

Data Policy Automation and Enforcement with BDM and Collibra

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

PRODUCT DESCRIPTION

Bluemetrix Data Manager (BDM) automates the creation of Collibra data governance and lineage and enforces Tag and Standard based policies on all pipelines created. Imagine the ease of data governance, secure data sharing and access, efficient data protection and provisioning at scale— it's all available within BDM.

BUSINESS CHALLENGES

Today, most enterprises understand the value of data governance and have programs designed specifically for how their data is accessed or processed.

Despite having numerous data policies produced and approved across the enterprise, these policies are not always enforced and implemented when data is accessed and used daily.

This happens due to two reasons: the tools are not available to enforce them, or the data user is not aware of the existence and detail of the policies. Without the ability to scale governance, enterprises run the risk of their security and governance controls falling short.

BDM SOLUTION: AUTOMATE & ENFORCE DATA POLICY FOR GOVERNANCE

BDM's full-service Spark based ETL solution integrates with Collibra Data Platform, enabling data users to create, capture and maintain the data governance state of all data assets that are processed in the pipeline, while also enforcing the policies that govern the data assets.

You'll grant more access to meaningful data confidently, quickly, and easily because you don't have to rely on the manual implementation of policies to ensure the right people get access to the data they need, in a secure and compliant manner.

BLUOMETRIX DATA USERS

Data users in a new era of data governance.

Data Owners

Data owners get an easy way to publish datasets and therefore gain a complete visibility into how the data is being used, accessed, and controlled.

Data Stewards/Governance

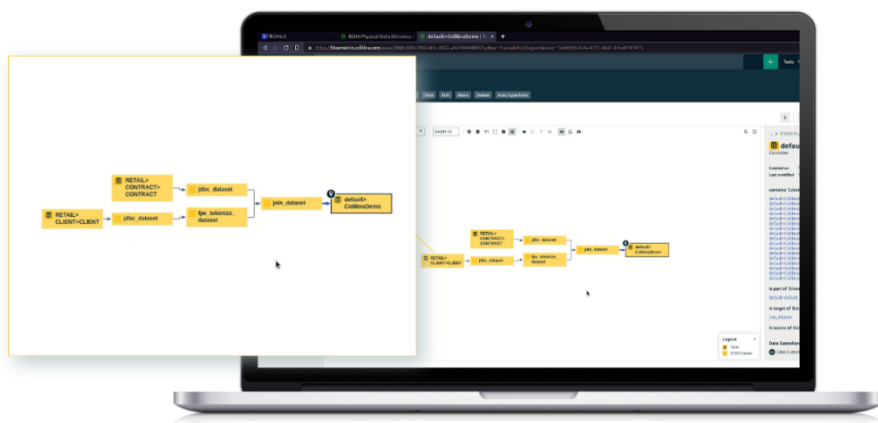
Data stewards automate away the manual tasks of implementing governance policy, while still being able to define and enforce access policies centrally at a fine-grained level, understand access levels, as well as audit historical data access activity.

Data Consumers

Data Consumers get empowered by the fast and secure self-service access to data for analytics success.

KEY BENEFITS

- **Improve underlying security of the data**
Automation of policy enforcement removes security risks brought about by non-compliance
- **Self-Service data capabilities**
Providing a data discovery solution in combination with a data authorisation solution
- **Unlocks the potential of existing data governance solutions**
Bluemetrix allows users to create a more automated and operational governance environment
- **Guarantee of policy enforcement**
The automation of policies guarantees enforcement and compliance with policies across the enterprise
- **Simplified data stack**
Integration of ETL and governance tools in one solution



PRODUCT DETAILS

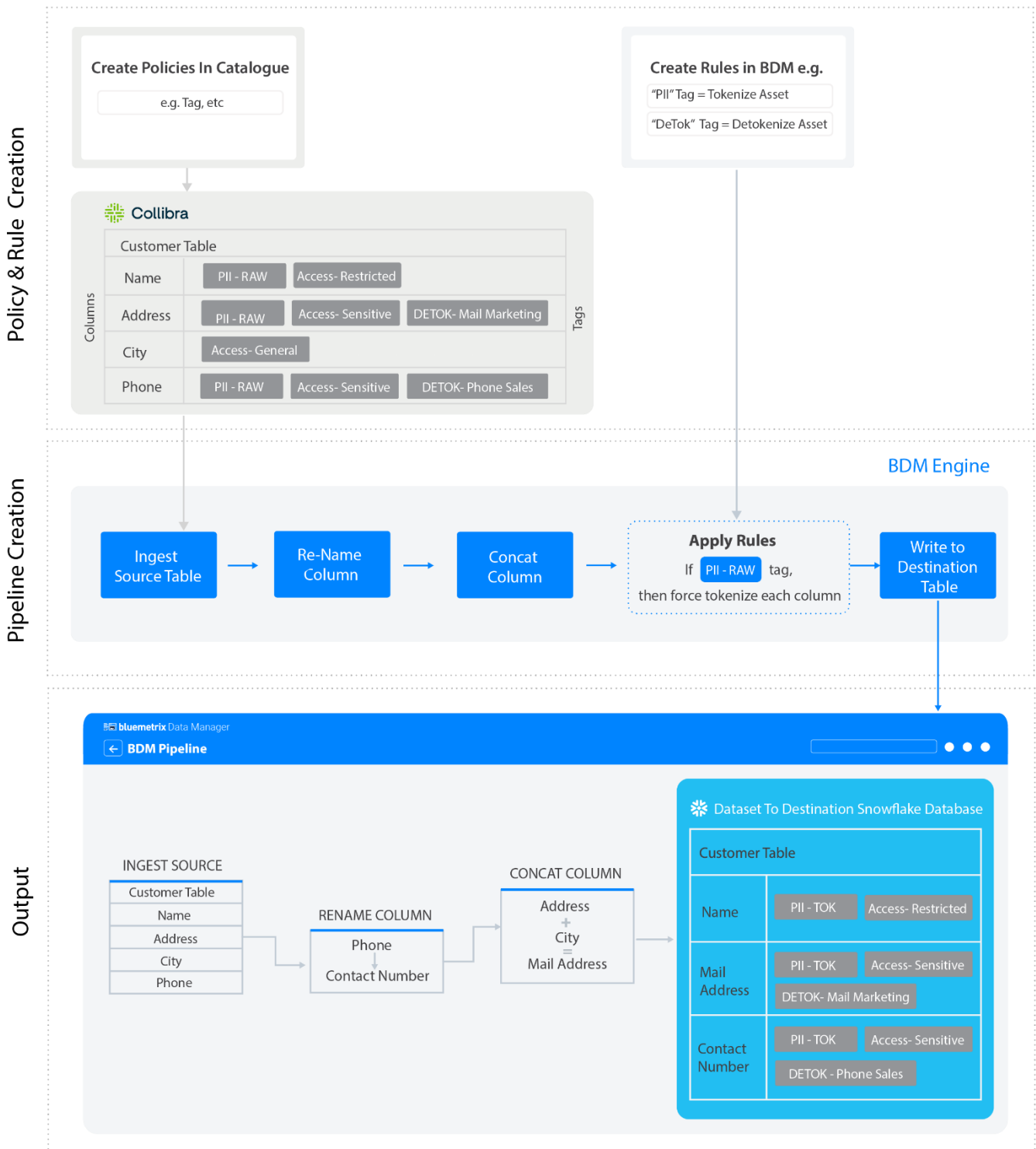
BDM enforces all Tag based and Standard based Data Policies that exist in Collibra Data Platform. It does this in the following manner:

- For all data sources in the pipeline, it checks for any Tags and Standards that govern the data source
- For each Tag and Standard found it determines if a corresponding rule exists in BDM e.g., all data that has a PII Tag must be tokenized on Write
- Before the Pipeline executes a Write, it applies all Tag or Standard based rules to the pipeline e.g., for all data assets

that are tagged PII, automatically apply a Tokenization function to the data before it is written to its destination.

In this way, BDM will automatically enforce and execute the Collibra Data Policies associated with each data source in the pipeline, with no input required from the data engineer.

BDM makes it easy for data platform teams to integrate it within their environment and gives them the choice to use other services as they see fit.



CAPABILITIES AT A GLANCE

DATA INGEST

- Simple template-based Connector system for all data sources
- Multiple Connectors available
- No need to develop any ingest code or select appropriate stack components
- New data sources can be deployed in hours rather than weeks or months
- Storage can be selected to suit the data type and processing requirement i.e., HIVE, HBase, etc.
- No extra code is developed, reducing the code release cycle time and complexity

DATA MASKING/TOKENIZATION

- Data Masking is available on ingest to the cluster.
- It can be carried out on a column or table basis
- Stateful and Stateless Tokenization solutions are available
- Different masking algorithms can be applied to suit the data i.e.
 - Complete removal of selected columns
 - Replace values with random data
 - Add a random value to each row in the table
 - Categorize data e.g., exact salary replaced with a range
 - Geolocation data – apply rotation methods to mask the data

DATA QUALITY & VALIDATION

- Data Consistency is guaranteed by applying checksums and other controls on the data
- Data Integrity is provided by Regular Expression and ML algorithms
- Easy integration with existing data quality and validation systems

DATA TRANSLATION

- All source schemas are translated into destination compatible schema's
- Control characters removed and changed as appropriate
- Data cleansed, formatted, and factored

DATA TRANSFORMATION

- Data transformations are coded and stored in a custom library deployed in Spark
- Data maps/flows can be created using a drag and drop interface
- Dramatic reduction in code developed and deployed
- Dramatic reduction in scripts developed
- Ability to write and create SQL transformations also available
- No requirement for Spark expertise to create transformations
- An API can be provided to the Spark library allowing client developers create and deploy their own Spark transformations

DATA CATALOGUE

- All data is automatically detected and catalogued on the Collibra system upon ingest
- ACL's -including tag-based policies – implemented and integrated with underlying Hadoop Stack where required
- Tags can be either technical or business, and can be propagated through the pipeline
- All data governance capabilities – Audit, Change Tracking, Lineage, etc. – are captured

DATA POLICY ENFORCEMENT

- Data Policies are read and interpreted for all data sources in the pipeline
- An extensible rule-based engine exists in BDM to capture all policy implementations
- Policy implementation occurs automatically during the execution of the Write phase of the data pipeline
- There is no requirement for the data engineer to access or implement the policies – it happens automatically within the system

TECHNICAL REQUIREMENTS

The combined solution requires Collibra Data Intelligence Cloud Platform and BDM release after June 1st 2022.

APPOINTMENT

For more assistance, you can schedule a one-to-one appointment via Skype, Zoom, WebEx or phone.

Bluemetrix

<https://www.bluemetrix.com/>

info@bluemetrix.com

EU: +353 21 421 2223

Solutions

- BDM Ingest
- BDM Control
- BDM Health

About Bluemetrix

Bluemetrix have 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. We guarantee delivery on all Hadoop projects we undertake.