

Data Quality & Trust

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

Validate the Consistency & Integrity of your data as it moves within your Data Lake

SECURE DATA

All validation checks happen in memory – there are no data copies left on disk

DATA IN TRANSIT

Corruption of Data in transit is detected by applying Consistency checks (checksums, etc.) on the data

EXTENSIBILITY

All validation algorithms are extensible and can be developed to suit the underlying data

SMART DASHBOARDS

All quality data is accessible through a dashboard which will provide a snapshot of the health of the data on the cluster

BUILT IN TRACEABILITY & LINEAGE

Prebuilt data traceability, tagging and lineage reporting capabilities for data policy adherence at all steps of the process

THE CHALLENGE

Poor data quality undermines business digital initiatives and weakens competitive standing and leads to customer distrust.

Given the growth in data that companies are seeing, which will continue to grow exponentially as 5G, and IoT Data becomes more mainstream – validating the quality of data has to be seen as a process. As data changes rapidly the definition of what is accurate may change from day to day (or hour to hour), and you need a system which will capture and identify the changes that are significant for your business.

Manual checking and spot checking of data will not deliver the accuracy and the timeliness of the results that your business requires. Data validation must be automated and part of your overall data preparation process, so that you can ensure the validations are carried out when required and are as up to date as your data itself.

THE SOLUTION

BDM in its simplest user-configurable mode moves data from a source to a destination, and as part of that movement process. BDM Data Validation carries out two distinct classes of validation on the data:

Consistency – checks the data was not corrupted in transit

Integrity – checks the value of data being moved is within expected values

These checks are carried out on the data as it is being moved between source and destination i.e. when it is in memory in the Spark cluster. This methodology is secure and optimal as it ensures that replicas of the data are not written to disk.

There is a base library of standard validations that can be applied to the data set i.e. Consistency (Checksum, null value, etc.) and Integrity (Numerical operators, Range Values, Limits, Regex, etc.). This library is extensible and new validation checks can be written and added as required.

CREATE A CULTURE OF DATA TRUST



Head of Data

- Manage data processing consistently and securely
- Reduce ingestion, access and reporting delays by weeks
- Secure buy in by increasing quality, security and trust in your data process
- Prove, Transparent data governance procedures



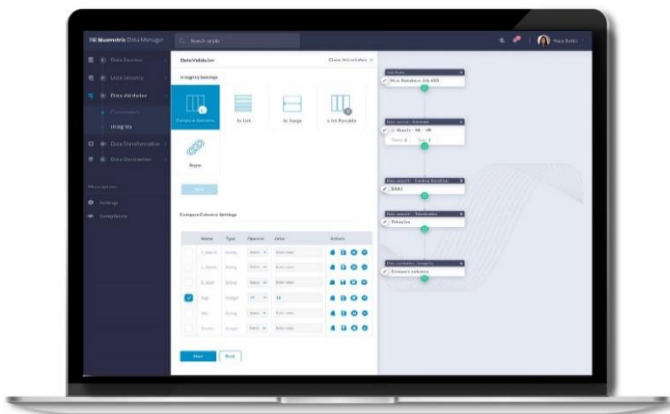
Head of Analytics & Data Science:

- Automatically create and execute your own data requests
- Put the controls you require around your pipelines
- Ensure your pipelines are getting the production ready data they need



Head of Department:

- Increase time to value on your data
- Reduce SLA and reporting delays
- Empower your teams to access the value in your data



LEARN MORE

To learn more about BDM Validation data sheet, the different BDM user interface metaphors, and the entire BDM data lake engine automation platform, please visit our website at www.bluemetrix.com. There are resources available to guide you on your journey and demo request available for you to see first-hand how the platform delivers for your data pipeline project needs.

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.