Educational Administration: Theory and Practice
2020,26(4) 965- 973
ISSN:2148-2403
https://kuey.net/

Research Article

Data Lineage Strategies – A Modernized View

Shakir Syed[1*], Rama Chandra Rao Nampalli[2],

[1*] Senior Solution Architect & SSBI Leader, Email: -shakir.syed.microsoft@gmail.com
[2] Oracle EBS and Workday Consultant, nampalli.ramachandrarao.erp@gmail.com

Citation: Shakir Syed (2020), Data Lineage Strategies – A Modernized View, Educational Administration: Theory and Practice, 26(4), 965- 973,
Doi: 10.53555/kuey.v26i4.8104

ARTICLE INFO

ABSTRACT

Data lineage refers to data sources and the data derived from them, along with the transformations that may be acquired from these sources. Data lineage is important to enterprises to understand how certain data was derived and if it is acceptable for use in particular analytical outputs. This could potentially lead to regulatory and compliance issues in many domains. We propose a modernized data lineage strategy that takes into account the modern approach to data management, inclusive of the current metadata stores and logical data structures. We will present specific approaches to modernize the existing data lineage strategies within the data warehousing and data virtualization paradigms. Finally, we will present research directions in the context of big data systems and data brokering for open data. Our modernized view on data lineage will benefit data management professionals, researchers, software and tools developers, and stakeholders who seek to promote open data.

Key words: Cloud-based software development, GitOps, Static application security testing, Container security, Dynamic application security testing, Azure DevOps, CI/CD pipeline, Security automation, Penetration testing, Cloud-agnostic solutions.

1. Introduction

A concept closely related to traditional data warehouses is data lineage, also known as data provenance. Here, the goal is to better understand the architecture, strategy, and best practices related to a robust solution. Data lineage is essential for banks to achieve regulatory compliance. Data lineage, data provenance, and data governance are critical elements in analyzing your application portfolios and data assets to be successful in the cloud. Each tool is capable of collecting different amounts of data and using diverse techniques to populate and identify data lineage. Some cloud companies charge you extra for viewing data lineage, and you'll want to monitor who is responsible for updates. You should have regular reviews within the organization to make sure all the parties involved are keeping the necessary data lineage accurate. Accurately display lineage; if network security is there, it's just a checkbox. If not, do you want to put external access in place?
The data lineage strategy must have a proper routine to index the data documentation and perform quality checks. The lack of lineage can represent a potential decrease in the accuracy of decision-making, in addition to a significant decrease in the time of obtaining information. The planned strategy is essential to improve the understanding and trust in our artificial intelligence models. IT, business, and data science must act together to define the rules and strategies that will guide data democratization. Data democratization is an inclusive approach that increases the impact of the data, facilitating access by less technical users and guaranteeing their democratization throughout the enterprise. This democratization process has been enhanced and leveraged by increasingly innovative tools, products, and techniques that lead the democratization of artificial intelligence and machine learning into the enterprise. Double-check your architecture components regularly with data lineage.

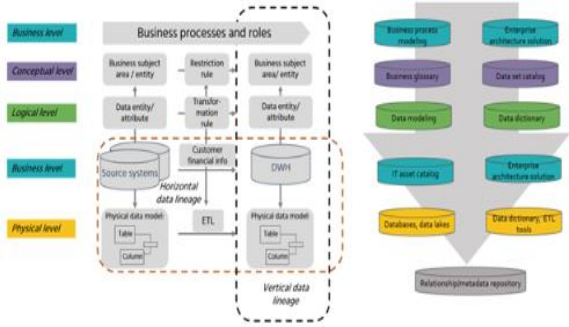

Fig 1 : Data Lineage: the Needs of and Benefits to Various Stakeholders

1.1. Background and Significance

The significance of data lineage cannot be overstated. Today's data management landscape is convoluted and complex, with data moving throughout an organization. Data lineage provides detailed risk management and compliance, control, and auditing capabilities. It can be used for all types of regulatory requirements and meet the needs of internal risk management. With data moving from a CRM to an office productivity tool and then into XML that is sent off somewhere, data lineage allows the data to be tracked. It showcases how the data has been transformed as it is moved about the organization. Data can move from the CRM into Excel, printed, edited, and then re-entered back into the system. Data lineage tracks all of these moves and demonstrates where the data entry was made and who made it. It provides risk management, control, and audit. Additionally, data lineage can be used in projects, such as impact analysis across the organization.

Understanding the impact sees the upcoming needs, and lastly, it is the right to link several items together to demonstrate a complete end-to-end flow of your data. For example, a data lineage solution integrating with a data quality tool while utilizing a business glossary can display a complete end-to-end flow of data reporting. It is a visual experience showing the data of the extractions from the source systems, explaining the transformations, displaying the final data fields, and linking the data to the business terminology of the data that flows into a financial report, hence explaining the source of the data presented in the final report. The underlying database of the financial report comprises a system of well-structured rows and columns of finance. The data layers solution will deliver a combined approach, including the connectivity and functionality of a data management tool with the data processing power pulling data from a host of sources.

1.2. Purpose of the Paper

The objective of this paper is to present a set of updated taxonomies for data lineage strategies. Specifically, this new framework focuses on the use of architectural constructs as a means of expressing those strategies. We show that these updated categories can be used to discuss, analyze, and select the most appropriate data lineage models given specific data management scenarios. The proposed model casts data lineage as a modular extension to the Enterprise Integration Architecture and thereby uses a standardized class of architectural viewpoints, models, and formal notations. This facilitates an explicit and convincing traceability link between a data-centric strategy and that of the related integration strategy. Such a link is often either missing or simply assumed. This can lead to non-classes of capable design, enterprise, and supply chain integration architectures, as well as failed, interrupted, or unintentional data exchanges.

In large organizations, data is stored disparately across various infrastructures inside and outside organizational borders. Where decentralized management of data sources occurs, this can lead to the assumption that data is either incomplete, incompatible, out-of-date, or soiled. In instances where data is withheld, under embargo, paid for, or of high value, stakeholders often rightly request verification or validation of the source as well as the integrity of the content. Data lineage has been researched and solved in the past mostly from a quality perspective. These researchers have suggested that data lineage should be exploited to support quality checks on the provenance of data to ensure the scientific method and to support integrated changes by automatically updating data based on data lineage and business service dependencies. Nonetheless, no comprehensive architectural model for data lineage has been proposed so far that faithfully represents the multitude of different data lineage strategies. This paper uses Enterprise Integration Architecture as the underlying architectural coverage.

Equation 1 : Data Lineage Completeness (DLC)

Represents the percentage of data assets that are being tracked across the full lifecycle.

$$DLC = \frac{T\_D\_Tracked}{T\_D\_Total} \times 100$$

Where:
T_D_Tracked = Total Data Tracked (in GB, TB, etc.)
T_D_Total = Total Data Available

2. Understanding Data Lineage

What is Data Lineage?
Data lineage is defined as a data management discipline that provides deep insights into the lifecycle of data. When a data management specialist talks about data lineage, the presenter primarily refers to the origin, history, processing, transformation, and flow of data sets from their early point of generation to the current point of consumption. Data lineage is further classified into two major categories: forward lineage and reverse lineage. Forward lineage is explained as the ability to trace the origin of data sets that are loaded to a system in question. Conversely, reverse lineage provides insights into the downstream flow of the data, right from the consumption touchpoints that are part of the system in question.
Understanding the relevance of Data Lineage in an organization's data journey:
In a modern enterprise, data is generated; it could be created intentionally or unintentionally by organizational stakeholders such as machines, users, robots, and antivirus software. While data generation is always interesting, its processing, storage, and consumption are challenging. The inability to effectively govern and manage data results in the following shortcomings: lack of data quality, incorrect record count, misleading results, the inflow of unforeseen business risks, regulatory impropriety, abnormal data processing overheads, improper disaster recovery testing and recall operations, excessive storage costs, and poor data security. A modest effort to pen this topic down to present data management styles has revealed that industry analysts, practitioners, and authorities have since failed to reconsider the values and dimensions presented in data lineage.

2.1. Definition and Concept

Before moving into the detailed discussion on data lineage strategies, it is important to first understand the definition of data lineage from a modernized view. A simplified description of data lineage can be that it represents the complete and detailed data flow paths beginning from where the data source comes until the multiple derived usages of these sources, expressing which dataset(s) served as information input, the process used to generate the data, and the final delivery of the data. Concepts such as data flow and data path exist in some domains like databases, data warehouses, and data integration. The idea of data lineage for database management came from the concept of forward recovery techniques that indicate how to rebuild the database if any future failure occurs. As newer data management concepts have evolved, new concepts are linked to the notion of data lineage, such as big data, logical data warehouses, and data lakes, to name a few. In simple terms, data lineage describes which datasets populated sources of a given data marketplace within a given date range come from a data sourcing, operational, and reporting setup environment until the final data delivery.

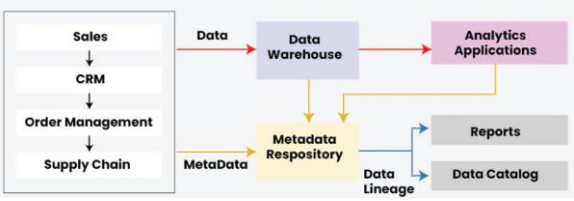

Fig 2 : Example of Data Lineage Process

2.2. Importance in Data Management

The typical golden thread in data lineage is one of "point A to point B and maybe C and D, etc." Flow is important, yes, but there's more to the data lineage story. A less prominent part, though critical to IT and architectural strategy, does not involve high-value lineage detail and is not of crucial importance in lineage event cataloging. This less obvious part is all about the sourcing and delivery of data to its destination, about latent versus hot artifacts that are the hookup points in the lineage event string. This is all about data writing and artifact management. The interconnections between the many producers of these data sourcing and delivery artifacts and the fulfillment that this word connects to in the most obvious foundation for data management is provided by the metadata. Beating the drum of data lineage importance is straightforward and almost cliché. How many data conversations have not ended in the refrain for "show me the lineage for this table or column"?
The compliance mandate means that data lineage has lifted the volume of its importance from far off in the background soundtrack to equal the crescendo of the action theme. But, as databases grow to gargantuan sizes and local data security and access concerns get louder and stricter, so grows the clamor for guarantees that data content is being efficiently warehoused and that legitimate users are being effectively serviced. Data lineage is rapidly evolving from a good idea to a crucially important practice. Data deployment and execution must be

guaranteed, as well as ETL and analytic events. In other words, lineage involves data creation, storage, and access.

Equation 2 : Data Lineage Depth (DLD)
Refers to the number of transformation and processing steps tracked between data origin and final destination.

$$DLD = \sum_{i=1}^{n} (S_i \times T_i)$$

Where: Si = Step i in the data pipeline
Ti = Complexity/importance of transformation at step i

3. Traditional Data Lineage Strategies

Data suffers the same issues. It's essential to have a comprehensive understanding of the data, especially when the data spans numerous distribution systems and is shared across the individual cycles and ultimately reported. In the financial services sector, this understanding of data is data lineage, and at the point of audit, it's not acceptable to have missing slots. In mainstream IT, data lineage information was held in the heads of highly knowledgeable personnel who, over time, moved from the business. This document shares strategies to capture this knowledge in a formalized way. Data lineage is of prime importance to data owners for both the technical track – we need to know where the data has come from so that we can trust it. In complex enterprise architectures, data spans many enterprise information ecosystems including life-cycle environments, retention systems, governance systems, and storage systems. Business track – we need to know not only where the data has come from and what's been done to it, but also what additional business logic has been applied and who requested changes, as our end business user trusted data includes not only accurate data points but also an understanding of why. The research for this report involved interviews with multiple professionals from leading financial services firms leading the data lineage field. This document compiles research colloquially shared with executives responsible for mid to senior technology decisions and is designed to help those leaders make strategic decisions around modernized data lineage functions in their organizations. The document outlines traditional data lineage strategies and highlights the differences encountered by enterprise data information owners versus other sectors.

3.1. Key Characteristics

As the process of managing and using data continues to grow by leaps and bounds in our tech-savvy society, the demand for a new strategy to efficiently allow for the increasing numbers of data sources, varying mounting techniques, and differing data uses while guaranteeing transparency and provenance grows as well. Modernizing our views on data will help organizations trust non-traditional data where provenance has not been historically documented and will address the need for data traceability in new applications, such as streaming data processes and microservices architectures. A modern view of data leads to embracing heterogeneities and pushing the frontier of what data management can handle. Satisfying these increased demands requires innovative and flexible data lineage strategies that go beyond the traditional view and meet these data integration challenges that are increasingly common in practice.
So, what makes data lineage flexible and modern? Herein lies the key characteristics of a modernized view of data lineage. In short, we want to start from an end user's question – something that the user encounters in everyday use – and work our way backward to find out how to introduce non-traditional data into a transformative visual discovery system. These nine key characteristics are progressive and build on top of each other. Because a modernized view of data lineage begins with an end user's question, the first key characteristic is the ability to generate lineage from query results. In practice, a user query provides knowledge of the needed results, and knowing the relationship of a result to an origin provides the prior knowledge needed to determine what data might be mounted given those particular semantics. The motive for documenting the lineage of the query results will determine what specific lineage is considered.

Equation 3 : Automated Data Lineage Tracking (ADLT)

Automating the lineage tracking process for real-time updates.

$$ADLT = f(R\_D, I\_P, M\_L)$$

Where: R_D = Real-time Data Flow
I_P = Integration of Platforms
M_L = Machine Learning-based automation for lineage detection

3.2. Advantages and Limitations

The advantages of this technique manifest themselves predominantly in two fundamentally important capabilities: the capability of covering expression interpretation and the capability of covering performance impacts. Expression interpretation is the capability of determining the derivations of the input attributes used in an expression, and it is extensively accomplished by the query semantics. Performance impacts, on the other hand, relate to the estimations of the major query plan actions, including the cost, cardinality, and selectivity of those actions being processed by the query optimizer. Lineage is a critical piece of information for query plan optimization, especially for the optimization of future queries. Supporting the benefits detailed above eliminates the main limitations of this system. The hybrid technique can also be coupled with real-time updates of column-level statistics, hence better query plans. Additional enhancements to the hybrid strategy will be investigated.

4. Modernizing Data Lineage Strategies

Treating data lineage as a traditional data management issue constrains its capabilities and results in impactful governance, compliance, risk, and privacy capabilities not being realized. This section presents strategies that bring data strategy planning into focus for modernized data management solutions. The proposals describe how data strategy planning identifies, evaluates, and selects data capabilities to satisfy the organization's business politics, economics, and technology. So far, traditional data management, with its back-office data models and implementations of data storage, access, and updates, has presented data lineage as an internal search for genealogy; like canon law, data lineage implementation has been hidden behind thick enigma codes. With data becoming recognized as an enterprise-wide, shared asset, the implementable redeployment of data lineage changes from being a long-term internal process to being a necessity in rating the quality of internal and externally acquired data, the availability of internal and external data services, and how other organizations interact and follow the shared data.

4.1. Technological Innovations

A combination of technological innovations has helped make it possible for data lineage to be implemented and rolled out on an enterprise scale. It has decentralization and a focus on providing deep analytics in real-time, thus enabling users to locate and fix production issues quickly. The new technologies facilitate improved efficiency and accuracy in the data lineage discovery process and help offset the risks associated with liberal access to sensitive data across the enterprise.

Enabled by all these innovations, the enterprise can better understand and manage its data and be assured about the accuracy of the processes used. In this section, I discuss how technological changes and the use of microservice principles have helped to modernize lineage discovery engines. They enable business users to leverage their data in production and meaningfully interpret the lineage they produce. This, in turn, allows for detecting faulty processing changes as they are deployed in real-time, thus protecting the user's digital business tirelessly without the need for excessive processing and organizational overhead.

Equation 4 : Data Lineage Accuracy (DLA)

The accuracy of data lineage mappings, i.e., how well the lineage reflects actual data flows and transformations.

$$DLA = \frac{Correct\_Mappings}{Total\_Mappings} \times 100$$

Where:

Correct_Mappings = Number of correct lineage mappings

Total_Mappings = Total lineage mappings recorded

4.2. Integration with Data Governance

Data governance and data lineage are tightly connected concepts. They complement each other, often to satisfy privacy and compliance regulations. While data governance ensures that an organization complies with regulatory standards and internal rules and policies, data lineage ensures that organizations can trace source data properties at any instance in its lifecycle. For this synergy to be effective, data lineage implementations generally rely on data governance insights to understand what data needs to be traced and to inform the lineage consumers about who in the organization is best positioned to consume it. Additionally, data governance best practices require keeping track of data quality, usage, and access, and data lineage is a valuable input for this.

Data governance is normally a managerial domain, with many associated formalities. Responsibility for data governance often lies with a governance board or data governance council, which may be considered one of the attributes of the governance process. They determine an organization’s policies by reviewing recommended policy definitions and ensuring that policy obligations are met. Data governance is implemented through data policies, based on compliance with those policies. This is where data lineage across a data management lifecycle, such as a data-centric process, which is known as compliant by design, by creating a trace of data

policy setting and runtime facts that are related to the data, is key. Organizations require data lineage to provide tangible data evidence supporting the data's compliance with the policies in place, by design.

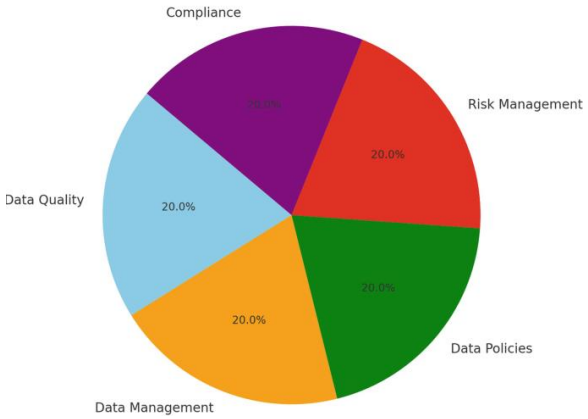

Fig 3 : Data Governance Components Distribution

5. Case Studies

Let's consider several real-world examples of implementing data lineage using critically important business dimensions. These are not fashioned as how-tos but rather what was achieved with what and why the choices help or hurt. It allows you to perceive which strategy would work best according to your specific requirements. Perceived data lineage use case: There is a metadata management tool. The client requires basic data lineage capture, a smaller scale than existing tool capability, and not on all SQL scripts; only specific files in predefined directories. Slowing the system performance is not an option, metadata must be captured under its profile name for audit reasons, and all metadata must be stored in the client's database.
Strategy observed: Target SQL Server database on the client's server. Scheduled jobs. In-house SQL scripts capture inherited from the tracking columns of source tables. An in-house shell script captures the profile name and target table name from the logs. The dashboard is minimized to four key pieces of information (Capture Start, elapsed time, Capture End, errors). The questions are based on only the most recent capture—two processes to update the data for the previous day and wrap for the names. A standard table is used, and a quick view.

5.1. Real-world Implementations

A model-driven approach for processing MDA-compliant data lineage strategy has already been explored in the implementation of the segregation of duties support in a project. The main goal of this project is to support the second European Directive/German GAAP-based reporting for banks. This task requires identifying related data and ensuring the non-comparability of the related mappings in the dashed enterprise models along the data processing steps. Dashed enterprise models are PIMs that represent PSMs in terms of technology-neutral domain-specific abstractions. Configuration-based data mappings integrate information about a) the business/sub-process performed by the given service implementation b) how to address potential disparities arising due to the account book used c) the address of a potential external system used when calling the external system d) the step-by-step data processing consistency rules used during the data migration process of the business activity service.
The services have been designed as API Assembly, which solves the issue of the guiding principle of the specification and possible effects of the services, considering that specific modules will be used to add specific business logic. In the future, an MDA approach will be applied to the project. However, the conceptual frameworks presented are in advance of the performed modeling, with an MDA that defines how to implement the key concepts of the proposed model from DashEDM to PIM and then from PIM to PSM. To handle this problem, Enterprise Data Flows, the first MDA-compliant methodology to define the Data Flow Metamodel about the business information while applying detailed PIMs and PSMs – each at a different level of data granularity – to represent PIMs. Enterprise data flows are applied using a set of tools to create consensus in complex scenarios involving large teams, highly technical decision-makers, and many enterprise data.
To design traceable data flow models, we apply the Object-Process Methodology. In our method, the final data flow model relies on conceptual diagrams of the business objects and class diagrams – allowing one to fully design the routing of the target data for tasks in more specific models. In our method, the routing of enterprise business models is performed using BPMN, UML Use Case Model, and the OWL-S Service profile for PIM. Finally, to define PSMs, Service-Oriented Enterprise Modeling is applied. In other words, organizational resources are business class diagrams and are provided by PSM using a common vocabulary for design that exploits our target architecture. Data collection from a practical project that is being analyzed from recent real-world experiences is presented and discussed.

6. Challenges and Future Directions

Several open issues remain in this space, not all of which are technological. It is, however, clear that in the years to come, a combination of growth pressures underscored by a series of regulatory developments that no longer makes a liberal data governance approach viable will together compel us to re-examine what we thought we knew about data management. Let us conclude our review of strategies and industry practices by examining the challenges and future directions that the data lineage landscape is likely to evolve towards as a result of these developments. One of the most pressing challenges ahead is likely to be that of offering more effective long-term sustainability, given evolving legal, operational, and technological landscapes. The funding modus operandi of most projects is not a viable approach to ensuring that the desired innovation is adequately funded today, tomorrow, and beyond, and that if some of the proposed roadmaps materialize, up to 80% of the funding comes as it is phased in at the end of the contracts, which is far from ideal. Ensuring effective user engagement is thus paramount – the lessons of organizations indicate that listening to and involving the user community is vital, and if a small connector strategy is to be pursued, the experience can also provide important pointers.

Equation 5 : End-to-End Lineage Visibility (E2ELV)

Ensuring visibility into all stages of the data lifecycle.

$$E2ELV = \frac{DLC \times DLD}{M}$$

Where: DLC = Data Lineage Completeness
DLD = Data Lineage Depth
M = Number of systems involved

6.1. Current Challenges in Data Lineage

Data lineage has been around for decades but is finally being recognized for its value. This resurgence is mainly due to organizations experiencing greater complexity in their data flows due to multi-cloud initiatives, data analytics platforms, regulatory compliance requirements, and the significance of AI models that have been developed and deployed. Alongside these challenges, governance and data compliance have to be factored in, as does the requirement of responsible AI. It is better to establish and document your overall strategy that includes the flow and impact of data in an AI system before embarking on this endeavor and developing documentation retrospectively.

Maintaining data lineage is fraught with challenges. New models often rely on some form of data before being updated. Business transformation or technical enhancements may require changes to existing models to achieve new or modified strategies. The lack of comprehensive model data access can lead to unnecessary data propagation for that model. Retraining an old model may require older forms of data. The availability and sourcing of this data could be more difficult in the future. Proper documentation and implementation of strategies ensure that up-to-date data sources are being used. Failure to follow through can result in data breaches and non-compliance, which could be very costly.

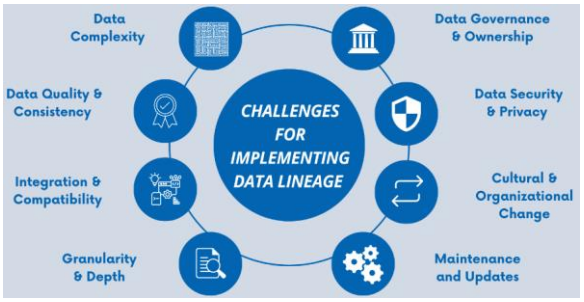

Fig 4 : Challenges for Implementing Data Lineage

7. Conclusion

The subject matter dealt with is related to Data, the tools associated with big data for creating data lineage, and different strategies for lineage pipeline with a set of line aging techniques. A thorough comparison between different modern-age strategies is also covered and analyzed. It touches on various aspects growing in this area comprising industries, academic research works, and the statistical domain. Reviewing the current groundbreaking developments and different strategies for creating comprehensive data lineage pipelines and associated techniques and the considerable research work done by different corporates through different representative products was truly an exciting journey. From the ground, we can see that more literature has

been developed from industry and few core academic research works have been contributed in the same domain for growing a more transparent, trustable, and compliant behavior of artifacts in a database system. Most academic works include a bit of a small portion of lineage, not from a generalized area, except a few excellent approach proposals, and an effective way to minimize the number of operations.

7.1. Summary of Key Points

To support modern functions for organizations, data must be accessed and interpreted seamlessly. A common problem that inhibits this capability is being unable to trust the integrity of authoritative providers for data that flows from such providers into consuming functions. Data lineage is commonly assumed to be a panacea for this issue, but known strategies for creating explicit and implicit data lineage are both technically challenging and have scalability and performance issues. There is also an implicit tradeoff between the expressiveness of a data lineage description and the associated governance and trust model. This chapter addresses these challenges by first differentiating explicit and implicit data lineage before proposing various legal and technical strategies to simplify and speed up both forms of data lineage. These strategies will be developed further to provide a process, methodology, and associated legal rights and responsibilities that both simplify and make more comprehensive explicit and implicit data lineage possible.

7.2. Importance of Modernized Data Lineage Strategies

With new tools and strategic principles, there are better ways of extracting value from modernized data lineage. Audience approaches that were focused on specific data lineage management activities in the past, which helped to answer only certain questions effectively, can give way to more comprehensive strategies that yield broader business value. Broken and incomplete lineage will become more apparent as the enterprise uses more analytical data to implement management decisions. New technologies can drive more preventative insurance strategies for keeping lineage intact. In this not-so-distant outcome, a much lower percentage of queries to your data lineage system will be less impacted by lineage preservation. This chapter describes these and more strategies for linking newer lineage into the enterprise line of business. It is much easier to think of the value of data lineage today than it was 30 years ago, or even 20 years ago. For many business projects, data lineage represented a low-value activity, and it had no immediate results payment. Much of the maximization of the return to data lineage work derived from understanding the lineage benefits as an extension of metadata management. At one point in the past, the plan was to use enterprise models for metadata to understand the impact of changes, and the models would evolve as things changed. Today, in truly integrated analytics, there is precious little time for enterprise modeling or data lineage.

Further, we expect many elements' values to an end user are not fungible. The benefit of modern data lineage, however, has broadened significantly. Classic data lineage, after all, was much more about an analytic model and the associated metadata. Satisfaction with just those models has become a reality that is disappointing to those who crave a much more enriched satisfaction with modernized data lineage.

Fig 5 : The Importance of Data Lineage Tools in Data Governance

8. References

1. Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219959
2. Mandala, V., & Surabhi, S. N. R. D. (2020). Integration of AI-Driven Predictive Analytics into Connected Car Platforms. IARJSET, 7 (12).
3. Kodanda Rami Reddy Manukonda. (2018). SDN Performance Benchmarking: Techniques and Best Practices. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219977
4. Mahida, A. Cross-Border Financial Crime Detection-A Review Paper.
5. Manukonda, K. R. R. (2020). Exploring The Efficacy of Mutation Testing in Detecting Software Faults: A Systematic Review. European Journal of Advances in Engineering and Technology, 7(9), 71-77.
6. Vaka, D. K. (2020). Navigating Uncertainty: The Power of 'Just in Time SAP for Supply Chain Dynamics. Journal of Technological Innovations, 1(2).
7. Chintale, P. (2020). Designing a secure self-onboarding system for internet customers using Google cloud SaaS framework. IJAR, 6(5), 482-487.

8. Manukonda, K. R. R. Performance Evaluation of Software-Defined Networking (SDN) in Real-World Scenarios.
9. Mandala, V. Towards a Resilient Automotive Industry: AI-Driven Strategies for Predictive Maintenance and Supply Chain Optimization.
10. Manukonda, K. R. R. (2020). Efficient Test Case Generation using Combinatorial Test Design: Towards Enhanced Testing Effectiveness and Resource Utilization. European Journal of Advances in Engineering and Technology, 7(12), 78-83.
11. Chintale, P. SCALABLE AND COST-EFFECTIVE SELF-ONBOARDING SOLUTIONS FOR HOME INTERNET USERS UTILIZING GOOGLE CLOUD'S SAAS FRAMEWORK.