redpoint™

DEEP DIVE

Customer Data with Redpoint Data Management™

Redpoint Data Management (RPDM) empowers organizations to ingest, cleanse, transform, and integrate data through a robust yet easy to use visual interface. From design to execution, data flows can be operationalized to enable business critical applications, fuel analytic teams and processes, and generate insights across the enterprise. All of this is done at unparalleled speed, scale and levels of productivity.

The solution provides a single point of control over all your data, connecting all types and sources of customer data – batch or streaming, structured or unstructured, 1st-, 2nd-, and 3rd-party – at a high speed and scale.

Further, RPDM enables business success through connected data and streamlined processes. It provides workflow and data flow; user controls for design, automation, and data management; and the detailed functionality needed to ingest, validate, cleanse, and merge customer information into consolidated and robust data views.

RPDM also helps to reduce operational costs while improving data management. It provides enterprise-level operational capabilities for handling sensitive customer information with appropriate compliance, security, and privacy while providing the performance, flexibility, and quality results needed to engage customers at scale in real time; either on premise or in the cloud.

This document details of the capabilities of Redpoint Data Management (RPDM) as it relates to customer data management within the rgOne platform. From this point forward, RPDM will be referred to as rgOne.

rgOne is a highly configurable platform that encompasses the following capabilities:

- Data Ingestion: Connections and ETL.
- Identity Processing: Match, merge, and integration.
- Data quality and enrichment: Validation, profiling, data hygiene, and geo-spatial analysis.
- Big Data Processing: Storage, access, and computation in Hadoop environments.
- Data Storage and Access: Persistent key management and updates.
- Process design: Macros, automations and notifications.
- Operational controls and reporting.
- Scalability, architecture, and usability.

Broad Function Data Management: Any Data, Any Source

rgOne offers an open garden approach by integrating all data sources, types, and formats, thanks to hundreds of standard, easily configured connectors. As a result, marketers and data scientists will achieve unprecedented speed, efficiency, and accuracy as they extract, transform, and aggregate any data. The rgOne platform goes beyond just bringing in data; it reconciles, cleanses, and validates customer details automatically, so marketers can focus on customer engagement.

A typical, automated data flow for ingesting data from multiple sources, normalizing and validating customer details, and creating customer "Golden Records" is shown in Figure 1. The interface provides an easy drag-and-drop model for defining data flows, and allows access to predefined flows, connections, and processing tools with direct configuration of controls – no developer coding or complex scripting is required.

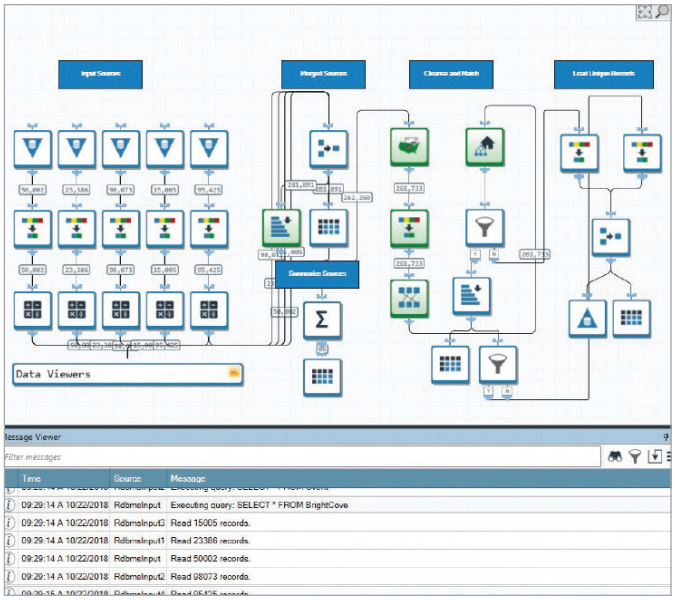

Figure 1: Redpoint Data Management Data Flows

Data Read/Write

Redpoint Data Management provides the ability to read and write single and multi-byte data with full Unicode support from a variety of sources:

- Basic read/write: access to flat and delimited text files, with support for single and multi-byte data with full Unicode support and EBCDIC, with or without EOL, and with any delimiter and surround character.
- Support for parsing structured files including XML and JSON.
- Traditional DBF files and proprietary DLD files, which provide a highly compressed file format for high throughput, check-point restart, and indexed query/append.
- Databases access: many database types are accessed using native drivers and most others using OLEDB or ODBC drivers. Supported databases include Microsoft SQL Server, Azure SQL Server, Azure SQL Data Warehouse, IBM Netezza, Teradata, Oracle, Sybase, Postgres/Greenplum, MySQL, DB2, AWS Redshift, AWS Aurora, Snowflake, Access/Excel, and others.
- EDI data formats: including X12, EDIFACT, Electronic Health Records and others.
- NoSQL / Document databases such as MongoDB and Cosmos DB.
- Hadoop databases and file types such as HBase, Hive, Avro, and Parquet (see Hadoop section below).
- Message queues including IBM MQ, Kafka, AWS SQS, and Azure Service Bus.
- Salesforce CRM.

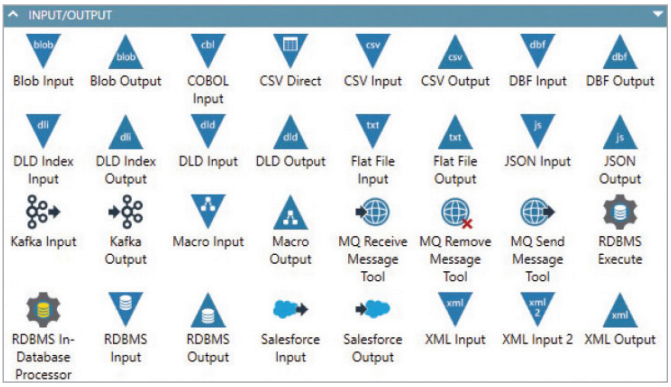

Figure 2: RPDM supports a wide range of the commonly used extract and load types

Transformation

rgOne includes an extensive set of data transformation capabilities. This set covers the full range of functions needed to transform and integrate data from any source for use in any target. The set of functions includes: :

- More than 40 date/time functions that can be used to transform or synchronize date/time attributes.
- More than 30 numeric and arithmetic functions.
- More than 60 string transformation functions for managing text, including managing unstructured block text, comment data, and "white-space" analysis.
- Extensive support for regular expressions (including regular expressions in string transformations) and for building/classifying tokens and pattern matching based on regular expressions.
- Logical processing functions including data filtering, contingent value processing, and data-based branching.
- Support for local and global variables that can be used in single data flow projects and across data flow projects.
- Conversion of data types and transliteration across Unicode code pages.
- Removal of non-printing characters (single and multi-byte).

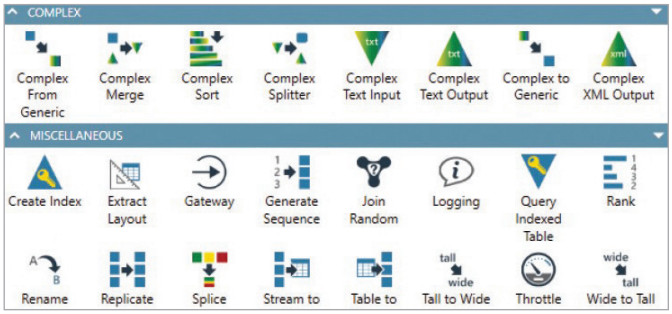

Figure 3: RPDM includes an extensive set of transformation capabilities

All of the transformation functions within rgOne are available in both automated expression building and in an expression language. This allows users to nest/stack functions as deeply as desired without sacrificing usability or performance.

rgOne also supports high performance sorting, joining and filtering (including single step file splitting and identification of unique records).

Web Services

Redpoint data management allows users to integrate the full range of ETL, data quality/hygiene and customer data integration functions with web-based data sources. RPDM web services capabilities allow for:

- Capturing session and cookie information to create secure multi-task (complex) service processes.
- Encoding session data in headers, URL parameters, XML, or JSON.
- Processing HTTP posts and gets, allowing for both push and pull exchanges.
- Full support for OAuth2.
- Defining execution and retry strategy.

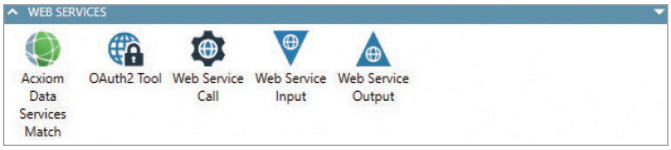

Figure 4: RPDM Web Services

The platform allows service calls within any project or data flow. Additionally, rgOne jobs can be deployed as SOAP-based web services with advanced web service load balancing capabilities.

Solve Customer Identity Challenges

Identity resolution helps marketers gain a better understanding of customers by building an accurate and usable representation of them (anonymous or known). This insight allows organizations to predict and shape customers' behavior, improve retention, increase sales, reduce friction in the customer experience, and maximize customer profitability. In addition, identity resolution through rgOne supports digital interactions that are primarily real time. Accurate recognition is the single most important step in providing a tailored and relevant experience for the customer in their moment of truth.

The customer data management capabilities within the rgOne platform:

- Provides the most powerful set of advanced data quality, identity resolution, matching, and master data management capabilities available on the market today. Using its advanced probabilistic and deterministic matching (with more than 375 built-in functions), marketers can easily identify, match, link, and de-duplicate files and standardize and correct data for more than 200 countries.
- Deciphers and relates individuals, households, cookies, IP addresses, IoT smart devices, and more to form a clear and complete picture of every customer. Using rgOne, marketers can create and enhance a customer "Golden Record" – a singular, accurate, and continuously updated view of each customer that is maintained with a persistent key, in minutes, seconds, or on demand.
- Includes a full range of data quality and cleansing capabilities. These are native to the product. Users need not separately license and integrate a data quality tool to perform normalization, validation, and probabilistic matching along with the ETL functions described above.

The example Identity Management data flow below includes normalization, validation, and probabilistic matching for name, address, email, and phone to produce accurate matches across a range of sources and interactions.

Identity Matching Approach

rgOne handles matching with a broad set of capabilities:

- **Out-of-the-Box Matching Tools** for matching at the household, person, address, email, URL, and account levels without coding or developing match algorithms.
- **Probabilistic and Deterministic Matching** that combines statistically based matching with customer data integration business rules to create highly accurate match results.
- **Completeness and Accuracy** with tools and reference data to cleanse and standardize names, addresses, phone numbers, and email addresses in North America and 240 other territories.
- **Alias Resolution** for name-variance mapping (e.g. Lewis, Louis, Lew, and Lou or Street, St, and Str).

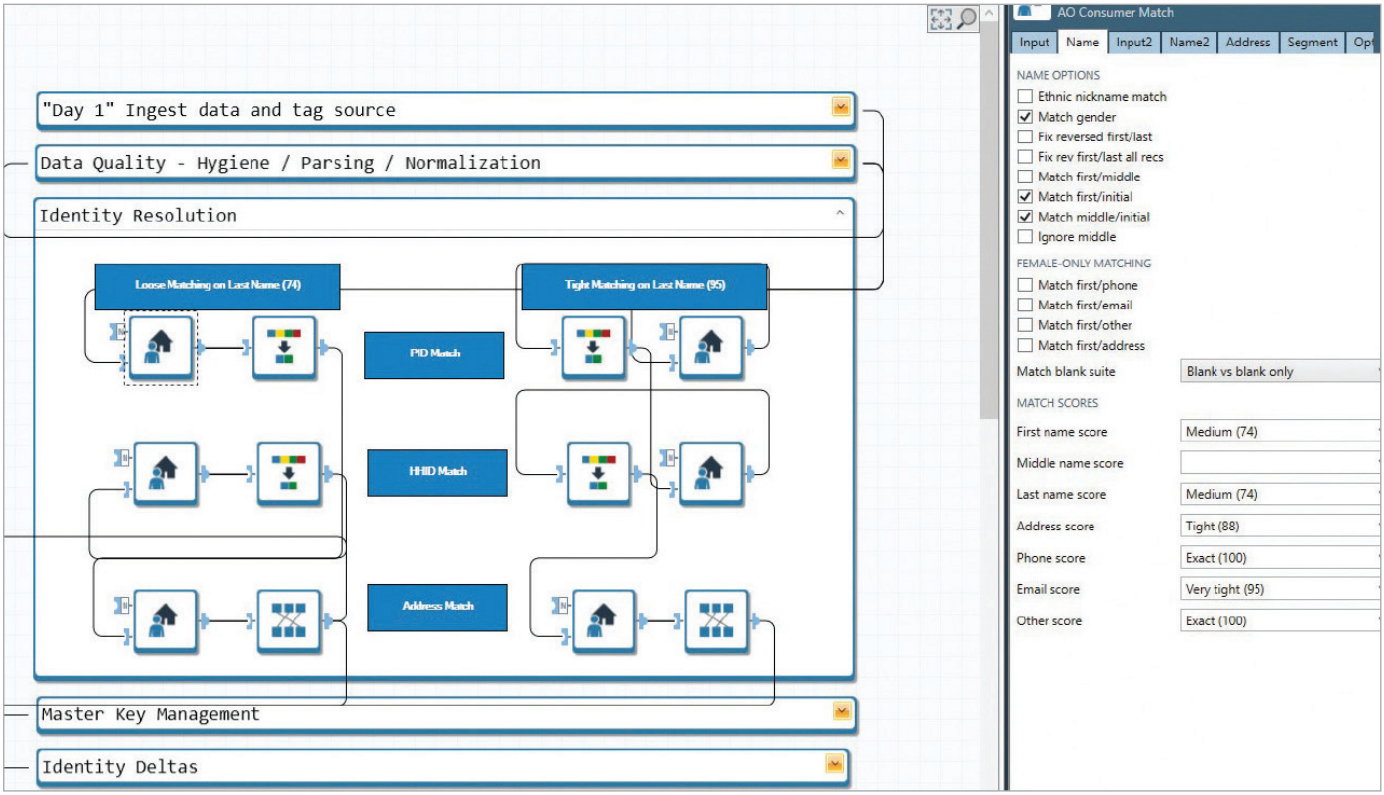

Figure 5: Matching data flow handles validation, normalization and deterministic and probabilistic matching

- **Iterative Cycles** mirror person/business natural changes and handle new data becoming available over time. This includes support for adding records or breaking apart existing person, household, or business groups.
- **Performance and Scalability** is a key requirement for the deep and iterative matching. Redpoint has been benchmarked to process high volumes of data 500 percent to 1,900 percent faster than leading alternatives.
- **Flexibility,** with multiple matching levels, from very tight matches with close to zero over-matching rates for compliance uses to simultaneous looser matching for marketing or fraud detection.

Address Standardization

rgOne includes tools to parse, standardize, correct, complete, and certify addresses from around the world. The following options are available:

- US address standardization that is certified by the United States Postal Service, with certifications in Coding Accuracy Support System (CASS), Delivery Point Validation (DPV), Locatable Address Conversion System (LACS), and SuiteLink.
- Appending of Extended Line of Travel (ELOT) data, also known as carrier walk sequence.
- Geocoding using US Census TIGER Geospatial directories.
- Canadian address standardization that is certified by Canada Post with a certification in the Software Evaluation and Recognition Program (SERP).
- International addressing standardization based on relevant local laws/regulations. .

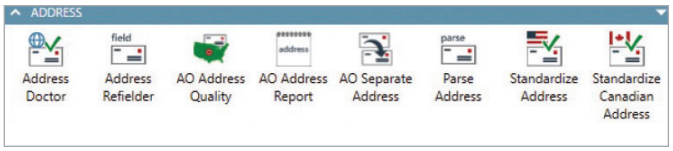

Figure 6: RPDM address analysis and transformation

Name Parsing

rgOne includes a full range of name parsing capabilities, including::

- Splitting full names into component parts (including name fields that have been overloaded with multiple persons such and John and Jane Smith).
- Standardization of name prefixes and salutations (e.g., Dr., Mr., Mrs., the Honorable).
- Probabilistic gender assignment (based on census name/gender assignment).
- Business name processing, including keyword and alternative name identification.
- Adding custom names files to account for cultural name differences and known aliases.

rgOne name parsing can be extended to include pattern matching for roles (e.g., guardian, beneficiary) and "extended data" (e.g., notes, deceased indicators, record references) that are sometimes overloaded into name fields. Name parsing can also be extended to account for name patterns and distribution for international locations. This allows users to customize the name processing to account for regional and ethnic variances and user specific data requirements.

Matching

rgOne uses a mix of probabilistic and deterministic techniques to easily identify, match, link, and de-duplicate files. The platform includes out-of-the-box consumer (B2C) and business (B2B) matching with highly configurable tools that combine flexibility and best practices into a single package.

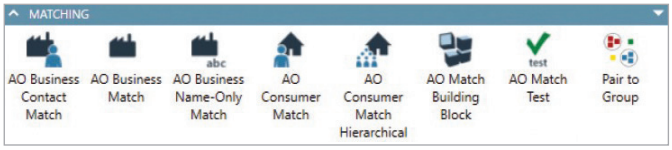

Figure 7: Redpoint matching automates keys, segments and reports

rgOne supports simultaneous matching across multiple match levels (tightness or looseness of matching) and multiple match types (e.g. name/address, name/phone number, name/account number). Simultaneous matching allows different match criteria and confidence factors to be specified for each pass/comparison, allowing for matches to layer without over-matching or having to write complex integration rules.

Duplication Handling Duplication Handling

rgOne can roll up duplicate records to master records based on data quality/completeness, data frequency, or user specified rules. These capabilities support cross-matching data from multiple sources with common entities (e.g., customers, suppliers) that do not have shared keys. Other uses include grouping account-level data at the person/ household/ group or business level to create "rollups" or common keys; creating unique records in files (de- duplication); and rolling up data across multiple records based on user specifications.

Data Classification and Standardization

rgOne provides out-of-the-box standardization for other field types, including:

- Phone (North American and international, area-code validity, area-code state, letter-number conversion, etc.).
- Social Security Number formatting and validation (sequence validity, area and group validity); assignment cannot be validated.
- Email format.
- Social media handle format (and validation via external web service as provided by the various social media channels).
- URL format and URL encoding.

Data Profiling

rgOne offers robust data profiling capabilities, including:

- Source validation tests input data against one or more specified formats (data layout, type, values, etc.) for acceptance testing and flow control; matching is based on the percentage of records that meet defined criteria and files can be accepted or rejected in whole or in part.
- Profiling input data sources for counts (by value and null), uniqueness (by field or record), data characteristics (min/max, longest/shortest, etc.), and compliance with pattern templates or masks (data type and data pattern).
- Table and column compliance with user-defined sets of rules (e.g., pattern templates and masks), range (by value or other column), specific values (by value or another column), etc.
- Key constraints (e.g., uniqueness within a table or value/key constraints across tables), including support for multi-column or compound keys.

Figure 8: RPDM Profiling

Geospatial Analysis

rgOne provides many spatial analyses and transformations, including:

- Shapefile import/export.
- MID/MIF (MapInfo) import/export.
- Point-in-polygon (market penetration, political analysis, heat- map generation) analysis.
- Spatial join (territory overlap, productivity, flood plains). Find nearest neighbors (store location, service areas).
- Transform polygons to/from raw data lists (custom region manipulation logic).
- Spatial object operations such as inflate, cut, intersect, union, convex hull.
- Spatial summarize (aggregation of spatial objects).
- Grid cell (mapping of areas onto uniform grids for analysis).

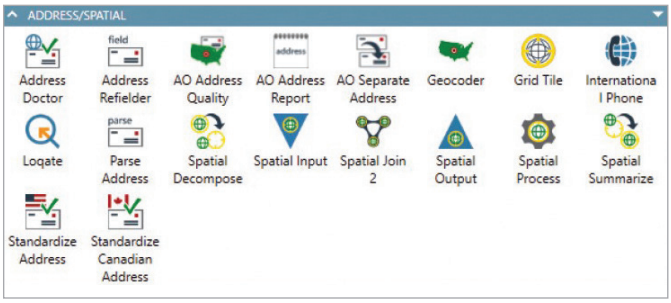

Figure 9: RPDM geospatial analysis

Persistent Key Management

rgOne includes tools that allow users to easily manage persistent keys – creating, updating, merging, and splitting as data changes over time.

The platform is designed to deal with the complexities of constant key persistence (new/additional records, record splits/ reassignment, historic key management, etc.) and householding (new/additional persons, household break-apart, maintaining head-of-house across match-groups, etc.). Persistent keys can also be assigned at the group level and source level, allowing customers to track performance at these levels.

Important capabilities and features relating to key management include:

- Handling new/updated data sources (new/additional data being added).
- Support for natural changes in household structure, including merging groups (such as marriage or cohabitation), new group members (such as births and adoptions), and splits (such as divorce or deaths).
- Managing persistent keys at the person, household, group, address, and business levels.
- Enforcing unique key constraints.
- Compiling master records from multiple sources based on completeness, frequency, and validation rules.

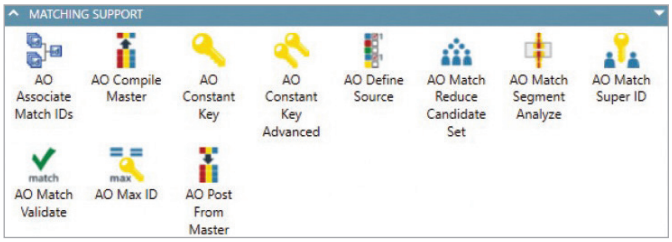

Figure 10: Persistent Key Support and Key Matching

Process Automation and Notifications

rgOne allows users to automate both the processing and monitoring of data management jobs and projects, resulting in highly streamlined and effective operations.

Automations

Within rgOne, users can build complex workflows (automations) that combine the various data-flow projects – ETL, data quality, customer data integration, and key management functions.

Other automation capabilities include::

- Scheduling jobs based on calendar, presence of files, or changes to a database.
- Creating webs of interdependent jobs that are managed from a single control point.
- Executing external programs.
- Waiting for user review and acceptance.
- Transferring files via FTP.
- Determining source format and validity.
- Notifying users and operations of job outcomes.

Automations are used to:

- Wait for files to appear in an upload directory, and automatically process them when they do.
- Break a long transformation process into smaller steps, so a single failure does not require a complete restart.
- Include other tools (e.g., compression, encryption, specialized file transfer tools, client proprietary data processes) in Redpoint processes.
- Integrate other tools through standard in-process data transfer (e.g., XML, data pipes and APIs, structured data files).
- Let the user enter parameters to control execution at startup such as filenames, filter options, or report options.
- Suspend execution midway and let the user review and edit data directly, for data stewardship approval, special-case corrections, and fuzzy-matching review.
- Loop over all files in a directory and run the same set of steps on each file/data feed.
- Validate a file against many possible formats, and when a format matches, take appropriate action.
- Remain "live" continuously, waiting for external factors to trigger execution (such as the appearance of a file, SQL query results, a calendar event, or an FTP data transfer).

Notifications

Automations use success/failure logic for error handling and can send notifications to operators via email or SMS. Automations can loop over files and sets of values, and they have broad support for variables and parameters. Finally, automations automatically support checkpoint and restart, so that if any steps fail in a multi- step procedure, only the failed steps need be restarted.

Operational Controls and Reporting

The rgOne platforms provides a broad range of functions to enable users and operators to manage development and execution.

Operational Controls

Redpoint users are empowered with many tools for maintaining operational control of data management functions:

- **Version control** – All data flows and automations can be managed through rgOne's integrated version control. Version control tracks user, update time, comments, and notes. Version control includes full rollback capabilities. Alternatively, rgOne jobs can be stored externally and managed by any tool capable of handling XML files.
- **Multi-server job distributio**n – Redpoint software can run in a centrally controlled environment with numerous processing nodes executing jobs.
- **Execution monitoring and logging** – All jobs can be monitored by administrators and operators to determine both overall progress and the actions of specific job steps. Developers and operators can drill down to the specific working load (e.g., CPU, memory allocation, disk use) of each component in a job for monitoring and optimization.
- **Command-line execution and web-service interfaces** – Allow rgOne to be integrated into other operational control systems.

Reporting

Redpoint users maintain oversight of data management functions with summarization and reporting capabilities, including:

- A project execution log that tracks records read, processed by various steps, and written to various outputs (including database load counts, non-processed record counts, etc.).
- Aggregate and summary functions (e.g., number of records by source or by criteria, min/max values) that can either be written to reports or out to databases/data files or merged back into the data processing stream.
- Creation of structured and cross-tab reports based on definable layouts and configurations, with support for concatenated or nested reports to create complex summaries of data at all phases from initial input to final output.
- Generation of reports and graphics based on Redpoint processing without having to reprocess data in BI/reporting tools, or creation of XML feeds that can be directly used by third party report tools.
- Specialty reports for data hygiene (e.g., address quality and address component reports) and customer data integration (e.g., match quality, match by source, duplication).

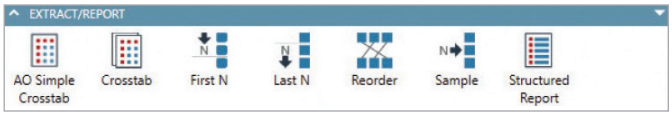

Figure 11: Reporting tools in RPDM

Support for Big Data Processing

rgOne performs data ingestion and integration across all sources and types of data – batch or streaming, internal or external, structured or unstructured. rgOne includes native support for Big Data infrastructure, operating in Hadoop distributions from Cloudera, HortonWorks, Amazon, Microsoft, and MapR. rgOne is certified across cloud infrastructures, Hadoop distributions, and leading databases.

Hadoop

rgOne works with Hadoop through three methods:

1. **Reading from and writing to the HDFS file system:** This allows users to ingest HDFS-based data into rgOne's sophisticated data transformation, cleansing, and matching algorithms. Redpoint 's HDFS file-system module lets users easily combine data stored inside and outside of HDFS, and enables high-performance extraction, transformation, loading, and format conversion between HDFS and external sources.

2. **Reading from and Writing to Hadoop Databases:** Reading from and Writing to Hadoop Databases: rgOne provides direct connections for a broad set of Hadoop and Cloud database types. These include:

   - Google Storage
   - Avro
   - Parquet
   - HBase
   - Hive
   - Google BigQuery
   - AWS S3
   - AWS Redshift
   - AWS EMR
   - Azure HDInsight

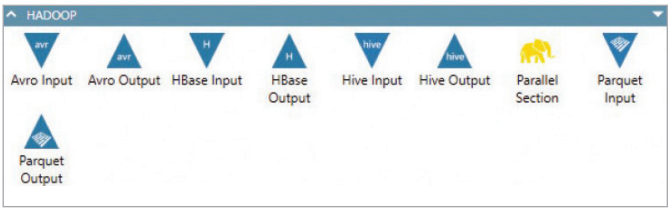

Figure 12: Redpoint handles Hadoop data in many formats

3. **Processing inside Hadoop as YARN tasks:** With this approach, Redpoint cooperates with Hadoop's YARN management system to create and run a series of parallel tasks to carry out its work. Redpoint projects are distributed to YARN worker nodes based on considerations such as tasks to be performed, record-splitting specifications, implicit parallelism, and HDFS data locality. rgOne's parallel data flow architecture is further extended on Hadoop to reap the benefits of task parallelism and data locality. Redpoint's inside-Hadoop architecture massively scales complex application-driven processes, such as data cleansing and hygiene, mapping and GIS functions, householding and advanced matching.

Scalability and Architecture

rgOne is architected for high performance, as well as for platform flexibility.

Scalability

rgOne is highly scalable and has been proven on very large volumes – i.e., billions of records. The platform is vertically scalable, taking advantage of multi-processor/multi-thread parallelism and is horizontally scalable in both OS mediated and ad hoc (application managed) clusters. As a result, users can create high performing data integration jobs that can combine hundreds of data sources and billions of rows of data in a single project, instead of having to create and maintain multiple separate jobs to accomplish the same task. This capability reduces complexity, so integration efforts are easier to maintain.

Platforms

The product is available in both Windows Server and Linux versions, and is one of the few applications that can work in blended Windows/ Linux settings.

Pure YARN Application for Hadoop

All of rgOne data quality and data integration capabilities are available for use with data stored in Hadoop. With rgOne, the same data analysts and database administrators already working with traditional databases can now work just as easily with data stored in Hadoop. No new coding skills are required. No MapReduce, no Hive, no Pig – thanks to rgOne's graphical user interface and pure YARN architecture.

With rgOne:

- All data quality and data integration functions can be performed in the Hadoop cluster.
- Absolutely no MapReduce is involved, and no MapReduce skills are needed.
- Data quality and integration processes execute as efficiently – and in many cases more efficiently – with Redpoint as with MapReduce-based solutions.
- Data doesn't need to be moved out of Hadoop for processing, analytics, reporting, or other action.
- No software needs to be installed in the cluster itself, and Redpoint respects YARN's task prioritization rather than competing for computing resources in the cluster.
- Users can manage data in both traditional and Hadoop data repositories with a single product, even bringing together data from separate environments or migrate data from one to the other.

Usability

rgOne offers the right level of power and simplicity to make life easy for any organization's data management teams. A visually oriented workbench enables developers to rapidly build and configure complex processes without writing code. Easy-to-use interfaces reduce the learning curve and accelerate deployment.

Summary

Organizations have lofty data quality and customer data integration expectations to meet. Relevance, personalization, timeliness – the list goes on. rgOne supports data-driven initiatives across the enterprise and is the core of Redpoint's Customer Data Platform, providing an accurate, always-updating Golden Record to drive customer analytics and personalized, omnichannel engagement.

With a solid foundation for customer data, engagement becomes easier to execute, less expensive, and more effective – and drives better business outcomes.

Redpoint's Data Management capabilities provides a connected data foundation for the rgOne platform, which also encompasses Automated Machine Learning and Intelligent Orchestration. Together, these capabilities are designed to help companies deliver highly relevant, personalized and omnichannel experiences to their customers and prospects across any point of interaction.

About Redpoint Global

With Redpoint's software platform, innovative companies are transforming their customer experiences across the enterprise and driving higher revenue. Redpoint's solutions provide a remarkably unified, single point of control where all customer data is connected and every customer touchpoint intelligently orchestrated. Delivering more engaging customer experiences, highly personalized moments, relevant next-best actions and tangible ROI—this is how leading marketers lead markets. To learn more, visit **redpointglobal.com**.

US Headquarters: Wellesley, MA | Tel: +1 781 725 0250 **EMEA Headquarters: London, UK | Tel: +44 (0)20 3948 8170**
www.redpointglobal.com

DD-DMUS0820-01