

Unified Data Lake for Healthcare Research and Analytics

Ingestion | Governance | Scheme Evolution | Masking | Tokenization | Transformations | Validations | Compliance

Enable rapid healthcare research and analytics with fully functional operational data lake

BENEFITS

DO MORE WITH EXISTING ENGINEERING TEAM

Skilled Data Engineers are one of the most valuable resources in IT. BDM automates the key coding requirements around creating data pipelines. Thus, your data engineers can focus on more critical requirement.

DATA GOVERNANCE

BDM captures and records all data access and movement activities, ensuring a high level of data quality, transparency, lineage, and auditability.

DATA LAKE ANONYMIZATION

BDM allows sensitive data to be anonymized as it is being moved into the data lake - as the whole process happens in memory, so there are never any copies of the data kept on disk.

FAST AND RELIABLE DATA PIPELINES

Enable automated pipeline creation in minutes and hours in comparison to traditional manual methods, which are error prone and take weeks and months to create.

DATA FEDERATION

Combining BDM's anonymization technology with Hadoop/ Spark's data access functionality allows different data users to access different views of the same data from one platform

DATA SELF SERVICE

With BDM's easy to use GUI, non-technical people can access the data they required directly on the data lake in a controlled and secure manner.

THE CHALLENGES

COVID-19 has exposed the lack of coherent and joined-up information systems across and within public healthcare systems worldwide. What data is available is siloed and is not easily integrated, which has resulted in public healthcare systems struggling to respond to the pandemic in an agile manner.

While European and global healthcare data is enormously fragmented, data is collected and controlled by a multitude of local, regional, research and national entities, making it difficult to get a complete view of healthcare-related issues.

In addition to the technical heterogeneity, datasets have different governance rules. Each organisation has its own policies for data access, research use, copying, and distribution; these vary from one dataset to another. Additional challenges are introduced by regional, national and EU legislation, of which GDPR and privacy laws are only the most prominent.

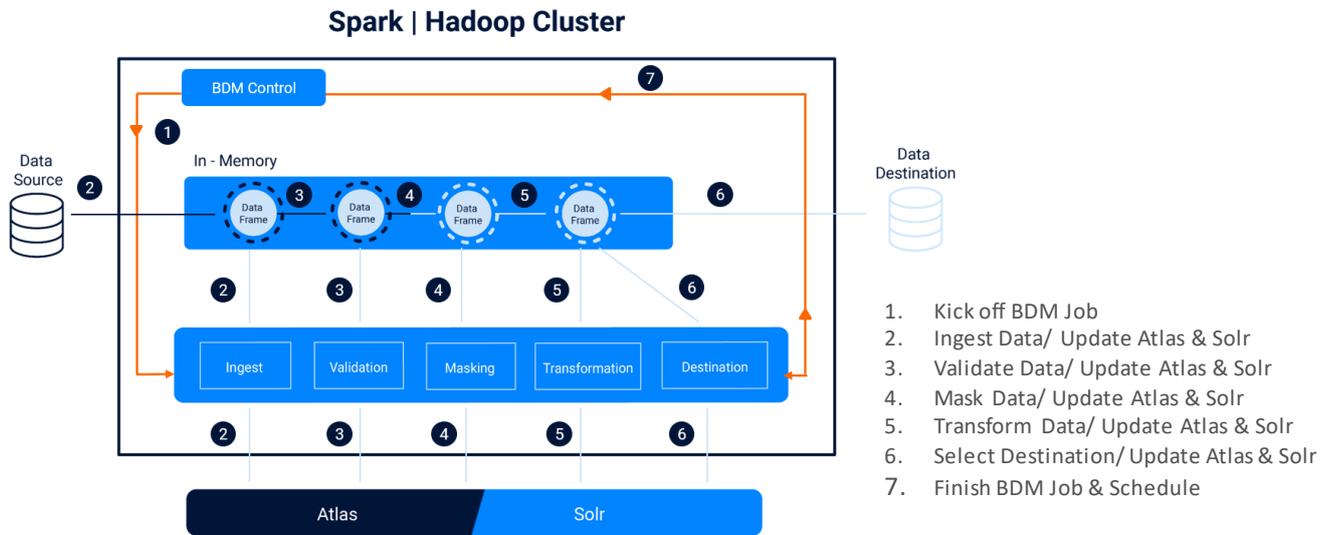
THE SOLUTION

To effectively understand healthcare insights, a data lake is optimal for predictive and advanced analytics when the volume and variety of data rapidly grow like the actual COVID situation. However, building a high-performance and fully functional operational data lake can take years before providing solutions.

Our Data Lake solution is built around the following two proven technologies:

- **Hadoop/ Spark:** is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage and computation power for any kind of dataset, enormous processing using parallel processing technology, and the ability to handle virtually limitless concurrent tasks and jobs. We are using an Enterprise version of Hadoop from Cloudera – which is used on 90%+ of all Data Lakes in Finance, Industry, Pharma, etc. today - on our solution as it ensures we always have access to the latest innovations and security features.
- **Bluemetrix Data Manager (BDM):** BDM is a suite of data automation functionality that has been developed by Bluemetrix for working with data in Data Lake environments in a rapid manner.

REFERENCE ARCHITECTURE: HOW DOES BDM PROCESS DATA PIPELINE?



BDM sits in the new and emerging market around Data Operations – Data Operations defines the movement of a data set (a file, a table in a DB, a stream of data) from a source to a destination, and the processing of this data and all of the steps in the processing that were carried out on the data. It includes traditional concepts such as ETL (Extract, Transform & Load data), but also includes data quality, data governance, data transformation and data anonymization.

BDM uses the Spark (an open-source software from the Apache foundation) processing engine to process the data. This means that the data is read into memory from its data source, the actions in the data pipeline are executed in memory, and then the final data is written out to the destination.

By implementing Bluemetrix’s BDM platform, you can guarantee secure, protected, and governed data before it was migrated to the data lake, which could be shared with the researchers and staff within the healthcare provider and with third parties. Also, when any stakeholders interacted with the data – from the healthcare provider to third parties – the integrity of the data remained constant and reliable.

Case Study: Data Lake Solutions for Healthcare

Learn how BDM Health help UK’s Largest Healthcare Provider harness the value of sensitive data to deliver improved patient care through analytics platform

[Get the details >](#)

ABOUT BLUEMETRIX

Bluemetrix has been working with Hadoop since 2009 and have experience in all areas of the Stack – Architecture, Infrastructure, Security, Application Development, Deployment, Operations and Data Science. We are leaders in developing and deploying innovative solutions to deal with problems on the Hadoop Stack, with a focus on developing real-world automation solutions that removes the need for Hadoop expertise. We guarantee delivery on all Hadoop projects we undertake.