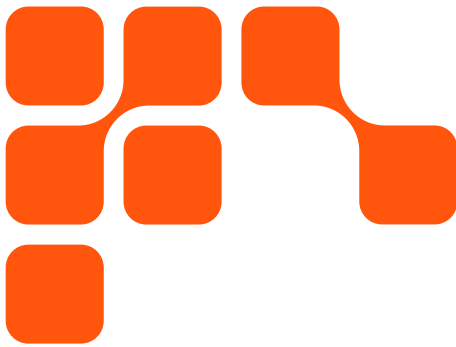


**CLOUDERA**

WHITEPAPER

# The Ultimate Guide to Data Lineage for the Finance Industry





## Introduction

Being a data professional in the finance industry is something like being in charge of the Information desk at Grand Central Station during rush hour.

You need to **handle many requests and demands**, from many people. No sooner do you help one person then someone else needs your assistance.

You work under **time pressure**. Everyone who asks is trying to get somewhere; they need your answer to do so, and they need it NOW.

Your answers need to be **accurate**. "Your train will depart from somewhere on the second floor" isn't helpful.

The **stakes are high**. The consequences of making a mistake or taking too long are that people are going to lose time and/or money.

The ability to **give quick, clear answers without compromising on accuracy** is the key to success in Grand Central Station – and no less so for success as a data professional in the finance industry.

This ebook will put those keys into your hands and enable you to lead your financial institution to success. We'll cover:

- **The 3 Key Data Professional Abilities**
- **How Automated Data Lineage Can Empower You with the 3 Abilities**
- **Key Areas Where these Abilities Have a Concrete Impact on Performance**

# Table of Contents

---

<b>The 3 Key Data Professional Abilities</b>	<b>4</b>
Here's why:	4

---

<b>The Secret to Gaining These 3 Key Abilities: Automated Data Lineage</b>	<b>5</b>
Accuracy	5
Speed	5
Explainability	5

---

<b>Let's take a closer look at the areas where these 3 key abilities have a concrete impact on your performance as a data professional:</b>	<b>6</b>
User Requests	6
Audit Preparations	6
Transitions and Migrations	7
Regulatory Compliance and Governance	7
Root Cause Analysis	8
Impact Analysis	8
Forward-Looking Business Insights	8

---

<b>How Farm Credit Services of America Completed complex Impact Analysis in Just One Day</b>	<b>9</b>
--	----------

---

<b>A Large European Investment Company Turned to Cloudera Octopai Data Lineage to Enable Easier Data Ingestion</b>	<b>10</b>
--	-----------

---

<b>Conclusion</b>	<b>11</b>
-------------------	-----------

---

<b>About Cloudera</b>	<b>12</b>
-----------------------	-----------



## The 3 Key Data Professional Abilities

There are three abilities critical to your data team's success within your financial institution:



**Speed**



**Accuracy**



**Explainability**

Having **all** three of these abilities is not just helpful; it's crucial. If you have fewer, even two out of three, you are going to fall short.

### Here's why:

#### If you have Accuracy + Explainability but no Speed:

Your business users will be paralyzed waiting for you to deliver answers. They will not be happy about it, and they won't hesitate to blame their lack of business responsiveness on your data team.

#### If you have Speed + Explainability but no Accuracy:

Quick turnaround time for responses to business users' questions will thrill them... until it turns out that the answers - and the underlying data - don't reflect reality all that well. And your ability to explain how you delivered the wrong answers to them so fast won't make them happier.

#### If you have Accuracy + Speed but no Explainability:

Business users may trust you implicitly and take you at your word. Upper management may not.

Auditors certainly won't. If you can't clearly show "where that figure came from," you'll be up a creek

As a data professional, you don't want to find yourself in any of these less-than-ideal situations. The solid trio of **Speed + Accuracy + Explainability backs you up for any situation** and displays your expertise for all to see.



# The Secret to Gaining These 3 Key Abilities: Automated Data Lineage



## Accuracy

Data lineage is your data's pedigree: where your data came from and what happened to it along the way.

While in the past data teams needed to manually trace data lineage, automated data lineage makes tracing your data's path exponentially faster and more accurate. It is the difference between using the Word Count function of your word processing program vs. manually counting words – for a 1000+ page document. In addition to the massive amount of time saved, you're also avoiding the inevitable human error.

With automated data lineage, all the information throughout your data landscape is comprehensively tracked. You will never lose track of individual data elements in the overall flow. Automated data lineage establishes a single source of truth, eliminating dirty data.

**The end result for you as a data professional?**  
**Trustworthy data that you, your team and your business users can rely on with confidence.**



## Speed

As noted above, no one would consider manually counting all the words in a 1000 page document to be a productive use of time.

When your business users request an answer, they want it as soon as possible... ideally right now. Performing data lineage manually slows down the business response and ties up your team's valuable time.

Automated data lineage frees your team from the time-consuming burden of reading through lines of code in order to understand what happened to make the data turn out the way it did, or what will happen to data if you make a change in your system. All you need to do is specify what data you want to trace – and look at the resulting comprehensive visualization of the data's flow through your system.

**The end result for you as a data professional?**  
**The ability to instantly locate and isolate the root cause of errors, quickly understand the impact of any changes, trace PII throughout your data systems in seconds, and use the time saved to focus on the other, more productive parts of your job.**



## Explainability

As a data professional in the finance industry, a good part of your time revolves around answering “how” questions.

**“How did we get this figure?”**

**“How can we confirm that all of this client's PII has been removed from our system?”**

**“How did this reporting error happen?”**

**“How will it impact our systems if we make a change?”**

To answer these questions, you need to be able to follow the trail of any given data point upstream and downstream within your data environment. The more systems that populate that environment, the trickier – and more confusing – it is to follow that trail.

Automated data lineage eliminates the confusion by creating a clear, visual, end-to-end map of your data environment. Place the “You Are Here” marker anywhere in your data landscape, and a comprehensive automated data lineage solution can show you exactly how your data got from its source to that point, precisely how it continued from that point all the way to its target – and what transformations it underwent along the way.

It's like putting on a pair of glasses. No more guessing, no more squinting: the answers to all your “how” questions are clear to see, easy to understand and simple to share with others.

**The end result for you as a data professional?**  
**The ability to quickly and confidently answer every question about your data thrown at you by business users, management or auditors.**

## Let's take a closer look at the areas where these 3 key abilities have a concrete impact on your performance as a data professional:

- Answers to business user requests
- Audit preparations
- Transitions and migrations
- Regulatory compliance and governance
- Root cause analysis
- Impact analysis
- Forward-looking business insights



### User Requests

Answers to Business as a data professional, you field an ever-growing list of demands from the business side.

“Where did the number in table Y, column Z come from?”

“Why don't these reports match up?”

“We need data about X.”

The business side relies on your responsiveness to make good decisions, meet regulatory requirements, and maintain the business's reputation.

Finance tops the list of industries where accuracy is critical. Business users who make calls that shape the company's decisions, investments and direction need data that perfectly reflects reality. Using “mostly accurate” data is like trying to cross a six-lane highway with “mostly good” visibility; what you can't see can really, really hurt you.

Fast, accurate answers are equally important for customer-facing business users. A customer-facing business user who cannot explain to an anxious (or irate) customer why he is seeing \$10K in his account when he thinks he should be seeing \$100K will cause irreparable damage to your institution's reputation. These business users need to be able to respond to questions from customers with accuracy and transparency - without taking days to do it.

Often it can seem like the number, type and frequency of business demands increase every quarter, leaving you and your team gasping for breath.

The strong handle on the **Speed + Accuracy + Explainability** trio provided by automated data lineage empowers you to satisfy your business users without being overwhelmed yourself. You can respond to business requests faster, providing more complete responses, with less effort and manpower invested in each request.



### Audit Preparations

Preparing for an audit is a headache and a half. You must show total accuracy; there cannot be mistakes. If you did discover errors while preparing for the audit (hopefully very few and very far between), each one needs to have a good explanation and reassurance of how that error will not reoccur in the future.

This preparation period can take up considerable human - and time - resources. And business users don't stop asking questions just because you're preparing for an audit, leaving you and your team spread thin and stressed.

When you possess the 3 key abilities of **Speed + Accuracy + Explainability**, audit preparations are **faster, calmer and not headache-inducing**.

You rarely discover errors, because your data and reporting were accurate in the first place. On the rare occasion that you might, it wouldn't take long to identify the root cause and understand how to fix it for the future. You can answer “where did this number come from” questions easily and quickly. And you have more time and patience to deal with all the non-audit data tasks that (for some inexplicable reason) don't disappear when you're preparing for an audit.



## Transitions and Migrations

Mergers and acquisitions, new technologies, changes in the financial landscape... the need for the data team's involvement in a transition or migration is so ubiquitous that you'll inevitably be part of one or more during your career as a data professional.

Like audits, transitions and migrations require LOTS of preparation. If you are migrating to a new ETL, database or reporting system, you need to lead your team in tracing every single data path, report, table, column... and making sure that everything is moved over correctly. If you are preparing for a regulatory or standards transition (like the LIBOR transition), you need to identify every point of impact that standard has throughout your data environment and prepare it for the new standard.

The strong handle on the **Speed + Accuracy + Explainability** trio provided by automated data lineage saves you immeasurable time and resources – both in the preparatory stages, and after the transition or migration is complete.

Taking the LIBOR transition as an example, you can quickly **identify all processes and reports that the LIBOR affects**. You can make the necessary changes and be confident that the new standard is cohesively integrated into your data landscape, without the concern of a workflow breaking or giving inaccurate data.

Migrating to a new ETL, database, or reporting system transforms from an onerous chore to an exciting opportunity when you can quickly trace data paths and only move the data and processes you need, pruning out old, unused, and resource-consuming workflows.



## Regulatory Compliance and Governance

The world of compliance is populated by little acronyms that make big demands of your institution: GDPR, BCBS 239, TRIM, or CPRA with PII. Keep awareness of the information in your systems – what it is, where it came from, and where it's going – is crucial in meeting those demands.

Common regulatory standards that affect financial institutions include:

- GDPR (**General Data Protection Regulation**)
- DCIA (Canada's **Digital Charter Implementation Act**)
- IFRS 17 (**International Financial Reporting Standard 17**)
- CCPA/CPRA (**California Consumer Privacy Act / California Privacy Rights Act**)
- BCBS 239 (**Basel Committee on Banking Supervision 239**, and specifically **Basel III and IV**)
- FRTB (**Financial Review of the Trading Book**)

As a data professional, the burden is often on your department both to enable the company to comply with regulations (e.g. to erase all the PII of a customer who requested) and to prove that they have complied with the regulations (e.g. to show that said customer's PII is no longer anywhere in your systems).

When you operate in a field as highly regulated as finance, the **Speed + Accuracy + Explainability** trio put into your hands by automated data lineage is what **gives you a firm hold on your data and the ability to pass compliance bars with confidence**.



## Root Cause Analysis

In the finance industry, many sources of data feed into your systems (e.g. interest rates, currency rates), many processes manipulate said data, and many people and systems use the results of those processes to make mission-critical decisions.

A broken process doesn't just mean an erroneous figure on a report; it can mean a short-term or long-term direct effect on your company's performance and profit.

It's no wonder the business side is frantic whenever they come knocking on your door for explanations of numbers that don't match up.

The powerful trio of **Speed + Accuracy + Explainability** enables you to promptly answer business users' questions and **correct errors and broken processes immediately**, facilitating short-term and long-term success. You will also be fielding fewer questions, thanks to more accurate data from the outset.



## Impact Analysis

Mergers and acquisitions, new technologies, changes in the financial landscape... the need for the data team's involvement in a transition or migration is so ubiquitous that you'll inevitably be part of one or more during your career as a data professional.

It shouldn't be a big deal to make a small change in a process or a system, right? Easier said than done.

No one knows as well as a data professional that small changes can have big impacts. Being the responsible employee that you are, you want to check that a proposed change will not have unintended upstream or downstream effects. If there will be an effect, you want to prepare for it before the change is made. But that can take days, weeks or months, severely limiting your agility and adaptability.

The mastery of the **Speed + Accuracy + Explainability** trio given to you by automated data lineage enables you to perform impact analysis in minutes instead of weeks and in hours instead of months. You will be able to **make system or process changes without a long preparatory period before and without unintended fallout after**.



## Forward-Looking Business Insights

To maximize your profit opportunities, you need top-quality data. In addition, your data must be organized in a way that you can quickly and easily access, manipulate and analyze it. If your data team is using their valuable time on cleaning data or unnecessarily drawn-out processes of root cause analysis and impact analysis, they won't have time to do the analysis of data trends that give forward-looking strategic insights to the business.

Mastering the **Speed + Accuracy + Explainability** trio with automated data lineage improves your data team's performance immeasurably. With maximum quality data available with minimal work, you will have the **resources and time available to deliver on your often neglected yet critical responsibilities**.



## CASE STUDY

# How Farm Credit Services of America Completed Complex Impact Analysis in Just One Day

When FCSA, a financial services company, realized that its data management vendor is about to upgrade its application and an estimated 500 fields would be deprecated or changed, it was decided to upgrade the supporting database and data processes and create new and shared knowledge that would be rooted in ongoing documentation and support.

However, the first phase of their process that involved manual data mapping took **four months** to complete and there were still plenty of unanswered questions.

During the next phase of the project which they estimated would take more than **100 days of work**, Kelly, in her position as Lead Financial System Analyst, leveraged Cloudera's Octopai Data Lineage for her impact analysis research. The **automated data discovery** tool helped her find which data they were looking for and to identify who owned each data set. The **automated data lineage** tool enabled her to visualize the data flow and see the impact of deprecation or alteration of the fields in question all the way through to reporting. Instead of seeking general input from a broad audience of potential users, Kelly was able to hone in on specific resources, allowing her to establish very concise communication to specific users.

As a result, Kelly completed the remainder of the impact analysis research in ONE day. What once had taken multiple analysts months of manual data mapping was accomplished by a single analyst in just a single day.

“

Gone are the days when we would have to ask multiple questions to multiple people. We try to target the root [of the issue] and fix it at the most upstream source so that it flows naturally through the systems. We now get the benefit of proliferated data quality.”

Andrew Stewardson,  
Data Risk Manager, FCSA



## CASE STUDY

# A Large European Investment Company Turned to Cloudera Octopai Data Lineage to Enable Easier Data Ingestion

With a significant percentage of their data ingested from external suppliers and source systems, this European investment firm was concerned about **data quality, consistency and trust**. Data lineage was critical to establishing confidence in the data, but the firm's lineage process often required manually wading through lines upon lines of code.

Cloudera's Octopai Data Lineage solution, with the ability to read code and produce a clear, detailed visualization of their entire data landscape, allows them to easily understand where data comes from in any of the 500 reports in their systems. Reports mandated by compliance regulations, which had always been a big-time drain for the data team, could be completed and delivered in a matter of minutes.

“

There is no area of the company that Cloudera Octopai Data Lineage has not affected positively. We have reduced the amount of ‘busywork’ that the data team must do while simultaneously empowering our department to deliver more direct business value. Confidence and trust in our data have increased a hundredfold – both within the data team and across the company.”

L.D.,  
Senior Enterprise Data Architect





## Conclusion

In the data departments of the finance industry, it's **always** like rush hour. Constant requests and demands. Time pressure. High stakes. As a data professional, it's in your hands to keep **your** company's data on the right track. Mastering **Speed + Accuracy + Explainability** is the ticket to success. And **automated data lineage** puts mastery of that critical trio right into your hands.

Navigating the intricate maze of data management is about striking the right balance: data should be accessible and accurate without breaking the bank. As data continues its upward trajectory in both volume and importance, platforms like Cloudera Octopai Data Lineage become the guiding star, illuminating the path for organizations in their data management journey.

To see a live product of how Cloudera Octopai Data Lineage can save you on data costs visit:

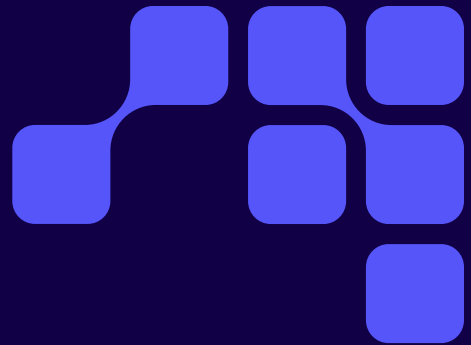
<http://www.ClouderaOctopaiDataLineage.com>

© Cloudera Octopai Data Lineage – Leveraging Data Lineage to Reduce Data Costs and Increase Efficiency .

## About Cloudera

Cloudera is the only true hybrid platform for data, analytics, and AI. With 100x more data under management than other cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable, trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to be able to use their data to solve the impossible — today and in the future.

To learn more, visit [Cloudera.com](https://cloudera.com) and follow us on [LinkedIn](#) and [X](#). Cloudera and associated marks are trademarks or registered trademarks of Cloudera, Inc. All other company and product names may be trademarks of their respective owners.



**CLOUDERA**

Cloudera, Inc. | 5470 Great America Pkwy, Santa Clara, CA 95054 USA | [cloudera.com](https://cloudera.com)