



# Rocket<sup>®</sup> Enterprise Data Intelligence Solutions

---

Data Lineage – Diving Deeper



---

# Contents

**03** The need

**04** What makes Rocket Software's solutions unique?

**05** Extensive data source coverage

**06** Data lineage between data stores

**07** Data lineage in application code

**09** Simple configuration and customization of metadata collection

**11** A comprehensive view of lineage-related information

**12** Conclusion



---

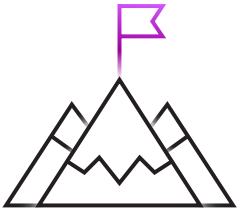
# The need

Knowing where data came from, how it moves through systems, and how it changes, is the most critical and most difficult task in any data management project. If that process—known as “tracing data lineage”—is not well executed, it is impossible to have confidence in the authenticity of the data supporting business operations and management. This paper describes how we deliver the fastest and most comprehensive data lineage mapping and visualization capability to govern and guide data projects. Data lineage seems simple: “I’m looking at a number on a report. Where did it come from?” However, since business users first identified the need, both they and technical users have recognized requirements that are more challenging:

- Tracing movement of data in Data Warehouse, Master Data, and Big Data environments
- Quickly identifying the source and quality of data
- Knowing how data is calculated and aggregated as it moves from source systems to target reports and inquiries
- Understanding data transformations, and tracing lineage implemented in ETL tools, PLSQL and other SQL variants, Java and .Net, and data lake environments to provide the broad understanding that is vital for many compliance projects
- Navigating backward or forwards in data transformation processes to understand data lineage and change impact
- Supporting understanding of lineage as data travels across multiple lines of business
- Providing a business context to lineage by linking it to business assets, data models, reference data, and data quality information
- Managing discrepancies in lineage to resolution with issue management and tracking for data stewards

---

## The issues of data lineage



### The challenge is great!

The number and diversity of data sources, the number of data items, and the complexity of applications makes Data Lineage analysis a daunting challenge.



### Data lineage is vital!

Data lineage is the key to understanding the relationships between data items across the enterprise or in individual lines of business.

The number and variety of information sources makes it difficult to meet this need. It is not always easy to locate all those sources. Even when you have found them, the problem can be overwhelming—many businesses have thousands, or even tens of thousands, of unique data elements. Refining the challenge and understanding data lineage at a macro level between data stores, and at a micro level within custom applications, adds yet another layer of complexity.

Lastly, data lineage is a valuable tool for understanding the information ecosystem—but it needs to be part of a suite of solutions that provide information about all the aspects of data involved in lineage.

## What makes Rocket Software's solutions unique?

### Our comprehensive data lineage solution delivers:

- Coverage of data sources across a wide range of applications, databases, data lakes, and enterprise modeling tools
- Macro-level analysis of data lineage between data stores
- Micro-level analysis of data lineage in application code
- Simple configuration and customization of metadata collection
- Simple scheduling and management of metadata collection processes
- Real-time monitoring of the scanning process
- Use-case appropriate visualization of data lineage
- Integration with Issue Management capability to manage the resolution of lineage problems, and track the history of issue resolution
- Flexible deployment in preferred enterprise architectures

Every element of the Data Lineage solution is required to meet the needs for complete data lineage analysis.

## Extensive data source coverage

**Out-of-the-box, the tool-agnostic Data Lineage solution consumes metadata from a broad range of sources. Source categories include:**

Business Process, data and application models	Custom and enterprise applications including SAP, JD. Edwards, and Salesforce
ETL, Data Warehouse, and Business Intelligence tools	Programming languages
Relational and non-relation Database Management Systems	Source code management tools
Data Quality Tools	Scheduling and scripting solutions
Spreadsheets and XML files	Data Lakes

This range delivers unique flexibility as the solution supports metadata captured and combined across mainframe, distributed, and cloud platforms.

The solutions include a toolkit that supports the creation of custom metadata interfaces for your enterprise-specific sources, allowing for the completion of a comprehensive data lineage knowledge base.

# Data lineage between data stores

A first key to mapping data lineage in an information system is to detect movement of information from one data store to another. A second is to understand how source data elements are changed and merged to create target elements.

To support this detection, we build a knowledge base by scanning metadata, reconciling that metadata into an integrated information model that supports full forwards and backward navigation. A transformation process uses a mapping schema to control aggregation of “raw” metadata from the source subject area into rationalized items used by the data lineage analysis.

This approach allows a combination of information about data stores, ETL technologies—including transformations “hard coded” in SQL—data warehouses and Business Intelligence tools to provide a complete and clear view of data lineage. As the process builds the data lineage knowledge base, the Metadata Repository Browser provides easy-to-use reporting functions that enable you to investigate and present metadata interdependencies.

The Data Lineage Solution provides wide ranging enterprise-wide lineage for classic “Data Warehouse”—style environments, showing flow from source systems through ETL tools to the Data Warehouse and BI tools, and for “Big Data” environments as external information flows into the data lake and out to analytic tools.

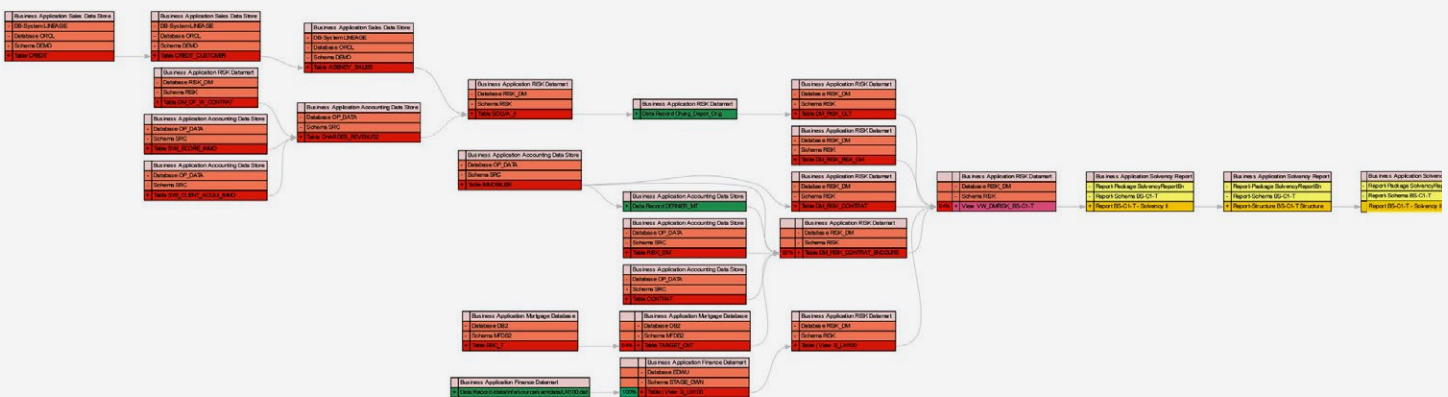


Fig 1: Decision support system data lineage

# Data lineage in application code

Auditors in compliance projects are demanding a deeper view of lineage. Application teams are redesigning systems to use web services to support business process flexibility. These considerations drive a requirement for understanding data lineage in custom applications.

The challenge in mapping data lineage in applications code is similar to the data store problem—but with another dimension. It is necessary to understand the code to understand and analyze data items and transformations. Rocket® Enterprise Data Intelligence solution contains a facility—the Common Analytic Engine (CAE)—that assembles the knowledge base to support data lineage analysis in application code.

The CAE uses a similar process to that used by compilers to understand how to execute programs. First, the code is broken down into constituent parts. Next, the process builds a model of the various elements—the parts of statements that constitute the program. This “static code analysis” process builds a model of program code and the sequence of code execution.

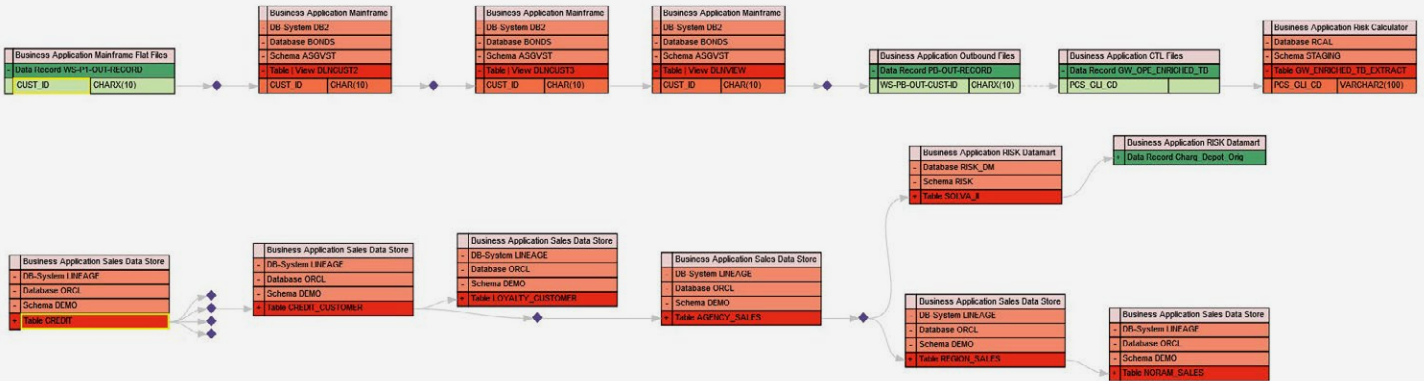


Fig 2: Application code lineage

**Now preparing for data lineage varies from the compiler process—instead of optimizing the model and translating the code into an executable form, the process builds the information required to support lineage queries:**



Information about lineage “steps”—source and target data items.



Information about transformations, the ways in which source items are combined and modified.



Analyzing program code adds levels of complexity to the data lineage challenge. A program component may use the same name for different data items. A data item may appear in many different paths in a program but every path will not be part of a given data lineage. The CAE accommodates these complexities.

Because of the numerous paths through even small applications, it is not possible to pre-calculate data lineage for every item that the application references. The CAE calculates lineage on demand, using the pre-built knowledge base.

Object-oriented (OO) programming languages are additionally challenging as they add programming constructs that are substantially different from those provided by “traditional” languages. Capabilities such as polymorphism, inheritance, and encapsulation make data lineage analysis more complex. We have extended CAE concepts to accommodate the challenge. In particular, the CAE considers the calling context when analyzing the target of a function call, keeping track of method invocation. “Instance Awareness” relates impacted instance variables to the instances that affected these data items.

The Data Lineage Solution has many ways of representing a clear view of lineage. In this instance, a complex lineage network involving Java, Enterprise Java Beans, WSDL, and COBOL is shown “Top Down,” allowing in-depth investigation of data lineage.



## Simple configuration and customization of metadata collection

The Metadata Repository Scanner Configuration Utility allows the creation of command and parameter files reusing shared parameters to simplify the implementation of metadata from a wide range of sources. Pre-built templates define the content of the scanner command and parameter files as well as required parameters.

### The utility supports:

- Creating and managing configurations
- Creating and managing templates for configurations
- Browsing logging information

## Simple scheduling and management of metadata collection processes

The Metadata Repository Process Monitor documents, archives, and validates import and update of metadata. The processes execute in a “batch” mode and send status information to the Monitor.

The Monitor supports the design of processes based on predefined templates and provides a view of the state of instances of processes as they execute and complete.

Process Monitor records the start time, duration, and state of tasks and individual task steps. It logs warnings and errors and gives a centralized view of the currency of the repository. A history view shows activity through time as well as warning of anomalies.

## Summary-to-detail views of data lineage

One of the most significant capabilities provided by the Data Lineage Solution is “out-of-the-box” flexibility in visualizing calculated lineage.

The default visualization provided by the Metadata Repository provides lineage at a “table-to-report” level. It answers the question “what data stores contribute to a given report?”

The summarized visualization provided by the Table Level view is ideal for clarifying system flow. There is often also a requirement to understand the derivation of a particular data element—either to investigate apparent data issues or to provide verification for compliance purposes. Metadata Repository provides a “detail” view giving a structured drill down starting with a multi-part key identifying metadata items.

The detailed data lineage view provides a data field and column level view of lineage. It displays transformations as empty diamonds to show simple data moves and filled diamonds to indicate the presence of underlying transformation logic.

Seeing lineage in the metadata management environment is an excellent way to see data flow in applications and decision support systems. Sometimes, however, when it would be useful to export data lineage information to other applications. Metadata Repository provides a simple and powerful approach through the export of graphics in SVG or PNG formats, for access through a web browser or incorporation in applications.

To provide detailed reporting for audit purposes and support further analysis of lineage, Metadata Repository also provides an export to spreadsheets.

Exported information provides details of sources and targets at each level through a lineage, helping understanding of traceability in third-party applications.



Fig 3: Table level data lineage

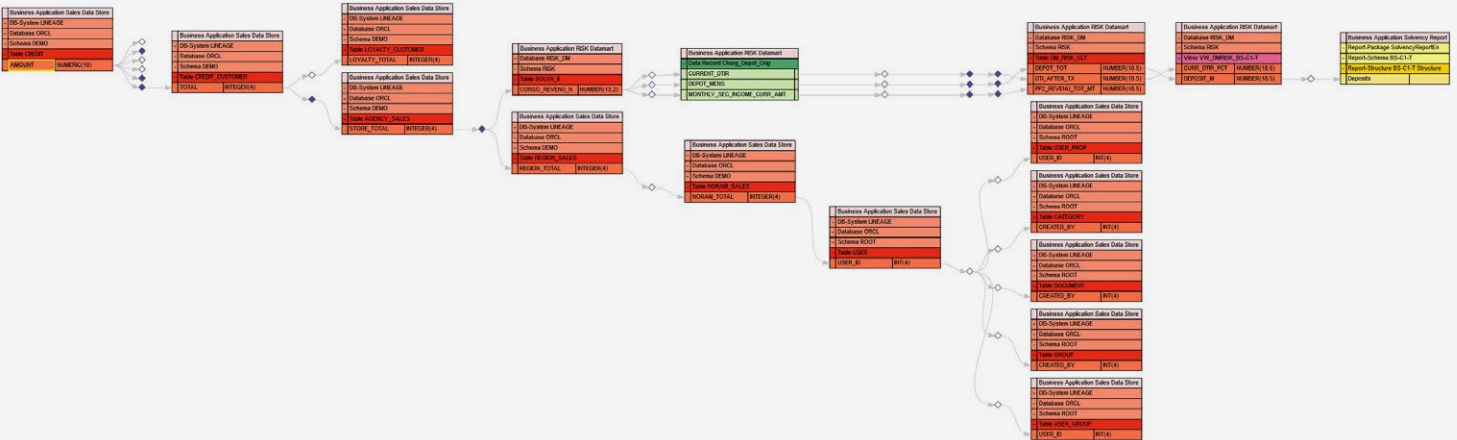


Fig 4: Detailed data lineage view

	A	B	C	D	E	F	G	H	I
	DB-System - Lvl 0	Database - Lvl 0	Schema - Lvl 0	Table   View - Lvl 0	Column - Lvl 0	ETL-System - Lvl 1	ETL-Task - Lvl 1	Transformation - Lvl 1	Transformation-Map - Lvl 1
1	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT				
2	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.oracle.tutorial.jdbc.CoffeesFrame.java	CoffeesFrame.java	CoffeesFrame.java
3	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.oracle.tutorial.jdbc.CoffeesTable.java	CoffeesTable.java	CoffeesTable.java
4	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.oracle.tutorial.jdbc.CoffeesTable.java	CoffeesTable.java	CoffeesTable.java
5	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.oracle.tutorial.jdbc.StoredProcedureJavaDBSample.java	StoredProcureJavaDBSample.java	StoredProcureJavaDBSample.java
6	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.oracle.tutorial.jdbc.WebRowSetSample.java	WebRowSetSample.java	WebRowSetSample.java
7	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	RAISE_PRICE	RAISE_PRICE	RAISE_PRICE
8	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	RAISE_PRICE	RAISE_PRICE	RAISE_PRICE
9	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
10	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
11	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
12	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
13	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
14	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
15	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
16	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
17	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
18	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
19	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
20	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
21	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java
22	LINEAGE	ORCL	DEMO	CREDIT	AMOUNT	LINEAGE	com.asg.tutorial.jdbc.CreditCustomer.java	CreditCustomer.java	CreditCustomer.java

Fig 5: Exporting data lineage information

## A comprehensive view of lineage-related information

Our metadata solutions tie together information about data lineage with business term definitions, reference data, and data governance information to show a complete view of a data item.

Figure 6 shows an entity “Permission Set” defined in an entity-relationship model, and its downstream impact on a Risk Datamart. Because of the way our metadata management environment integrates data lineage, the solution also provides much other useful information. Exploiting the ability to pull metadata from modeling tools, we can see that Permission Set is part of the “Salesforce” model. We can see the lineage that runs from the model to the Risk Datamart. Following the connection to our Business Glossary, Rocket-metaGlossary, shows the business rule associated with the entity, and the Data Steward responsible for the term. The solution could also provide links to Reference Data and Data Quality information, providing a complete perspective on the entity to support problem resolution or Governance and Compliance requirements.

## Flexible deployment in preferred enterprise

**Our Metadata Management solutions offer a broad range of deployment options.**

- Linux, Unix, Windows, and z/OS are platforms for the Metadata Repository Server
- The Metadata Repository application can use a choice of RDBMS platforms
- A cloud/SaaS deployment is available for selected elements of the solution

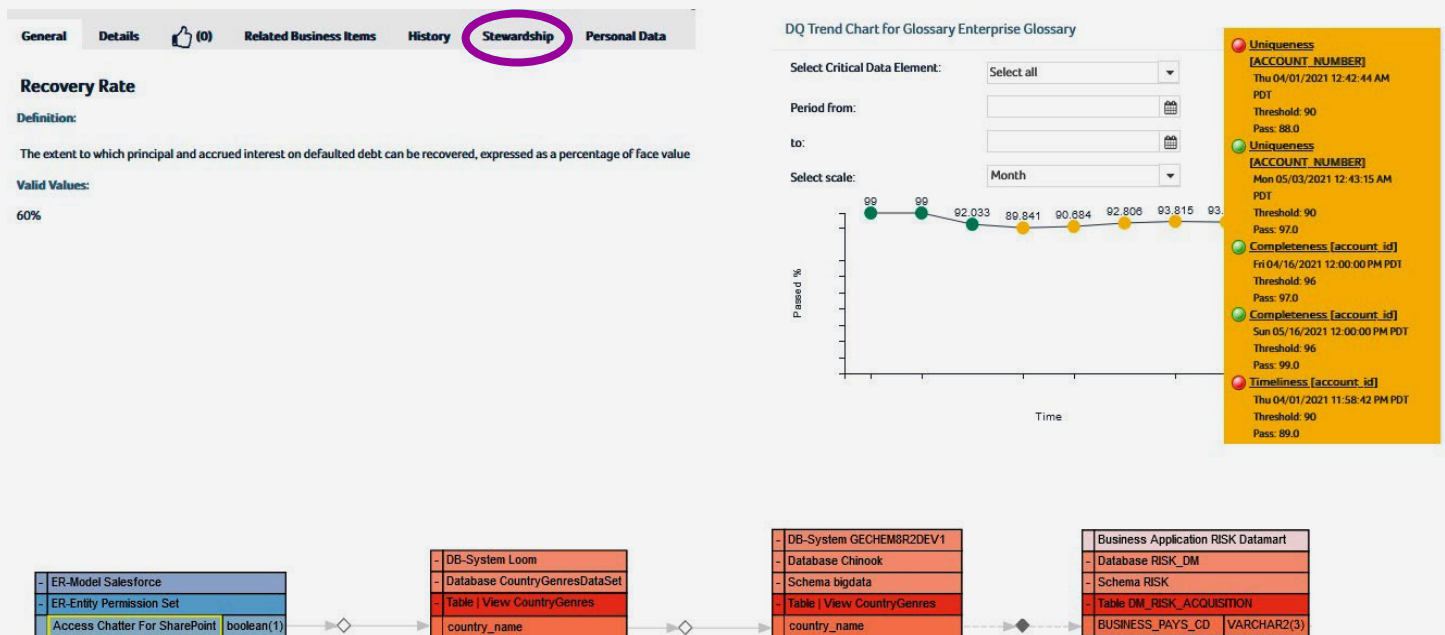


Fig 6: Data lineage in Enterprise Information Management

---

# Conclusion

Rocket Enterprise Data Intelligence Data Lineage solution delivers a faster, more accurate view of how your business is really operating. With it, you can accelerate, simplify, and improve the accuracy of responses to requests for information from auditors about how numbers are calculated, moved, and changed. You can understand the data required to build new package offerings for customers. You can speed the gathering of data to support decisions about new marketing campaigns. Use the Data Lineage Solution to see clear paths through complex networks of applications and data.



# About Rocket Software

Rocket Software partners with the largest Fortune 1000 organizations to solve their most complex IT challenges across Applications, Data and Infrastructure. Rocket Software brings customers from where they are in their modernization journey to where they want to be by architecting innovative solutions that deliver next-generation experiences. Over 10 million global IT and business professionals trust Rocket Software to deliver solutions that improve responsiveness to change and optimize workloads. Rocket Software enables organizations to modernize in place with a hybrid cloud strategy to protect investment, decrease risk and reduce time to value. Rocket Software is a privately held U.S. corporation headquartered in the Boston area with centers of excellence strategically located throughout North America, Europe, Asia and Australia. Rocket Software is a portfolio company of Bain Capital Private Equity. Follow Rocket Software on [LinkedIn](#) and [Twitter](#).

**The future won't wait—modernize today.**

Visit [RocketSoftware.com](https://RocketSoftware.com) >

**Book a demo**



© Rocket Software, Inc. or its affiliates 1990–2022. All rights reserved. Rocket and the Rocket Software logos are registered trademarks of Rocket Software, Inc. Other product and service names might be trademarks of Rocket Software or its affiliates.

MAR-5932\_Brochure\_DataLineage\_V2

