



Data Lineage

The Path to Understanding Your Applications



90% of the data in the world today has been created in the last two years alone, at a rate of 2.5 quintillion bytes of data created per day¹.

Corporations manage a staggering volume of data that continues to grow at a rapid rate. Strong consumer demand for personalization through smartphone applications, social media, and IoT drives it. Companies are more than willing to meet the growing demand, destroying barriers with cheaper storage and better and faster software to parse all the data that comes with it, ensuring that data growth will continue for years to come.

With bigger and more complex data portfolios, companies are increasingly encountering the need to trace data along its complete journey - from input to output - for development, keeping track of business rules and security.

The need for security might be the biggest challenge yet. According to a recent Cisco survey, most IT and security pros say they aren't able to afford the minimum amount of security they need to properly protect their data. The survey also finds that most of them have been hit with major security compromises, and realize that they need to do more.

Companies seeking to gain control of all of this new data are finding they need specific expertise along with a dizzying array of tools and techniques to get the job done to keep track of it, derive value from it, and protect it. That's where **CM evolveIT Data Lineage** comes in.



Researchers with Cisco surveyed 80 IT decision-makers on security budgeting and planning and found that more than half (56%) had experienced a significant security event - a breach, intrusion, or malware infection - in the past year, and 94% admit they have further to go to implement effective security practice.



Contents

- 2 Introduction
- 3 Enter Data Lineage
- 4 Data Lineage with CM evolveIT
- 7 Summary

¹ IBM Marketing Cloud, "10 Key Marketing Trends For 2017"

Enter Data Lineage

Data Lineage includes the data's origins, what happens to it, and where it moves over time.

- Wikipedia

Data lineage methodologies are commonly applied to **data-focused development**, **identifying business rules**, and **meeting regulatory requirements**.

Data Focused Development

Development efforts that involve the management of large amounts of data benefit greatly from a disciplined application of Data Lineage techniques, especially for an application that experiences a high growth-rate of collected sensitive data. When a field is expanded in development, tracing the data can help address all of the other instances of the field that are impacted by the field expansion. Data lineage can also help answer key business questions that arise when data shows up in management reports.

Identify Business Rules

Data Lineage techniques can be instrumental in identifying, tracing, and cataloging business rules in large enterprise applications: Identify business processes based on application data interfaces, filter out unwanted current architecture, identify data flow and decisions relevant to business rules, or store data flow and decisions in a user-definable folder structure.

Regulatory Requirements

As the "data firehose" grows stronger and injects more sensitive data into your system, your business liabilities grow along with it. Employing data lineage techniques to protect data from security breaches and reduce risk becomes more critical.

Using Data Lineage to know where your data is at any given point in its journey is also your best weapon to survive audits and keep regulatory risk at a minimum. With increasing financial regulations (Sarbanes Oxley, State Regs, and Banking/Financial Services), data privacy regulations (GDPR, HIPAA, etc.), and any other regulations that we can expect to be enforced in the future, having a solid, manageable audit trail is more important than ever.

Data Lineage with CM evolveIT

CM evolveIT's Data Lineage functionality analyzes mainframe applications to identify where and how sensitive data is processed – in screens, reports, files and databases. Its fast and complete analyses help application owners determine what code changes must be made to ensure all instances of data use are known, appropriate and secure.

Data can be traced across programs, subsystems, and even full applications. Data can be traced **across programs** in several ways. A called program with data can be passed from the calling program, or read from files/databases. Data can also be processed by language statements via data moves or calculations.

Tracing data **through an application** is much more complicated and CM evolveIT Data Lineage rises to the challenge, setting it apart from other solutions. CM evolveIT Data Lineage maps a path through all impacted programs within the application, including file read/write, database read/write, and traces through the program calls, elegantly stitching together all of the application's moving parts to build a highly valuable and actionable overview of the data's origin and eventual journey to its destination.

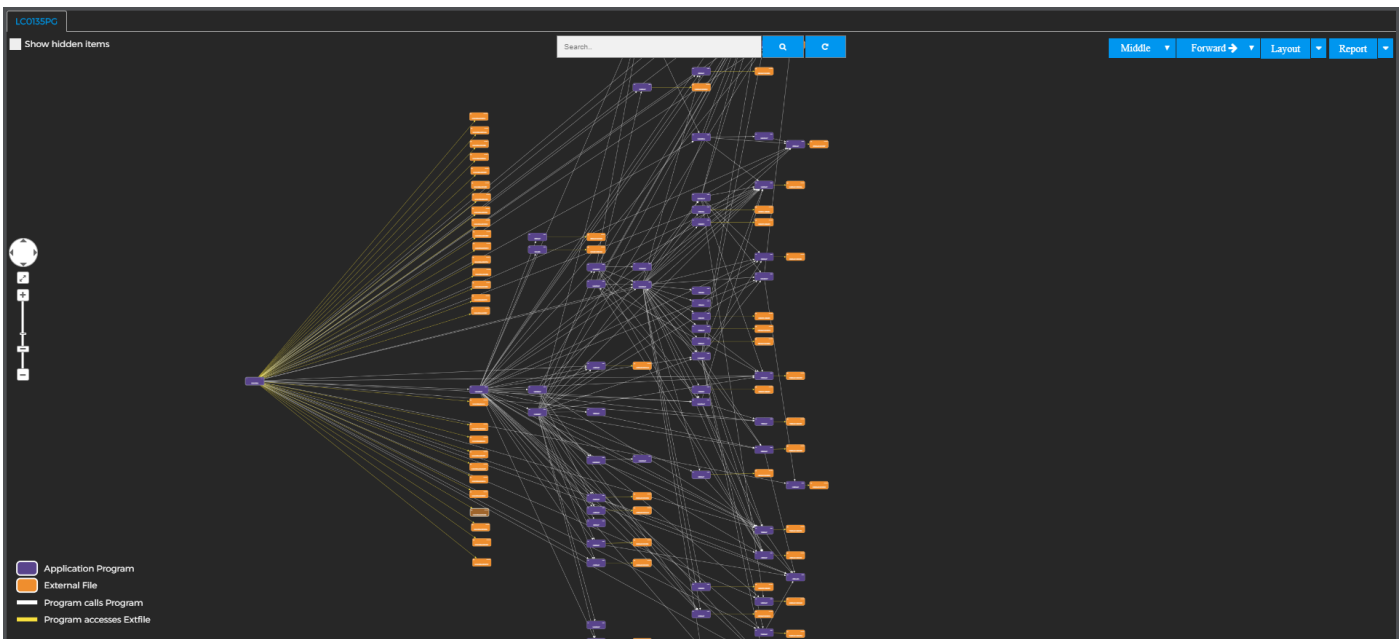


Figure 1: CM evolveIT traces data through a large application, capturing and storing complex data paths and relationships.

Starting Points: Places to Look

Data Lineage begins at your targeted data's starting point, which can be in any of several places. The first step is to identify them:

- Inputs / Outputs (Screens / Reports)
- Any program where a Field / Column is used
- Programs
- Copybooks
- Database
- Files

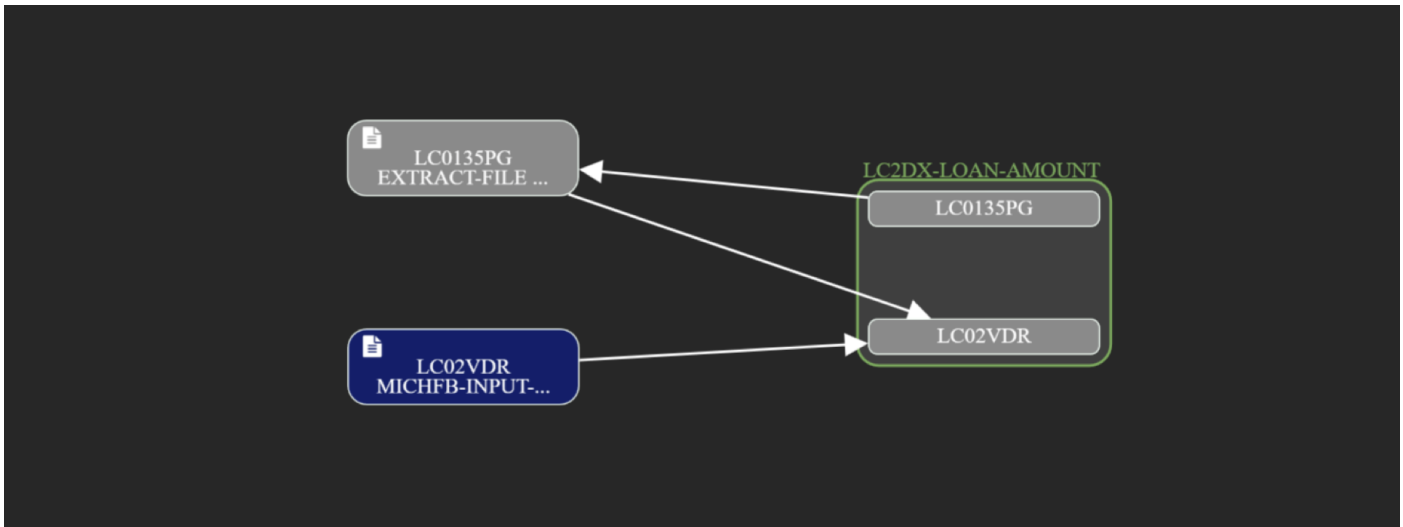


Figure 2: Start where field(s) are used within programs of interest.

Data Flow Paths

Once you've originated the data, the next step is to identify and catalog **Data Flow Paths**. To accomplish this, "pull the string" on a field to find the data lineage. As you pull the string, you can see what happens to data as it's processed. Once you've "pulled the string", you'll see the data's complex journey throughout your application (Look back at Figure 1 for an example of data complexity).

Once the data flow paths have been traced, you can identify and target specific data flows to meet your objectives.

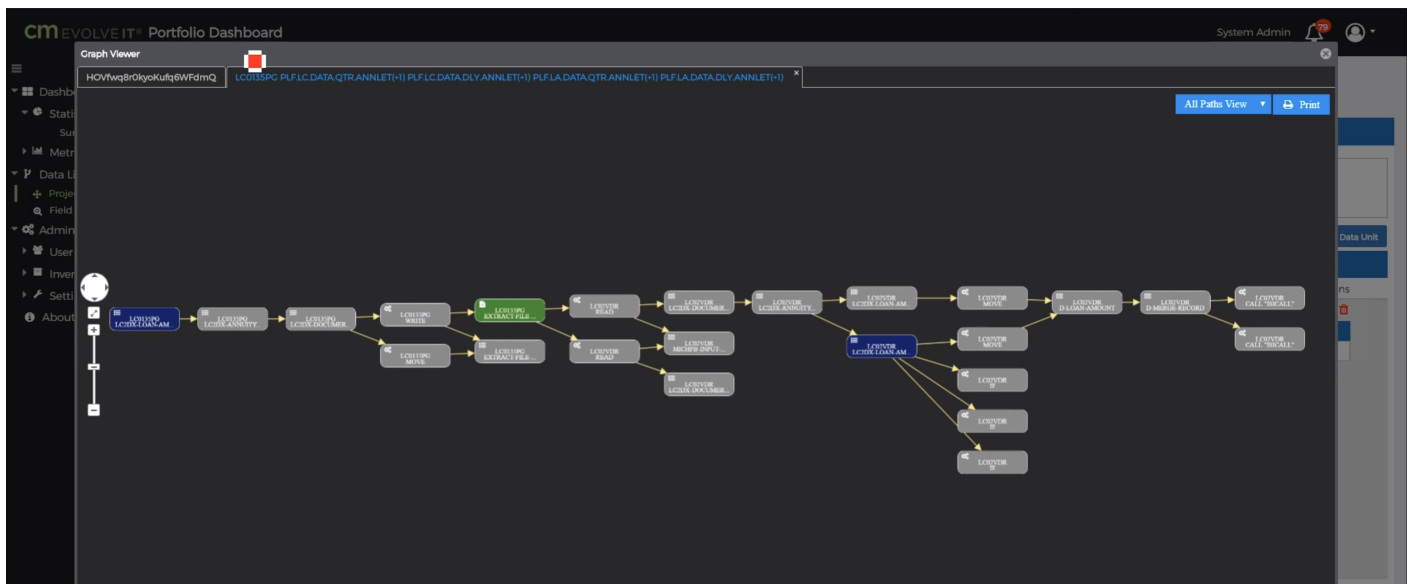


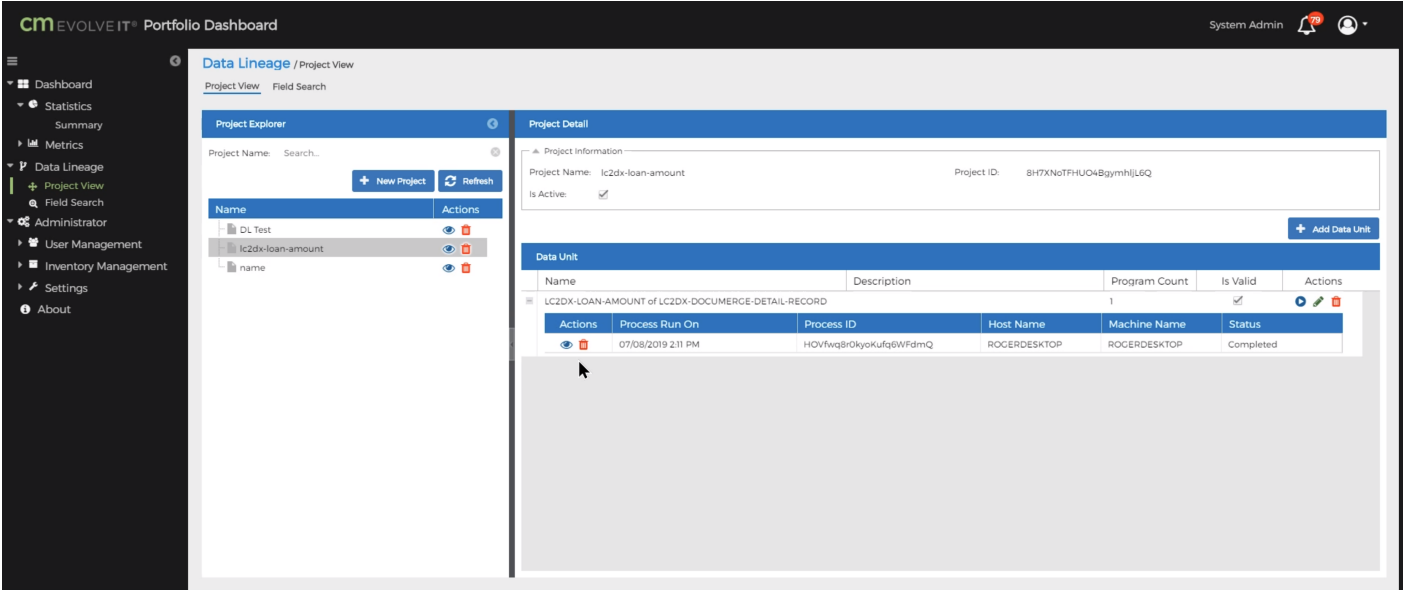
Figure 3: Drill down and target specific data flows.

End Points

Once the data source has been identified, and the path traced, the last key step is to locate the data end points - where the data lands, whether in files, databases, or storage. Your detailed path analysis with all statements is now complete.

Detailed Path Analysis



Once your data is traced from origin to endpoint, your detailed path analysis with all statements is now complete and can be stored for future reference.



The screenshot displays the 'Data Lineage / Project View' interface. On the left is a navigation sidebar with options like Dashboard, Statistics, Metrics, Data Lineage, Project View, Field Search, Administrator, User Management, Inventory Management, Settings, and About. The main area is divided into two panels: 'Project Explorer' and 'Project Detail'.

Project Explorer: Shows a tree view of project components. The selected project is 'lc2dx-loan-amount', which contains sub-items 'DL Test', 'lc2dx-loan-amount', and 'name'. Buttons for 'New Project' and 'Refresh' are visible.

Project Detail: Displays project information for 'lc2dx-loan-amount' (Project ID: 8H7XN0TFHU04B9ymhLL6Q). It includes an 'Add Data Unit' button and a 'Data Unit' table.

Name	Description	Program Count	Is Valid	Actions
LC2DX-LOAN-AMOUNT of LC2DX-DOCUMERGE-DETAIL-RECORD		1	<input checked="" type="checkbox"/>	 



Actions	Process Run On	Process ID	Host Name	Machine Name	Status
 	07/08/2019 2:11 PM	HOVfwq8r0kyokufq6WfFdmQ	ROGERDESKTOP	ROGERDESKTOP	Completed

Figure 4: Project view with detailed path analysis saved for future reference.

About CM First Offerings

CM First's powerful automation tools, augmented by professional services staff with many decades of software engineering and DevOps experience, ensure successful outcomes for even the most demanding modernization projects. Our products and expertise have helped over 400 customers in the public and private sectors reach their desired future state faster and more cost effectively than by using conventional approaches.

CM First software quickly analyzes, documents and re-platforms legacy code bases with minimal errors and rework, including those that are too large and complex for humans to tackle in any reasonable timeframe. The output is immediately usable by all team members, regardless of experience and knowledge of legacy software languages, accelerating application maintenance and modernization projects.

For more information, visit cmfirstgroup.com

Request a Demo Today

Contact us for more information or to schedule a demo. Call 888-866-6179 or email us: info@cmfirstgroup.com



CM First Group
888-866-6179
cmfirstgroup.com

7000 North Mopac Expressway
Plaza 7000, 2nd Floor
Austin, Texas 78731