PHEMI

PHEMI Central Big Data Warehouse

Product Guide

PHEMI Central — Lorem Ipsum for Secure Big Data

For the first time, organizations that need to protect sensitive information can take advantage of big data technology to access and analyze diverse digital assets. [Content TBD — this is an introductory page to whet the appetite of your reader before the TOC]

+PHEMI

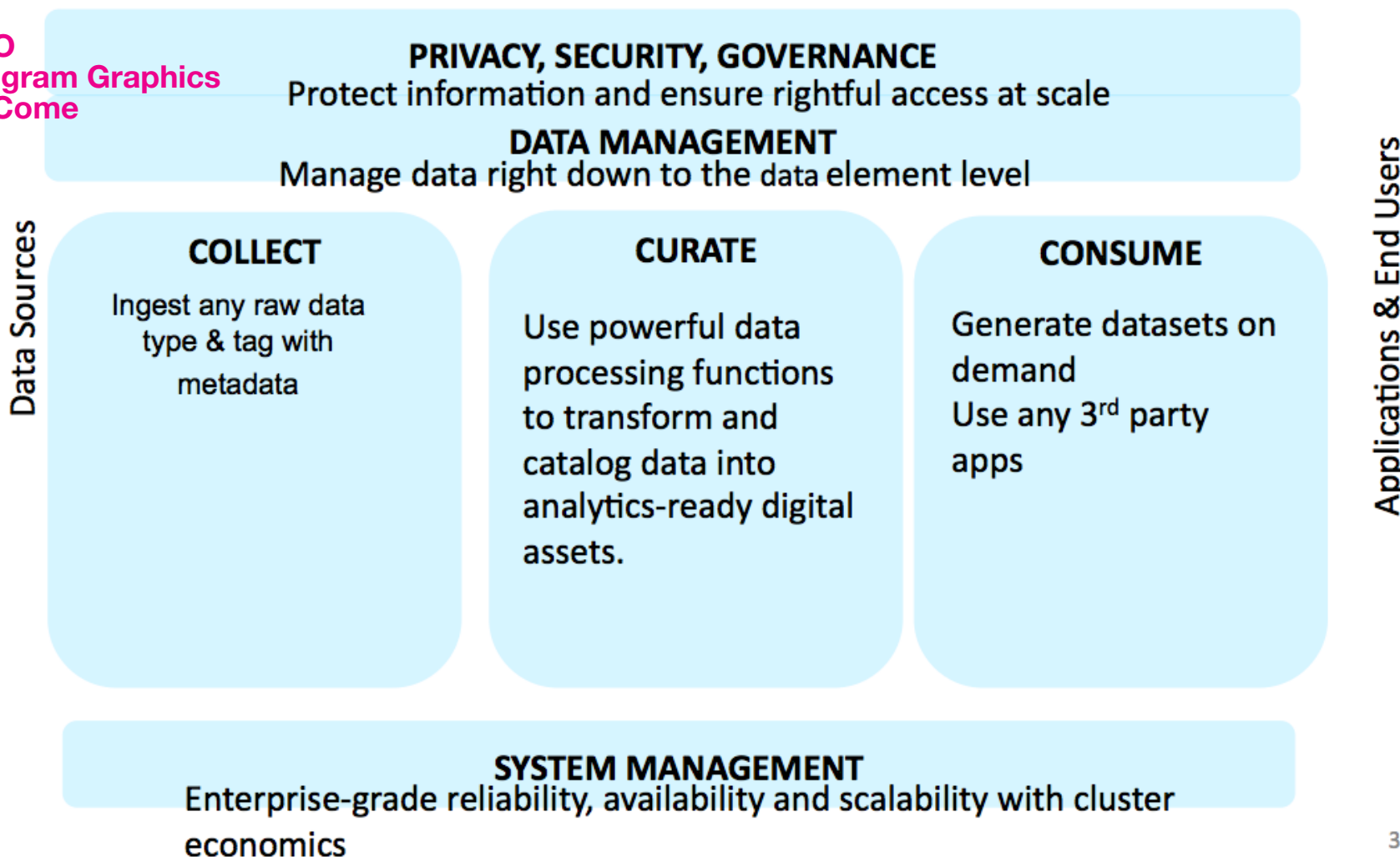# Table of Contents

The three applications won't necessarily appear in the TOC, particularly if "Applications" is only one page. (Note:Having them in the TOC gives them extra emphasis.)

PHEMI Central Big Data Warehouse

PHEMI Central is a new class of data warehouse that uses big data technologies to store and retrieve any volume of data with fine-grained data management features and an industry-pioneering approach to privacy, security and governance.

FPO Diagram Graphics to Come

Data Sources

PRIVACY, SECURITY, GOVERNANCE
Protect information and ensure rightful access at scale

DATA MANAGEMENT
Manage data right down to the data element level

COLLECT

Ingest any raw data type & tag with metadata

CURATE

Use powerful data processing functions to transform and catalog data into analytics-ready digital assets.

CONSUME

Generate datasets on demand
Use any 3rd party apps

Applications & End Users

SYSTEM MANAGEMENT
Enterprise-grade reliability, availability and scalability with cluster economics

3

PHEMI Central — Unlocking the Power of Your Data

PHEMI Central is a fully integrated system so you can turn your focus from collecting and storing data, to using it strategically to derive insights.

Designed to adapt to constantly evolving business demands, PHEMI Central unlocks your data silos and transforms structured and unstructured data into analytics-ready digital assets for users to gain better insights, faster. PHEMI Central gives organizations the agility to aggregate data sources, catalog and curate a powerful inventory of digital assets, conceive new business applications and rapidly build new solutions to support your strategic objectives.

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. But PHEMI Central goes further by providing sophisticated and unique privacy, security, governance and cataloging capabilities, so you can easily find and control your digital assets, making PHEMI Central ideal for organizations dealing with large-scale private and sensitive data.

PHEMI Central is a fully integrated system that leverages Apache Hadoop, YARN, Accumulo, and several other leading software projects to deliver an enterprise-grade system for storing, managing, and accessing data. Users and external applications interact with it through either the robust REST based web-services interface for transmitting data, or via SQL querying capability provided through Hive.

- **Aggregate data from a wide variety of data types: Excel, Word, PDF, text, image files, machine data without predefined schemas, databases and information systems**

- **Transform unstructured, structured, and machine data into analytics-ready digital assets**

- **Manage and protect the privacy, security, and governance of digital assets with a Privacy by Design* architecture**

- **Index and link data across multiple datasets to find exactly the information you need at petabyte scale.**

- **Implement and manage data sharing agreements and consent management**

- **Publish data to off-the-shelf or custom analytics and reporting tools, rapidly adapting to new applications and evolving organizational needs**

- **Scale to hundreds of petabytes at 60% lower cost of ownership than a traditional enterprise data warehouse As a reader I am still asking myself what this means.**

Graphics or Callouts TBD

PHEMI

Collecting Data

Aggregate any type of data and tag it…

**FPO
Diagram Graphics
to Come**
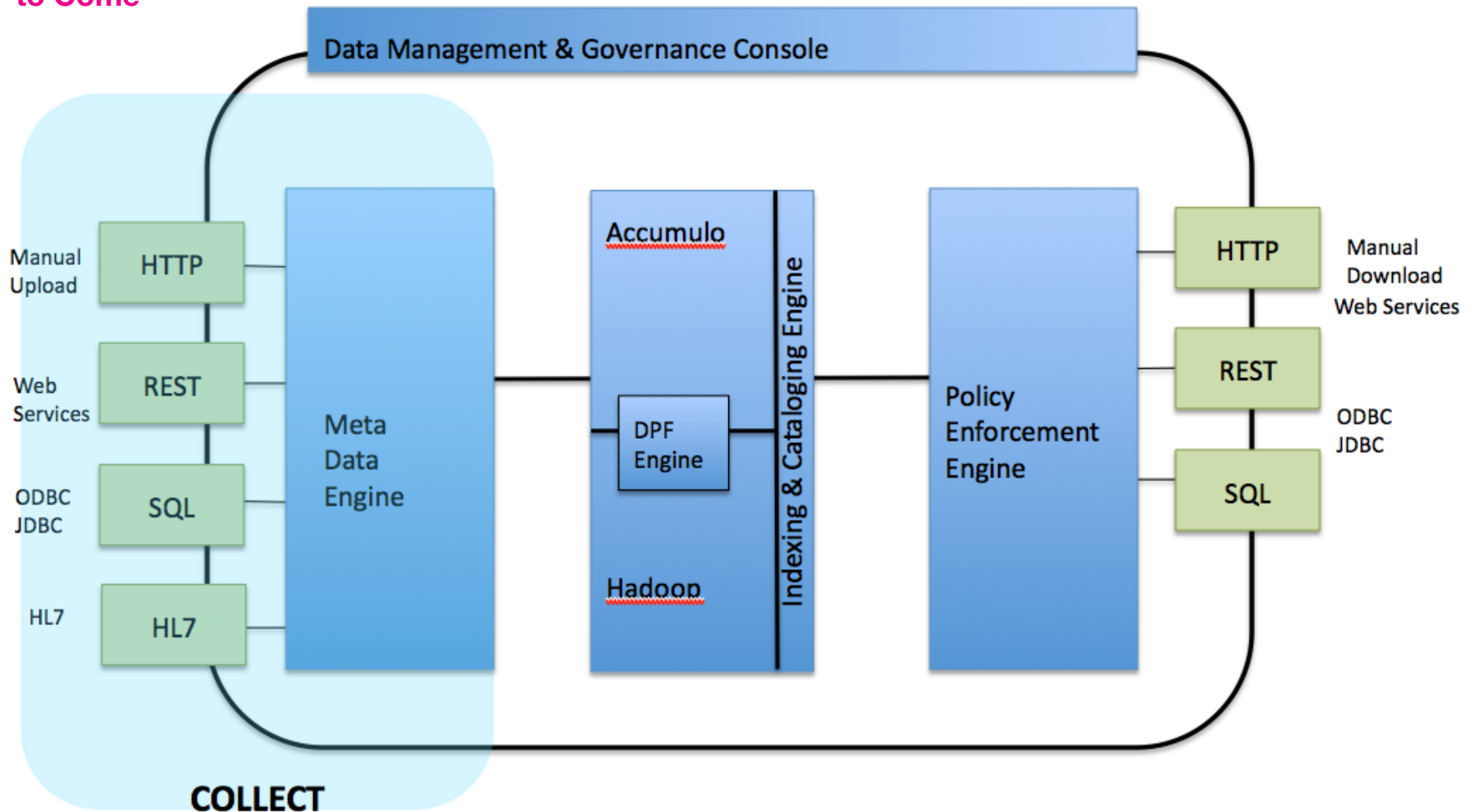
Goal of this image is to convey what pieces in the logical architecture fit into the function "COLLECT"...whether its this teal bucket to contain it, and grey everything else out...or whether its just this teal bucket and ditch the rest...diagram

PHEMI

More "Collecting" Content TBD; likely to include:

Paragraph on Data Sources
Paragraph on Data import options
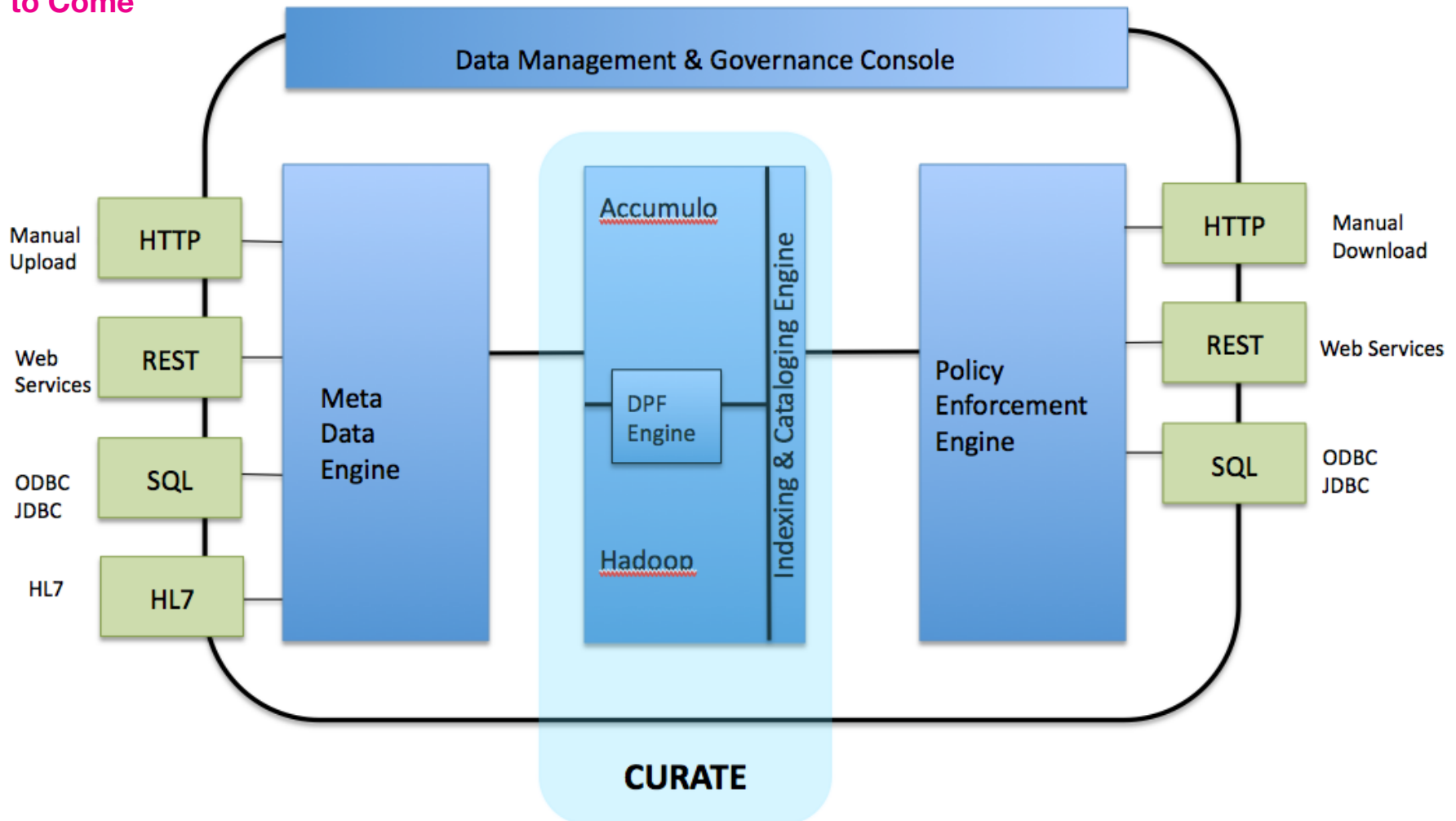**Introduction to Metadata (meta data engine)**
Introduction to Schemaless
One short definition of Digital Asset
…

PHEMI

Curating Data

Transform data into analytics-ready digital assets…

**FPO
Diagram Graphics
to Come**

Goal of this image is to convey what pieces in the logical architecture fit into the function "CURATE"…whether its this teal bucket to contain it, and grey everything else out…or whether its just this teal bucket and ditch the rest…diagram

PHEMI

More "Curating" Content TBD; likely to include:

2-3 Paragraphs on DPF's (including DPF engine & framework)

1 small chart of PHEMI Library of standard DPF's

Discussion on Storage

Introduce Hadoop platform including Accumulo, Ambari, Hive and how you don't need to know MapReduce

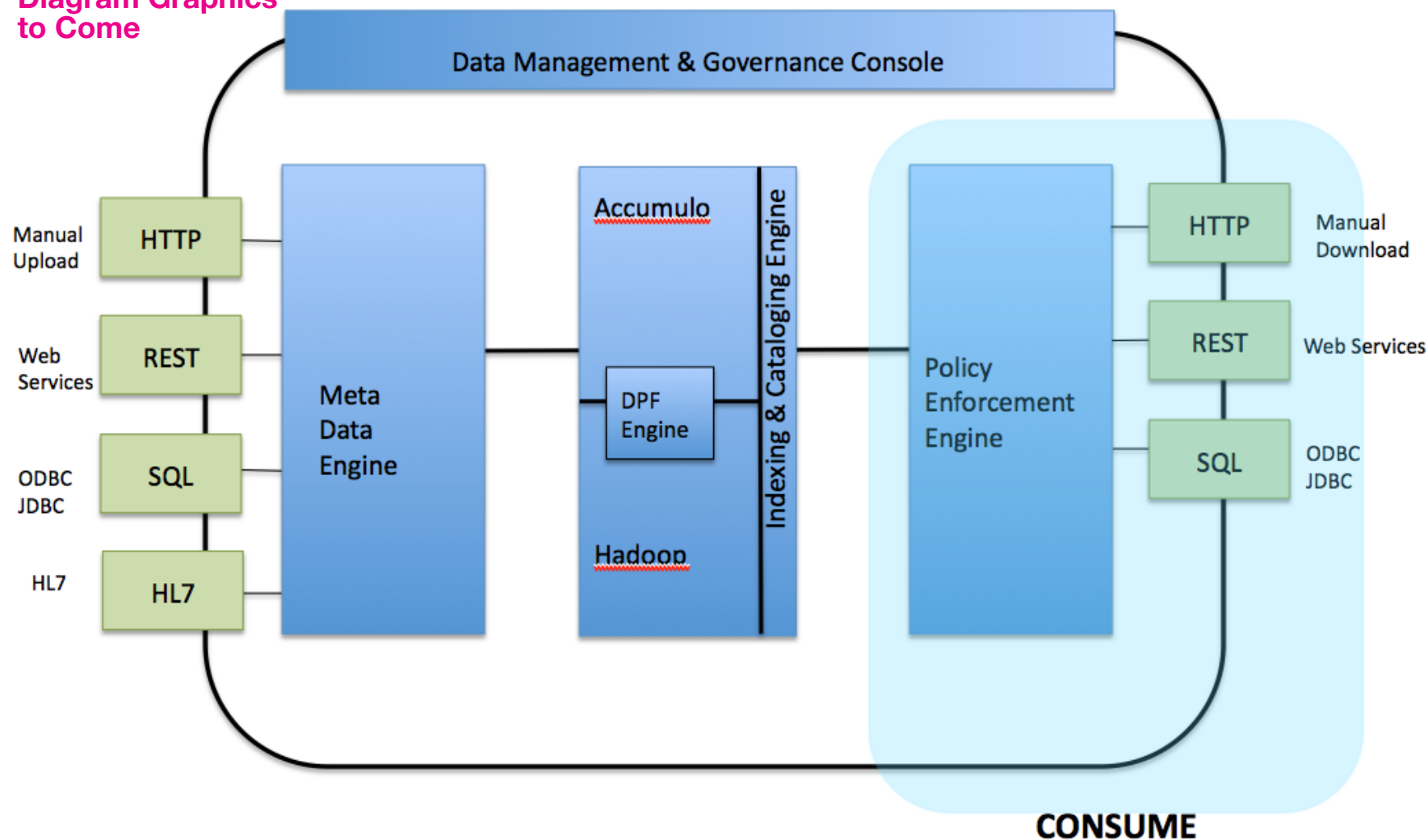**1 Paragraph on our view of traditional ETL tools (we are compatible)**

1 Paragraph on Cataloging (indexing & cataloging engine) then Linking

1 Paragraph on Store by Reference and Action

Consuming Data

Design Innovative New Applications While using Your Existing Analytics Tools & Software…

**FPO
Diagram Graphics
to Come**

Goal of this image is to convey what pieces in the logical architecture fit into the function "Consume"…whether its this teal bucket to contain it, and grey everything else out…or whether its just this teal bucket and ditch the rest…diagram

More "Consuming" Content TBD; likely to include:

Paragraph on how datasets can be accessed (excel, csv, REST…)

Paragraph on on-demand datasets

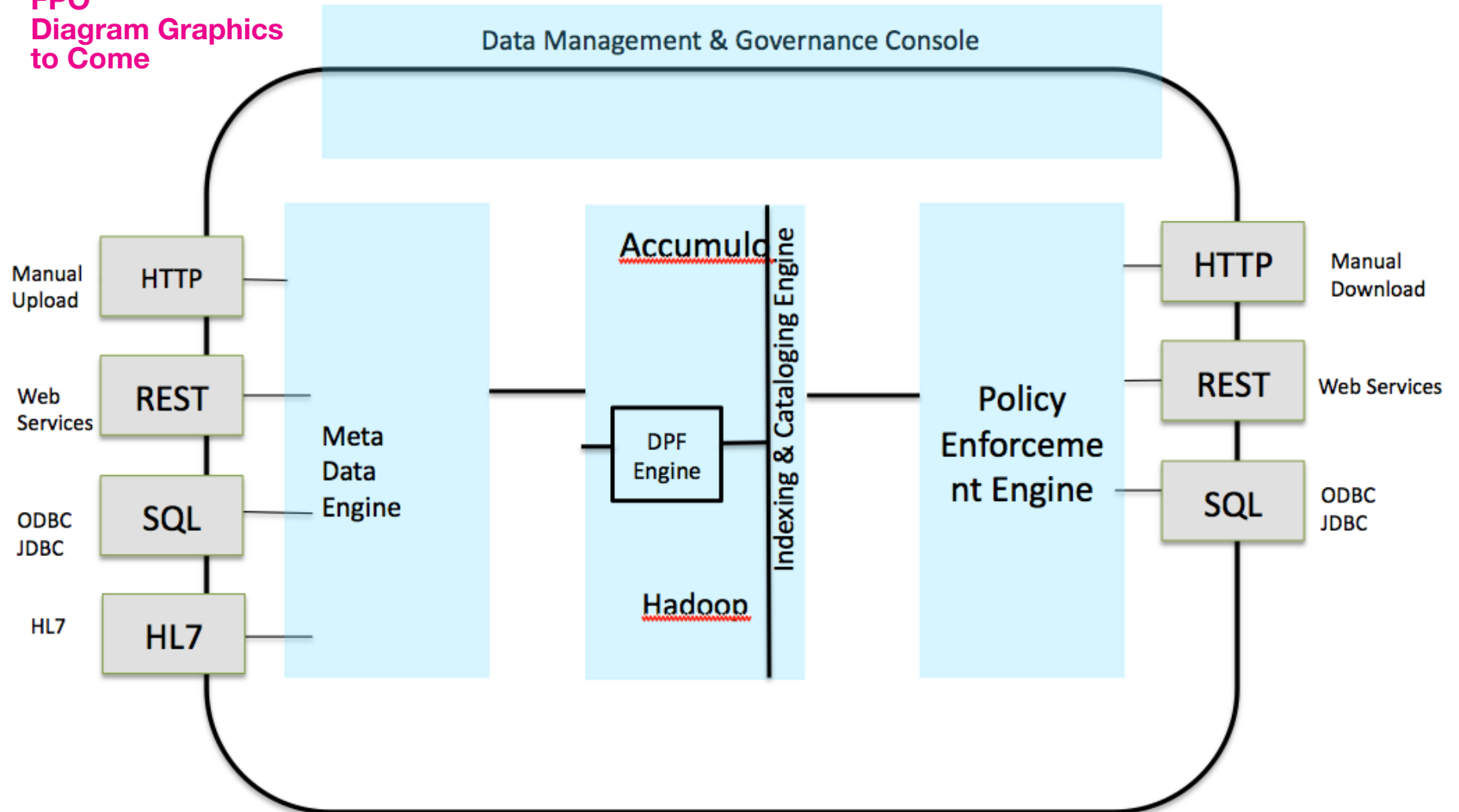**Paragraph on Data Analysis tools, including SQL**

Paragraph on Custom Applications

There may need to be some discussion on the role of the Policy Enforcement Engine, since all requests go back through it.

Privacy, Security, and Governance

Protect sensitive information at scale…

**FPO
Diagram Graphics
to Come**

Goal of this image is to convey what pieces in the logical architecture fit into the function "PSG"…Highlighting the different components….keeping grey in? Or out?

Content TBD; could include:

Maybe a few screen shots of how to define access policies, etc..

Intro paragraph

Paragraph on Privacy By Design

Graphic on PbD, how PHEMI Central implements, why its cool

Paragraph on Access Controls

Paragraph on enforcement & how we do data Sharing Agreements

Paragraph on Anonymization & De-identification

Paragraph on Encryption – AT Rest & In Motion

Paragraph on Audit Logs
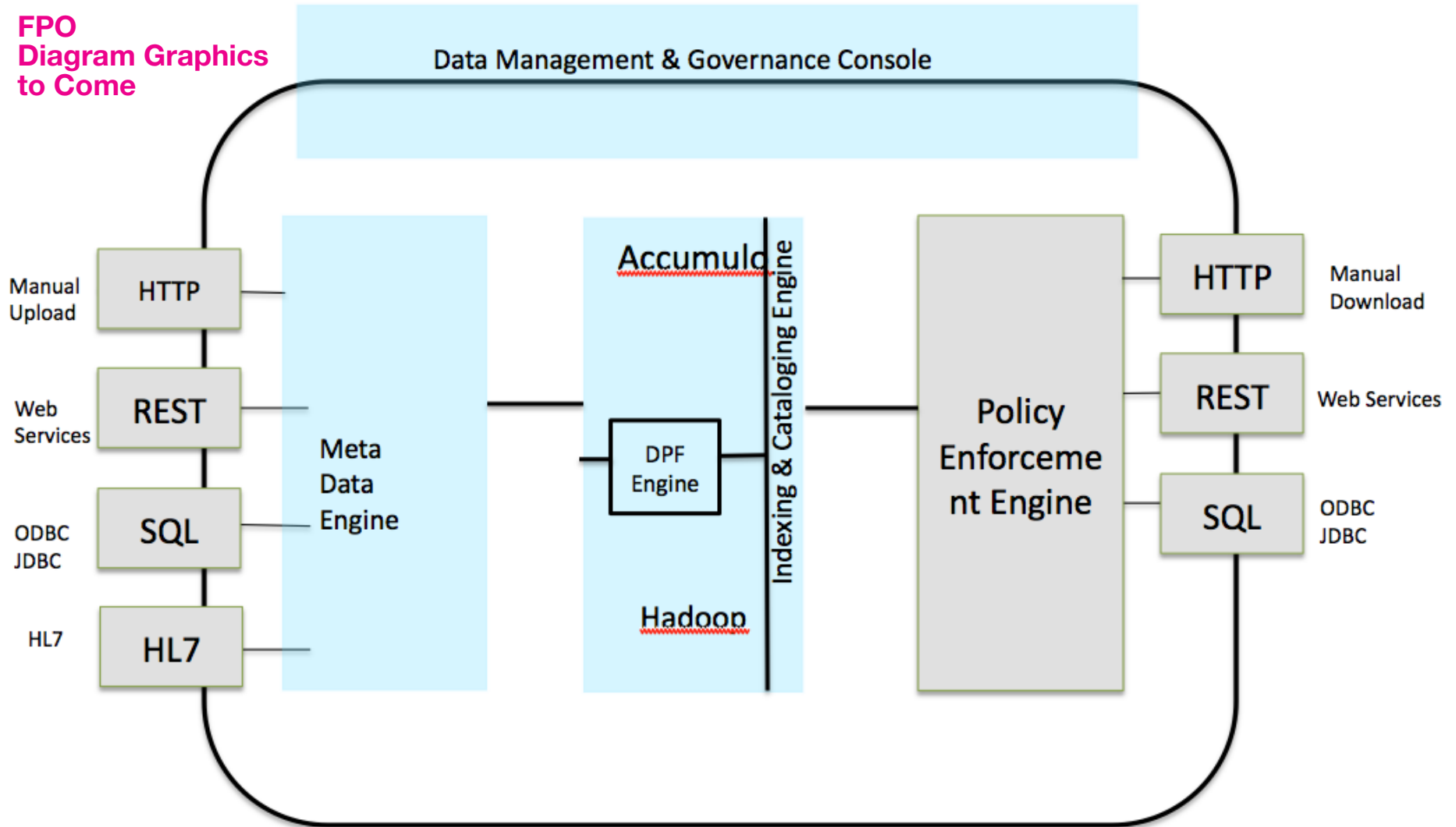
Paragraph on Data Immutability

Paragraph on Version Control

Paragraph on how powerful separating data protection from data use.

Data Management

Fine-grained management. Manage your data at the data element level.

**FPO
Diagram Graphics
to Come**

Goal of this image is to convey what pieces in the logical architecture fit into the function "Data Management"…Highlighting the different components….keeping grey in? Or out?

PHEMI

Content TBD; could include:

Paragraph about managing the execution of dpfs across all data. Enriched pieces at scale (Managing the DPF Framework)

Paragraph on Life Cycle Management

Standards Plus Approach to Data management (unlike other hadoop, we strictly manage psg & cataloging. Users can't bypass)

* Hmm…Paragraph on Meta Data Framework (manage data at the data element level) or is this a repeat of Collect

* Paragraph on the value of schema-less – how it helps in data collection/curation. Or is this a repeat of Collect

System Management

Enterprise Grade Reliability, Availability and Scalability with Cluster Economics.

**Diagram
ROUGH SKETCH
NEEDED**

PHEMI

Content TBD; could include:

Diagram of the hardware/nodes, etc… (get from Roy)

Paragraph on Scalability and performance

Paragraph on Reliability and Availability performance

Paragraph on Apache Hadoop Platform….??
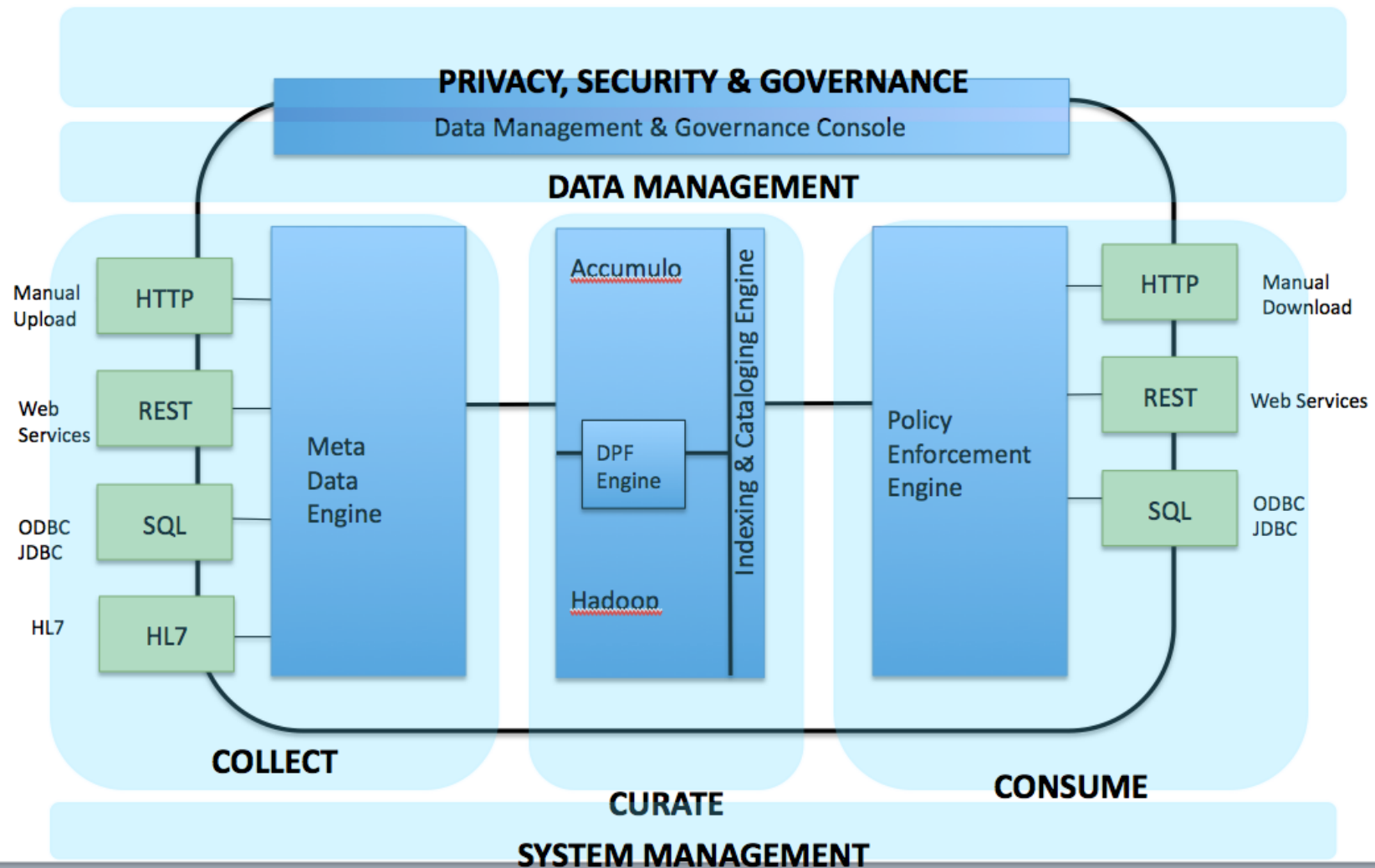
Paragraph on Commodity Hardware

Paragraph on LDAP integration

Paragraph on Maintenance & Support (skill sets and types of activities that a user would need to get the system up & running)

Lorem Ipsum…

Content TBD; could include:

3 columns of Applications, each with a short paragraph
Enterprise Information(5)
Healthcare Industry (3)
Public Sector (3)

Lorem Ipsum

Content TBD; could include:

2 buckets (on premise/in the cloud)_

PHEMI

Ideal for Organizations dealing with Large Scale Private and Sensitive Data

Content TBD; could include:

Intro Paragraph

1 Paragraph/Column on Co-exist with your existing data warehouse" (bullets)

1 Paragraph/Column about how PHEMI Central augments/is better than a DIY hadoop (bullets)

PHEMI

Content TBD; could include:

1 to 2 paragraphs about PHEMI and how to reach us/book a demo/start a trial…

Questions and Issues from AB

There will be some debate on where the subjects of "schema-less" and "meta data" should reside. (Data Management Vs Collect)

Screen shots? Which sections for most value?

LifeCycle Management – does it belong in Data Management or in PSG?

Not enuf Data Management! Feels like the logical home for Meta data, take care of execution of dpfs across all data. Enriched pieces at scale.

Feels like we need more content on what governance is?

Where does discussion about Hadoop/Accumulo appear?

Logical Warehouse…no fit in this document.