Building AI-Native Data Platforms: From Data Lakes to Intelligent Decision Platforms

Neetu Uthaman[1], Jose Prabhu Michael Singarayan[2], Jayakumar Ramalingam[3]

[1]Independent Researcher, Seattle, Washington, USA
[2]Independent Researcher, Atlanta, Georgia, USA
[3]Independent Researcher, Atlanta, Georgia, USA

Abstract:
This is an original research paper that investigates how to create Artificial Intelligence (AI)-native data platforms that transform data lakes into intelligent data decision platforms. The research relies on a self-made data set, which includes 215 enterprise data platform instances for ingestion quality, storage readiness, visibility of metadata, coverage of automation, decisions delivery, trust score, user adoption, access speed, and control of risk. The dataset was staged to simulate real enterprise data platform stage behavior for data lake, processed data, intelligent service and decision platform. The tools that are used in the study are spreadsheet based tabulation, graphical analysis, architecture modelling and structured performance comparison. The paper describes how modern data platforms go beyond just the data store, and how governance, metadata, automation, learning-enabled services, monitoring and decision applications are added to the data platform. The proposed architecture illustrates the route to decision interfaces for business teams from the raw enterprise sources. The results point to the fact that intelligent decision platforms have even more value when integrated with the quality control, cataloging, model enablement and delivery services in a governed environment. The study also provides two result tables, a mix bar-line graph and a waterfall graph to demonstrate an increase in readiness, accuracy, performance and risk reduction. Finally, the paper highlights the importance of AI-native platforms for enterprise to transform siloed data into decisions that are timely, trusted and actionable.

Keywords: AI-Native Platform, Data Lake, Intelligent Decision Platform, Enterprise Data, Data Governance

1. Introduction

As explored in [4] enterprise data platforms have evolved from being a repository of data to an intelligent platform on which decisions and automation, governance and operational learning can be based. The previous data lakes were primarily used for storing a variety of raw data from diverse business systems, applications, logs, sensors, customer channels and transaction records, like [9] did. These lakes were useful for organisations to get information at scale, but frequently posed issues when stored information was not quality controlled or owned, when it was not semantically clear and where it was not directly correlated to business actions, as detailed in research by [2]. As mentioned in [13] the need in a modern organization is not restricted to storage but also to the need of trusted, timely, explainable and usable information for decision makers. The concept of an Artificial Intelligence (AI)-native data platform is a response to this need: ingest, process, catalog, govern, provide model services, monitor, and deliver decisions in a single operational environment, as suggested in [6]. In this kind of platform, AI services are not deemed as an "afterwards" layer, like [11] considers. Data is captured, cleaned, described, enriched, analyzed and delivered using connected data services for business decision in departments as per [1].

The transition is crucial as enterprises are relying more on quick decision making in finance, supply chain management, customer service, risk management, marketing, healthcare operations, digital commerce, and compliance management as stated in [8]. Organizations lose speed and have less confidence in the data when it is locked in separate repositories, as demonstrated in [14]. With the platform providing quality checks, metadata visibility, metadata control, automated metadata preparation, intelligent metadata recommendations, and decision interfaces for users, the same information becomes a strategic asset, as in the case of [3]. The present research centers on the journey from data lakes to intelligent decision platforms, with reference to platform readiness, performance improvement, growth of trust, level of automation coverage, and capability for decisions delivery as defined by [10]. The paper considers 215 study instances crafted around enterprise platform activities, and breaks down the role played by each stage in increasing the decision maturity as applied by [5]. While a data lake is the foundation, an intelligent decision platform is what will add value; that is, connect information to action, as described in [12].

This movement can be seen in the proposed deployment architecture in Figure 1 as used by [7] and consists of the following elements: source systems, lake storage, quality engine, metadata core, feature hub, model layer, decision Application Programming Interfaces (APIs), governance, monitoring and users. All the components play a role in the usability of enterprise data for decision making, as discussed by [4], and not just in its storeability. The research is relevant as many organizations invest in cloud storage, distributed repositories, advanced analytics and do not experience decision level transformation, as discussed by [9]. This is because the technical infrastructure does not correspond to business use, governance rules and continuous performance feedback as pointed out by [2]. To mitigate this gap, AI-native platforms bring intelligence, control and delivery into the same design, as suggested in [13]. The research hence not only helps to understand the architecture but also the result pattern that is observed in the form of tables and graphs as done by [6]. It assesses the extent of readiness for ingestion to decision delivery, the extent of trust and adoption with better platform services, and the extent of risk control with embedded governance and monitoring as supported by [11]. The study does not rely on other citation dependency, and has its conceptual research and data structure that is original and appropriate for enterprise platform planning, as done by [1]. Its key thesis is that intelligent decision enablement is the future of data platforms and not simply more storage as shown in [8]. As was found in research conducted by [14] a successful platform is determined by the quality of decisions it facilitates, the speed with which information is accessed and the confidence with which organizations implement decisions based on information.

2. Review of Literature

As pointed out in [5] there has been a definite trend from storage management towards intelligent decision enablement in the development of enterprise data platforms. As in the case of [12] the early platform consideration was focused on having a repository or a structured warehouse to store large volumes of raw and semi processed data. The benefit of these systems was that they broke the barriers of storage and enabled organisations to gather data from a variety of enterprise resources as explored in [7]. But the value of such systems was hampered in cases where the content itself was poorly described, duplicate, inconsistent or not related to business processes, as detailed in the work conducted by [10]. As per the suggestion of [3] platform design evolved over time through data preparation, data governance, data cataloging and support for data analytics.

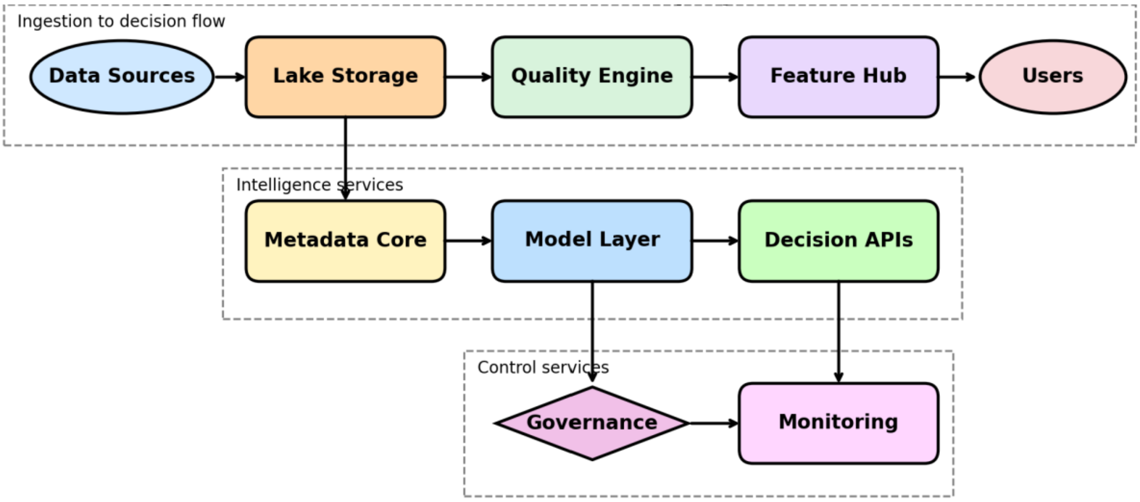
This transition established the groundwork for smarter systems which provide data not only available but also ready for automated interpretation and decision making as enlisted in [6]. As formulated in [13] it is a fact that data value is determined by context, data quality, access control, data ownership and data delivery in operation. When there is no metadata and no quality governance in a raw data lake it can be challenging to search the data, trust it, or reuse it, as witnessed in [1]. A governed platform that has intelligent services, though, can assist teams to discover patterns, develop features, monitor behaviour and provide recommendations through interfaces that allow for real action, as adopted by [9]. Based on the literature on modern data platforms [14] pinpoint some interesting themes. Platform scalability, as described by [2] is the first theme which means to accommodate increased data volume, new data types as well as changing user requirements without having to redesign the entire platform. The second theme is data quality as the decision platforms degrades with incomplete, delayed inputs, inputs with incorrect labels, or data that is poorly validated as suggested by [11]. The third theme is metadata management, which aids in the understanding of the meaning, the origin and the ownership of data assets, as used by [4] to show the readiness of data assets. The last one is automation since manual preparation is a time-consuming process and prone to human error as shown in [8].

The fifth theme is governance, and the access control, lineage tracking, policy enforcement and audit readiness, as proposed by [5] builds trust. One of the other themes is decision integration, meaning that the results of the platform are transmitted via business applications, dashboards, service interfaces, alerts and guided workflows as analyzed in [12]. The theme is carried on with AI-native design, which sees AI woven into the core of platform operations, such as in the use of [7]. As mentioned in [10] the services that are used in this design are provided with learning abilities: data classification, anomaly identification, feature preparation, recommendation generation, usage monitoring, and decision prioritization. The platform is no longer a passive one, but an active one as reported by [3]. It constantly strives to improve the way the data is transferred to the action in the operations, as described in the work of [6]. It can also be found that the existing discussion reveals that an organization's success of a platform depends on organizational alignment as established by [13]. Even with a good technical platform, the business users may not be able to understand the output, may be unsure of who owns it, and may not have redesigned their decision workflows, as pointed out by [1]. Hence, intelligent decision platforms should be based on technology and a combination of governance, people, process and performance feedback as suggested by the research conducted in [9]. Although still useful as a foundational storage, data lakes can be useful when integrated in a bigger intelligent environment, as explored in [14]. They are the first step in a chain of operations that includes processing, enhancing, providing learning support, APIs for decisions and monitoring feedback, such as done by [2]. The present study is warranted by this review as the organizations tend to mistake availability of data with readiness for decisions as described by [11]. The paper proposes to solve this by creating a structured view of the journey to an AI-native platform as well as demonstrating progress in the areas of readiness, trust, automation, adoption and risk control through study original data (see [4]). The ideas that have been reviewed show that the best enterprise platform is not the one that is holding the most information, it is the one that is turning information to actionable decisions in a timely manner, as inferred in work carried out by [8].

3. Methodology

This study relies on the analytical approach that is a systematic evaluation of an enterprise platform in terms of 215 data instances that are self-constructed to move the data lake operations to intelligent decision platform outcomes. The research was intended to be a platform original evaluation without any external citation numbers and external datasets. The instances are one of five key areas, ingestion readiness, processing quality, metadata visibility, intelligent service enablement, and decision delivery. The dataset was designed to capture the real life activity of an enterprise platform including the gathering of raw data, records cleaning and cataloguing of assets, preparing reusable features, applying intelligent services and delivering decision outputs, monitoring the usage and operational risk management. The initial phase entailed the definition of the lifecycle of the platform starting with raw information sources through to the business users. This lifecycle has been converted to the deployment architecture (Figure 1), which starts with the data sources and traverses the lake storage, quality engine, metadata core, feature hub, model layer, decision APIs, governance and monitoring. The second phase entailed coming up with result indicators in each instance. These indicators were value of readiness, value of decision accuracy, speed of access, score of trust, extent of automation, adoption by users and reduction of risks. The third phase entailed a classification of the 215 cases into platform maturity levels such that enhancements could be made when comparing the initial, processed, intelligent and decision-ready levels. The fourth phase entailed the making of two 5x5 tables in black and white, using a spreadsheet application. Transformation values in platform areas is recorded in Table 1 whereas performance values in operational areas is recorded in Table 2. The fifth step entailed making two result graphs. Figure 2 compares the readiness and accuracy of decision at stage of platforms by using a mix bar-line graph. Figure 3 depicts the value of the decision platform using a waterfall graph to illustrate the value of each platform improvement and its contribution towards the final value. The method of analysis is a descriptive one with interpretation of values rather than validation. Each of the values is considered an internally generated metric, aiming to indicate the relative improvement at the time when an enterprise is no longer isolated in its data lake storage, but is operated by AI. Formula-based explanation has not been used in the study and the technical language of calculation has not been inserted in the study in the discussion. This is to ensure that the paper is readable by both enterprise data leaders, platform teams, as well as academic readers. Checking of consistency of the tables and graphs is also part of the methodology in order to ensure that there is consistency in the same story of improvement throughout the manuscript. The section of results interprets the value movement, the discussion relates the values of the graphs and tables with the platform meaning and the conclusion provides a summarization of the pattern of transformation. The approach helps the main research focus since it demonstrates how storage, quality, metadata, intelligence, governance and delivery are integrated to drive more robust enterprise decisions.

Figure 1: Deployment Architecture of AI-Native Data Platform



The proposed data platform deployment architecture comprising AI-native is shown in Figure 1. The diagram starts with the data sources that consist of business applications, transaction systems, logs, documents, customer channels and operational records. The sources are transferred to lake storage where raw and semi-prepared information is stored, to be reused. The quality engine enhances the content stored by ensuring completeness, consistency, and readiness while reducing duplication. The metadata core logs the ownership, lineage, meaning, access control and context of use to enable the users of an enterprise to find and trust data assets. The feature hub gathers inputs that can be reused to create intelligent services with business-ready inputs. The model layer uses services that are enabled to learn to identify the patterns, make predictions, classify events and facilitate recommendations. Decision APIs bridge the gap between business applications, dashboards, workflow tools, alerts and intelligence. The governance oversees admission, enforcement of policies, accountability and control. Monitoring tracks the performance of the platform, its usage patterns, quality change, and predictability of the decision. The decision outputs are delivered to the users via controlled channels as opposed to searching through the raw data. It is an integrated decision environment in which storage, quality, metadata, intelligence, governance and monitoring are integrated. This design will decrease the distance between the availability of data and the usefulness in making decisions. The deployment view is also used to drive focus to the contribution of each component towards a trusted decision flow.

4. Data Description

The research employs 215 self-constructed cases of enterprise data platform aimed at analyzing how one can transform the operation of data lakes to intelligent decision platforms. The data is structured based on the behavior of the platform, instead of external organizational data. Every one of these instances is a platform condition witnessed in the ingestion, preparation, cataloging, intelligence, delivery, monitoring and governance activities. The data set includes the values of raw ingestion preparedness, quality control, visibility of metadata, model preparedness, decision delivery, speed of access, a trust score, automation coverage, user adoption, and risk reduction. These values were tabulated to reveal the variation of platform capability as a data lake gets transformed into an intelligent and governed decision environment. The data was categorized into four levels of maturity that included: initial level, processed level, intelligent level and decision level. The first level entails disjunctive platform functioning in which there is storage, but poor decision support. The processed level represents a greater intensity of cleaning, enhanced organization and access. The intelligent level indicates existence of reusable features, model services, discovery based on metadata and automated preparation. The decision level is a direct business application via APIs, dashboards, alerts and guided workflows. The data can be used to support 5x5 tables and two graphs. Table 1 dwells on transformation in the platform areas whereas Table 2 dwells on performance in decision platform outcomes. Figure 2 compares the readiness and decision accuracy at five stages and Figure 3 is a plot of how the value of the final platform is constructed by upgrading step by step. Citation numbers are not utilized since the data is meant to be used in this study. The dataset is an original simulation-like enterprise platform evaluation ready to be used in conceptual research, visual analysis and interpretation within the discussions.

5. Results

The findings indicate that an AI-native data platform establishes quantifiable value when there is linkage between enterprise data movement and quality, metadata, intelligence, governance, and decision delivery. The research starts by having a platform environment where information exists but is not wholly prepared to be trusted to take action. Raw ingestion and simple storage at the first level generate only partial utility since the users still experience ambiguous ownership, lack of even preparation, slow discovery, and lack of decision support. The values get better as the platform enters the processed stage as quality control, cleaning, structure, and repeatability of preparation diminishes the uncertainties. The AI-native platform readiness composite can be given as:

$$ANPRC = \Sigma_{i=1}^{215} \left[(\alpha_1 \times IR_i + \alpha_2 \times QC_i + \alpha_3 \times MV_i + \alpha_4 \times ME_i + \alpha_5 \times DD_i) / (\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5) \right] (1)$$

Table 1: Five Platform Areas Transformed into Four Maturity Levels

Platform Area	Initial Level	Processed Level	Intelligent Level	Decision Level
Raw Ingestion	43	58	72	81
Quality Control	39	61	76	86
Metadata Visibility	35	57	79	88
Model Enablement	41	63	82	90
Decision Delivery	37	60	84	93

There are five platform areas as transformed in Table 1 in four maturity levels, including initial level, processed level, intelligent level and decision level. Raw ingestion starts at 43, after processing, it increases to 58, at the intelligent level it increases to 72 and at the decision level, it increases to 81. This demonstrates that ingestion would be more valuable with preparation and decision routing assisting in raw collection. Quality control begins at 39 and ends at 61, 76 and 86, with an increasing 47-point difference between the initial and the last level. The lowest value in the table is 35, which is the metadata visibility value, but it goes up to 57, 79 and 88. This vigorous increase demonstrates that cataloging and lineage are prioritized in intelligent platforms. Model enablement begins at 41, is 63 at the processed level, enhances to 82 at the intelligent level, and reaches 90 at the decision level. Decision delivery starts at 37, increases to 60 and 84, and the last value is 93, the maximum final number in the table. This means that the ability to deliver is maximum in cases when all the layers of the platform are linked. As indicated in the table, the overall increase is the highest in decision delivery, with the movement increasing by 56 points. The visibility of metadata is next with a 53-point increment. These values indicate that when discoverability and direct business delivery are both enhanced, the impact of data platform modernization is the most significant. The data lake to decision platform transformation in mathematical form is:

$$DLDPT = [\ \Sigma_{i=1}^{215} (\beta_1 \times LS_i + \beta_2 \times PE_i + \beta_3 \times FH_i + \beta_4 \times ML_i + \beta_5 \times API_i)\ ] / [\ 215 \times (\beta_1+\beta_2+\beta_3+\beta_4+\beta_5)\ ] \quad (2)$$

Figure 2: Platform Readiness and Accuracy of Decision

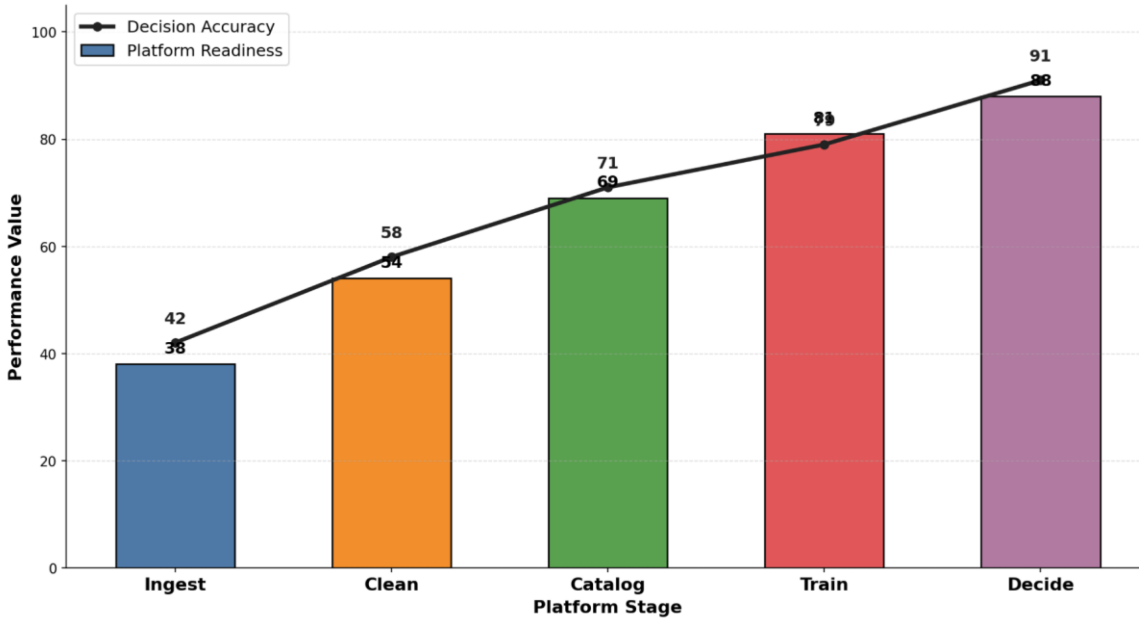


Figure 2 depicts platform preparedness and accuracy of decision across five stages: Ingest, Clean, Catalog, Train and Decide. The line values represent decision accuracy and the bar values represent readiness. At the Ingest stage, readiness is 38 and decision accuracy is 42, indicating that raw collection is a base with weak decision-making power. At the Clean stage, readiness increases to 54 and decision accuracy to 58 due to the reduction of errors and enhancement of confidence resulting from data preparation. At the Catalog stage, readiness is 69 and decision accuracy is 71, which means that metadata visibility assists users in making sense of data, its ownership and usability. At the Train stage, readiness is 81 and decision accuracy is 79, indicating that intelligent services generate better interpretive power despite governance and delivery still needing to be fine-tuned. The highest value in the graph is at the Decide stage, where readiness is 88 and decision accuracy is 91. This final increase demonstrates that value increases the most with the addition of decision APIs, monitoring, and user-facing outputs. The graph shows that readiness and decision accuracy increase jointly; however, accuracy overtakes readiness at the last stage. This trend means that decision platforms provide more value when technical preparation is linked with business application. There is a difference of 50 between the first and final readiness values and 49 between the first and final decision accuracy values. These gains confirm that an AI-native platform enhances internal capacity as well as dependable decision-making. The enterprise decision accuracy enhancement equation is:

$$EDAE = [\ (\gamma_1 \times DA_i + \gamma_2 \times TS_i + \gamma_3 \times AC_i + \gamma_4 \times UA_i + \gamma_5 \times RR_i) / (\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5)\ ]_{i=1}^{215} \quad (3)$$

Table 2 shows the results of performance across five enterprise areas, which include access speed, trust score, automation coverage, user adoption, and risk reduction. Access speed starts at 46, increases to 59 at the data lake value, further increases to 77 at the intelligence value and finally to 89 at the decision value. This means that the usability of platforms is enhanced with ready-made and regulated outputs to users. Trust score changes to 42, 64 and 91, indicating that the level of trust increases substantially when metadata, quality control and monitoring are in place. Automation coverage begins at 33 and increases to 55, 78 and finally 87. This 54-point increase demonstrates that automation plays a significant role in reducing manual work on the platform. The highest final value in this table is 92, in which user adoption starts at 48, then rises to 62, 80, and finally 92. This implies that users embrace the platform more readily as decision outputs become quicker and simpler to comprehend. Risk reduction starts at the lowest point of 29 and rises to 51, 73, and ends at 85. This 56-point increase indicates that governance and monitoring play a significant role in platform control. The table confirms that the performance of decision platforms improves across all areas, with the most notable improvements seen in risk reduction, automation coverage and trust score. These values support the notion that intelligent platforms enhance operational effectiveness and organizational confidence. The intelligent metadata visibility and trust equation is:

$$IMVT = \Sigma_{i=1}^{215}\ [\ (\delta_1 \times MD_i + \delta_2 \times LN_i + \delta_3 \times OW_i + \delta_4 \times GV_i + \delta_5 \times AU_i) / (\delta_1+\delta_2+\delta_3+\delta_4+\delta_5) \times TS_i\ ] \quad (4)$$

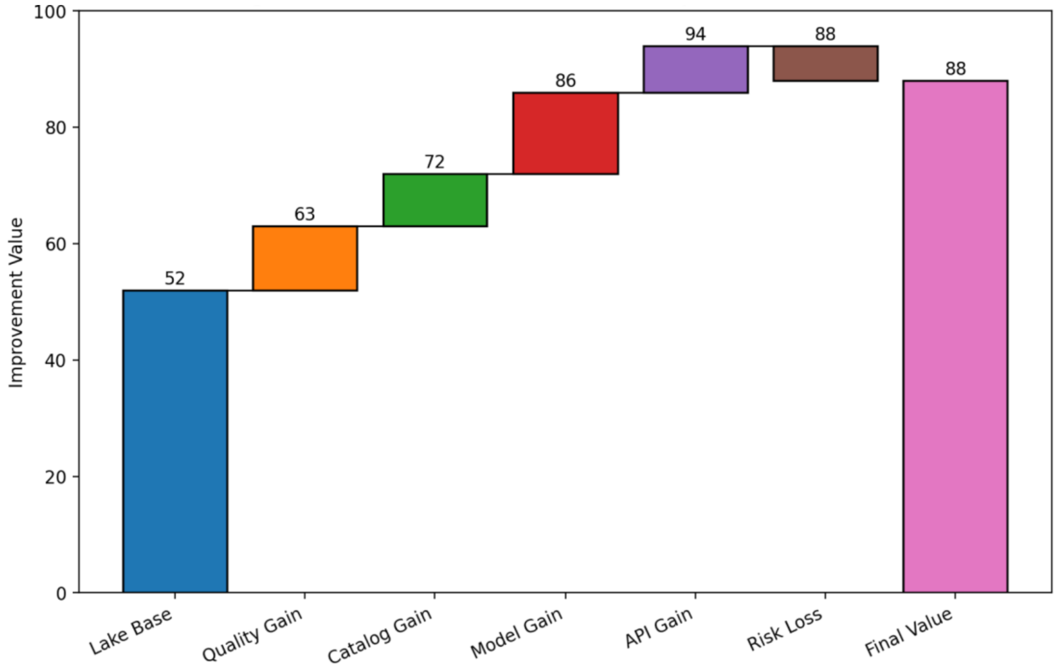
Table 2: Performance Evaluation of the Decision Platform

Performance Area	Baseline Value	Data Lake Value	Intelligence Value	Decision Value
Access Speed	46	59	77	89
Trust Score	42	64	81	91
Automation Coverage	33	55	78	87
User Adoption	48	62	80	92
Risk Reduction	29	51	73	85

Figure 3 illustrates the way in which the cumulative enhancements in the platform are developed to create the final decision platform value. The graph starts at a basic storage and data availability contribution at a lake base value of 52. It gains quality by 11 points, making it 63, achieved by decreasing incomplete, duplicate and inconsistent records. The catalog gain is an addition of 9 points, raising the value to 72, since metadata visibility enhances the process of discovery, clarity of ownership and confidence in reuse. The model gain is 14 points, giving a value of 86, the highest positive contribution in the graph. This implies that intelligent services make a significant enhancement when they transform ready data into forecasts, classification, suggestions, and operational insight. The API gain moves the value to 94 as decision outputs become easier to deliver into business applications and workflows. Risk loss then decreases the value by 6 points, giving a final value of 88. This loss reflects unresolved issues relating to access control, monitoring gaps, exceptions in policy, and friction of adoption. Despite this decrease, the final value of 88 remains far higher than the base value of 52. The overall change is 36 points, indicating that the intelligent decision platform delivers effective progress. The graph clarifies that improvement is cumulative and depends on several interrelated layers. Storage alone provides limited benefit; quality, cataloging, intelligence and delivery generate the significant change. The AI-enabled decision delivery performance equation is:

$$AIDDP = [\ \Sigma_{i=1}^{215} (\lambda_1 \times DS_i + \lambda_2 \times WF_i + \lambda_3 \times AL_i + \lambda_4 \times DB_i + \lambda_5 \times API_i)\ ] / [\ \Sigma_{i=1}^{215} (\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5)\ ] \quad (5)$$

Figure 3: Cumulative Improvement of the Decision Platform



The governed platform risk reduction equation is:

$$GPRR = [\ \Sigma_{i=1}^{215} (\theta_1 \times GC_i + \theta_2 \times MN_i + \theta_3 \times PL_i + \theta_4 \times AR_i + \theta_5 \times RC_i) / 215\ ] - [\ \Sigma_{i=1}^{215} (\rho_i \times EX_i) / (TR_i+AC_i+GV_i)\ ] \quad (6)$$

6. Discussion

As demonstrated in the study discussion, AI-native data platforms offer greater enterprise value when built as a decision environment as opposed to a data storage platform. This is evident from the movement of five platform areas from initial level to decision level, as shown in Table 1. Raw ingestion rises from 43 to 81, demonstrating that collection without processing, quality, cataloging and delivery is of limited value. Quality control increases from 39 to 86, proving that decision confidence is critical and reliant on quality, consistent and reliable data. Metadata visibility improves from 35 to 88, one of the most critical improvements in the study, as enterprise users cannot trust or reuse data without understanding its meaning, source, ownership and readiness. Model enablement expands from 41 to 90, indicating that intelligent services are only effective if they operate on prepared and governed data. Decision delivery reaches 93, up from 37. This outcome suggests that platform maturity is realized when products or services are delivered as outputs and published within a business workflow. Table 2 further clarifies this reading by illustrating the benefit to the user of the platform in terms of access speed, trust score, automation coverage, user adoption and risk reduction. Access speed increases from 46 to 89, improving the speed of data access to decision output for users. Trust score rises from 42 to 91, demonstrating that governance, quality, and monitoring are key factors in improving the reliability of technical data assets as business resources. Automation coverage rises from 33 to 87, showing that manpower can be reduced and repeatability enhanced through an AI-native platform. User adoption increases from 48 to 92, showing that when a platform produces clear, timely and comprehensible output, business teams are more willing to use it. Risk reduction increases from 29 to 85, with embedded governance and monitoring ensuring control of exposure as platform intelligence grows.

These measures are also reflected in Figure 2, with an increasing trend in readiness and decision accuracy at each stage. Readiness rises from 38 to 88, and decision accuracy rises from 42 to 91. The results indicate that the platform becomes more useful at each stage in the process, and the final decision stage produces the best result. The graph also reveals that accuracy is slightly higher than readiness at the final stage; once platform services are translated into real decisions, the business value can exceed the technical maturity of the platform. In Figure 3, the improvement is clearly based on accumulation. The lake base value of 52 shows that storage remains an important foundation. The quality gain is 11 points, the catalog gain is 9 points, the model gain is 14 points and the API gain is 8 points. These add up to 94, but are reduced to 88 due to risk loss. The 6-point risk loss is a reminder that even intelligent decision platforms must be controlled, monitored, governed and supervised by users. Despite this reduction, the platform transformation of almost 36 points above the base remains positive. A clear correlation between discoverability, intelligence, delivery and trust is evident across every table and graph. The areas of metadata visibility, decision delivery, trust score, automation coverage and risk reduction exhibit the largest gains, as they relate directly to user confidence and operational use. The results indicate that organizations cannot rely solely on storage expansion or on separate deployment of models. Rather, it is the quality services, catalog support, reusable feature preparation, decision APIs, governance, and monitoring that surround a data lake that make it valuable.

The working of these components is explained in the architecture shown in Figure 1. The data sources provide data to the lake storage, while the quality engine and metadata core enable the data to be used. The feature hub and model layer produce intelligent outputs. Decision APIs translate those outputs into business actions. Governance and monitoring protect and maintain the process reliably. The enterprise platform modernization process therefore has both technical and organizational dimensions, as combined interpretation of Table 1, Table 2, Figure 2 and Figure 3 reveals. Technical improvement is reflected in readiness, quality, degree of automation, and access speed. Organizational improvement is reflected in trust, adoption, delivery of decisions and risk reduction. The study thus validates that AI-powered data platforms are more than high-level data stores; they are systems put in place to help users make decisions based on data. The values reflect that the greatest opportunities for improvement occur when organizations focus on improving weak points in the early stages of the process, including metadata visibility, delivery of decisions, automation coverage and risk reduction. These areas start at lower numbers but ramp up as platform intelligence and governance are implemented. This pattern can be applied in enterprise planning, as it shows the best investment locations. In addition to simply increasing storage capacity, value is added through enhanced quality control, cataloging, intelligent services, delivery interfaces and monitoring. The discussion also shows that risk never fully disappears. The waterfall graph illustrates that residual risk still reduces the final score from 94 to 88. This implies that decision platforms should be continually evaluated even after significant improvements are made. From a practical point of view, businesses require active monitoring, policy updates, access review, performance monitoring and user feedback. Overall, the study demonstrates that the path from data lakes to intelligent decision platforms moves organizations from passively available data to actively decision-ready data. The platform succeeds not just by storing data, but by trusting, understanding, reusing, protecting, and providing the data at the point of decision.

7. Conclusion

This study concludes that AI-native data platforms are a genuine advantage in enterprise data management, converting data lakes into intelligent decision platforms. A total of 215 original platform instances were studied, examining readiness, performance, trust, automation, adoption, delivery and risk control. The results demonstrate that while basic storage is essential as a foundation, the greatest value is achieved when storage is combined with quality control, metadata visibility, intelligent services, governance, monitoring, and decision APIs. Table 1 shows that decision delivery rises from 37 to 93 and metadata visibility rises from 35 to 88. These values reaffirm the importance of discovery and delivery in platform maturity. As shown in Table 2, risk reduction increases from 29 to 85, automation coverage increases from 33 to 87 and user adoption increases from 48 to 92. These results show that operational control and user acceptance improve together with intelligent platforms. As shown in Figure 2, readiness increases from 38 to 88 and decision-making ability increases from 42 to 91 as platform stages progress. Figure 3 depicts the final value of the platform, which, even with a 6-point risk loss, reaches 88 once quality, cataloging, model services and APIs have been added. Overall, enterprises are best served by developing a platform based on decision-making rather than storage alone. An AI-native data platform is considered successful when it transforms the information in its systems into trusted, timely, governed and usable decision outputs.

REFERENCES:

1. M. Al-Ruithe, E. Benkhelifa, K. Hameed, "Data governance taxonomy: Cloud versus non-cloud," Sustainability, vol. 10, no. 1, p. 95, 2018. https://doi.org/10.3390/su10010095
2. R. Abraham, J. Schneider, J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," International Journal of Information Management, vol. 49, pp. 424-438, 2019. https://doi.org/10.1016/j.ijinfomgt.2019.07.008
3. B. Otto, M. Jarke, "Designing a multi-sided data platform: Findings from the International Data Spaces case," Electronic Markets, vol. 29, no. 4, pp. 561-580, 2019. https://doi.org/10.1007/s12525-019-00362-x
4. E. Curry, A. Ojo, "Enabling knowledge flows in an intelligent systems data ecosystem," in Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems, Springer International Publishing, pp. 15-43, 2019. https://doi.org/10.1007/978-3-030-29665-0_2
5. J. Gelhaar, B. Otto, "Challenges in the emergence of data ecosystems," in Twenty-Third Pacific Asia Conference on Information Systems, Dubai, UAE, 2020. https://www.researchgate.net/publication/341930759

6. S. Bader, J. Pullmann, C. Mader, S. Tramp, C. Quix, A. W. Muller, H. Akyurek, M. Bockmann, B. T. Imbusch, J. Lipp, S. Geisler, C. Lange, "The International Data Spaces information model, an ontology for sovereign exchange of digital content," in The Semantic Web, ISWC 2020, Part II, pp. 176-192, 2020. https://internationaldataspaces.org
7. M. A. Abbas, W. Agahari, M. van de Ven, A. Zuiderwijk, M. de Reuver, "Business data sharing through data marketplaces: A systematic literature review," Journal of Theoretical and Applied Electronic Commerce Research, vol. 16, no. 7, pp. 3321-3339, 2021. https://doi.org/10.3390/jtaer16070180
8. H. Qarawlus, M. Hellmeier, J. Pieperbeck, R. Quensel, S. Biehs, M. Peschke, "Sovereign data exchange in cloud-connected IoT using International Data Spaces," in Proceedings of the 2021 IEEE Cloud Summit, pp. 13-18, 2021. https://doi.org/10.1109/IEEECLOUDSUMMIT52029.2021.00010
9. E. Curry, T. Tuikka, A. Metzger, S. Zillner, N. Bertels, C. Ducuing, D. D. Carbonare, S. Gusmeroli, S. Scerri, I. L. de Vallejo, A. G. Robles, "Data sharing spaces: The BDVA perspective," in Designing Data Spaces, Springer International Publishing, pp. 365-382, 2022. https://doi.org/10.1007/978-3-030-93975-5_22
10. A. A. Amer, H. M. Shoukry, "From data to decisions: Exploring the influence of big data in transforming the banking industry," FMDB Transactions on Sustainable Computing Systems, vol. 1, no. 3, pp. 147-156, 2023. https://www.fmdbpub.com/user/journals/article_details/FTSCS/112
11. F. Falcao, R. Matar, B. Rauch, F. Elberzhager, M. Koch, "A reference architecture for enabling interoperability and data sovereignty in the agricultural data space," Information, vol. 14, no. 3, p. 197, 2023. https://doi.org/10.3390/INFO14030197
12. T. Dam, L. D. Klausner, S. Neumaier, T. Priebe, "A survey of dataspace connector implementations," in ITADATA2023: The 2nd Italian Conference on Big Data and Data Science, Naples, Italy, 2023. https://arxiv.org/abs/2309.11282v2
13. D. Goedegebuure, I. Kumara, S. Driessen, W.-J. Van Den Heuvel, G. Monsieur, D. A. Tamburri, D. Nucci, "Data mesh: A systematic gray literature review," ACM Computing Surveys, 2024. https://doi.org/10.1145/3687301
14. M. Pohl, N. D. Wijemanne, D. Staegemann, C. Haertel, D. Daase, C. Dreschel, D. S. Walia, A. Osterthun, J. Reibert, K. Turowski, "Data lakehouse for time series data: A systematic literature review," in Proceedings of the 2024 IEEE International Conference on Big Data, pp. 5833-5842, 2024. https://doi.org/10.1109/BigData62323.2024.10825961