Accelerating Clinical Informatics with Data Virtualization in the MedBus Service-Oriented Architecture

Amitava Shee, James Estill, Brian Rafferty, John Brussolo
University of Michigan, Medical School Information Services, Ann Arbor, MI

Problem : Disparate Dynamic Data

Clinical research informatics requires flexible, scalable and rapid solutions for delivering contextually tailored data sets in secured systems that protect patient data.

- Clinical research often requires data integration from heterogeneous sources with distributed governance.
- Research ready data resources need to be provided to multiple analytical platforms with competing standards.
- Research relevant data from the UM Medical School exist in over 300 enterprise repositories that need to be combined with multiple external data resources.

Solution : Service Oriented Data Integration

MedBus is a Service Oriented Architecture that we have developed to provide secure data services that are findable, accessible, interoperable and reusable.
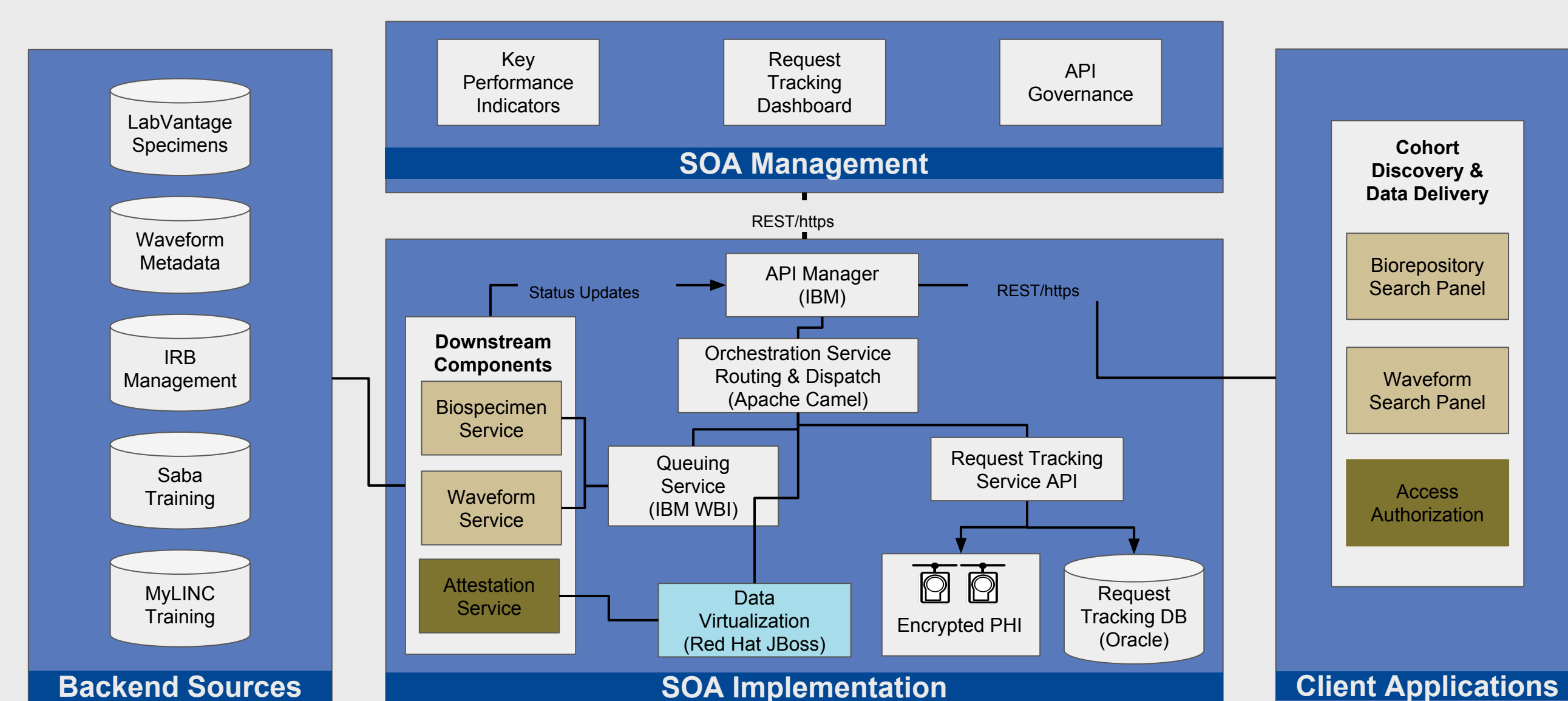
- Since a March 2015 launch, MedBus has responded to >1000 requests from a cohort discovery application that integrates MedBus services with EMR data.
- Embraces open science and collaborative research using github as a collaborative platform.
- With a growing portfolio of over 300 data sources, we are developing strategies to accelerate development and deployment of our integrated data services.
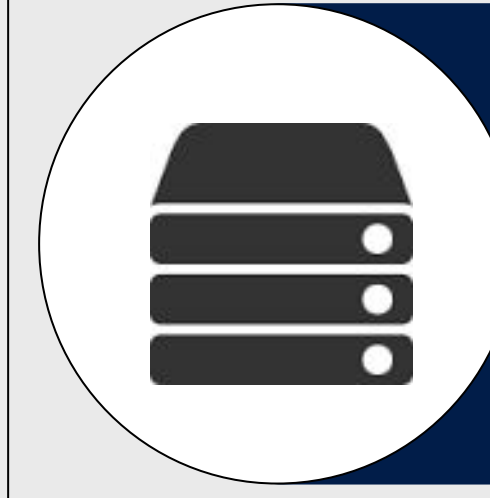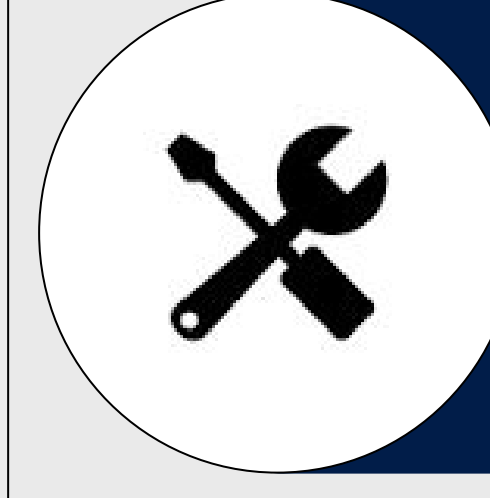
Data Virtualization
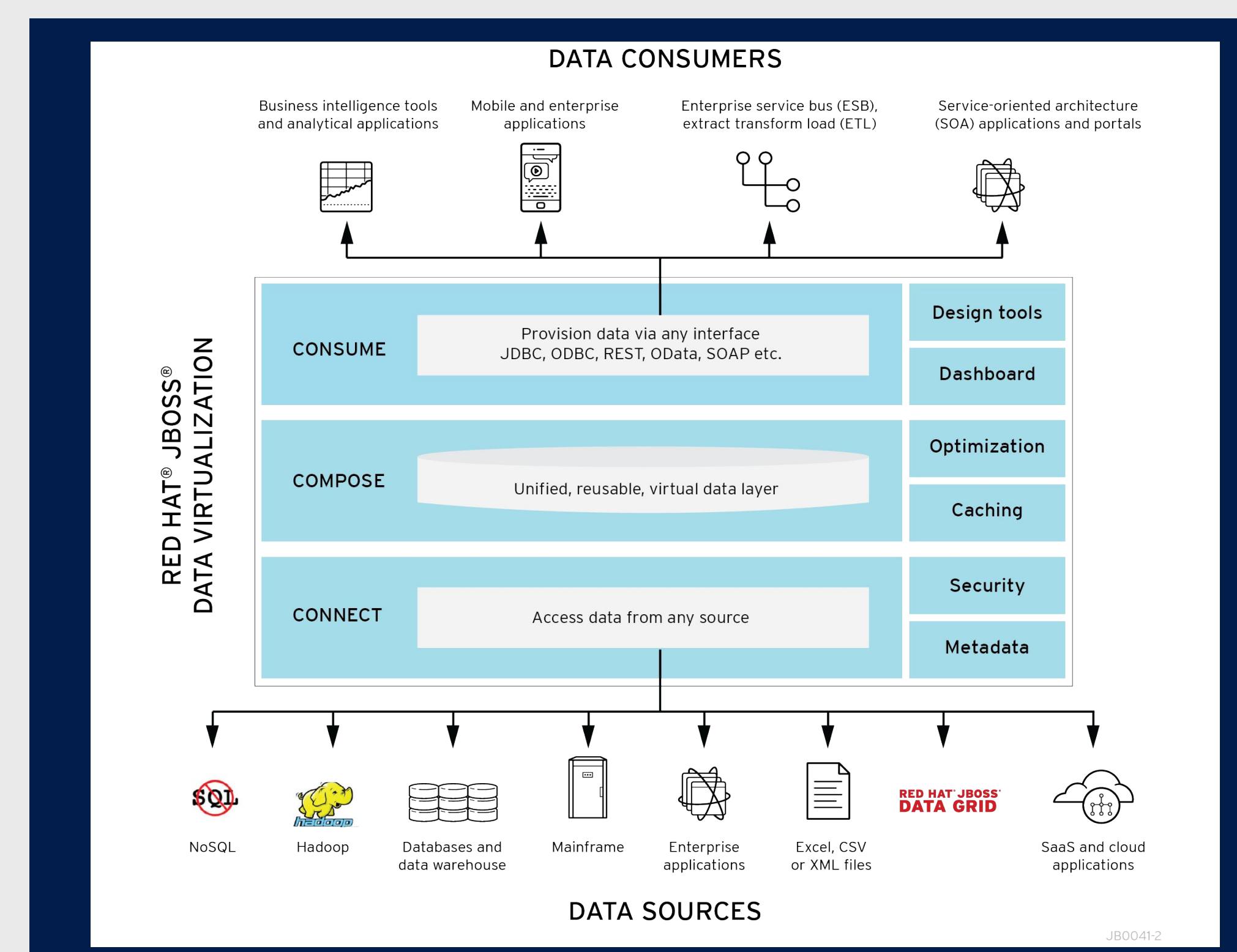
Key Capabilities

HIPAA compliant - implements tight security, auditing and access controls

Rapid delivery - real time data integration with no custom coding effort required

Enterprise Scale - industrial grade technology that is scalable by clustering commodity machines

Customizable - open source java platform allows for highly customized data source adaptors

Acceleration: Virtualized Data Integration

JBoss Data Virtualization is a data integration solution that allows us to rapidly transform data from disparate platforms into real-time virtual views that are consumable via ODBC, JDBC or as web services.
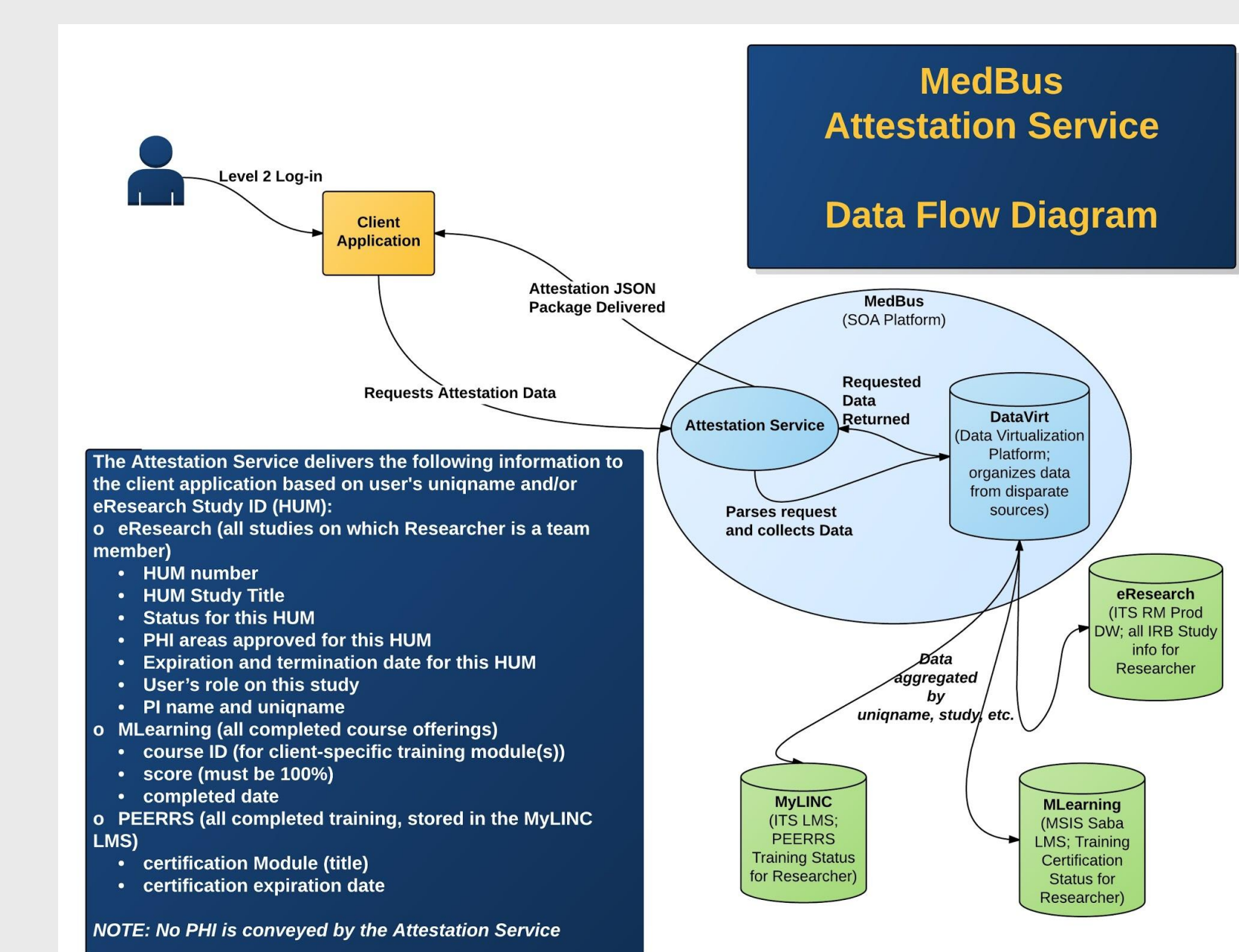
- Centralizes ePHI data access & auditing while also providing fine-grained access control that ensures we deliver only the data a user is authorized to access.
- We developed an access authorization service in two weeks that combines data from 3 disparate sources.

2016 Roadmap

- ❏ **Streaming data capabilities** added to MedBus will allow TB scale queries of >10 million of records.
- ❏ **Epic service translation** will use data virtualization to translate our SOAP web services into JSON.
- ❏ **Genomics data services** will query metadata from genetic variant assays and exome sequencing.
- ❏ **A big-data enabled** platform will use virtualization to accelerate provisioning of data models such as the PCORNet CDM in a financially sustainable manner.

More information available at:
http://medbus-umich.github.io