

### Data Analysis

PHEMI Central supports standard tools, including R, SAP, SAS, SPSS, Stata, Tableau and more, to let organizations leverage their existing software.

### SQL Support

SQL remains the primary method for analyzing and accessing data in many logical data warehouse architectures. PHEMI Central supports batch and interactive SQL queries.

### Custom Applications

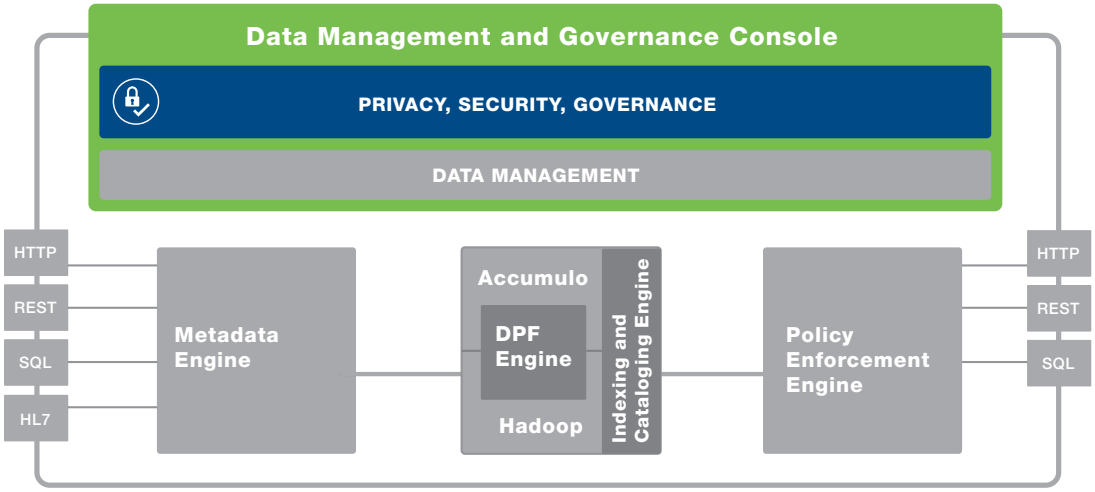
In-house developers, PHEMI developers, or third-party developers can develop custom applications using the PHEMI REST web services interface. The REST interface provides a rich set of statements that can access core system functionality, including flexible queries for data. REST queries, like all other queries are subject to the access controls enforcing information privacy, security and governance. Organizations can quickly trial innovative new applications, putting information in the hands of their users. This development agility means that applications that gain traction can be hardened and expanded and others retired.

PHEMI Central integrates with existing infrastructure, applications, and analytics tools to let you immediately take advantage of the big data warehouse.

# Privacy, Security, and Governance

## Protect Sensitive Information at Scale

Information governance is about controlling an organization's data. The data may be sensitive; or perhaps it is important that the data be absolutely accurate; or perhaps the organization must achieve legislative and compliance targets. Data governance includes the process and policies around the protection, curation, and access to data. Data governance encompasses all of privacy protection, data security, and data audit. PHEMI Central helps organizations achieve compliance objectives by providing an industry-pioneering set of capabilities to manage the privacy, security, and governance of data. These capabilities are fully configurable and are automatically enforced throughout the data lifecycle.




## Privacy is Built Right Into PHEMI Central Design

| Privacy by Design Principles | PHEMI Central Implementation |
|---|---|
| Proactive, not reactive | Metadata enables policies to define data access |
| Privacy as default setting | Assets are immutable. Policies required to access data |
| Privacy embedded in design | Metadata and computational access are the core of the system |
| Full functionality — positive sum, not zero sum | Data governance policies enable data use/analysis and do not create restrictions |
| End-to-end security — full lifecycle protection | Digital assets self-specify how they are managed and handled |
| Visibility and transparency — keep it open | Metadata and auditing provide accountability |
| Respect for user privacy — keep it user-centric | Data steward defines and sets policies on use |

**PHEMI Design Innovation**

**Data firewalls protect data internally, not just externally**

**Rely on automated operational policies, instead of manual processes**

**Proper management and control enables positive use of private data**

### Privacy by Design

PHEMI Central was built from the ground up on an **innovative Privacy by Design framework** to define, manage, and enforce data sharing agreements and privacy policies. Because PHEMI Central's privacy, security, and governance features are one coordinated design across the system, you don't have to rely on a cobbled-together mish-mash of security mechanisms to protect your organization's sensitive data.

# Privacy, Security, and Governance continued

## Build Your Access Policy Quickly and Easily

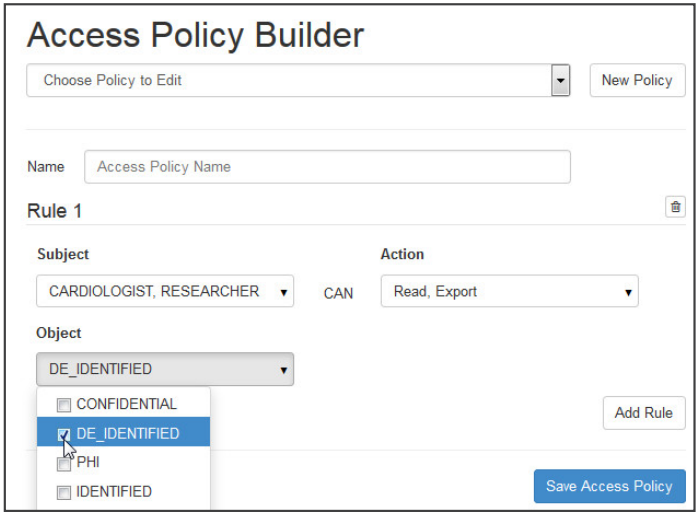
PHEMI Central tags sensitive data to identify its visibility, captures user authorizations, and combines them in simple, powerful access rules for attribute-based access control.

### Role Based Access Control

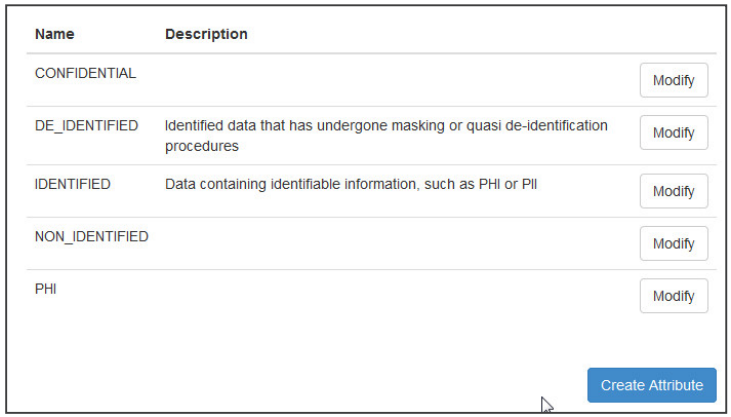
User roles determine what operations a user can perform. For example, only users with a role of administrator can configure the system, while only users with a role of data analyst can execute or export a dataset.

### Attribute Based Access Control

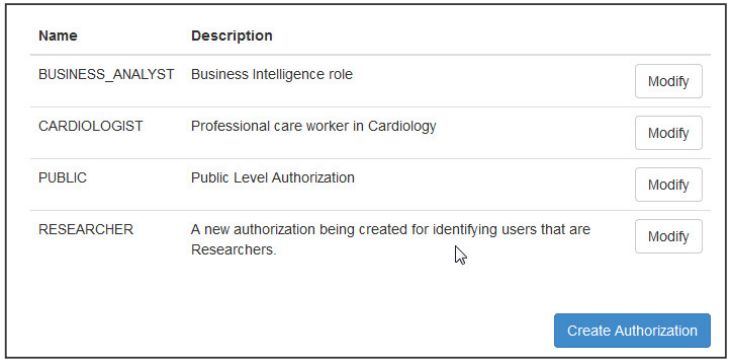
Users can be tagged with attributes that describe their level of authorization. Data can be tagged with attributes that describe its level of sensitivity or its requirements for privacy. Together, these two attributes can be combined to allow sophisticated access privileges to identified, unidentifiable, de-identified, or anonymized data.




Access Policy Attribute Table




Access Policy Authorization Table




### Audit Log

PHEMI Central maintains complete audit logs of system and user operations. They include all create/modify/delete operations, along with a record of all queries made to the system through the REST interface. These log files are completely tamperproof for all users. Approved users can filter log files and export the information for downstream analysis and compliance reporting.

### Encryption at Rest

For performance reasons, it is usually unnecessary to encrypt all data. Instead, encryption of only personally identifiable information is advised. PHEMI Central allows you to specify what data must be encrypted when at rest within the system.

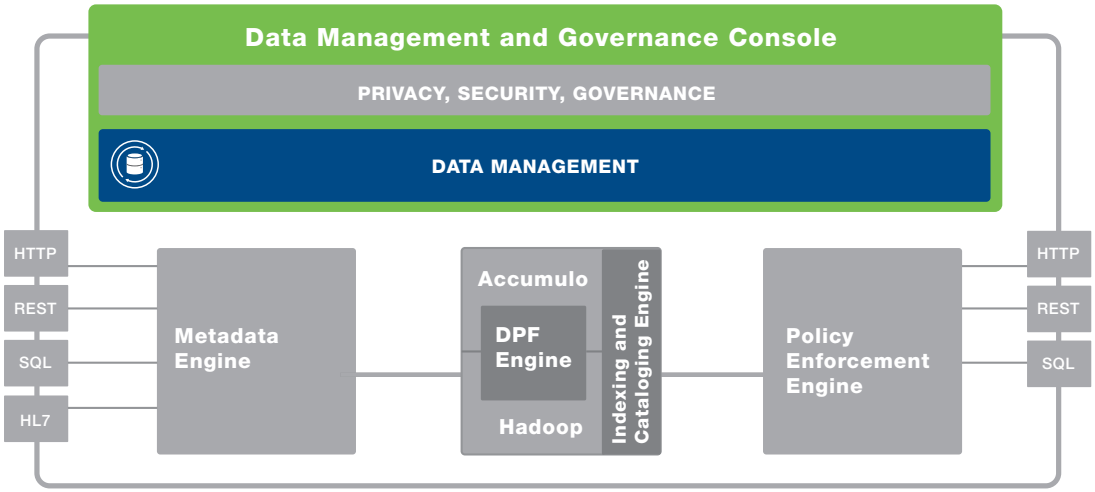
### Encryption in Motion

PHEMI Central can encrypt links from data sources and to consuming applications and analytics tools using either Secure Sockets Layer (SSL) or Transport Layer Security (TLS).

# Data Management

## Manage your data right at the data element level

Proper management of data is critical as volumes grow and variety increases. The PHEMI Central Data Management and Governance Console provides the primary interface for data management, including data management features for privacy, security, and governance.




### Powerful Metadata Framework

The power and sophistication of PHEMI Central's data management capability arises from its powerful metadata framework, which extends from end to end across the system. Metadata is applied on ingestion and enriched by cataloging, indexing, and invoking data processing functions. The result is data description at the element level embedding the rules and policies governing the element, as well as configured properties such as the data source ownership, retention time (time to live), and what visibility the element should have. For example, de-identification, encryption, and masking, along with other privacy restrictions can be enforced per data item, at the data element level.

In a traditional system, a user would have to plan a file-system hierarchy or a database schema. Data would be forced to comply with this rigid hierarchy, and everyone just has to hope that the design scales and that requirements do not change. When PHEMI Central's metadata framework is deployed with its scalable distributed key-value store (based on Accumulo), users no longer need worry about how to structure the system. PHEMI Central structures data automatically, on the fly. Data scales to large volumes at minimal cost while still providing fast access, and changes to requirements do not necessitate changes to design of the data store.

Users and integrated applications benefit from the metadata because they can use simple web-service calls based on the properties of the data, rather than having to navigate complex directories or schemas to find the data they seek.

### Governance Rule Enforcement

Most organizations have governance rules and data sharing agreements that stipulate how data may be used and shared. Governance rules are instantiated in PHEMI Central through configuration, using the PHEMI Central Data Management and Governance Console. PHEMI Central manages and enforces data sharing agreements by flagging the sensitivity of individual digital assets, tracking the retention period, recording rules around version control, and specifying de-identification, encryption, and data access permissions. Automating this function at scale across a variety of data sources and types is critical to managing privacy, security, and governance.

### Flexible, Attribute-Based Access Policies: Decoupling Data Protection from Data Use

PHEMI Central can automatically de-identify, encrypt, or mask personal information and enforce privacy based on sophisticated user access privileges and fine-grained sharing and consent rules. PHEMI Central stores the fully identified data but strictly controls the rightful use of all digital assets. When the user's access privileges and the recorded data sharing agreements dictate, PHEMI Central can invoke a data processing function to de-identify or anonymize any information information. Anonymization and de-identification may include disallowing access to personally identifiable information, masking certain information, redacting content, or may involve more sophisticated data dependency algorithms to reduce the risk of re-identification. Centralizing anonymization and de-identification helps reduce data sprawl and reduces the risk of data consistency errors.

# Data Management continued

### Lifecycle Management

PHEMI Central uses organization-specific retention rules to manage digital assets throughout their entire life cycle, from data creation through curation, usage, and end-of-life. Retention rules are captured in the Data Management and Governance Console, and the system calculates a time to time for every data element based on the retention policy and the time of ingestion. It also prevents users from deleting data during a configured retention period and automatically de-identify, delete or otherwise process information when the retention period expires.

### Data Immutability

PHEMI Central stores all data in a write-only data system that is never modified. Data may only be deleted at when its predetermined "time to live" expires, as specified by the organization's retention policy. This approach provides assurance of data integrity for audit and compliance requirements.

### Version Control

PHEMI Central has robust version control and rollback capabilities to ensure data is never lost, corrupted, or overwritten. The system keeps a history of data revisions and allows administrators to trace changes over time, including the ability to audit who made changes and when, and the ability to roll back changes if necessary. This design provides a complete history for audit and compliance requirements. For example, if your organization develops a DPF and discovers that it has been operating incorrectly, you can revert the data back to the original state without having to revert the DPF (which might impede development).

### DPF Framework

Just as the metadata framework manages the tagging of data items throughout the lifetime of data in the system, the PHEMI Central DPF framework manages DPF deployment and execution. The DPF framework is very simple and easy for programmers to learn and use: code libraries are uploaded into the system as simple ZIP files; PHEMI Central manages DPF execution across all datasets and data elements.

The power and sophistication of PHEMI Central's data management capability arises from its powerful metadata framework, which extends from end to end across the entire system.

# System Management

## Enterprise-Grade Reliability, Availability and Scalability with Cluster Economics

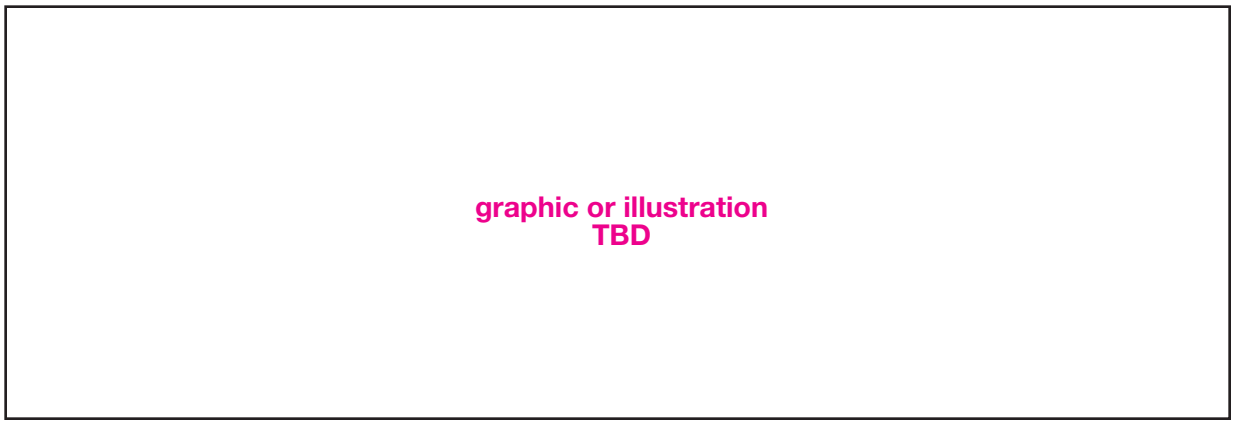
PHEMI Central uses well-established and industry-leading big data technologies to reliably store the curated digital assets at scale. PHEMI Central leverages this base "operating system" capability to build powerful features that index, protect, and transform data for business use.

### Beyond Plain Old Hadoop

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.




### 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 are 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.

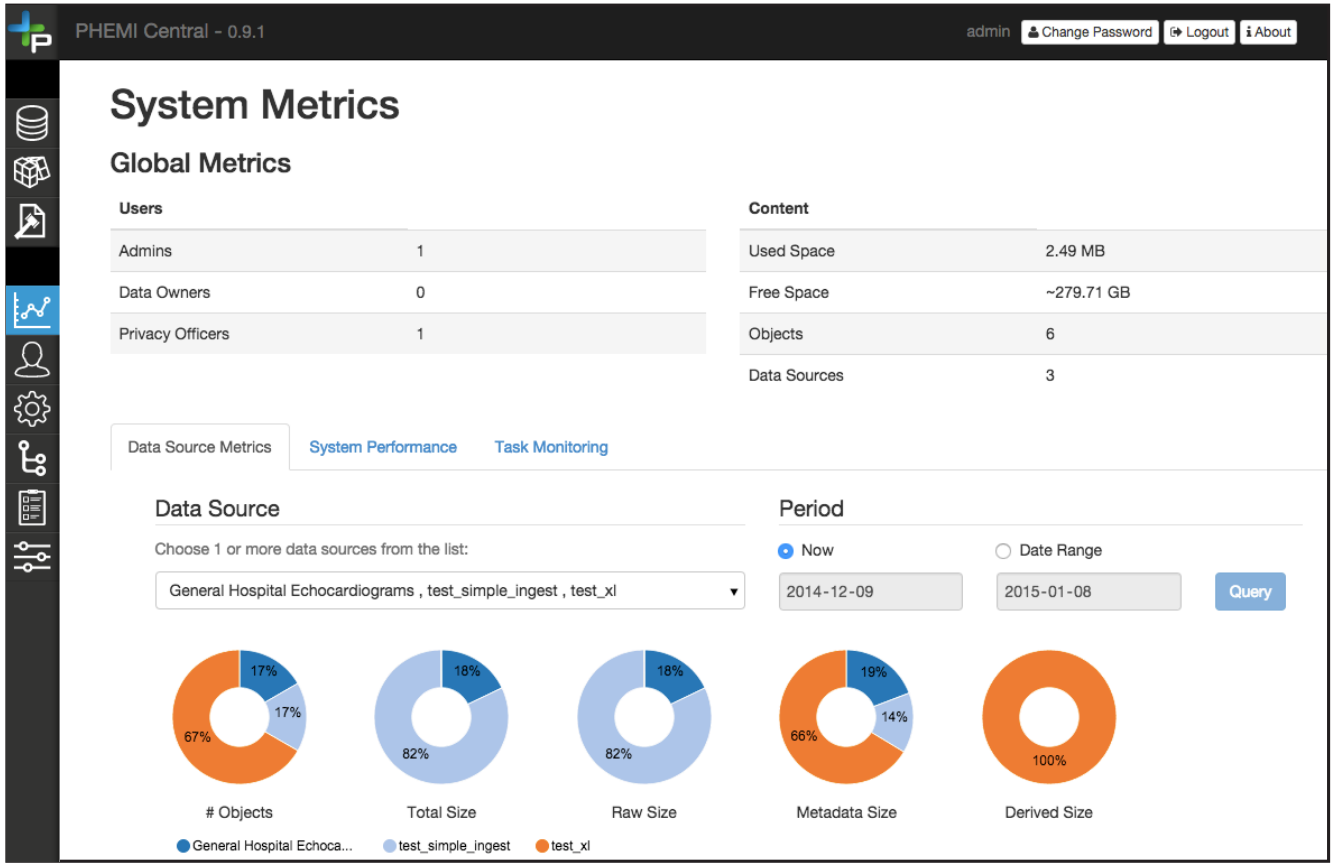
### Maintenance and Support

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?



# Deployment Options for PHEMI Central

PHEMI Central is available in either a cloud or on-premise environment and can typically be operational within 30 days.

## On-Premise

Run PHEMI Central at your site. Choose between a virtual machine or physical server deployment.

## In the Cloud

Run PHEMI Central in a cloud-based deployment, using industry-leading Amazon Web Services.

PHEMI Central uses a cluster architecture for low-cost scaling and built-in availability, redundancy, and load balancing. There's a management node to handle data management and governance, and cluster nodes for storing your digital assets and executing data processing and compute functions.

# Why PHEMI Central?

**Collect, Curate and Consume at Scale — PHEMI Central is perfect for a variety of organizations** <-- EXPANDED SUBHEAD TO INTRODUCE THE CONTENT BELOW; PLEASE EDIT AS REQUIRED

## PHEMI Central — Big data that's perfect for organizations not ready to let go of their current EDW Co-exist strategy

- **Aggregate New Data Sources Faster**
  PHEMI Central uses a schemaless architecture to quickly aggregate new data sources, lowering startup cost and complexity.
- **Integrate Any Data Type**
  PHEMI Central easily works with structured, semi-structured and unstructured data.
- **Lower Cost of Ownership by 60%**
  PHEMI Central uses low cost commodity hardware, eliminating expensive storage and server hardware.
- **Scale to Petabytes**
  PHEMI Central is built on big data technology - proven to scale to petabytes in production environments worldwide.
- **Enforce Privacy, Security and Governance Across Your Organization**
  Governance rules for privacy and security are enforced within the PHEMI Central big data warehouse, rather than at the application layer, ensuring data custodians control privacy and security, not application developers.

## PHEMI Central — Big Data that's perfect for organizations who need better than Plain Old Hadoop

- **Curate Digital Assets**
  PHEMI Data Processing Functions present a common Python, Java or C++ programming environment for data scientists abstracting away the complexities of working with MapReduce and YARN jobs.
- **Catalog All Digital Assets**
  PHEMI Central maintains a catalog of all digital assets in the system so that analysts can quickly build datasets across very large and varied datasets.
- **Enforce Fine-Grained Privacy, Security and Governance Across Your Entire Organization**
  Governance rules for privacy and security are enforced within the PHEMI Central big data warehouse, rather than at the application layer, ensuring data custodians control privacy and security, not application developers.
- **Focus on Mining Digital Assets**
  PHEMI Central is an enterprise-grade, fully integrated system so you don't have to worry about the confusing world of Apache Hadoop, MapReduce, YARN, Pig, HIVE, Sqoop, HBase or Accumulo. Instead, you can focus on unlocking your silos and discovering new insights with your digital assets.

DESIGN OF TABLE ABOVE TO BE REFINED

# Enterprise Information Applications

### Build New Applications

Consolidate data silos and create a warehouse of curated digital assets.

Quickly build innovative new and experimental applications for stakeholders. Those applications that gain traction may be hardened and expanded while others are retired.

Outsource application development without compromising privacy, security or governance.

### Retire Legacy Systems

Reduce database license costs by retiring and consolidating legacy systems.

Migrate data and point the retired application to the PHEMI Central Big Data Warehouse with commodity hardware economics.

Manage privacy and security of data.

Provide read-only access to data for archive and historic purposes.

### Machine Data Analysis

Collect billions of machine readable messages from telemetry and Internet of Things endpoints.

Convert messages into digital assets, build aggregates and monitor data for anomalies, trends and out-of-range readings.

Conduct predictive analysis, trending and ad hoc analysis via a third party analytics tool.

### Curate Documents

Aggregate Microsoft Word, Excel, PDF, image and text documents with commodity hardware economics, applying privacy, security and governance rules on all documents to control rightful access.

Perform Optical Character Recognition where necessary and index text in all files.

Allow users to search for key words, exposing documents based on their access privileges.

### Self Serve Data

Aggregate and curate an inventory of analytics-ready digital assets, applying privacy, security and governance rules.

Provide a portal for users to authenticate themselves and gain access to appropriate identified and de-identified data based on their access privileges.

Empower developers, data scientists and data analysts to explore and innovate.

STYLE TO BE DEVELOPED FOR TABLE above to provide better separation between columns and to improve type spacing, etc.

# Healthcare Applications

## Personalized Medicine

Aggregate genotype and phenotype data at scale with commodity hardware economics. Integrate mircoarray, Whole Genome Sequences, and microbiome data.

Add clinical data from the hospital EMR and claims data, supplementing it with datasets from patients, researchers and clinicians.

Apply and enforce data sharing agreements.

Convert all raw data into fine-grained analytics-ready data and perform a cluster analysis across petabytes of de-identified data in seconds.

Integrate advanced bioinformatics and visualization tools to navigate, annotate and discover insights.

Search for biomarkers to select the best drug and treatment for individuals based on their genetic profile.

STYLE TO BE DEVELOPED FOR TABLE above to provide better separation between columns and to improve type spacing, etc.

## Population Health

Investigate undiagnosed, untreated and at-risk patients by integrating information from various hospital information systems.

Convert medical reports into SNOMED, LOINC, RxNorm and ICD codes.

Link the data and write data processing functions to flag patients for screening.

## Quality & Outcomes

Identify and close gaps in care. Integrate information from various hospital information systems.

Apply and enforce data sharing agreements.

Convert medical reports into SNOMED, LOINC, RxNorm and ICD codes.

Link the data and analyze treatment plans and outcomes by provider.

Use your existing analytics tools to generate de-identified reports for each provider showing peer benchmarks compared to registry and other in-network providers.

Conduct business analytics on-demand to build a clearer view of financial and operational performance and measure the impact of operational changes in real time.

# Public Sector Applications

### Open Data

Aggregate data sources.

Apply and enforce data sharing agreements.

Build an inventory of digital assets.

Mark the sensitivity of data such as social insurance number (Public, Classified, Secret, Top Secret, etc).

Link data across various sources.

Allow self-service access to some data and accept dataset applications when users request more sensitive data, de-identifying data when appropriate.

### Program Review

Collect data from government information systems.

Apply and enforce data sharing agreements.

Write Data Processing Functions to convert Microsoft Excel, database and unstructured data into structured digital assets.

Use your existing analytics tools to conduct ad hoc analysis, drilling in to program effectiveness, identifying opportunities to improve program efficiency.

### Citizen Services

Build new applications to automate citizen services such as permitting and licensing.

Manage the privacy, security and governance of applicant data and de-identify information based on the reviewer identity.

Draw on information from various government information systems to supplement the application and apply rules as part of the process workflow to validate or approve the applications.

STYLE TO BE DEVELOPED FOR TABLE above to provide better separation between columns and to improve type spacing, etc.

# About PHEMI

**Call to Action, etc.**

**content TBD**

**PHEMI was founded in 2013 by a team of proven entrepreneurs and indus experts. Headquartered in Vancouver, Canada, the PHEMI team has exten experience bringing innovative technologies to enterprise-class customer Industry expertise, including healthcare and security, drives PHEMI Centra features, while networking and high performance computing technology expertise drive PHEMI architecture to meet the challenges of big data.**

**PHEMI Central gives organizations the agility to seamlessly collect data sources, catalog and curate a powerful inventory of secure digital assets, conceive new business applications and rapidly build new solutions to su strategic objectives.**

**PHEMI partners with best-in-class technology and service providers to de complete solution to meet any organization's needs.**

**Visit www.phemi.com for more information.**

www.phemi.com info@phemi.com twitter.com/PHEMIhealth linkedin.com/company/PHEMI



PHEMI