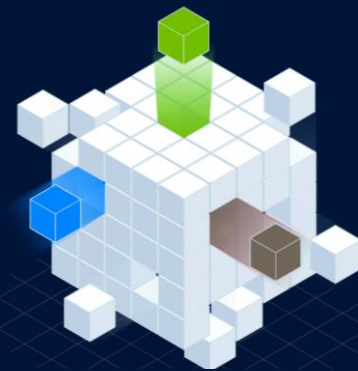**bluemetrix** | **IDERA** | **Collibra**

How Bluemetrix & IDERA Accelerate JSON Data Extraction into Snowflake

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

PRODUCT DESCRIPTION

Together, Bluemetrix, IDERA and Collibra provide the enterprise-class analytic engineering automation that allows your data team export data from nested JSON files into relational tables, simplifying the process and expediting access to your data. With powerful and seamless integration, your data team can now scale analytics projects as needed and deliver end-to-end ETL pipelines with the highest level of compliance and security standards.

BUSINESS CHALLENGES

Snowflake offers a powerful way to store and process semi-structured data like JSON, allowing data analysts to derive valuable insight from the data using standard SQL. However, for data analysts to access the value in this data, data engineers must first extract the nested JSON data into relational tables to process it.

Managing a large amount of JSON data is a manual process that takes time away from data projects, introduces risks, and slows down Snowflake's use and adoption.

In addition, with more stringent data privacy regulations arising from the need to ensure every approval for data access and sharing requests adheres to governance policies, the process can be daunting and challenging.

INTEGRATION SOLUTION: EFFORTLESSLY LOAD NESTED JSON INTO SNOWFLAKE

Now with this solution, we provide enterprises of all sizes the assurance and confidence that extracting data from JSON files has never been more effortless and secure allowing your data team to take full advantage of the scale, speed and flexibility of Snowflake Engine.

PUT BLUEMETRIX AND IDERA TO WORK FOR YOU

Data Engineers/Data Operations
Are freed from manual pipeline creation and governance implementation and therefore devote more time to real business problems and accelerate safe, secure data provisioning to data consumers.

Data Architects
Design a fully functional data architecture that leads to better allocation of resources at lower costs – while maintaining processing needs.

Data/BI Analysts
Get empowered by the fast and secure self-service access to data – accelerating query execution to analytics success
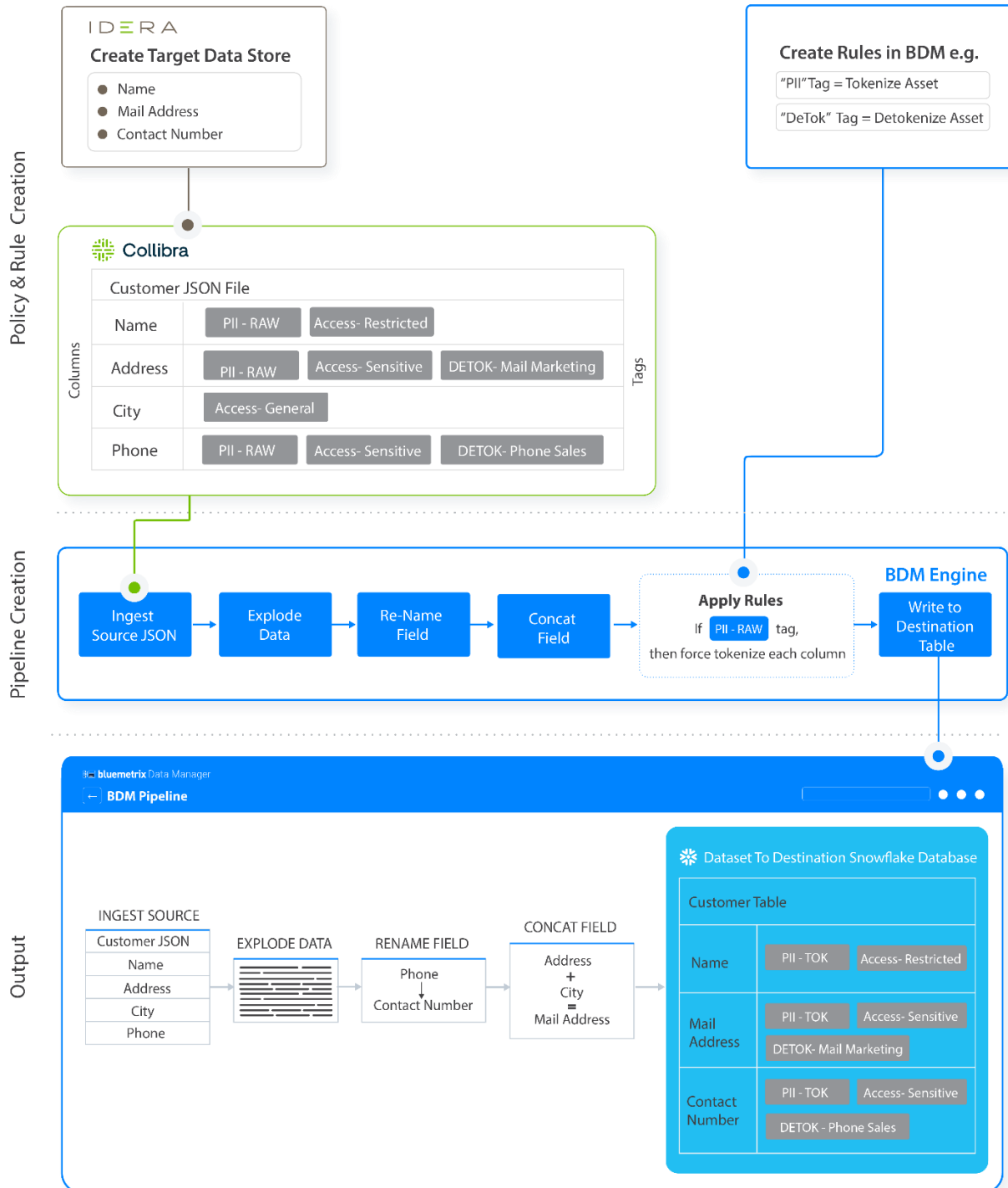
KEY BENEFITS

- **Automatically apply governance in BDM**
  Data policies in Collibra will be automatically implemented and enforced on the JSON data
- **Streamline data access and sharing**
  Bluemetrix makes it easy for data owners to change policies in Collibra, safe in the knowledge that they will be implemented in BDM
- **No-code engineering built for Snowflake**
  Manage end-to-end ETL pipelines and automate the integrity of your Snowflake datasets at every stage of the data lifecycle – without writing a single line of code
- **Advanced security to PII data**
  Secure data tokenization and masking capabilities enable scalable and efficient protection for sensitive, regulated data
- **Automatic schema conversion**
  Use ER/Studio to convert JSON hierarchical schema to a relational database design for most technologies

PRODUCT ARCHITECTURE DETAILS

Using automation from ER/Studio, Collibra and BDM, we can simplify the extraction of data from nested JSON files and make it available for analysts by carrying out the following steps:

- Create target relational databases schemas and models for Nested JSON files
- Store these models and schemas in Collibra along with glossary information highlighting sensitive and PII data

- Read these models from Collibra into BDM
- Read the JSON files and use the models to explode the data, and transform it into relational tables in Snowflake
- Automatically apply data policies to the data as it is being transformed
- Capture and record all of this activity in Collibra

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.

bluemetrix.com   #datacontrol