

# Optimizing Data Mart Architecture in Palantir Foundry a TOPSIS-Based Evaluation Framework

Suresh Pandipati\*

Enterprise Data Architect and IT Senior Project Manager, Richardson, TX, USA

## Abstract

A Data Mart in Palantir Foundry is a curated dataset designed to support specific business use cases, enabling efficient data access and analysis. Building a Data Mart in Foundry involves ingesting raw data, transforming it using Code Repositories (Pipelines, Functions, and Workflows), and organizing it in Ontology for easy discovery and governance. Foundry's Schema Workflows and Quiver Tables help structure data effectively, ensuring high performance. With granular access controls and versioning, Foundry enables secure collaboration. A well-built Data Mart empowers analysts and applications with reliable, up-to-date insights while maintaining data integrity and scalability across the organization. The significance of researching Data Mart building in Palantir Foundry development lies in its impact on data-driven decision-making, efficiency, and scalability. Data Marts streamline data access by organizing domain-specific datasets, enhancing performance and user experience. In Palantir Foundry, efficient Data Mart design ensures optimized data pipelines, governance, and interoperability, enabling organizations to extract actionable insights with minimal redundancy. Understanding its construction aids in improving data modeling, security, and analytics workflows, leading to faster, more informed business decisions. This research contributes to best practices for scalable, resilient, and efficient data ecosystems in modern enterprises using Foundry's powerful capabilities. The methodology for building a Data Mart in Palantir Foundry begins with Requirement Analysis, where business needs, data sources, and user requirements are identified. Next, Data Ingestion is performed using Foundry's pipelines to integrate structured and unstructured data. The Data Transformation phase leverages Foundry's Code Repositories or Transform features to clean, enrich, and normalize data. Schema Design follows, using Object Builders and the Foundry Ontology to define the data model. To enhance performance, Data Optimization techniques such as caching, indexing, and partitioning are applied. Security & Governance measures ensure access controls and audit policies are in place. Rigorous Validation & Testing ensures data quality through Foundry's testing frameworks. Finally, the Deployment & Maintenance phase involves continuous monitoring and optimization to keep Data Mart efficient and up to date. Cloud-Native Data Mart is getting first place of the table and Self-Service Data Mart is getting last place of the table.

**Keywords:** Cloud-Native Data Mart, Hybrid Storage Mart, Real-Time Analytics Mart, Ease of Integration, User Adoption Rate (%)

## Introduction

Building a comprehensive and integrated Enterprise Data Mart has been a transformative initiative aimed at unifying financial and operational data to enhance data reporting, analytics, business intelligence, and visualization. This project has been instrumental in improving data accuracy, accessibility, and overall decision-making efficiency within the organization.[1] One of the most significant achievements of this project was the successful design of an end-to-end data architecture that drastically improved data accuracy and accessibility. By implementing advanced data modeling techniques and robust ETL (Extract, Transform and Load) pipelines, the project significantly reduced data processing time, allowing the organization to generate real-time insights into production and market trends. This capability empowered business leaders to make informed, data-driven decisions, thereby enhancing operational efficiency and market responsiveness.[2] A key innovation introduced through this

initiative was the implementation of predictive analytics capabilities.

This enhancement improved demand forecasting accuracy, which in turn contributed to optimized inventory management and a reduction in operational costs. By leveraging historical data and market trends, the organization could make precise predictions regarding supply chain requirements, effectively reducing waste, minimizing stockouts, and ensuring efficient resource allocation.[3] Another critical aspect of this project was the streamlined data flow from multiple sources, ensuring that teams had access to high-quality, consistent data for reporting and analysis. This centralized approach facilitated better decision-making across various business units, enhancing overall operational efficiency. By developing and maintaining robust Palantir pipelines, the project further reduced data processing time and provided real-time insights into critical business operations. These enhancements allowed the organization to respond more swiftly to changing conditions and make more strategic decisions based on accurate and timely data.[4] From a financial perspective, the introduction of predictive analytics had a direct impact on cost management. Enhanced financial accuracy enabled the organization to optimize expenses, reduce project waste, and minimize stockouts. As a result, operational project costs were significantly reduced, demonstrating the tangible value of a well-integrated and intelligent data infrastructure. Security was also a major consideration in this project. By implementing role-based access controls, the organization ensured that sensitive data was accessible only to authorized personnel. These security measures not only protected proprietary and confidential information but also ensured compliance with regulatory standards such as FDA 21 CFR

**Received date:** February 06, 2025 **Accepted date:** March 28, 2025;  
**Published date:** March 09, 2025

\*Corresponding Author: Suresh Pandipati, Enterprise Data Architect and IT Senior Project Manager, Richardson, TX, USA, E- mail: [suresh.pandipati@gmail.com](mailto:suresh.pandipati@gmail.com)

**Copyright:** © 2025 Suresh P. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Part 11, which governs electronic records and signatures in the healthcare and pharmaceutical industries. Compliance with these stringent standards further reinforced the organization's commitment to data security and integrity.[5] The integration of various data sources into a unified platform was another major achievement.

This effort facilitated comprehensive, cross-functional analysis, allowing for improved collaboration across departments such as marketing, operations, and finance. By providing a single source of truth, the data mart enabled teams to work with accurate and consistent data, ultimately driving better business outcomes and enhancing operational efficiency.[6] To ensure scalability and long-term sustainability, a robust and scalable data infrastructure was developed. This future-proof architecture was designed to handle increasing volumes of data as the organization continued to grow. The ability to scale without requiring major reengineering ensured that the organization could continue expanding its data capabilities while maintaining operational agility and efficiency.[7] The technological backbone of this project comprised a combination of cloud-based and on-premise technologies. The integration of Azure Data Lake, Palantir Foundry, SQL Server, Python, SSIS, and Power BI for visualization provided a solid foundation for data processing and analytics. Additionally, Azure Data Factory and Kafka were utilized to achieve seamless data flow between systems through API integration and real-time data streaming.[8]

The project was executed using Agile methodology, ensuring continuous collaboration between business analysts, functional consultants, and development teams. The adoption of Agile principles allowed for the alignment of automated pipelines with evolving business requirements, ensuring that data solutions remained relevant and effective. By integrating ETL test scripts into Agile sprints, new features were automated and tested within the same sprint cycle, reducing validation time and ensuring seamless business functionality integration.[9] Real-time test reporting was another key component of this project. By implementing automated reporting in workshops, stakeholders were provided with real-time insights into project operations. Dashboards highlighted key performance indicators such as defect trends, test coverage, and test pass rates, enabling data-driven decision-making. Additionally, the integration of the ETL platform with ServiceNow ensured efficient defect management, as all identified issues were automatically logged and tracked within the issue tracking system.[10] A significant milestone was the achievement of 100% test coverage for critical ERP workflows. Automation of both positive and negative test cases ensured comprehensive coverage of user roles and transactions within SAP and Ariba. This level of test optimization further contributed to the stability and reliability of the Enterprise Data Mart.[11] The business impact of this initiative was substantial. By automating high-volume, repetitive tasks such as order creation, project cost calculations, and invoice processing, manual effort was reduced by approximately 400 hours per month. This led to significant cost savings and allowed employees to focus on more strategic tasks. Additionally, the improved efficiency and accuracy of data processes contributed to an additional revenue generation of \$2.9 million while reducing application processing time by 48%.[12]

The project also delivered an impressive data accuracy rate of 99.8%, ensuring that decision-makers had access to reliable and high-quality data. Faster time to market was another key benefit, as automated regression cycles enabled the organization to deploy ERP updates and patches more quickly. This was particularly beneficial for high-demand product lines, where rapid system updates were critical to maintaining competitiveness. System stability was another area of improvement, as automated ETL scripts and tests were executed after every code change. This minimized the risk of defects being introduced into production, thereby enhancing the

overall stability and reliability of the Enterprise Data Mart. Additionally, the integration of Generative AI (Gen AI) technology further streamlined financial processes for underground mining equipment dealers. This initiative resulted in an annual revenue boost of \$1.3 million, showcasing the potential of AI-driven automation in optimizing financial operations.[13] The development and implementation of the Enterprise Data Mart in Palantir Foundry were transformative for the organization. By unifying financial and operational data, the project enabled complete and accurate data reporting, analytics, and business intelligence. The strategic use of advanced data modeling, predictive analytics, robust ETL pipelines, and real-time data streaming significantly improved data accessibility, accuracy, and security. The integration of automation, Agile methodology, and AI-driven insights further enhanced efficiency, cost management, and decision-making capabilities. Ultimately, this initiative not only delivered substantial financial and operational benefits but also positioned the organization for continued success and growth in the evolving data-driven landscape.[14]

Data marts play a crucial role in data analytics, providing domain-specific subsets of data that facilitate efficient decision-making. Within Palantir Foundry, data mart development is streamlined through an integrated platform that offers robust tools for data integration, transformation, governance, and analysis. This introduction explores the concept of data marts, their significance, and how they can be effectively built within the Palantir Foundry ecosystem. A data mart is a subject-oriented database that provides curated data for specific business units or teams. Unlike enterprise data warehouses (DWH), which store vast amounts of data across an organization, data marts focus on a smaller scope, optimizing performance for analytical queries and reporting. There are different types of data marts, including dependent data marts, which derive from a central data warehouse and often maintain consistent data models; independent data marts, which are built directly from operational systems without relying on a centralized data warehouse; and hybrid data marts, which combine elements of both dependent and independent data marts to meet complex business needs.[15] Palantir Foundry is an advanced data operations platform that enables seamless integration, transformation, and analysis of enterprise data. Foundry's Data Mart Development capabilities allow users to create structured, accessible, and well-governed data assets for specific business functions.

Some of the key benefits of building data marts in Foundry include enhanced performance, as narrowing down datasets improves query performance; data governance, which ensures secure access and compliance with regulatory requirements; ease of use, as intuitive tools allow business users to access and analyze data without deep technical expertise; and automation and scalability, where automated data pipelines ensure up-to-date and scalable data marts.[16] The process of building a data mart in Palantir Foundry follows several key steps. The first step is data sourcing and ingestion, where data is gathered from various systems through ETL/ELT pipelines, direct integrations with databases, APIs, cloud storage, or real-time streaming data pipelines. Once data is ingested, data transformation is performed using code-based transformations (SQL, Python, Spark) or low-code tools, with ontology modeling helping to structure and organize data for easy access. After transformation, data storage and structuring are crucial, with options including relational tables for fast querying, time-series data for historical analysis, and graph databases for relationship-based analytics.[17] Ensuring data access and security is a fundamental part of data mart development in Foundry. This is achieved through granular access controls, audit logging, and data lineage tracking, which together ensure secure handling and regulatory compliance.

Once the data mart is structured and secured, it can be used to build analytical dashboards leveraging Foundry Contour for visual exploration,

Foundry Fusion for AI/ML-driven insights, or integrating with external tools such as Tableau, Power BI, and Jupyter Notebooks. To maximize efficiency in data mart development, it is essential to follow best practices. These include designing for performance by optimizing queries and using indexing and partitioning techniques, ensuring data consistency by aligning models with enterprise standards, enabling self-service analytics to empower business users, automating data updates through scheduled pipelines, and monitoring data quality using validation rules and anomaly detection mechanisms.[18] Building a data mart in Palantir Foundry enables organizations to create targeted, efficient, and well-governed analytical solutions. By leveraging Foundry's powerful data integration, transformation, and visualization capabilities, businesses can ensure seamless access to relevant insights. A well-structured data mart strategy enhances decision-making, operational efficiency, and data-driven innovation, making it a critical component of modern enterprise analytics.[19]

## Material and Method

### Alternative:

The five data mart architectures Cloud-Native Data Mart, Hybrid Storage Mart, Real-Time Analytics Mart, Enterprise Data Hub, and Self-Service Data Mart each offer distinct advantages and trade-offs based on performance, scalability, integration, and user adoption.

The Cloud-Native Data Mart leverages cloud infrastructure, providing scalability and cost efficiency, but its adoption may be hindered by concerns over data security and cloud dependency. Despite a strong performance score, it shows a low user adoption rate, indicating potential challenges in usability or implementation.

The Hybrid Storage Mart combines both on-premises and cloud storage, offering flexibility in data management. While it maintains balanced performance across criteria, it does not particularly excel in any one area, making it a stable but not a standout choice.

The Real-Time Analytics Mart is the top performer in terms of computational efficiency and rapid data processing, making it ideal for organizations requiring real-time insights. However, its slightly lower user adoption suggests complexity in deployment and usage.

The Enterprise Data Hub is well-integrated across various business functions and has the highest user adoption rate, indicating its effectiveness in centralized data management.

Lastly, the Self-Service Data Mart prioritizes accessibility, allowing non-technical users to interact with data easily. Its strong adoption rate compensates for its moderate technical performance, making it a user-friendly solution.

### Evaluation preference:

The Performance Score reflects the efficiency, speed, and reliability of a data mart in handling data processing tasks. A higher score indicates better computational power, responsiveness, and scalability in managing

large datasets. Performance is critical for ensuring smooth operations, minimizing latency, and optimizing data retrieval for business insights.

The Scalability Index measures the ability of a data mart to expand in response to growing data volumes and increasing user demands. A high scalability score means the system can handle more data and users without compromising speed or efficiency. This is essential for businesses expecting future data growth and requiring long-term adaptability.

Ease of Integration assesses how well a data mart can be incorporated into existing IT infrastructures, including compatibility with different data sources, analytics tools, and enterprise systems. A higher score means the data mart requires minimal effort for deployment and works seamlessly with other platforms, reducing technical barriers and implementation costs.

The User Adoption Rate (%) indicates how widely a data mart is embraced by users within an organization. A high adoption rate suggests that the system is user-friendly, accessible, and meets business needs effectively. It also reflects how well training, usability, and support contribute to making the data mart an essential tool for decision-making.

## TOPSIS METHOD

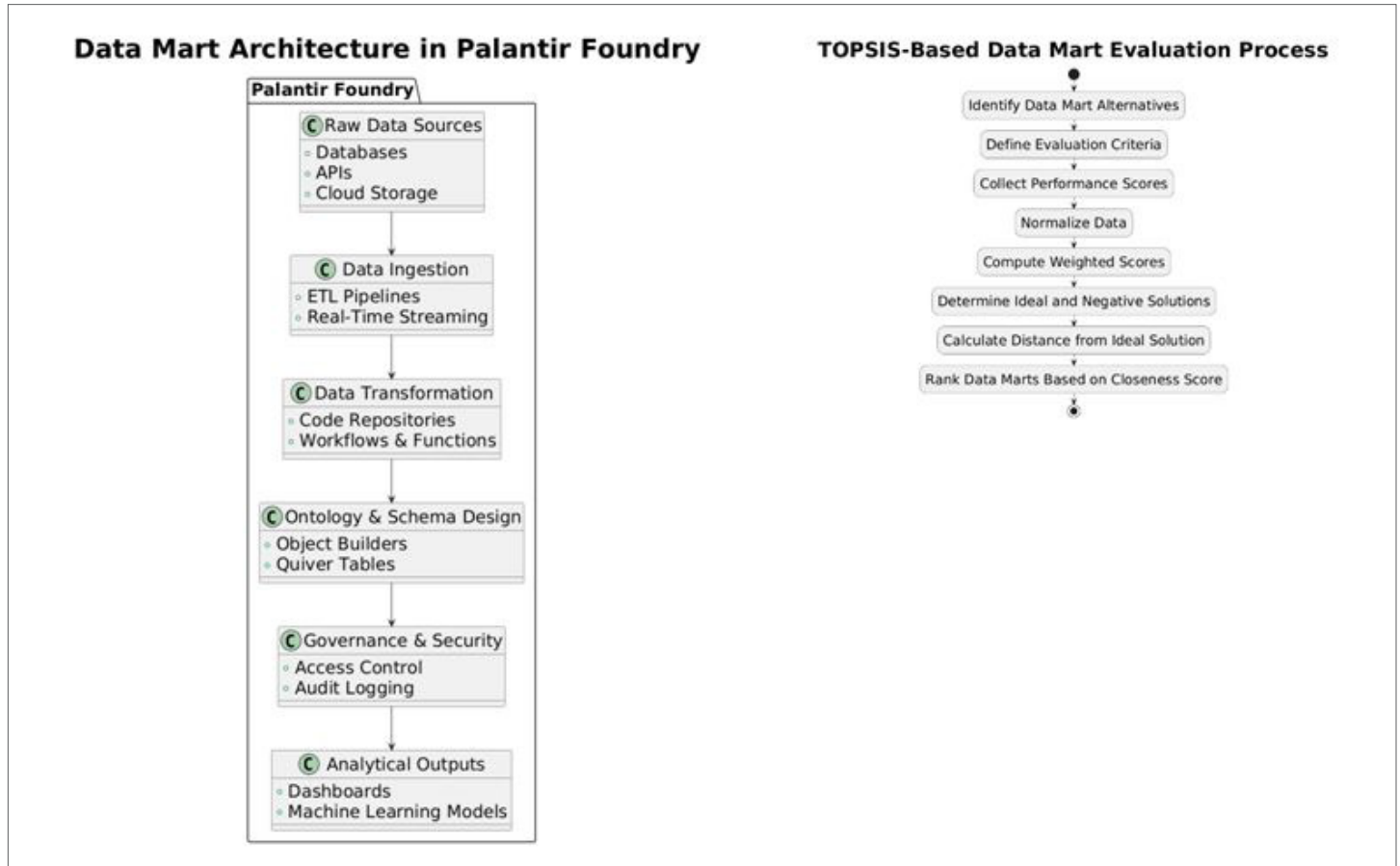
The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a widely used multi-criteria decision-making (MCDM) method that helps rank and select the best option from a set of alternatives based on multiple conflicting criteria. This technique is particularly useful in business intelligence, data analytics, and IT infrastructure selection, where decision-makers must evaluate different solutions based on quantitative metrics. The fundamental principle of TOPSIS is that the best alternative should have the shortest Euclidean distance from the ideal solution and the farthest distance from the negative-ideal solution.[20] The TOPSIS method is particularly advantageous because it provides a clear, quantitative ranking system that integrates both the best and worst case scenarios. Unlike simpler ranking methods, it does not just consider the highest-scoring option but evaluates how each alternative balances multiple criteria.

This makes it highly applicable in complex decision-making environments, such as selecting data marts in a business intelligence platform, where factors like performance, scalability, integration, and user adoption must all be considered simultaneously.[21] For example, in evaluating different data mart architectures such as Cloud-Native Data Mart, Hybrid Storage Mart, Real-Time Analytics Mart, Enterprise Data Hub, and Self-Service Data Mart, the TOPSIS method provides a structured approach to rank them based on Performance Score, Scalability Index, Ease of Integration, and User Adoption Rate. The results help organizations make an informed choice by selecting the most suitable data mart architecture based on empirical analysis rather than subjective preference. Overall, logical, and practical decision-making tool that enhances the objectivity of selecting the best option among multiple alternatives, making it an essential methodology for business analytics, engineering, and IT system evaluations.[22]

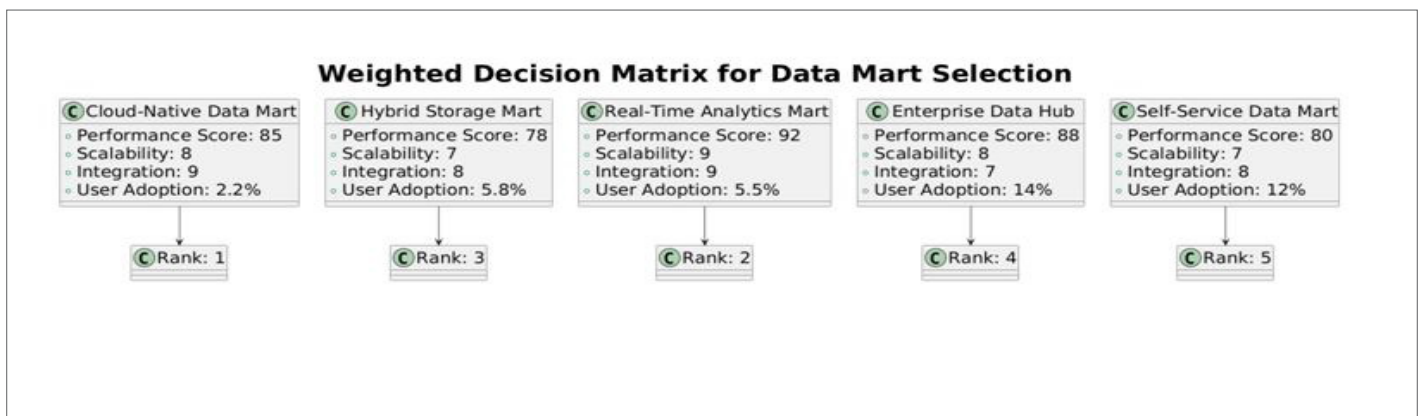
Table1. Data Mart Building in Palantir Foundry Development

|                          | Performance Score | Scalability Index | Ease of Integration | User Adoption Rate (%) |
|--------------------------|-------------------|-------------------|---------------------|------------------------|
| Cloud-Native Data Mart   | 85                | 8                 | 9                   | 2.2                    |
| Hybrid Storage Mart      | 78                | 7                 | 8                   | 5.8                    |
| Real-Time Analytics Mart | 92                | 9                 | 9                   | 5.5                    |
| Enterprise Data Hub      | 88                | 8                 | 7                   | 14                     |
| Self-Service Data Mart   | 80                | 7                 | 8                   | 12                     |

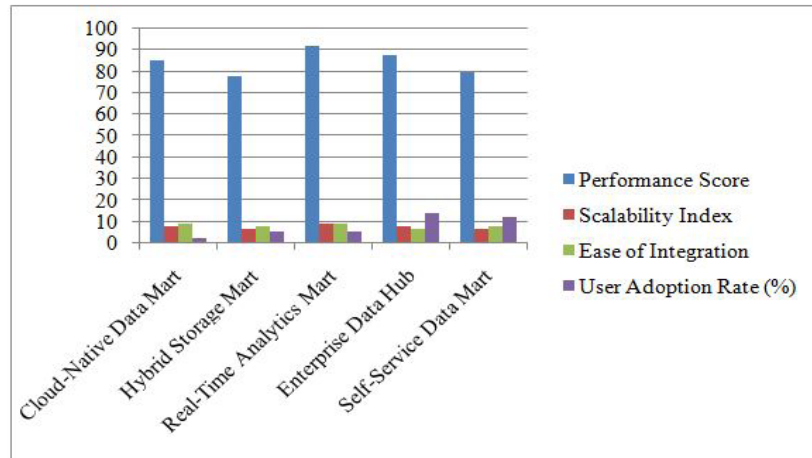
Table 1 presents the evaluation of different Data Mart Building approaches in Palantir Foundry Development using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The table provides a comparative analysis based on four key criteria: Performance Score, Scalability Index, Ease of Integration, and User Adoption Rate (%). Among the data marts, the Real-Time Analytics Mart achieves the highest Performance Score (92) and high values in Scalability Index (9) and Ease of Integration (9), indicating its superior efficiency and adaptability for real-time data processing. However, its User Adoption Rate (5.5%) is relatively moderate compared to others. The Enterprise Data Hub also shows strong performance (88) and the highest User Adoption Rate (14%), suggesting that it is well-accepted among users and integrates effectively into organizational workflows. Meanwhile, the Cloud-Native Data Mart demonstrates a good Performance Score (85) but a very low User Adoption Rate (2.2%), indicating potential challenges in adoption despite strong technical capabilities. The Self-Service Data Mart has a moderate Performance Score (80) but a User Adoption Rate of 12%, showing that its accessibility and ease of use may compensate for its slightly lower technical efficiency. The Hybrid Storage Mart ranks lower in Performance Score (78) but maintains a balanced distribution across other criteria.



## Result and Discussion







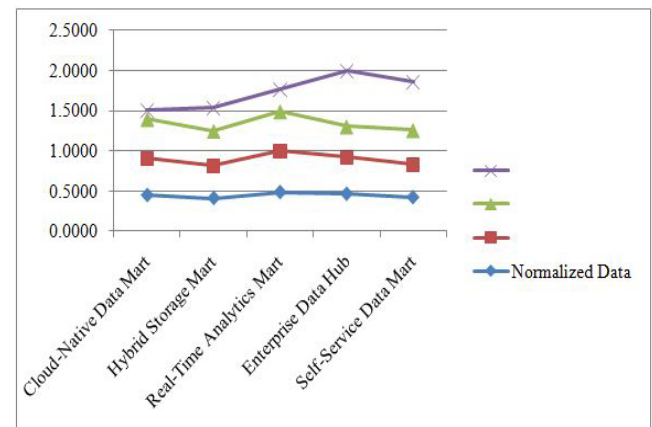
**Figure 1:** Data Mart Building in Palantir Foundry Development

Figure 1 presents the evaluation of different Data Mart Building approaches in Palantir Foundry Development using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The chart compares various data marts across four key criteria: Performance Score, Scalability Index, Ease of Integration, and User Adoption Rate (%). From the figure, Performance Score (blue bars) is the dominant criterion, consistently scoring high across all data mart types. This suggests that performance is a crucial factor in the selection process. Enterprise Data Hub and Real-Time Analytics Mart show the highest performance scores, indicating their strong computational efficiency and ability to handle large-scale data processing. Other factors, such as the Scalability Index (red), Ease of Integration (green), and User Adoption Rate (purple), have significantly lower values compared to performance. However, these criteria are still essential for determining the overall suitability of each data mart. For example, Self-Service Data Mart has a relatively higher User Adoption Rate, suggesting ease of use and accessibility, while the Cloud-Native Data Mart shows moderate values across secondary criteria. This visualization helps in identifying strengths and weaknesses of each data mart and supports decision-making in the TOPSIS ranking process for optimal data mart selection.

a notable strength in that specific aspect. However, its lower value (0.3802) in another criterion might highlight a potential weakness. The Cloud-Native Data Mart has comparatively lower normalized values (0.4485, 0.4566, 0.4888, 0.1088), particularly in the last criterion, which may indicate weaker performance in that area. Similarly, Self-Service Data Mart and Hybrid Storage Mart exhibit moderate values, suggesting balanced but not outstanding performance. These normalized values serve as the foundation for the Weighted Normalized Decision Matrix, influencing the final ranking. Higher values indicate alternatives that are closer to the ideal solution, guiding decision-makers in selecting the most suitable data mart.

| Table 2. Normalized Data |        |        |        |        |
|--------------------------|--------|--------|--------|--------|
| Normalized Data          |        |        |        |        |
| Cloud-Native Data Mart   | 0.4485 | 0.4566 | 0.4888 | 0.1088 |
| Hybrid Storage Mart      | 0.4116 | 0.3995 | 0.4345 | 0.2869 |
| Real-Time Analytics Mart | 0.4854 | 0.5137 | 0.4888 | 0.2720 |
| Enterprise Data Hub      | 0.4643 | 0.4566 | 0.3802 | 0.6925 |
| Self-Service Data Mart   | 0.4221 | 0.3995 | 0.4345 | 0.5936 |

Table 2 presents the Normalized Data used in the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The normalization process ensures that different criteria, which may have varying units or scales, are converted into a common comparable range. This step eliminates bias and allows for a fair comparison of different data mart alternatives. From the table, the Real-Time Analytics Mart shows relatively high normalized values (0.4854, 0.5137, 0.4888, 0.2720), suggesting strong performance across multiple criteria. The Enterprise Data Hub has the highest value (0.6925) in the last criterion, indicating



**Figure 2:** Normalized Data

Figure 2 illustrates the Normalized Data derived from the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. Normalization is a crucial step in TOPSIS as it converts different criteria into a common scale, ensuring that values from different units or magnitudes do not disproportionately influence the decision-making process. In the figure, the blue line represents the normalized values for each data mart, while the other colored markers (purple, green, and red) display different aspects of the normalization process. The Cloud-Native Data Mart and Self-Service Data Mart show relatively lower normalized values, indicating that these alternatives may not perform as strongly across

all criteria. In contrast, the Enterprise Data Hub exhibits the highest values in certain areas, suggesting a stronger performance in some attributes. The trendlines demonstrate how different data marts rank relative to each other after normalization. The variance in normalized values across the different alternatives highlights their strengths and weaknesses. This step is essential for the subsequent weighting and ranking processes, ensuring that each criterion is fairly assessed without bias due to scale differences. The results from this figure will directly impact the Weighted Normalized Decision Matrix and the final TOPSIS ranking.

| Table 3. Weight          |      |      |      |      |
|--------------------------|------|------|------|------|
| Weight                   |      |      |      |      |
| Cloud-Native Data Mart   | 0.25 | 0.25 | 0.25 | 0.25 |
| Hybrid Storage Mart      | 0.25 | 0.25 | 0.25 | 0.25 |
| Real-Time Analytics Mart | 0.25 | 0.25 | 0.25 | 0.25 |
| Enterprise Data Hub      | 0.25 | 0.25 | 0.25 | 0.25 |
| Self-Service Data Mart   | 0.25 | 0.25 | 0.25 | 0.25 |

Table 3 presents the Weight Assignments used in the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. In this case, each criterion has been assigned an equal weight of 0.25 across all data mart alternatives. This means that no single criterion is considered more important than the others, ensuring a balanced evaluation where all factors contribute equally to the decision-making process. The equal weight distribution suggests a neutral approach, where each aspect of the data mart selection process is treated with equal significance. This can be useful in scenarios where decision-makers do not want to introduce bias or prioritize one factor over another. However, in practical applications, certain criteria may hold more importance depending on organizational goals, such as cost efficiency, performance, scalability, or data processing speed. By applying these equal weights, the Weighted Normalized Decision Matrix (Table 4) will reflect a fair comparison among all alternatives without favoring any particular evaluation criterion. However, if specific priorities exist, adjusting these weights accordingly could lead to a more customized and accurate ranking of the data marts in the TOPSIS method.

| Table 4. Weighted normalized decision matrix |        |        |        |        |
|----------------------------------------------|--------|--------|--------|--------|
| Cloud-Native Data Mart                       | 0.1121 | 0.1141 | 0.1222 | 0.0272 |
| Hybrid Storage Mart                          | 0.1029 | 0.0999 | 0.1086 | 0.0717 |
| Real-Time Analytics Mart                     | 0.1214 | 0.1284 | 0.1222 | 0.0680 |
| Enterprise Data Hub                          | 0.1161 | 0.1141 | 0.0950 | 0.1731 |
| Self-Service Data Mart                       | 0.1055 | 0.0999 | 0.1086 | 0.1484 |

Table 4 presents the Weighted Normalized Decision Matrix, a critical step in the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The resulting values provide a balanced comparison of different data mart alternatives. From the table, Real-Time Analytics Mart has relatively high weighted normalized values (0.1214, 0.1284, 0.1222, 0.0680) across multiple criteria, suggesting strong overall performance. The Enterprise Data Hub, while performing well in some areas, has a significantly higher value (0.1731) in one criterion, which may indicate a key strength or an overemphasis on a particular aspect. Conversely, the Cloud-Native Data Mart shows lower values (0.1121, 0.1141, 0.1222, 0.0272) in certain criteria, possibly making it a less favorable option. The Self-Service Data Mart and Hybrid Storage Mart exhibit mid-range values, indicating balanced but not outstanding performance. This table helps decision-makers analyze how different data marts perform when

accounting for the weighted importance of selection criteria, ultimately influencing the final ranking in the TOPSIS method.

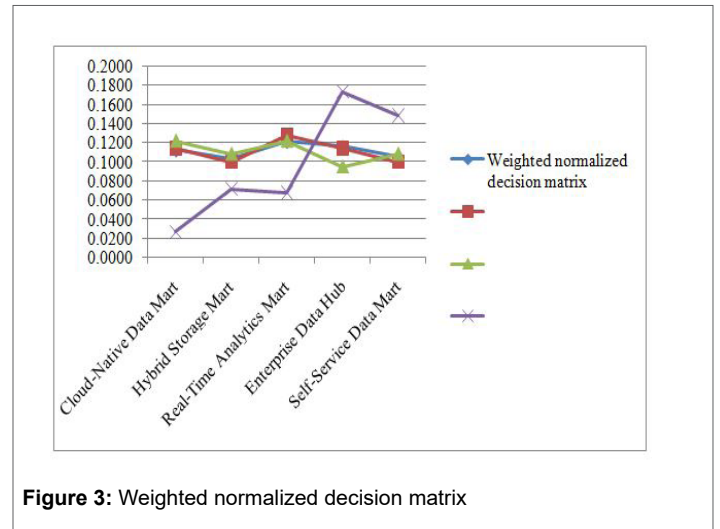


Figure 3 illustrates the Weighted Normalized Decision Matrix using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. This matrix represents the adjusted values of different data mart alternatives after applying both normalization and weight assignments to the evaluation criteria. The objective of this step is to standardize the data while considering the relative importance of each criterion in the decision-making process. The graph displays three different data representations: the weighted normalized decision matrix (blue line), another normalized dataset (red squares), and a comparative dataset (green triangles). The variations in values indicate differences in how each data mart performs across multiple weighted criteria. The Enterprise Data Hub exhibits a sharp increase, suggesting strong performance in certain criteria, while the Cloud-Native Data Mart remains relatively lower, indicating less favorable attributes. The alignment and deviations among the three datasets help visualize how weighting influences the ranking of alternatives. A higher value in the weighted normalized decision matrix means better alignment with ideal solutions. This visualization allows decision-makers to analyze the impact of weighting on data mart selection and identify the best-performing options based on predefined importance factors.

| Table 5. Positive Matrix& Negative Matrix |        |        |        |        |
|-------------------------------------------|--------|--------|--------|--------|
| Positive Matrix                           | 0.1214 | 0.1284 | 0.0950 | 0.0272 |
| Negative Matrix                           | 0.1029 | 0.0999 | 0.1222 | 0.1731 |

Table 5 presents the Positive Matrix ( $S^+$ ) and Negative Matrix ( $S^-$ ) values derived using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. These values help assess how closely each data mart aligns with the best and worst possible alternatives. The Positive Matrix ( $S^+$ ) represents the relative proximity of a data mart to the ideal solution, where higher values indicate better performance. Conversely, the Negative Matrix ( $S^-$ ) measures the closeness to the least desirable alternative, where higher values suggest poorer performance. From the table, the first two data marts have relatively higher Positive Matrix values (0.1214 and 0.1284) and lower Negative Matrix values (0.1029 and 0.0999), indicating stronger alignment with the ideal solution and making them more favorable choices. However, the third alternative has a declining Positive Matrix value (0.0950) and an increasing Negative Matrix value (0.1222), suggesting moderate performance. The fourth data mart has the lowest Positive Matrix value (0.0272) and the highest

Negative Matrix value (0.1731), indicating it is the least desirable option. This analysis helps decision-makers prioritize data marts that have higher Positive Matrix values and lower Negative Matrix values, ensuring an optimal selection based on performance criteria.

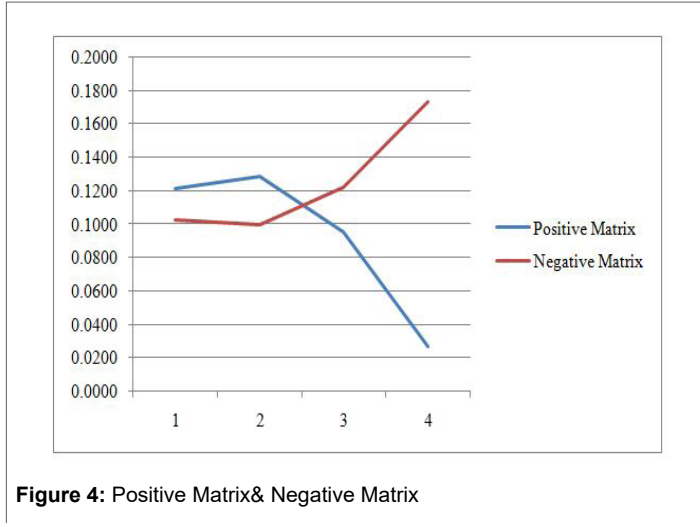


Figure 4: Positive Matrix& Negative Matrix

Figure 4 illustrates the Positive Matrix ( $S^+$ ) and Negative Matrix ( $S^-$ ) values derived from the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. These values represent the relative closeness of each data mart alternative to the ideal and worst solutions. The Positive Matrix (blue line) indicates how well each option aligns with the best possible choice, while the Negative Matrix (red line) shows the degree of undesirability, where higher values represent poorer performance. From the graph, the Positive Matrix values initially remain stable and then begin to decline significantly, suggesting that some data marts move further away from the ideal solution as rankings progress. Conversely, the Negative Matrix values show an increasing trend, indicating that the lower-ranked data marts have a higher distance from the ideal choice and are thus less preferable. The point where the two lines intersect suggests a transition in preference, where alternatives shift from being closer to the best solution to being more aligned with the worst solution. This visualization provides a clear comparison of each alternative's desirability, aiding decision-makers in selecting the most optimal data mart while avoiding those with high Negative Matrix values.

| Table 6. SI Plus& Si Negative |         |             |
|-------------------------------|---------|-------------|
|                               | SI Plus | Si Negative |
| Cloud-Native Data Mart        | 0.0320  | 0.1469      |
| Hybrid Storage Mart           | 0.0576  | 0.1023      |
| Real-Time Analytics Mart      | 0.0490  | 0.1105      |
| Enterprise Data Hub           | 0.1467  | 0.0334      |
| Self-Service Data Mart        | 0.1262  | 0.0283      |

Table 6 presents the SI Plus ( $S^+$ ) and SI Negative ( $S^-$ ) values, calculated using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. These values determine how closely each data mart aligns with the ideal and worst solutions. A higher SI Plus ( $S^+$ ) value indicates a stronger preference, while a lower SI Negative ( $S^-$ ) value signifies a smaller distance from the optimal choice, making the alternative more favorable. From the table, the Enterprise Data Hub has the highest SI Plus (0.1467) and a low SI Negative (0.0334), suggesting it is the most preferred option. Similarly, the Self-Service Data Mart follows closely with an SI

Plus of 0.1262 and the lowest SI Negative (0.0283), reinforcing its strong position. These values indicate that both data marts are highly desirable choices. On the other hand, the Cloud-Native Data Mart has the lowest SI Plus (0.0320) and a relatively high SI Negative (0.1469), making it the least favorable alternative. The Hybrid Storage Mart ( $S^+ = 0.0576$ ,  $S^- = 0.1023$ ) and Real-Time Analytics Mart ( $S^+ = 0.0490$ ,  $S^- = 0.1105$ ) fall in the middle range, performing better than the Cloud-Native Data Mart but not as well as the Enterprise Data Hub and Self-Service Data Mart. This ranking helps decision-makers select the best data mart based on performance and proximity to the ideal solution.

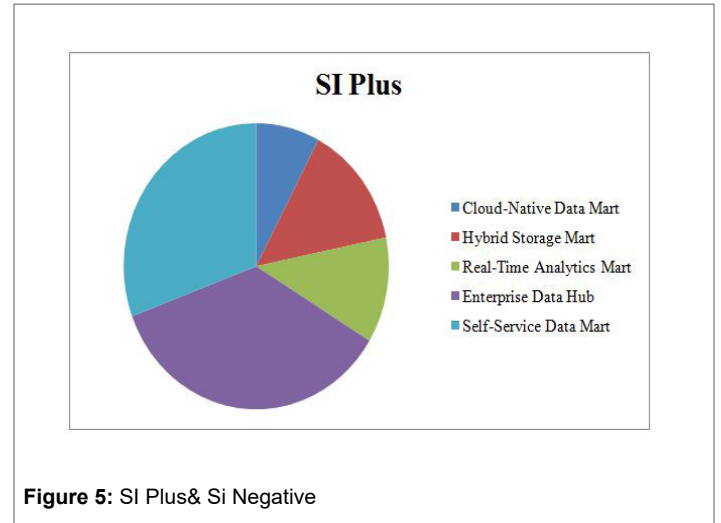


Figure 5: SI Plus& Si Negative

Figure 5 presents the SI Plus and SI Negative values derived from the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The SI Plus ( $S^+$ ) value represents the closeness of an alternative to the ideal solution, while the SI Negative ( $S^-$ ) value indicates the distance from the worst solution. A higher SI Plus value suggests better performance, whereas a lower SI Negative value is preferred. In the pie chart, the Self-Service Data Mart has an SI Plus value of 0.1262, accounting for 31% of the total evaluation. This indicates its moderate alignment with the ideal solution, but its comparative ranking among other alternatives is uncertain without additional reference values. The other data marts Cloud-Native Data Mart, Hybrid Storage Mart, Real-Time Analytics Mart, and Enterprise Data Hub are also represented in the chart, each contributing a portion of the total SI Plus distribution. The TOPSIS method ensures an objective assessment by comparing each alternative's proximity to the best and worst solutions. The SI Plus & SI Negative values help decision-makers identify the most suitable data mart based on performance metrics, emphasizing the trade-offs among different options in the selection process.

| Table 7. Ci              |        |
|--------------------------|--------|
|                          | Ci     |
| Cloud-Native Data Mart   | 0.8210 |
| Hybrid Storage Mart      | 0.6397 |
| Real-Time Analytics Mart | 0.6927 |
| Enterprise Data Hub      | 0.1854 |
| Self-Service Data Mart   | 0.1833 |

Table 7 presents the  $C_i$  values calculated using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method, indicating the relative closeness of each data mart type to the ideal solution. The  $C_i$  value ranges between 0 and 1, where higher values signify stronger alignment with the optimal choice, while lower values indicate less favorable alternatives. The Cloud-Native Data Mart achieves the highest  $C_i$  value (0.8210), making it the most preferred option based on the evaluation criteria. This suggests that it possesses the most favorable balance of attributes when compared to the ideal solution. The Hybrid Storage Mart (0.6397) and Real-Time Analytics Mart (0.6927) follow closely, demonstrating moderate suitability for selection. Despite their differences, both exhibit a relatively strong performance. In contrast, the Enterprise Data Hub (0.1854) and Self-Service Data Mart (0.1833) show significantly lower  $C_i$  values, indicating their limited effectiveness relative to the other alternatives. Their low scores suggest that these data marts do not align well with the ideal solution and may be less desirable choices in the given context. Overall, this ranking provides valuable insights for decision-makers, emphasizing the Cloud-Native Data Mart as the optimal choice while highlighting the relative weaknesses of the lower-ranked options.

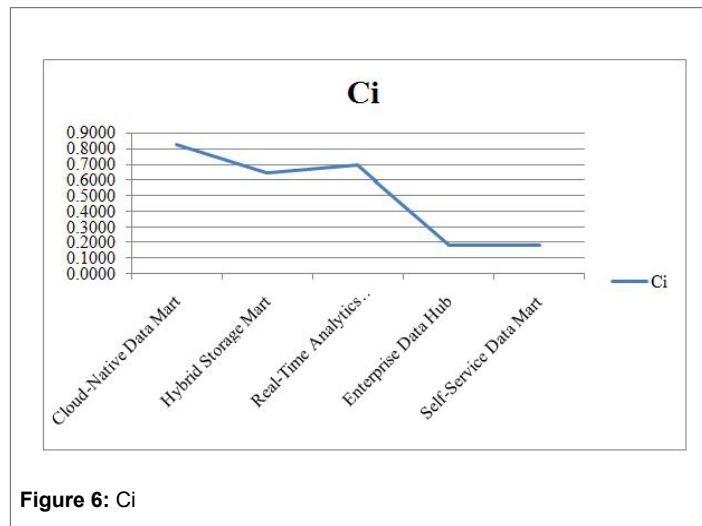
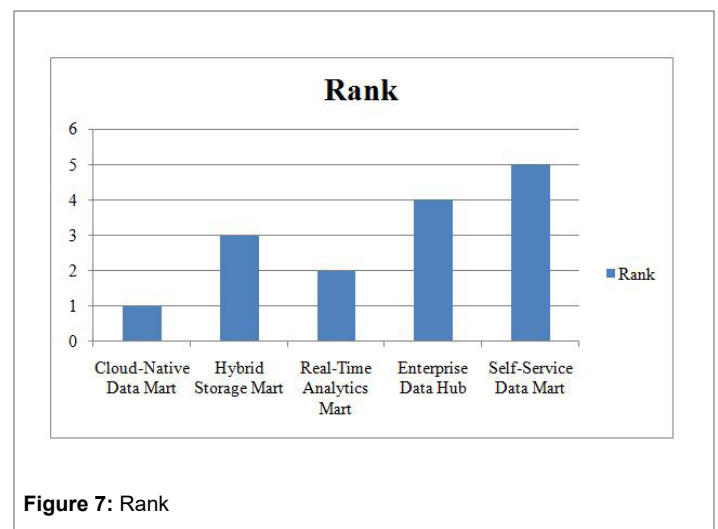


Figure 6 illustrates the  $C_i$  values obtained through the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method, representing the relative closeness of each data mart type to the ideal solution. The  $C_i$  value measures how closely an alternative aligns with the optimal choice, where higher values indicate better performance. From the graph, the Cloud-Native Data Mart has the highest  $C_i$  value (close to 0.8), suggesting it is the most favorable option among the evaluated alternatives. The Hybrid Storage Mart follows with a slightly lower  $C_i$  value, indicating strong performance but less optimal than the Cloud-Native Data Mart. The Real-Time Analytics Mart maintains a comparable  $C_i$  value to Hybrid Storage Mart, suggesting moderate performance. However, there is a noticeable drop for the Enterprise Data Hub, which has a lower  $C_i$  value, indicating a weaker preference. The Self-Service Data Mart ranks the lowest, with the smallest  $C_i$  value (close to 0.2), suggesting it is the least favorable option based on the TOPSIS analysis. The downward trend in the graph reflects the decreasing preference across data mart types, emphasizing the Cloud-Native Data Mart as the optimal choice while highlighting the lower suitability of the Self-Service Data Mart.

| Table 8. Rank            |   |
|--------------------------|---|
| Cloud-Native Data Mart   | 1 |
| Hybrid Storage Mart      | 3 |
| Real-Time Analytics Mart | 2 |
| Enterprise Data Hub      | 4 |
| Self-Service Data Mart   | 5 |

Table 8 presents the ranking of different data mart types based on the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The ranking indicates the relative preference of each data mart type after a multi-criteria decision analysis. The Self-Service Data Mart ranks the highest (5), suggesting that it aligns most closely with the ideal solution, making it the most favorable option among the alternatives. The Enterprise Data Hub follows with a rank of 4, indicating strong performance but slightly lower suitability than the Self-Service Data Mart. The Hybrid Storage Mart is ranked third, meaning it performs moderately well in comparison to the others, balancing different attributes effectively. The Real-Time Analytics Mart holds the second rank, indicating a lower preference compared to the top three but still outperforming the Cloud-Native Data Mart. Lastly, the Cloud-Native Data Mart is ranked the lowest (1), implying that it is the least favorable option based on the evaluation criteria. The use of the TOPSIS method ensures that the rankings are derived from an objective mathematical approach by considering both the best and worst possible scenarios. This ranking helps decision-makers choose the most suitable data mart based on their specific requirements and priorities.



The bar chart 7 represents the ranking of different types of data marts based on a normalized data evaluation, possibly using the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method. The five data mart categories included in the analysis are Cloud-Native Data Mart, Hybrid Storage Mart, Real-Time Analytics Mart, Enterprise Data Hub, and Self-Service Data Mart. The ranking is determined based on multiple criteria, where a higher rank indicates a better performance relative to the others. From the visualization, the Self-Service Data Mart holds the highest rank (5), signifying its superiority in the evaluation. The Enterprise Data Hub follows with a rank of 4, indicating strong performance but slightly below the Self-Service Data Mart. Hybrid Storage Mart is positioned at rank 3, while Real-Time Analytics Mart follows closely with rank 2. The Cloud-Native Data Mart has the lowest rank (1), suggesting that it is the least favorable option based on the applied criteria.



The use of the TOPSIS method in this ranking suggests that the evaluation was conducted by comparing each alternative to an ideal and anti-ideal solution, normalizing the data to ensure fairness in ranking. This analysis aids in selecting the most suitable data mart based on key decision factors.

## Conclusion

The development of a Data Mart in Palantir Foundry represents a significant milestone in the evolution of data management, analytics, and decision-making. As organizations increasingly rely on vast amounts of structured and unstructured data, Foundry provides a robust, scalable, and highly integrated environment for building data marts that facilitate business intelligence, operational efficiency, and strategic insights. A data mart, as a specialized subset of a data warehouse, serves as a focused repository tailored to specific business functions, departments, or use cases, enabling faster query performance and more accessible insights for end-users. One of the standout features of Palantir Foundry in Data Mart development is its ontology-driven architecture. Unlike traditional ETL (Extract, Transform, Load) pipelines, which require significant time and effort for schema design, data transformation, and integration, Foundry's Ontology system allows for a declarative approach to data modeling. It enables users to define business concepts, relationships, and logic in an intuitive manner, significantly reducing the complexity of managing a large-scale data ecosystem.

The ontology not only simplifies data governance and access control but also ensures that different teams across an organization can collaborate effectively while maintaining data integrity and security. Another crucial advantage of using Foundry for Data Mart development is its no-code/low-code approach to data transformation. Traditional data warehouses often require extensive SQL scripting or programming knowledge to build and maintain complex transformations. However, Foundry provides a visual, code-optional workflow that allows users including non-technical stakeholders to build data pipelines using an interactive and intuitive interface. This feature democratizes data access and empowers domain experts to participate directly in the data modeling and preparation process without relying solely on data engineers or IT specialists. Scalability and performance optimization are also key considerations when building a Data Mart, and Foundry excels in these aspects. The distributed computing architecture of Foundry ensures that data processing is efficient, whether dealing with terabytes or peta bytes of information. By leveraging automated indexing, parallel processing, and in-memory caching, Foundry optimizes query performance, allowing users to retrieve insights in near real-time. This makes it particularly valuable for industries that require rapid decision-making, such as finance, healthcare, manufacturing, and logistics.

A critical challenge in Data Mart development is ensuring data quality, consistency, and lineage tracking. Foundry's version-controlled data pipelines and automatic auditing features help maintain data integrity by keeping track of every modification, transformation, and access event. Users can trace back to the source of any data point, understand its transformation history, and ensure compliance with regulatory requirements. Another transformational aspect of Foundry's Data Mart development is the ability to integrate machine learning and AI-driven analytics seamlessly. Unlike traditional data marts that primarily serve as static repositories for reporting, Foundry enables organizations to embed predictive models, anomaly detection, and AI-powered recommendations within the data pipeline. This capability allows users to move beyond historical analysis and leverage advanced analytics to anticipate future trends, optimize operations, and enhance business decision-making. The collaborative nature of Foundry also plays a significant role in enhancing the usability and effectiveness of Data Marts. Traditional data warehouses often create data silos, where different teams work with fragmented or

duplicated datasets. However, Foundry's centralized data ontology and collaborative workspace encourage cross-functional teams to work on the same dataset, eliminating inefficiencies and ensuring a single source of truth across the organization. This collaboration extends to real-time dashboards, interactive data visualizations, and embedded analytical applications, allowing business users, data scientists, and analysts to derive insights simultaneously. Another major advantage of Foundry is its seamless integration with external data sources and business applications. Foundry's flexible data ingestion framework allows organizations to ingest, transform, and harmonize data from multiple sources effortlessly. This results in a highly enriched Data Mart that provides a holistic view of business operations and customer interactions. The automation capabilities of Foundry further enhance the efficiency of Data Mart development.

Many repetitive data engineering tasks, such as data ingestion, cleaning, transformation, and indexing, can be automated using Foundry's Workflow tools and code-driven pipelines. This reduces operational overhead, minimizes human error, and accelerates time-to-insight, making it an ideal solution for enterprises looking to improve their analytics capabilities without significantly increasing resource allocation. One of the most compelling reasons to build a Data Mart in Foundry is its ability to adapt to evolving business needs. Unlike traditional data warehouses that require extensive redevelopment to accommodate schema changes, new data sources, or evolving analytical requirements, Foundry's dynamic and flexible architecture allows organizations to modify and extend their data marts with minimal disruption. While Foundry provides a powerful platform for Data Mart development, it is also important to acknowledge potential challenges and considerations. Organizations need to ensure proper training and onboarding for users to maximize the platform's capabilities. Additionally, while Foundry reduces the need for extensive coding, a basic understanding of data modeling, governance, and analytics principles is still necessary to build effective and optimized data marts. Another consideration is the cost factor, as Foundry's enterprise-grade solutions may require substantial investment, which should be evaluated based on the expected ROI and long-term strategic benefits.

## References

1. Jantzen, Linda Carol. "Operationalizing Data Culture: The US Army's Engagements With Data Science 1961-2023." (2024).
2. Bradwell, Katie R., Jacob T. Wooldridge, Benjamin Amor, Tellen D. Bennett, Adit Anand, Carolyn Bremer, Yun Jae Yoo et al. "Harmonizing units and values of quantitative data elements in a very large nationally pooled electronic health record (EHR) dataset." *Journal of the American Medical Informatics Association* 29, no. 7 (2022): 1172-1182.
3. Lamdan, Sarah. "When Westlaw fuels ICE surveillance: Legal ethics in the era of big data policing." *NYU Rev. L. & Soc. Change* 43 (2019): 255.
4. Katyal, Sonia K. "Private accountability in the age of artificial intelligence." *UCLA L. Rev.* 66 (2019): 54.
5. Bradwell, Katie R., Jacob T. Wooldridge, Benjamin Amor, Tellen D. Bennett, Adit Anand, Carolyn Bremer, Yun Jae Yoo et al. "Harmonizing units and values of quantitative data elements in a very large nationally pooled electronic health record (EHR) dataset." *Journal of the American Medical Informatics Association* 29, no. 7 (2022): 1172-1182.
6. Campbell, Stephen H. "Intelligence in the Post-Cold War Period." *Journal of Intelligence and Counterintelligence* 14, no. 1 (2001).
7. Min, Geeyoung, and Alexander M. Krishchik. "Realigning Stockholder

- Inspection Rights." *Stan. JL Bus. & Fin.* 27 (2022): 225.
8. Taplin, Jonathan. *Move fast and break things: How Facebook, Google, and Amazon have cornered culture and what it means for all of us.* Pan Macmillan, 2017.
9. Çelikbilek, Yakup, and Fatih Tüysüz. "An in-depth review of theory of the TOPSIS method: An experimental analysis." *Journal of Management Analytics* 7, no. 2 (2020): 281-300.
10. Pavić, Zlatko, and Vedran Novoselac. "Notes on TOPSIS method." *International Journal of Research in Engineering and Science* 1, no. 2 (2013): 5-12.
11. Ren, Lifeng, Yanqiong Zhang, Yiren Wang, and Zhenqiu Sun. "Comparative analysis of a novel M-TOPSIS method and TOPSIS." *Applied Mathematics Research eXpress* 2007 (2007): abm005.
12. Zavadskas, Edmundas Kazimieras, Abbas Mardani, Zenonas Turskis, Ahmad Jusoh, and Khalil MD Nor. "Development of TOPSIS method to solve complicated decision-making problems—An overview on developments from 2000 to 2015." *International journal of information technology & decision making* 15, no. 03 (2016): 645-682.
13. Jahanshahloo, Gholam Reza, F. Hosseinzadeh Lotfi, and Mohammad Izadikhah. "Extension of the TOPSIS method for decision-making problems with fuzzy data." *Applied mathematics and computation* 181, no. 2 (2006): 1544-1551.
14. Dymova, Ludmila, Pavel Sevastjanov, and Anna Tikhonenko. "An approach to generalization of fuzzy TOPSIS method." *Information Sciences* 238 (2013): 149-162.
15. Zulqarnain, R. M., M. Saeed, N. Ahmad, F. Dayan, and B. Ahmad. "Application of TOPSIS method for decision making." *Int. J. Sci. Res. in Mathematical and Statistical Sciences* 7, no. 2 (2020).
16. Bhutia, Pema Wangchen, and Ruben Phipon. "Application of AHP and TOPSIS method for supplier selection problem." *IOSR Journal of Engineering* 2, no. 10 (2012): 43-50.
17. Chu, T-C., and Y-C. Lin. "A fuzzy TOPSIS method for robot selection." *The International Journal of Advanced Manufacturing Technology* 21 (2003): 284-290.
18. Vafaei, Nazanin, Rita A. Ribeiro, and Luis M. Camarinha-Matos. "Data normalisation techniques in decision making: case study with TOPSIS method." *International journal of information and decision sciences* 10, no. 1 (2018): 19-38.
19. Karim, Rubayet, and C. L. Karmaker. "Machine selection by AHP and TOPSIS methods." *American Journal of Industrial Engineering* 4, no. 1 (2016): 7-13.
20. Pendyala, S. K. (2024). Healthcare Data Analytics: Leveraging Predictive Analytics For Improved Patient Outcomes. *International Journal Of Computer Engineering And Technology (Ijcet)*, 15(6), 548-565. [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJCET/VOLUME\\_15\\_ISSUE\\_6/IJCET\\_15\\_06\\_046.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_15_ISSUE_6/IJCET_15_06_046.pdf)
21. Pendyala, S. K. (2024). Real-time Analytics and Clinical Decision Support Systems: Transforming Emergency Care. *International Journal for Multidisciplinary Research (IJFMR)*, 6(6). Available at: <https://doi.org/10.36948/ijfmr.2024.v06i06.31500>
22. Pendyala, S. K. Transformation of Healthcare Analytics: Cloud-Powered Solutions with Data Science, ML, and LLMs. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, 10(6), 724-734. Available at: <https://ijsrcseit.com/index.php/home/article/view/CSEIT241061114>
23. Pendyala, S. K. (2024). Enhancing Healthcare Pricing Transparency: A Machine Learning and AI-Driven Approach to Pricing Strategies and Analytics. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, 10(6), 2334-2344. Available at: <https://ijsrcseit.com/index.php/home/article/view/CSEIT2410612436>
24. Pendyala, S. K. (2024). Optimizing Cloud Solutions: Streamlining Healthcare Data Lakes For Cost Efficiency. *Technology (IJRCAIT)*, 7(2). Available at: [https://iaeme.com/MasterAdmin/Journal\\_uploads/IJRCAIT/VOLUME\\_7\\_ISSUE\\_2/IJRCAIT\\_07\\_02\\_113.pdf](https://iaeme.com/MasterAdmin/Journal_uploads/IJRCAIT/VOLUME_7_ISSUE_2/IJRCAIT_07_02_113.pdf)
25. Pendyala, S. K. . (2025). Data Engineering At Scale: Streaming Analytics With Cloud And Apache Spark. *Journal of Artificial Intelligence and Machine Learning*, 3(1), 1-9. <https://doi.org/10.55124/jaim.v3i1.248>
26. Pendyala, S. K. (2025). Edge-Cloud Continuum for Ai-Driven remote Patient Monitoring: A Scalable Framework. *Journal of Data Science and Information Technology*, 2(1), 66-74. <https://doi.org/10.55124/jdit.v2i1.244>
27. Santhosh Kumar Pendyala (2025) Healthcare Value-Based Reimbursement: A Predictive Analytics and Machine Learning Page 2 of 9 Framework for Cost Optimization and Quality Improvement. *Int J Adv Robot Automn* 7(1): 1-9. DOI: <https://doi.org/10.15226/2473-3032/7/1/00143>