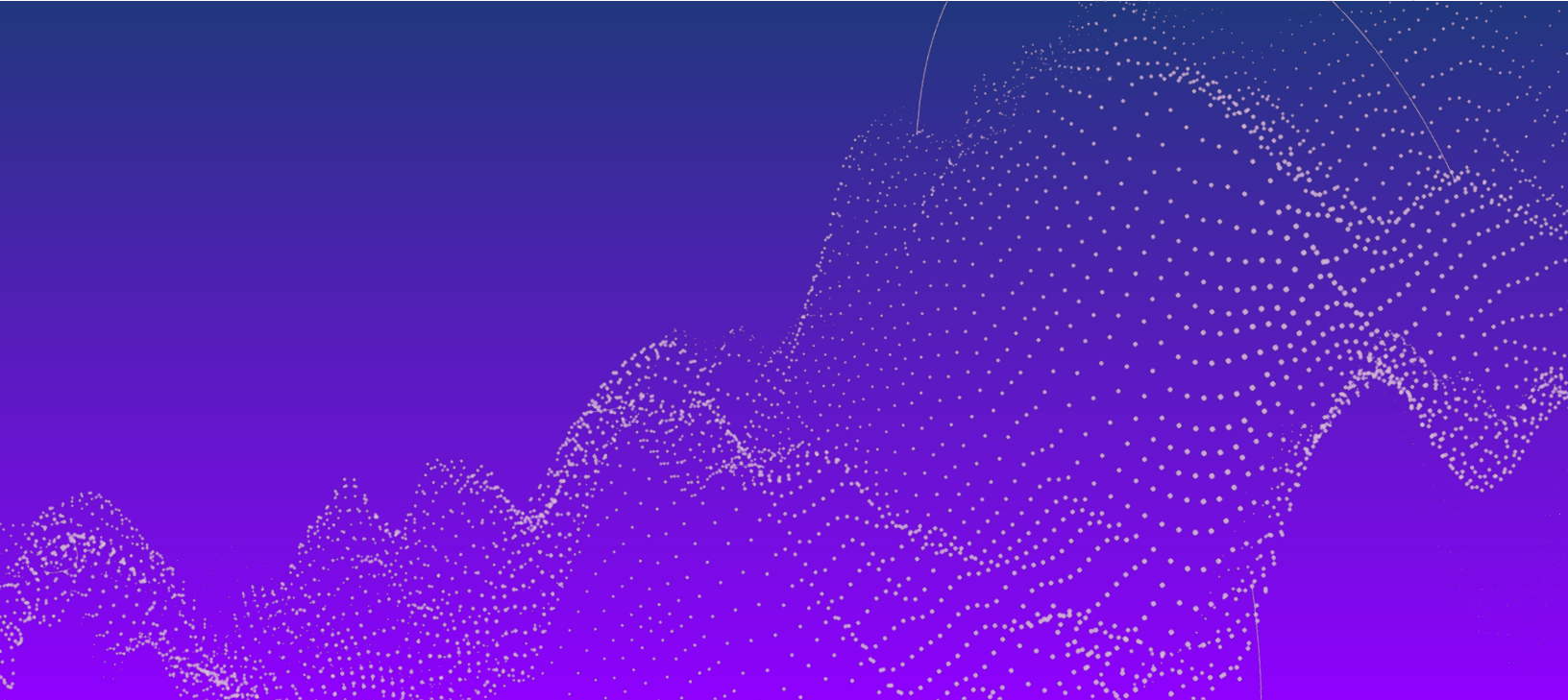


Datameer®



Datameer for Data Preparation:

Explore, Profile, Blend, Cleanse, Enrich, Share,
Operationalize

Datameer for Data Preparation

Datameer offers the complete set of capabilities you need for collaborative, operational data preparation pipelines to serve the needs of both your data and business analysts. With its data preparation functionality, Datameer allows you to:

- Explore and profile data visually
- Cleanse data to ensure validity
- Blend a variety of data formats
- Enrich and transform data
- Shape and organize datasets
- Share and collaborate datasets
- Operationalize data pipelines
- Govern with an open-secure model



Explore and Profile Data Visually

Explore all types of data

Reach the next level in data discovery. With Datameer, you can access and use any type of data, anywhere in the world.

Freely upload and parse any file formats for analysis, including:

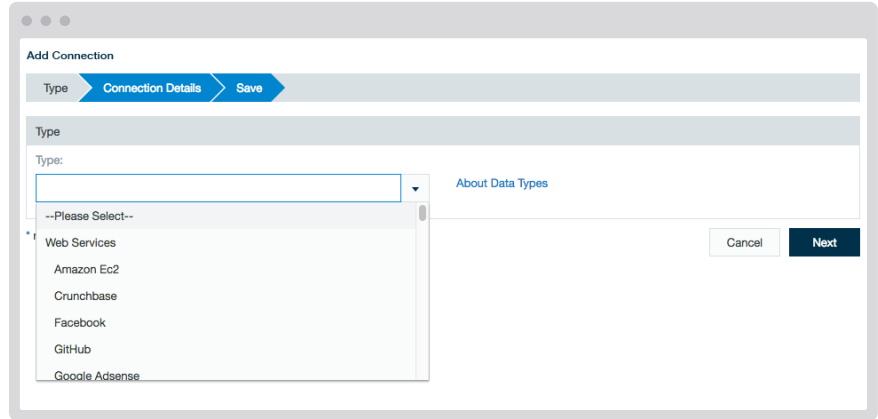
- HTML
- JSON
- XML
- CSV
- Text documents
- Apache logs

Pre-built data connectors and parsers

- Extract from 70+ data sources and a variety of pre-built data formats
- REST API allows any new file formats that suit your needs

Extract custom structures

- Examples include custom data parsing, tokenizing for text parsing and time zone support
- Extract matching patterns with advanced parsing through regular expressions

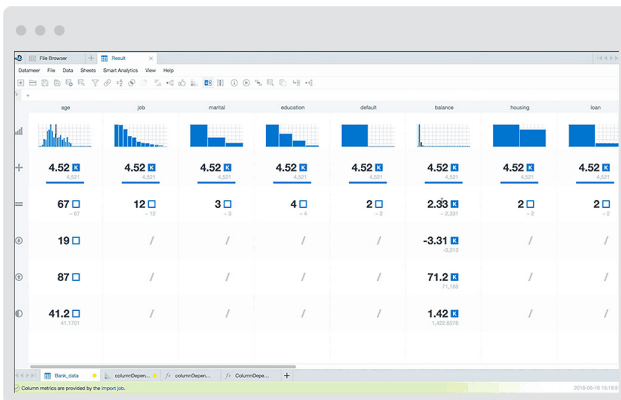


Profile data visually

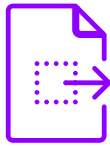
Gain a new level of understanding. With Datameer Flipside™, you have instantaneous statistical data profiling of your data at any step in the preparation and analytic process. Now you can quickly determine the next preparation steps for your data — which means a more accurate view of the data.

Flipside™ also allows you to uncover hidden patterns. And with better profiling, you'll have a better sense of the data's accuracy.

More accuracy and more recently refreshed data provides better insights.



age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	TimeSpent	NumOfCalls	Campaign
38	unemployed	married	primary	no	1,787	no	no	cellular	19	oct	79	1	-1	8	unknown
33	services	married	secondary	no	4,709	yes	yes	cellular	11	aug	188	1	209	4	failure
35	management	single	tertiary	no	1,358	yes	no	cellular	16	apr	185	1	308	1	failure
38	management	married	secondary	no	1,478	yes	yes	unknown	3	jan	199	4	-1	8	unknown
39	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	206	1	-1	8	unknown
35	management	single	tertiary	no	187	no	no	cellular	23	feb	242	2	208	3	failure
36	self-empL	married	tertiary	no	387	yes	no	cellular	14	may	342	1	308	2	error
39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	8	unknown
41	entrepreneur	married	tertiary	no	221	yes	no	unknown	18	may	17	2	-1	8	unknown
43	services	married	primary	no	488	yes	yes	cellular	17	apr	283	1	147	2	failure
31	services	married	secondary	no	9,274	yes	no	unknown	28	may	273	1	-1	8	unknown
43	adminL	married	secondary	no	264	yes	no	cellular	17	apr	113	2	-1	8	unknown
36	technician	married	tertiary	no	1,299	no	no	cellular	13	aug	308	2	-1	8	unknown
38	student	single	secondary	no	182	no	no	cellular	28	apr	262	1	-1	8	unknown
31	blue-collar	married	secondary	no	368	yes	yes	cellular	25	jan	89	1	245	1	failure
48	management	married	tertiary	no	194	no	yes	cellular	29	aug	189	2	-1	8	unknown
36	technician	married	secondary	no	4,873	no	no	cellular	27	aug	239	5	-1	8	unknown
37	adminL	single	tertiary	no	2,837	yes	no	cellular	28	apr	134	1	383	2	failure
23	blue-collar	single	primary	no	-252	yes	no	unknown	23	mar	258	1	-1	8	unknown
31	services	married	secondary	no	132	no	no	cellular	7	jul	148	1	152	1	error
38	management	divorced	unknown	no	0	yes	no	cellular	18	nov	96	2	-1	8	unknown



Cleanse Data to Ensure Validity

Improve data quality

Who said data cleansing had to be tedious? With Datameer, you can quickly identify outliers, duplicates, inconsistent values and more.

Identify outliers

Use Flipside™ data profiles to gain visual feedback .

Handle duplicates

Use a formula for de-duplication.

Remove inconsistencies

Filter missing values, nulls, blanks and missed cardinalities.

Custom rules for cleansing

Use the formula builder for any advanced patterns in the dataset. It identifies edge cases or unique patterns in the data and whether they need to be corrected.



Blend a Variety of Data Formats

Use additional data or explore it in different ways. After completing the data preparation workflow, you can blend datasets or embed meaningful data elements for future analysis.

Join multiple datasets

- Join disparate sources without much data lineage
- Blend data easily with inner, outer, self and range JOINS. Multiple JOINS can be chained in one step (e.g. multi-lookups).

Blend a variety of data formats

- Easily blend structured with unstructured data
- Union feature helps you simply append datasets to each other without a cardinality requirement. Many multi-structured data sources do not have any obvious matching join keys and the formula builder can help perform fuzzy joins in such scenarios.



Enrich and Transform Data

Understand data relationships and the meaning of data. For the analyst, data conditioning and enrichment are an important part of the cycle.

Use sophisticated transformations

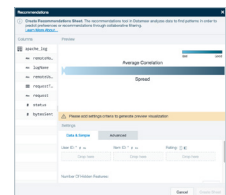
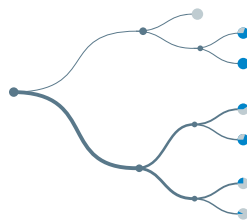
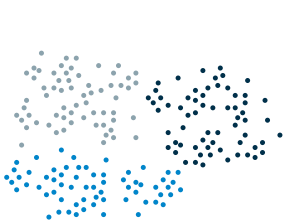
- Time saving
- Column splitting
- Column and row pivoting
- Statistical grouping
- Advanced text parsing
- Working with lists (concatenate fields to lists and expand to rows)
- Path construction
- Powerful if and comparison functions
- Date, time and text manipulation functions
- Enrich datasets inline with support for expressions

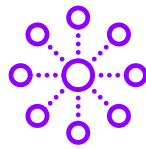
Analytic enrichment

Create valuable insights from data to complete the journey of data prep and data discovery. Datameer provides a rich array of capabilities to enrich datasets with analytics including:

- Path analysis
- Graph analytics
- Statistical functions
- Clustering
- Correlations
- Decision trees
- Text mining
- Sentiment analysis

These are all available via Datameer Smart Analytics™





Shape and Organize Datasets

Go beyond standard slicing-and-dicing of data

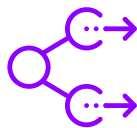
Find hidden patterns and trends in the data. Organize data during the data preparation phase in different ways than with typical dimensions and metrics used by standard analytics. The Datameer function API allows users to write functions for domain specific transformations.

Use built-in functions to organize and shape data

- Sessionization
- Custom binning
- Time windowing
- Advanced statistical grouping
- Function API to create domain-rich datasets

Drive analytical use cases

- Fraud
- Preventative maintenance
- Buying patterns
- Clickstream analysis
- Time-series analysis
- And more



Share and Collaborate Datasets

Browse through files

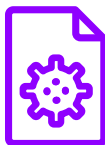
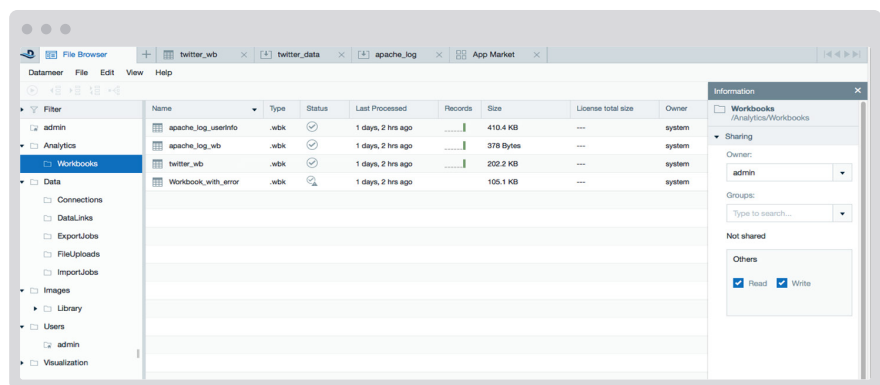
Organize and search for different data artifacts used across the enterprise

Organize files

- Allow access to various folders across the organization, depending on projects and teams working on it
- Manage access to these artifacts for files or entire folders

Reusable data artifacts

- Can be any data prep component such as:
- Connection to a data source
- Import job with raw data
- Workbook with data prep logic or a prepared dataset
- Chart or visualization published for sharing with any consumers of data
- Entire data prep workflow with chained workbooks



Operationalize Data Pipelines

Answering modern business questions isn't a one-time affair. As new data arrives, answers to the same question need to be continuously fed to business teams so they can perform timely, accurate actions.

Automate custom data preparation steps

Datameer gives you a full suite of capabilities to operationalize your data preparation pipelines to deliver results to downstream analysts and business teams. This includes:

- Sophisticated job scheduling and management
- Adaptable data retention and management policies
- A scalable, intelligent execution framework
- Full lineage, auditing and logging
- Complete security, including encryption and masking



Govern With an Open-Secure Model

For effective data-preparation pipelines and deeper analytic insights, organizations need a highly collaborative environment. Sharing both analytic datasets and content enables analysts to cast wide question nets while ensuring data and results are in line with corporate accuracy standards and compliance requirements.

Maintain compliance and consistency

Datameer maintains full lineage, auditing the definition of and any changes to artifacts and calculations, and supporting even the most rigid compliance processes.

Datameer also includes a deep suite of encryption and obfuscation features, allowing you to keep data such as Personally Identifiable Information (PII) safe and secure using SHA1 plugins for obfuscation. Migration tools are bundled from development to production.

Fine-grained, role-based security

Datameer ensures full control over the data and models. Datameer also keeps a full catalog of datasets, models and derived content, and enables sharing with virtual team members across the organization.

Irrespective of your analyst team's organization — centralized, virtual or distributed — Datameer has the complete suite of data preparation capabilities needed for cleansing, modeling, enrichment, collaboration, governance and operationalization. To learn more, please visit our website at www.datameer.com.

Datameer Enterprise

- User Management
- LDAP / AD Integration
- SAML / Single Sign-On
- Permissions and sharing
- Role-based security (with custom roles)
- Kerberos Integration
- Sheet dependency graph
- Audit & data volume logs
- Data retention policies
- Secure impersonation to HDFS, YARN, Hive (Sentry), HBase
- Basic Sentry integration
- Encrypted metadata (with key rotation)