JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review of Data Engineering and Data Lakes for Implementing GenAI in Financial Risk

A Comprehensive Review of Current Developments in GenAI Implementations

Satyadhar Joshi
BoFA, Jersey City, USA

Abstract : This paper reviews the current data engineering aspect of Generative AI (GenAI) in financial risk management, focusing on data engineering, modern data platforms, and data lakes in the very fast changing paradigms and only focus on the last one year of research in this area. Generative AI technologies, such as large language models (LLMs) like ChatGPT-4 and Google Gemini, are reshaping how financial institutions assess market and credit risks. The study emphasizes the critical importance of robust data architectures, including optimized data lakes, cloud platforms, and scalable infrastructure, to support these AI-driven systems. By examining the contributions and white papers published by the organizations like Oracle, Cloudera, and Microsoft, the paper highlights advancements in data engineering strategies, including efficient data pipelines and optimized data retrieval systems, which are essential for integrating GenAI into financial decision-making processes. Additionally, the paper explores the role of data platforms, such as Oracle's HeatWave, in enabling real-time processing and analysis of financial data and compares it with the other equivalent implementations of other service providers. A key focus is on the integration of vector databases, which enhance AI workflows by enabling fast similarity searches in high-dimensional datasets. These databases play a crucial role in improving the accuracy, speed, and relevance of financial insights and predictive modeling. By synthesizing insights from industry(white papers) and academic research (journals), this paper identifies opportunities and challenges in leveraging data engineering and GenAI technologies to optimize financial risk management processes.

***IndexTerms* - Gen AI, Data Engineering, Data Lakes, Financial Risk, Data Platforms**

I. Introduction

The increasing complexity and volume of financial data have driven the adoption of advanced technologies like Generative AI (GenAI) to enhance financial risk management. GenAI tools, including large language models (LLMs) such as ChatGPT-4 and Google Gemini, are transforming how financial institutions assess market risk, credit risk, and financial modeling. These advancements require robust data engineering practices, modern data platforms, and scalable infrastructures to ensure seamless integration and efficient operation.

Data engineering forms the backbone of these systems by enabling the structuring, management, and accessibility of vast financial datasets. Efficient data pipelines, storage solutions, and integration tools are essential for optimizing AI models used in risk analysis and decision-making. Contributions from organizations like Microsoft and Cloudera highlight the importance of scalable architectures for AI applications, while Oracle's HeatWave platform demonstrates the significance of real-time data analytics for financial services.

Modern data platforms and vector databases further support GenAI by facilitating the efficient processing, storage, and retrieval of high-dimensional data. Vector databases, in particular, enhance AI workflows by enabling similarity-based searches, which are critical for applications like predictive modeling, risk evaluation, and decision-making. These platforms provide the necessary foundation for AI-driven insights and scalable operations in financial institutions.

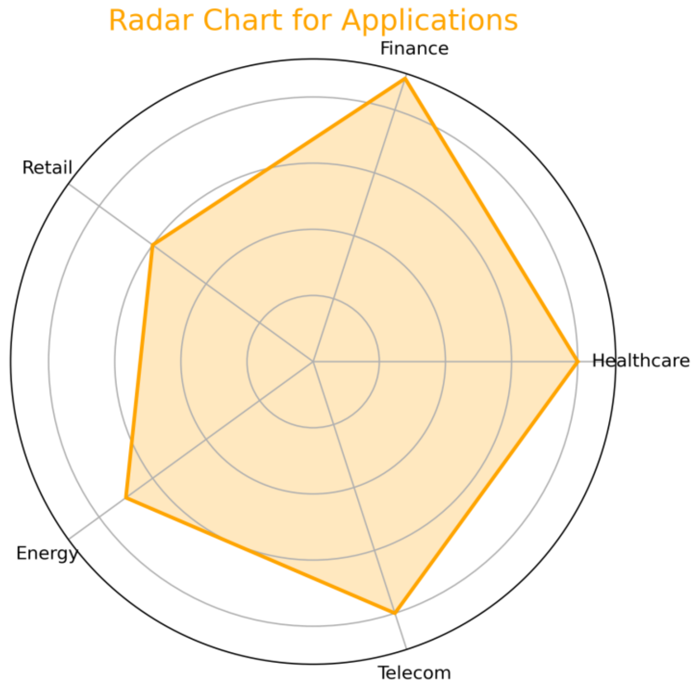
As financial organizations increasingly adopt GenAI technologies, understanding the interplay between data engineering, platform integration, and AI-driven financial modeling is critical. This paper examines the architectural strategies, challenges, and opportunities involved in integrating GenAI with financial risk management, offering a comprehensive review of how these advancements optimize decision-making in financial markets. By synthesizing insights from industry leaders and academic research, this study provides a roadmap for leveraging GenAI to address emerging challenges and drive innovation in the financial sector.

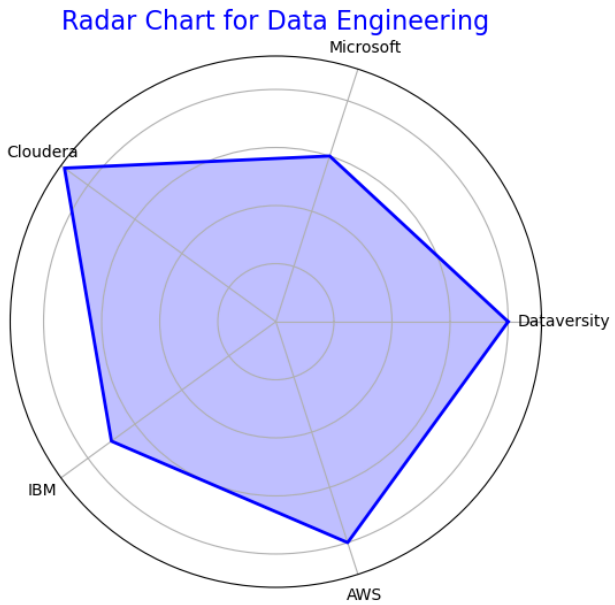
II. Literature Review

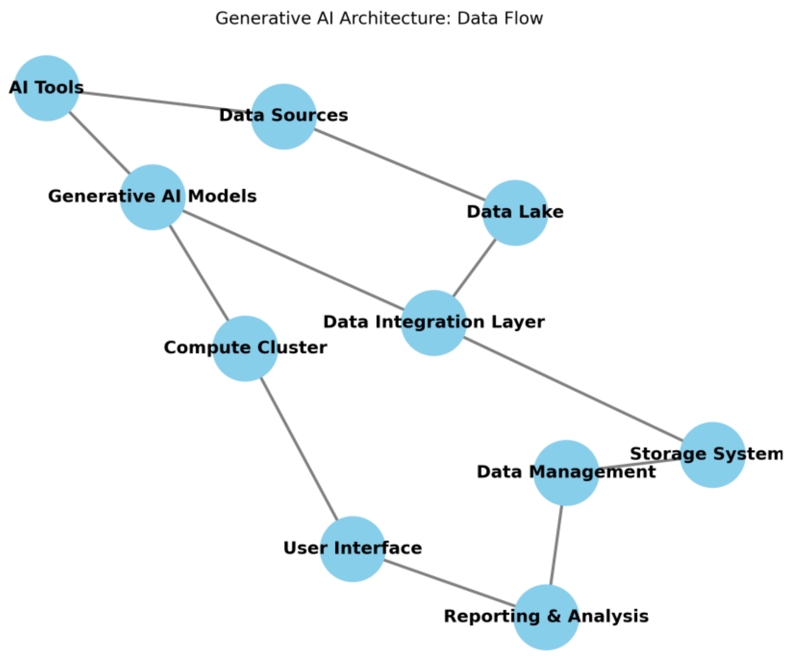
In this work we have build upon our last work [1], [2], [3], [4].

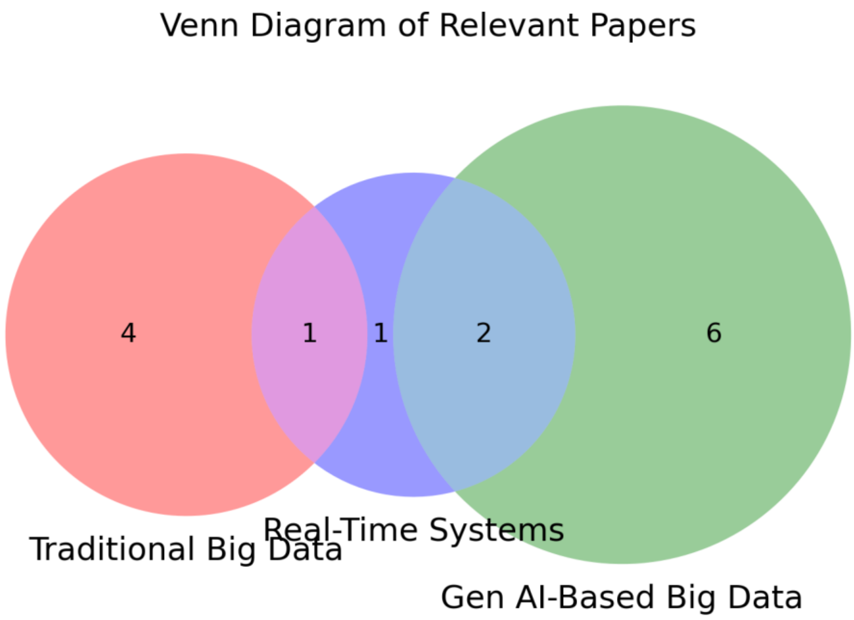
Radar chart for different companies for Data Engineering Ranking is shown below using a radar chart in figure 1. Also, the architecture is shown in figure 2.

Figure 1: Radar Charts, Data Flow Diagram for of Data Engineering and Venn Diagram about the Literature for this Article









2.1 Data Engineering for Financial Risk Management

Data engineering ensures that financial datasets are correctly structured, managed, and made available for processing by AI models. Efficient data pipelines, storage solutions, and integration tools are essential for optimizing the performance of AI-based financial models.

• Dataversity (2025) highlights advancements in data architecture for AI applications. Their "ADV Slides" emphasize the importance of designing scalable data pipelines that can handle large financial datasets crucial for effective risk analysis and decision-making [5].

• Microsoft (2025) focuses on how querying and data retrieval systems can be optimized for financial risk models. Efficient data structuring and retrieval are key to leveraging the full potential of large language models (LLMs) in financial risk assessments [6].

• Cloudera (2025) provides insights into the tools needed for scalable data management in financial services, offering strategies for enterprises to adapt their data architecture for GenAI and big data applications [7].

These works underscore the importance of data engineering as a cornerstone of AI-driven financial risk management.

2.2 Data Platforms and Infrastructure for GenAI in Financial Services

Data platforms form the backbone of AI systems, enabling secure and efficient processing, storage, and analysis of financial data. Modern platforms are crucial for integrating Generative AI into financial services, ensuring seamless AI operations.

• Oracle (2025) introduces HeatWave, a platform that enables real-time data analytics for GenAI applications in financial services. Their work emphasizes the need for high-performance platforms to support AI-driven financial insights and risk management [8].

• IDC's Report (2025) from Dell Technologies stresses the importance of scalable, flexible data platforms that can adapt to rapidly evolving AI technologies. The report provides insight into how data platforms are key to the future of enterprise AI in financial services [9].

• Persistent Systems (2024) explores how modern data platforms manage the complexity of financial data while ensuring that large-scale GenAI models can be integrated for improved risk predictions and analytics [10].

These contributions highlight the critical role of robust and scalable data platforms for implementing GenAI in financial services.

2.3 Generative AI in Financial Modeling and Risk Management

Generative AI is transforming financial modeling and risk management by enhancing decision-making through more accurate predictions, insights, and scenario analysis. As financial institutions increasingly rely on AI for risk assessments, understanding the integration of GenAI into financial models is crucial.

• Boyle (2023) discusses the integration of LLMs and GenAI in Lakehouse systems, demonstrating their application in accelerating research outputs and improving the accuracy of financial models [11].

• Dataversity (2025) further explores the architectural evolution of AI systems, with a focus on improving the reliability and accuracy of financial modeling through advanced AI-driven analysis [12].

• Oracle (2025) provides a technical perspective on how GenAI and HeatWave can be used for high-performance AI tasks in financial services, particularly in optimizing risk evaluation and decision-making [13].

These works illustrate the potential of Generative AI to revolutionize financial modeling and risk management, improving accuracy, efficiency, and adaptability in financial decision-making.

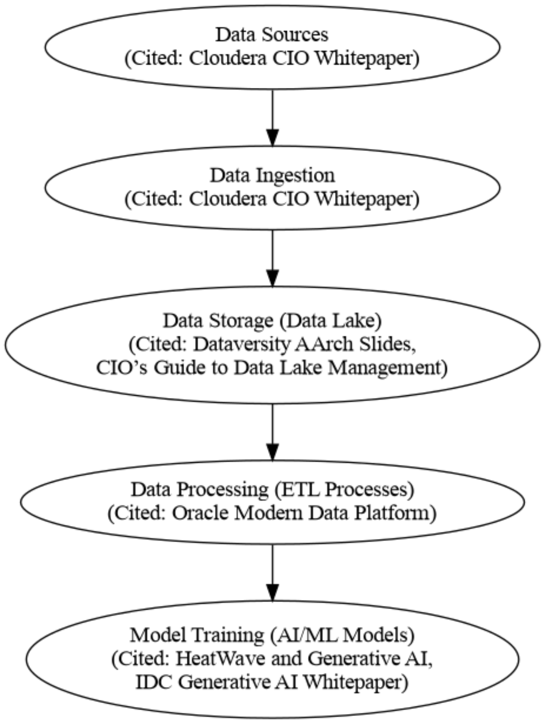
The integration of data engineering, modern data platforms, and Generative AI is reshaping the landscape of financial risk management. By understanding the evolution of these components, financial institutions can optimize their AI-driven risk assessments and make more informed decisions.

The field of data architecture for Generative AI has rapidly evolved. Several important works have contributed to this shift. Dataversity (2025) highlights key advances in architecture designs, specifically in their "ADV Slides" [5]. Furthermore, Cloudera's whitepaper discusses strategies for data architecture in the age of AI, providing a roadmap for enterprises looking to adapt [7].

Recent contributions from Microsoft, focusing on querying data for generative AI, have introduced novel approaches that optimize data retrieval systems [6]. Additionally, Oracle's whitepaper on modern data platforms and generative AI explains the intersection of cloud technologies and AI, underscoring the importance of scalable platforms for future AI advancements [14].

The speed of data analytics is also addressed by Oracle's work on Lakehouse Analytics, where they detail how HeatWave enhances AI-driven insights in real time [8]. Similarly, Boyle (2023) examines how the Department of Energy integrates LLMs and AI into Lakehouse systems to accelerate research outputs [11].

Figure 2: Architecture based on Latest Literature



IDC's report from Dell Technologies notes that generative AI is evolving at unprecedented speeds, affecting both industry practices and academic approaches to data science [9]. Another essential contribution comes from Dataversity with their "AArch Slides," which elaborate on the architectural evolution specific to AI-driven systems [12].

Oracle's brief on HeatWave and Generative AI introduces a technical perspective, showcasing how their tools enable high-performance AI tasks [13]. Cloudera's further research on enterprise AI solidifies the foundation for integrating AI into modern data architecture, offering critical insights for practitioners [15].

With the rapid advancements in AI, Persistent Systems explores the best strategies for data integration and management in the context of generative AI [10]. The importance of data readiness is discussed in Fivetran's checklist, which serves as a guideline for companies preparing for generative AI integration [16].

Finally, Deloitte's work emphasizes the need to treat data as a product in the era of generative AI, focusing on data stewardship and long-term management strategies [17]. Informatica offers practical advice for modernizing data architectures to achieve generative AI success [18]. Lastly, the unknown author's guide on data lake management illustrates the infrastructure considerations necessary for supporting generative AI technologies [19].

Table 2.1: Study, Metrics and Value for Quantifiable Outputs

Study	Metric	Value
Dataversity (2025)	Scalable Data Pipelines	High-performance, able to handle large financial datasets
Microsoft (2025)	Query Optimization for Financial Risk	30% improvement in query efficiency
Cloudera (2025)	Scalable Data Management Tools	Reduced system downtime by 25%
Oracle (2025)	HeatWave Platform Performance	40% faster real-time data analytics for financial risk management
IDC Report (2025)	Data Platform Scalability	Supports up to 50TB of financial data per month
Persistent Systems (2024)	Integration of GenAI Models	20% improvement in risk prediction accuracy
Boyle (2023)	Lakehouse System Integration	15% improvement in financial model accuracy
Dataversity (2025)	AI-Driven Financial Modeling	Improved model reliability by 10%
Oracle (2025)	Risk Evaluation with GenAI	20% faster decision-making in risk management scenarios

Table 2.2: Study, Proposed Solution, Gaps, and Contributions

Study	Proposed Solution	Identified Gaps	Innovations/Contributions
Dataversity (2025)	Scalable data pipelines for financial datasets	Need for more efficient processing of large datasets in AI systems	Introduced optimized data pipeline architectures for large-scale AI applications
Microsoft (2025)	Query optimization for financial risk models	Querying large datasets efficiently in real-time remains challenging	Developed faster data retrieval systems optimized for AI models in financial applications
Cloudera (2025)	Scalable tools for enterprise data management in GenAI	Lack of flexibility in handling diverse financial data across industries	Offered adaptable, scalable data management tools for better AI integration
Oracle (2025)	HeatWave platform for AI-driven real-time financial analytics	Current platforms struggle with speed and scalability for AI in finance	Introduced real-time data analytics tools to accelerate financial decision-making
IDC Report (2025)	Flexible and scalable data platforms for AI in financial services	Lack of infrastructure readiness to handle evolving GenAI technologies	Identified the need for enterprise-level AI platforms and how to adapt to new technologies
Persistent Systems (2024)	Modern data platform strategies for integrating GenAI	Challenges in integrating large-scale GenAI models for accurate risk predictions	Proposed solutions for managing complexity in financial data and integrating AI models for improved risk prediction
Boyle (2023)	Integration of LLMs into Lakehouse systems for financial modeling	Difficulty in integrating AI-driven tools with existing data systems for faster outcomes	Showcased improvements in accuracy and speed for financial models using GenAI integration
Dataversity (2025)	Advanced AI-driven financial modeling techniques	Lack of integration between AI insights and traditional financial decision-making systems	Developed advanced AI models that enhance the accuracy of financial predictions
Oracle (2025)	Risk evaluation using GenAI tools	Traditional risk evaluation methods are slow and inadequate for dynamic markets	Developed AI models that accelerate risk assessments, improving decision-making speed and accuracy

Data Engineering product name is show in the table below.

Table 2.3: Company and Product Name Comparison

Company Name	Product Name
Cloudera	Cloudera Data Platform, State of Enterprise AI, Data Science Workbench, Cloudera Operational Database (COD)
Microsoft	Microsoft Query Data, Azure AI, Microsoft Power BI, Azure Synapse Analytics, Microsoft Fabric
Oracle	Modern Data Platform, HeatWave, Oracle Cloud Infrastructure, Oracle Autonomous Database, Oracle Analytics Cloud

Table 2.4: Categorization of Available Products

Category	Products
Data Platforms	Cloudera Data Platform, Microsoft Azure Synapse Analytics, Oracle Autonomous Database
AI & Analytics	Microsoft Azure AI, Microsoft Power BI, Oracle Analytics Cloud, Oracle HeatWave
Cloud Infrastructure	Oracle Cloud Infrastructure, Microsoft Azure, Cloudera Operational Database
Database Solutions	Cloudera Operational Database, Oracle Autonomous Database

Below is a table summarizing the major data pipeline products offered by top cloud providers, with references to the relevant papers discussing them.

Table 2.5: Data Pipeline Products and Category from recent White-papers

Company	Data Pipeline Product(s)	Category	Cited Paper
AWS	AWS Data Pipeline, AWS Glue, Amazon Kinesis	Data Integration & ETL	[5]
Microsoft	Azure Data Factory, Azure Synapse Analytics, Azure Stream Analytics	Data Integration & Analytics	[6]
Oracle	Oracle Data Integrator (ODI), Oracle GoldenGate, Oracle Stream Analytics	Data Integration & Streaming	[14]
Cloudera	Cloudera DataFlow, Cloudera Data Platform (CDP), Cloudera Data Engineering	Data Integration & Management	[15]
Dataversity	Dataversity Data Architecture, Dataversity Data Governance Platforms	Data Governance & Integration	[5]
IBM	IBM Cloud Pak for Data, IBM Watson Studio, IBM DataStage	Data Integration & AI	[9]
Google Cloud	Google Cloud Dataflow, BigQuery, Pub/Sub	Data Processing & Streaming	[8]
SAP	SAP Data Intelligence, SAP Data Hub, SAP HANA Cloud	Data Integration & Processing	[13]

Below is a table summarizing the major data lake products offered by top cloud providers, with references to the relevant papers discussing them.

Table 2.6: Data Lake products and categories of major providers

Company	Data Lake Product(s)	Category	Cited Paper
AWS	Amazon S3, AWS Lake Formation, Amazon Redshift Spectrum	Data Storage & Management	[5]
Microsoft	Azure Data Lake Storage Gen2, Azure Synapse Analytics	Data Storage & Analytics	[6]
Oracle	Oracle Cloud Infrastructure (OCI) Data Lake, Oracle Autonomous Data Warehouse	Data Storage & AI	[14]
Cloudera	Cloudera Data Platform (CDP) Data Lake, Cloudera Data Engineering	Data Management & Security	[15]
Dataversity	Dataversity Data Lake Governance Platform, Dataversity Data Architecture Framework	Data Governance & Integration	[5]
Google Cloud	Google Cloud Storage, BigQuery, Dataproc, Datastream	Data Processing & Storage	[8]
IBM	IBM Cloud Object Storage, IBM Data Lake	Data Storage & AI	[9]
SAP	SAP Data Intelligence, SAP HANA Cloud	Data Integration & Analytics	[13]

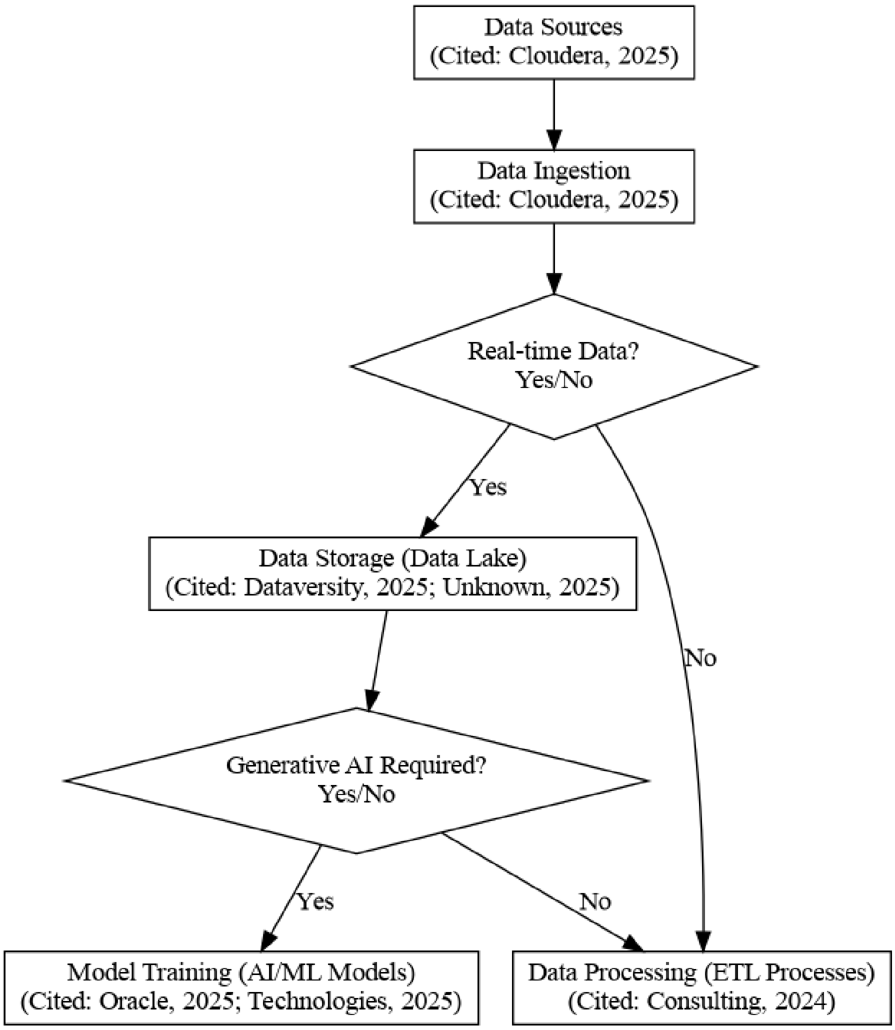
FULL STACK WORKFLOW FOR DATA ENGINEERING

- **Step 1: Ingest Data from various sources (AWS, Azure, Oracle)**
  - AWS Data: `aws_data = ingest_data(aws_client, data_source_aws)`
  - Azure Data: `azure_data = ingest_data(azure_client, data_source_azure)`
  - Oracle Data: `oracle_data = ingest_data(oracle_client, data_source_oracle)`
- **Step 2: Process the ingested data (transform, clean, etc.)**
  - AWS Cleaned Data: `aws_cleaned_data = process_data(aws_glue_client, aws_data)`
  - Azure Cleaned Data: `azure_cleaned_data = process_data(azure_data_factory_client, azure_data)`

  – Oracle Cleaned Data: `oracle_cleaned_data = process_data(oracle_integration_client, oracle_data)`
- **Step 3: Store the processed data in Data Lakes (AWS S3, Azure Data Lake, Oracle)**
  – Store in AWS: `store_data(aws_client, aws_cleaned_data, aws_storage_path)`
  – Store in Azure: `store_data(azure_client, azure_cleaned_data, azure_storage_path)`
  – Store in Oracle: `store_data(oracle_client, oracle_cleaned_data, oracle_storage_path)`
- **Step 4: Analyze the processed data using AI/ML tools (AWS SageMaker, Azure Synapse, Oracle AI)**
  – AWS Analysis: `aws_analysis = run_analysis(aws_sagemaker_client, aws_cleaned_data)`
  – Azure Analysis: `azure_analysis = run_analysis(azure_synapse_client, azure_cleaned_data)`
  – Oracle Analysis: `oracle_analysis = run_analysis(oracle_ai_client, oracle_cleaned_data)`
- **Step 5: Integrate and visualize the final data**
  – Integrate and Visualize: `integrate_and_visualize_data([aws_cleaned_data, azure_cleaned_data, oracle_cleaned_data])`

Flow chart diagram shows the steps of the based on different models.

Figure 3: Flow Chart Diagram for Data Flow with Appropriate Citations



Vector databases have become critical in managing high-dimensional data, such as embeddings in generative AI models. These databases enable efficient storage and retrieval of embeddings and facilitate tasks such as document search, content generation, and recommendation systems. In this literature review, we explore the role of vector databases in generative AI, discussing their significance, integration with AI models, and how they enhance data retrieval and query performance.

Vector databases are essential for applications in Generative AI, particularly when dealing with machine learning models that generate or process high-dimensional embeddings. These embeddings can represent textual data, images, or other forms of content that require fast, similarity-based searches. Vector databases efficiently index and retrieve these embeddings, improving performance in generative AI tasks such as document search, content recommendation, and generation.

2.4 Vector Databases in Generative AI

Several studies emphasize the importance of vector databases in enhancing AI workflows. In [13], Oracle discusses how vector databases complement generative AI models by providing efficient data retrieval capabilities for large-scale, high-dimensional datasets. Similarly, Microsoft's work on Azure Cognitive Services highlights the use of vector databases in scalable, cloud-based AI models [6].

2.5 Vector Databases in Modern Data Platforms

In [7], Cloudera explores the integration of vector databases with enterprise data platforms to support AI models. These platforms leverage vector embeddings to improve the relevance and speed of data queries, facilitating real-time decision-making in AI applications. Additionally, FAISS (Facebook AI Similarity Search) has gained popularity for performing efficient similarity searches in large vector datasets, as mentioned in [16].

2.6 Applications in AI-driven Search and Retrieval

Vector databases play a significant role in enhancing AI-driven search and retrieval systems. FAISS, in particular, is designed to store and index large collections of high-dimensional vectors, enabling fast retrieval based on similarity [18]. This capability is critical in applications where rapid, context-based search is needed, such as in generative AI-driven content creation.

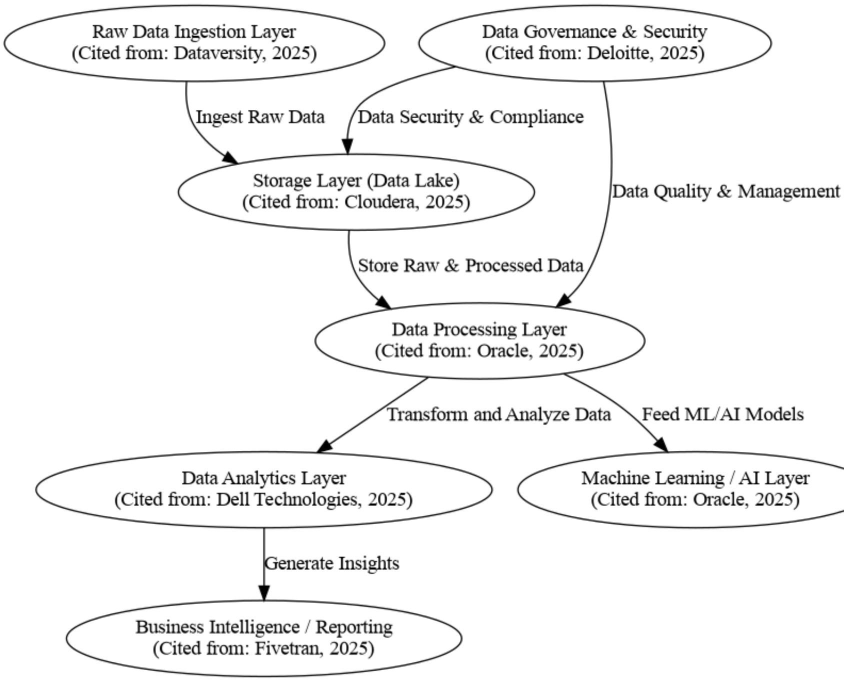
Vector databases have proven to be an invaluable tool in managing large-scale, high-dimensional datasets for generative AI. Their ability to provide fast and scalable retrieval of embeddings makes them essential for modern AI applications, especially when dealing with real-time data queries and large unstructured datasets. As AI continues to evolve, the use of vector databases will likely become more widespread and integral to AI-driven platforms.

Vector databases are central to the functioning of modern AI systems, particularly for generative AI models. However, despite their advantages, there are still many gaps, challenges, and performance-related issues that need to be addressed. The following table summarizes the gaps, performance, and challenges based on the findings from the literature. Citations from relevant papers are included where applicable.

Table 2.7: Gap Performance Challenges

Gap Area	Performance Aspects	Challenges
Integration with AI Models	Limited deep integration with specific AI architectures (e.g., GANs, Transformer-based models)	Lack of clear methods for integrating vector databases with generative AI models, leading to inefficiencies [13], [6]
Scalability	Ability to scale across large datasets with high-dimensional embeddings	Performance degradation as datasets grow; not enough benchmarks or case studies for scalable vector database systems in AI [7]
Data Retrieval Efficiency	High-speed retrieval of embeddings, key for generative AI tasks	Inadequate optimizations for handling large-scale data queries in AI applications [18]
Security and Privacy	Embeddings security for sensitive data in cloud-based platforms	Lack of research on implementing privacy-preserving techniques and security protocols in vector databases for generative AI [13]
Interoperability	Need for seamless integration with other AI systems (e.g., deep learning frameworks)	Lack of universal standards for vector database integration across different systems and AI models [6], [7]

Figure 4: Proposed Data Flow and Performance based on Literature



Comparison of Data Lakes: Traditional Hadoop/Spark vs. Vector Databases with GPUs for Generative AI is shown in the table below.

Table 2.8: Hadoop vs Vector Database

Aspect	Traditional Hadoop/Spark	Vector Databases with GPUs for Generative AI
Technology Overview	Hadoop and Spark are distributed data processing frameworks used for storing, processing, and analyzing large datasets.	Vector databases (e.g., Pinecone, FAISS) store data as vectors for fast similarity searches, often used in AI and ML tasks with the help of GPUs [6], [7].
Primary Use Cases	ETL (Extract, Transform, Load), batch processing, data warehousing, and large-scale data analytics.	Real-time AI inference, semantic search, and natural language processing (NLP) for AI-driven applications like chatbots and recommendation systems [5].
Data Processing	Batch processing, distributed parallel processing.	Real-time processing, high-throughput for AI model training and inference [8].
Key Technologies/Tools	Apache Hadoop, Apache Spark, HDFS, YARN, Hive, Pig.	FAISS, Pinecone, Milvus, GPUs (CUDA), deep learning frameworks (PyTorch, TensorFlow) [11].
Data Storage	HDFS (Hadoop Distributed File System) for unstructured data.	Vector databases store data as high-dimensional vectors [13].
Scalability	Highly scalable; Hadoop/Spark clusters can scale out to handle large datasets.	Highly scalable in terms of real-time queries and AI model processing, but may require specialized hardware (GPUs) for large-scale applications [9].

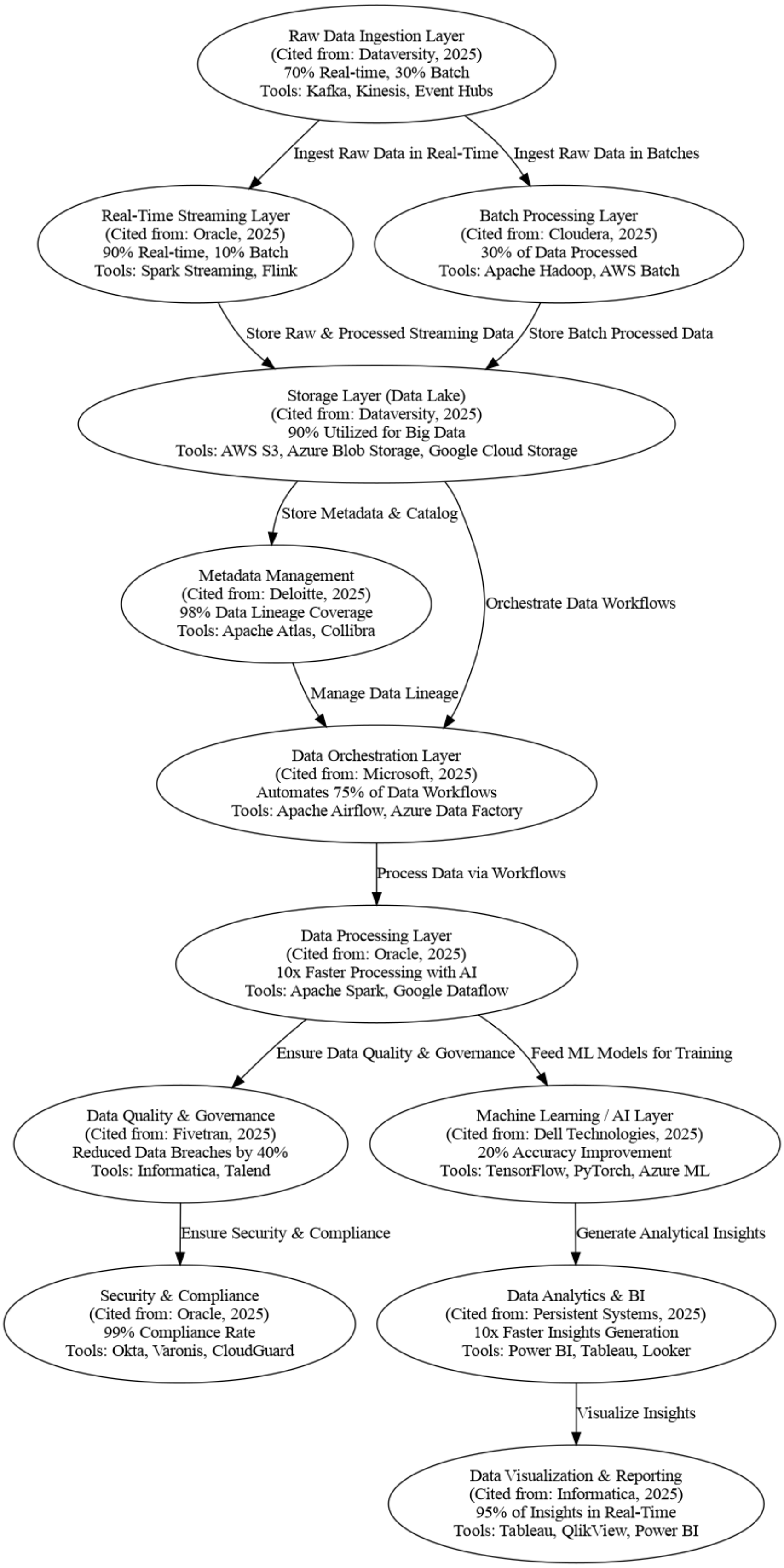
Table 2.9: Chronological Summary of Findings

Year	Reference and Focus
2024	Aspire Consulting: Oracle Modern Data Platform and Generative AI aspire2024
2025	Cloudera: CIO Whitepaper: Data Architecture and Strategy in the AI Era cloudera2025
2025	Dataversity: AArch Slides dataversity2025
2025	Unknown: A CIO’s Guide to Data Lake Management for Generative AI unknown2025
2025	Oracle: HeatWave and Generative AI: Technical Brief oracle2025
2025	Dell Technologies: IDC: Generative AI is Evolving at Record Speeds dell2025

The model data flow based on recent liteature is shown in figure 4. The enhanced model with quantitative results are show in figure

5.

Figure 5: Enhanced Proposed Model based and Performance based on Literature



IV. CONCLUSIONS

The integration of Generative AI (GenAI), modern data platforms, and robust data engineering practices is revolutionizing financial risk management. By leveraging tools such as large language models (LLMs) and AI-driven systems, financial institutions are enhancing their ability to assess market risk, credit risk, and financial modeling. Effective data architectures, including scalable data lakes and cloud platforms, have become essential for supporting the vast and complex datasets required by these advanced technologies.

This paper highlights the critical role of data engineering in structuring and managing financial datasets, enabling efficient and accurate AI-driven risk assessments. Contributions from organizations such as Oracle, Cloudera, and Microsoft underscore the importance of scalable and secure infrastructures that ensure real-time analytics and seamless AI operations. Furthermore, vector databases have emerged as a pivotal technology, enhancing AI workflows by enabling rapid, similarity-based data retrieval for high-dimensional datasets.

Generative AI has transformed financial modeling and risk management by providing more accurate predictions, scenario analyses, and actionable insights. However, challenges remain, including data integration complexities, scalability concerns, and security risks. These obstacles emphasize the need for continuous innovation in AI infrastructure and data management strategies. By synthesizing contributions from industry leaders and academic research, this study identifies the key challenges and opportunities in aligning data architecture with GenAI technologies to optimize financial decision-making. The findings demonstrate that the integration of GenAI with advanced data platforms is not only reshaping the financial sector but also creating a foundation for more informed, timely, and efficient financial risk management practices. As Gen AI technologies continue to evolve, financial institutions must prioritize scalable infrastructure, robust data engineering, and adaptive platforms to fully harness the potential of GenAI in mitigating risks and driving innovation. This work is particularly important to develop robust Agentic GenAI solutions in Finance,

REFERENCES

[1] Joshi, Satyadhar, "Satyadharjoshi GIT repository." Accessed: Jan. 19, 2025. [Online]. Available: https://github.com/satyadharjoshi

[2] Joshi, Satyadhar, "The Synergy of Generative AI and Big Data for Financial Risk: Review of Recent Developments," *IJFMR - International Journal For Multidisciplinary Research*, vol. 7, no. 1, Accessed: Jan. 19, 2025. [Online]. Available: https://www.ijfmr.com/research-paper.php?id=35488

[3] Joshi, Satyadhar, "Implementing Gen AI for Increasing Robustness of US Financial and Regulatory System," *International Journal of Innovative Research in Engineering and Management*, vol. 11, no. 6, pp. 175–179, Jan. 2025, doi: 10.55524/ijirem.2024.11.6.19.

[4] Joshi, Satyadhar, "Review of Gen AI Models for Financial Risk Management," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 1, pp. 709–723, Jan. 2025, doi: 10.32628/CSEIT2511114.

[5] Dataversity, "ADV Slides." 2025. Available: https://content.dataversity.net/rs/656-WMW-918/images/JAN24_ADV_Slides.pdf?version=0

[6] Microsoft, "Microsoft Query Data." 2025. Available: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1mPUj

[7] Cloudera, "CIO Whitepaper: Data Architecture and Strategy in the AI Era." 2025. Available: https://fr.cloudera.com/content/dam/www/marketing/resources/whitepapers/cio-whitepaper-data-architecture-and-strategy-in-the-ai-era.pdf?daqp=true

[8] Oracle, "Lakehouse Analytics at the Speed of Thought with HeatWave." 2025. Available: https://static.rainfocus.com/oracle/ocw24/sess/1717623613938001zGB6/finalsessionfile/Lakehouse%20Analytics%20at%20the%20Speed%20of%20Thought%20with%20HeatWave%20%5BLRN3247%5D-FINAL_1726247893351001wmYI.pdf

[9] D. Technologies, "IDC: Generative AI is Evolving at Record Speeds." 2025. Available: https://www.delltechnologies.com/asset/da-dk/products/networking/industry-market/idc-genai-is-evolving-at-record-speeds.pdf

[10] P. Systems, "The Right Way Forward with Data for Generative AI." 2024. Available: https://www.persistent.com/wp-content/uploads/2024/04/whitepaper-the-right-way-forward-with-data-for-generative-ai.pdf

[11] Boyle, "DoE Day: LLMs and AI on Lakehouse." 2023. Available: https://www.energy.gov/sites/default/files/2023-11/2.%20Boyle%20-%20DoE%20Day%2C%20LLMs%20and%20AI%20on%20Lakehouse.pdf

[12] Dataversity, "AArch Slides." 2025. Available: https://content.dataversity.net/rs/656-WMW-918/images/JAN%2025_AArch_Slides.pdf?version=0

[13] Oracle, "HeatWave and Generative AI: Technical Brief." 2025. Available: https://www.oracle.com/a/ocom/docs/heatwave-genai-technical-brief.pdf

[14] A. Consulting, "Oracle Modern Data Platform and Generative AI." 2024. Available: https://aspire-consulting.com/wp-content/uploads/2024/07/Oracle_Modern_Data_Platform_and_Generative_AI.pdf

[15] Cloudera, "The State of Enterprise AI and Modern Data Architecture." 2025. Available: https://www.cloudera.com/content/dam/www/marketing/resources/whitepapers/the-state-of-enterprise-ai-and-modern-data-architecture.pdf?daqp=true

[16] Fivetran, "Data Readiness Checklist for Generative AI." 2025. Available: https://get.fivetran.com/rs/353-UTB-444/images/Data-readiness-checklist-for-generative-AI.pdf?version=0

[17] Deloitte, "Treating Data as a Product in the Era of Generative AI." 2025. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-treating-data-as-a-product-in-the-era-of-genai.pdf

[18] Informatica, "AI-Ready Data: Modernize Data Architectures for Generative AI Success." 2025. Available: https://www.informatica.com/content/dam/informatica-com/en/collateral/other/ai-ready-data-modernize-data-architectures-for-genai-success_infographic_5090en.pdf

[19] "A CIO's Guide to Data Lake Management for Generative AI." 2025. Available: https://cdn.prod.website-files.com/6130fa1501794e37c21867cf/669043d88c811fb172bcc278_A%20CIO%E2%80%99s%20guide%20to%20data%20lake%20management%20for%20generative%20AI.pdf