

PHEMI Engagement Framework

Five Steps to Sharing Data Responsibly

Many organizations have taken their first steps to deploy a big data system and are successfully taking advantage of Hadoop for its ability to economically store any file type at scale. Whether your organization is new to big data solutions or you're looking to take your implementation to the next level, PHEMI can help.

It's a privacy and security obligation that organizations protect their sensitive data. Yet at the same time, it's a competitive imperative that information is put in the hands of customers and decision-makers. PHEMI uses a five-step process to help organizations share data responsibly. This checklist shows a general PHEMI engagement. Of course, each project is unique, and we look forward to addressing your specific needs.

PHEMI Engagement Checklist

1. Scope and Define
2. Plan and Approve
3. Install and Secure
4. Load and Analyze
5. Handoff and Operationalize

Step 1: Project Scoping

The project is defined in business terms to confirm the investment and desired outcomes meet management priorities. This may be accomplished through a series of meetings or workshops.

Customer Resources: Business Sponsor, Technical Sponsor

Timeline: 2-8 weeks

- ☐ Business and technical sponsors are identified
- ☐ Clear business problem statement for initial use-case
- ☐ Inventory of data sources available to be used
- ☐ Description of data preparation steps
- ☐ Type of analysis to be performed
- ☐ High-level solution architecture, timeline and budget

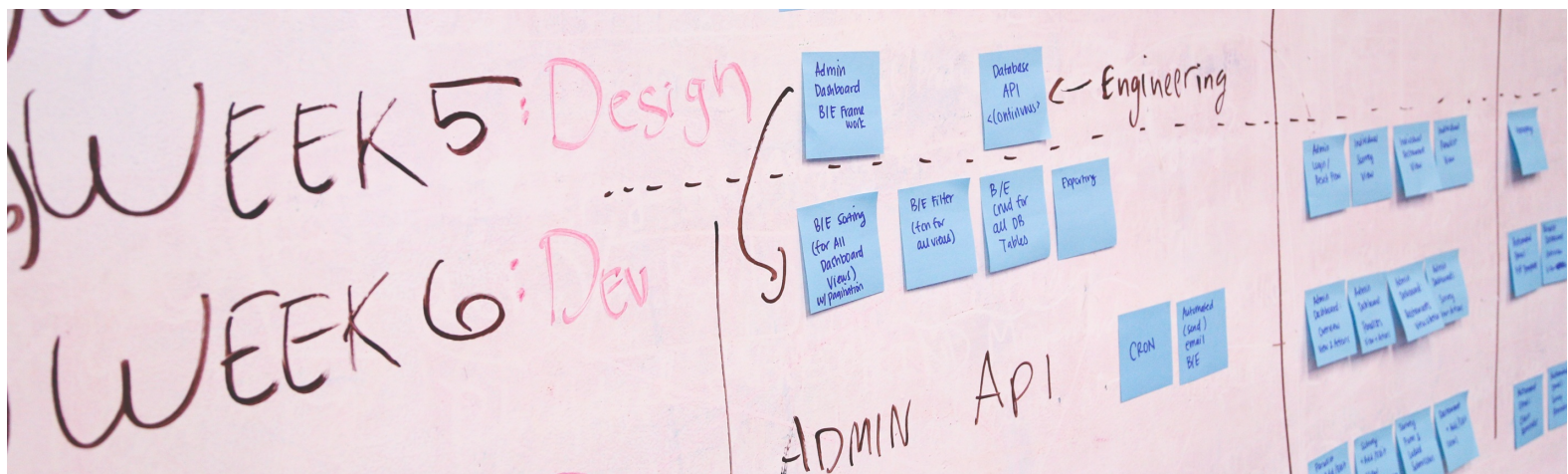
Step 2: Project Planning

The implementation plan is formalized and resource allocation is approved.

Customer Resources: Business Sponsor, Technical Sponsor, Enterprise Architect

Timeline: 4-12 weeks

- ☐ Audit existing hardware, software, systems infrastructure
- ☐ Review sample data extracts
- ☐ Review enterprise privacy, security and governance requirements
- ☐ Conduct Hadoop Privacy Assessment*
- ☐ Identify any partners and internal/external contributors
- ☐ Joint project team identified and RACI roles/responsibilities agreed
- ☐ Build project plan, budget and statement of work
- ☐ Documented Acceptance Test Plan
- ☐ Contracts in place



Step 3: Project Implementation: PII-Ready*

The PHEMI Central software is installed on the Hadoop cluster and the system is tested end-to-end to ensure it is functioning properly. Before sensitive data is loaded, a full penetration test is performed and the customer signs off the system as PII-ready.

Customer Resources: Project Sponsor, Project Manager, Enterprise Architect, Application Owner, Security Lead

Timeline: 2-6 weeks

Deliverables:

- ☐ Close any hardware, software or system gaps identified in the audit.
- ☐ Configure enterprise Active Directory and Security components
- ☐ Install PHEMI hardware or cloud resources
- ☐ Install PHEMI Central software
- ☐ Complete System Validation Testing
- ☐ Conduct Penetration Testing
- ☐ Complete part one of the Acceptance Test Plan
- ☐ Complete PII-Ready Sign-off from customer

Step 4: Project Implementation: Analytics-Ready*

Data is loaded into the system, processed and analyzed.

Customer Resources: Project Sponsor, Data Steward, Application Owner, Data Engineer/Analyst

Timeline: 2-8 weeks

- ☐ Configure users, collections, access policies and other system parameters
- ☐ Load sample data and build/validate data preparation workflow
- ☐ Validate pseudo-identification, de-identification algorithms and access policies
- ☐ Load data
- ☐ Perform analysis
- ☐ Conduct PHEMI Central training
- ☐ Complete Acceptance Test Plan

Step 5: System Handoff and Ongoing Service and Support*

The customer team is fully trained and ready to take over day-to-day operations.

Customer Resources: Project Sponsor, Application Owner, Data Engineer/Analyst

Timeline: 1 week

- ☐ Establish help desk accounts
- ☐ Measure outcomes
- ☐ Plan system expansion to add new data sources and analytics
- ☐ Ongoing technical support and maintenance upgrades.

For more information, please contact your PHEMI representative or visit www.phemi.com

1: Request contact(s) for discussing HDFS data	4
5: Generate checklist of libs/packages added to	
5: Request 5 temporary VM(s) to act as Master	
1.6.1: Request IP addresses and network nam	
1.6.2: Request DNS for temp mas and mgr vm	
1.6.3: Firewall changes for all master and mgn	
7: Document workflows	
3: Agreement of responsibilities for data migra	
3: Request additional equipment for Cloudera	
/KMS servers ready	
L0: Request backup location	
L1: Determine Repository for software installa	
L2: Check disk health on all nodes, replace unf	
L3: Active Directory and Kerberos changes	
1.13.3: Test access to Active Directory for crea	
1.13.1: Service principals created for temp ma	21
1.13.2: dsci user and tenant user	21
L5: Request VPN access for all new VMs.	
L4: Request admin/root access to cluster mach	
L6: Request a test plan / validation exercise fr	
L7: Access Cards for PHEMI employees travelin	
L8: Confirm a project manager	
L9: Determine any custom KMS ACL configurat	
20: Confirm and resolve any space issues in the	
21: Run PHEMI de-duplication and hard delete	25
22: Create golden image (ideally ready for dev	25,24

