denodo

accenture

Multi Cloud Data integration made simple.

Written by

Sambrita Chakrabarti: Sushil Mishra: Sonia Mawandia: Vijay Krishna Nandhini NR, Managing Director and Executive Sponsor

CONTENTS

1. Cloud Evolution

Introduction

Cloud computing has transformed the landscape of modern technology, revolutionizing how businesses and individuals' access, store, and manage their data and applications. This paradigm shift from traditional on-premises infrastructure to virtualized, scalable, and on-demand resources has been a remarkable journey that has spanned several decades. Let us delve into the evolution of cloud computing and how it has shaped the digital world we live in today.

Precursors to Cloud Computing

The seeds of cloud computing were sown in the 1960s with the development of time-sharing systems. As technology advanced, the concept evolved into client-server architecture, where powerful servers delivered resources to connected client devices.

Birth of the Internet and Early Cloud Concepts

The birth of the internet in the 1990s provided the infrastructure necessary for the expansion of cloud computing. Amazon Web Services (AWS) played a pivotal role by launching Amazon Elastic Compute Cloud (EC2) in 2006, allowing users to rent virtualized computing resources over the internet. This marked the beginning of Infrastructure as a Service (IaaS), one of the key models of cloud computing.

The success of early cloud providers prompted other tech giants, including Microsoft, Google, and IBM, to enter the cloud computing arena. This led to intense competition and innovation, with providers continuously enhancing their offerings in terms of features, scalability, and security.

Hybrid and Multi-Cloud Architectures

Businesses with existing on-premises systems faced the challenge of transitioning to the cloud. Hybrid cloud solutions emerged, enabling seamless integration between on-premises infrastructure and public cloud services. Furthermore, multi-cloud strategies gained popularity, allowing organizations to distribute workloads across multiple cloud providers to mitigate risks and vendor lock-in.

Advancements in Virtualization and Containerization

Virtualization and containerization technologies like VMware and Docker revolutionized how applications were deployed and managed within cloud environments. Containers facilitated consistent application delivery across various environments, from development to production.

As-Is state.

From its humble beginnings as time-sharing systems to its current dominance as the backbone of modern technology, cloud computing has come a long way. Its evolution has been marked by groundbreaking innovations, fierce competition, and a relentless drive to provide more accessible, scalable, and efficient computing resources. As we look ahead, cloud computing is poised to shape the future of technology and reshape industries across the globe.

2. Overview of cloud and key vendors

Cloud computing is a technology that allows individuals and organizations to access and use computing resources (such as servers, storage, databases, networking, software, and more) over the internet, rather than owning and maintaining their own physical infrastructure. Cloud computing offers numerous benefits, including scalability, cost-efficiency, flexibility, and the ability to access resources from anywhere with an internet connection. There are several deployment models and service models in cloud computing:

Deployment Models:

1. Public Cloud: Services are offered to the public by cloud providers, and resources are shared among multiple customers. Examples of public cloud providers include Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).
2. Private Cloud: Resources are dedicated to a single organization and can be hosted on-premises or by a third-party provider. Private clouds offer more control and security but can be more expensive to set up and manage.
3. Hybrid Cloud: Combines both public and private cloud resources, allowing data and applications to be shared between them. This provides flexibility and can help organizations balance cost and performance.

Service Models:

1. Infrastructure as a Service (IaaS): Provides virtualized computing resources over the internet. Customers can rent virtual machines, storage, and networking infrastructure. Example: AWS EC2, Azure Virtual Machines.
2. Platform as a Service (PaaS): Offers a platform that allows developers to build, deploy, and manage applications without worrying about the underlying infrastructure. Example: Google App Engine, Heroku.
3. Software as a Service (SaaS): Delivers software applications over the internet on a subscription basis. Users can access the software via a web browser. Example: Microsoft Office 365, Salesforce, Gmail.

Key Cloud Vendors:

1. Amazon Web Services (AWS)
2. Microsoft Azure
3. Google Cloud Platform (GCP)
4. IBM Cloud
5. Oracle Cloud
6. Alibaba Cloud
7. Salesforce

These cloud vendors compete in a rapidly evolving market, constantly innovating and expanding their service offerings to meet the diverse needs of businesses and organizations worldwide. Customers choose cloud providers based on factors such as their specific use cases, geographic reach, pricing, and the services they offer.

3. Data Integration and Data Virtualization

In traditional data integration methods, physical data is mostly transformed, moved and/or replicated into a physical repository such as a data warehouse or data lake, by means of technologies such as ETL.

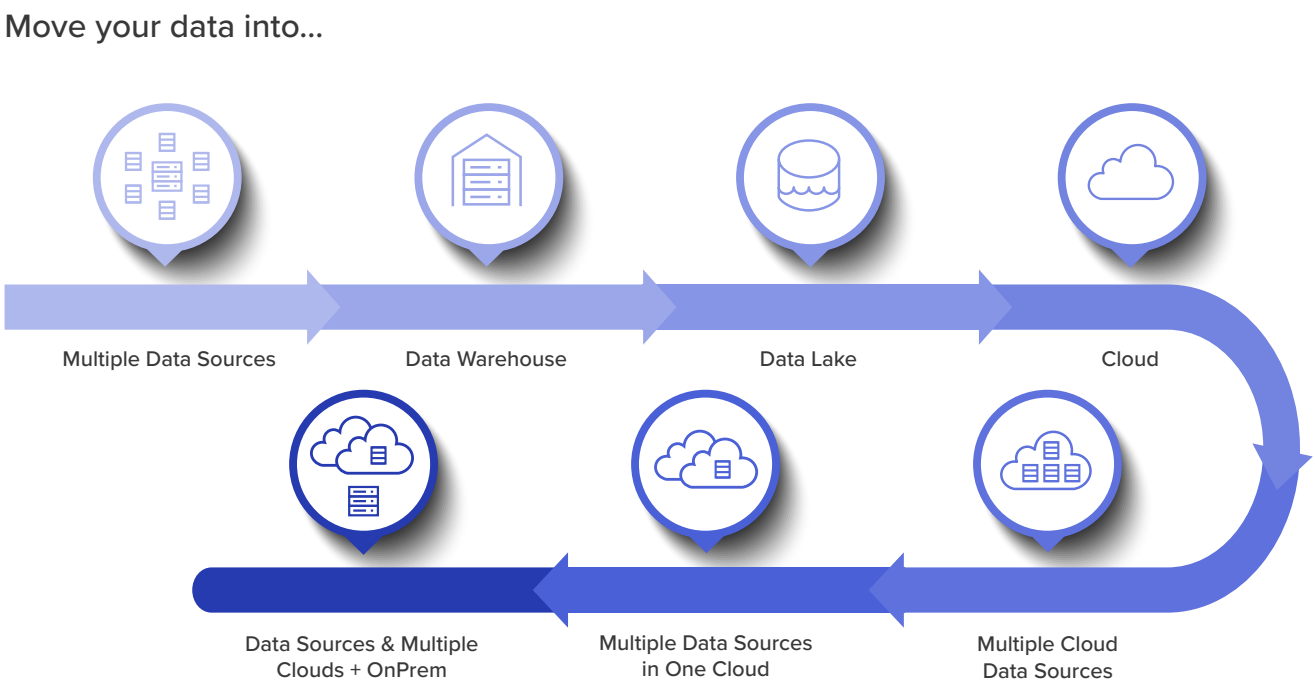
Data integration began with the creation of data cubes or small operational data stores for reporting purposes and would include only structured data. With enterprise data warehouses, data was physically copied into one central location. As the internet emerged, unstructured data started flowing in as well and data lakes were introduced as a solution to capture all kinds of data including structured and unstructured data. With advent of cloud, focus shifted towards core application development and avoid logistics and infrastructure management such as O/S patching, security software management etc. However, the core problem of endless copying of data and challenges of monolithic/central data management remained.

As a solution, data virtualization was first introduced two decades ago. Data virtualization focuses on **virtual data management instead of physical data copying**. Since then, the technology has evolved considerably. Data virtualization has helped to:

- match the changing pace of data integration requirements.
- overcome limitations of legacy infrastructure.
- effectively manage the massive amounts of structured and unstructured data that organizations are collecting.
- enable the shift from merely integrating the data to comprehensive data management.

In this paper we shall study data virtualization technology and how it is solving challenges of cloud and modern data management.

The diagram below shows a rough evolution of data integration technologies:



4. Challenges Of Data Management in Cloud

Multi-cloud environments, where an organization uses multiple cloud service providers to host their applications and data, offer numerous benefits like flexibility, cost optimization, and resilience. However, they also come with several challenges that need to be carefully managed:

Complexity: Managing multiple cloud providers can be overly complex. Each provider has its own set of tools, APIs, and management interfaces, making it challenging to maintain consistency and interoperability.

Data Integration: Moving data between different cloud providers and ensuring data consistency can be challenging. Data integration and synchronization between different clouds can lead to latency and data consistency issues.

Vendor Lock-In: Avoiding vendor lock-in is one of the reasons organizations opt for multi-cloud. However, it can be challenging to design applications and systems in a way that makes them truly portable across different cloud providers.

Cost Management: Managing costs across multiple providers can be complicated. Different providers have their own pricing models, and it can be difficult to track and optimize spending effectively.

Security and Compliance: Security policies and compliance requirements may differ between cloud providers. Ensuring that all providers meet the necessary security and compliance standards can be a complex task.

Performance Optimization: Optimizing the performance of applications and workloads across multiple clouds can be challenging. Different cloud providers have different network and infrastructure characteristics that may require custom configurations.

Skillset and Training: IT teams need to be proficient in multiple cloud platforms, which may require additional training and resources. This can strain the human resource aspect of managing a multi-cloud environment.

Service-Level Agreements (SLAs): Each cloud provider has its own SLAs, and it can be complex to manage and meet SLAs across multiple providers. Downtime in one provider can affect the overall availability of the application.

Interoperability: Ensuring that different cloud services work seamlessly together can be challenging. Integrating services from multiple providers may require custom development and integration efforts.

Governance and Management: Maintaining consistent governance and management policies across multiple clouds is crucial for security, compliance, and efficiency. Managing identities, access control, and policies can be complex.

Data Sovereignty: Compliance regulations and data sovereignty issues can make it difficult to manage data across multiple regions and cloud providers, requiring careful planning and data management.

Monitoring and Visibility: Getting a unified view of the entire multi-cloud environment can be complex. Organizations need comprehensive monitoring and visibility tools to ensure performance and security.

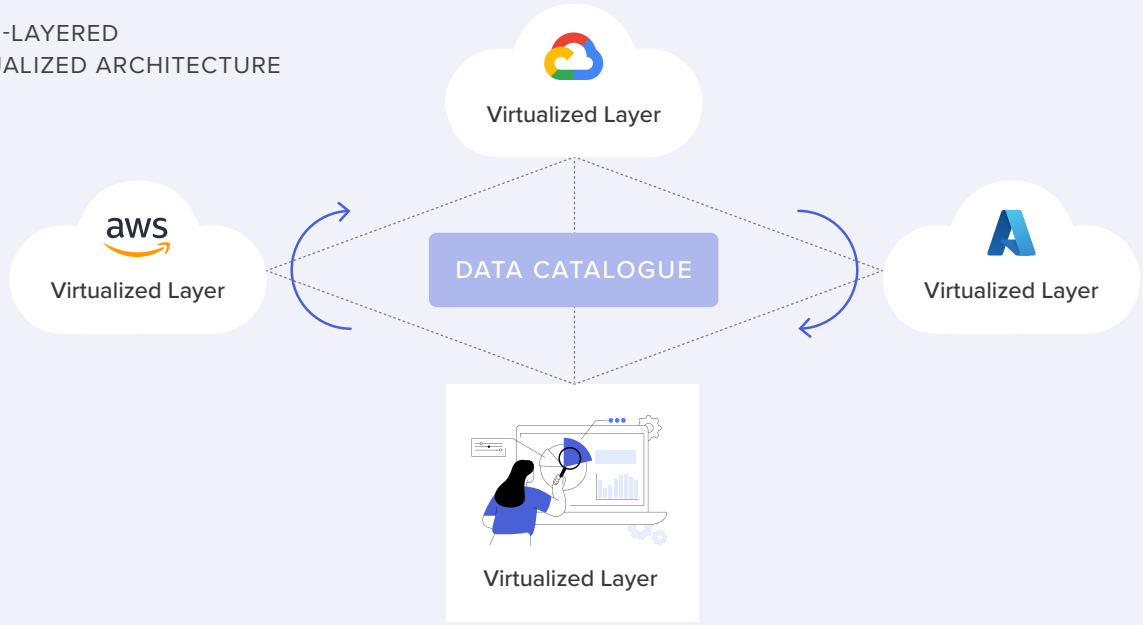
Disaster Recovery: Planning for disaster recovery and business continuity across multiple cloud providers can be challenging. Ensuring data backup and recovery strategies work seamlessly is crucial.

To overcome these challenges, organizations need a well-thought-out multi-cloud strategy, robust governance frameworks, automation, and the right mix of tools and expertise. Additionally, ongoing monitoring and optimization are essential to maximize the benefits of a multi-cloud approach while mitigating its challenges.

5. Multi cloud data virtualization: solving data management challenges in cloud

Data virtualization is a technology that can be immensely beneficial in a multi-cloud environment by addressing several key challenges related to data access, integration, and management. Here's how data virtualization can help in a multi-cloud environment:

MULTI-LAYERED
VIRTUALIZED ARCHITECTURE



Central Data Access Layer

- Data virtualization provides a unified, virtual layer that abstracts the underlying data sources, making it easier to integrate data from various locations and formats without physically moving or copying the data. This simplifies data integration and reduces data redundancy.
- Data virtualization abstracts the complexity of underlying data sources, allowing applications and users to access data using a common interface. This abstraction layer provides a consistent view of data, regardless of where it is stored, making it easier for developers and data consumers to work with the data.

Centralized Security & Governance Layer

- Security architecture is simplified as there is no need to open so many ports.
- Authorization rules can be defined in a specific location simplifying the security architecture.
- Localized security control implementation would help to manage different user organizations.
- Can help enforce security and compliance policies by providing a centralized access point where security controls can be applied. This ensures that data is accessed securely and in accordance with regulatory requirements, regardless of the underlying data sources.

Agility and Flexibility

- Provide real-time access to data by connecting to multiple data sources in real time and providing up-to-date information to applications and users. This is crucial for decision-making and analytics in a multi-cloud environment where data is continuously changing.
- Organizations can easily add or switch data sources without making significant changes to applications, reducing the impact of vendor lock-in, and allowing for rapid adaptation.
- The multi-layer architecture allows access to data in other locations (with proper permissions

Performance Optimization

- Data virtualization can optimize query performance by intelligently routing queries to the most suitable data sources. This can include choosing the closest data centre or selecting a cloud provider with lower latency for specific queries, improving overall application performance.

Cost Optimization

- Data virtualization can help reduce data transfer costs in a multi-cloud environment by minimizing the need to move data between cloud providers. Instead of duplicating data across providers, data can be accessed virtually, reducing unnecessary data egress charges.

Simplified Management

- Data virtualization simplifies data management by providing a single point of control for data access and governance. This streamlines data cataloguing, metadata management, and data quality assurance processes.

Enabling Data Federation

- Data virtualization allows organizations to create a unified view of data that spans multiple cloud providers and on-premises systems. This federation enables holistic analytics and reporting without the need to consolidate data into a specific location.

In summary, data virtualization plays a crucial role in simplifying data access and integration in a multi-cloud environment. It provides a flexible, real-time, and secure way to access and manage data from diverse sources, enhancing the agility, efficiency, and effectiveness of multi-cloud strategies.

6. Strategy to Integrate multi cloud with virtualization:

Native cloud support for better efficiency and faster integration

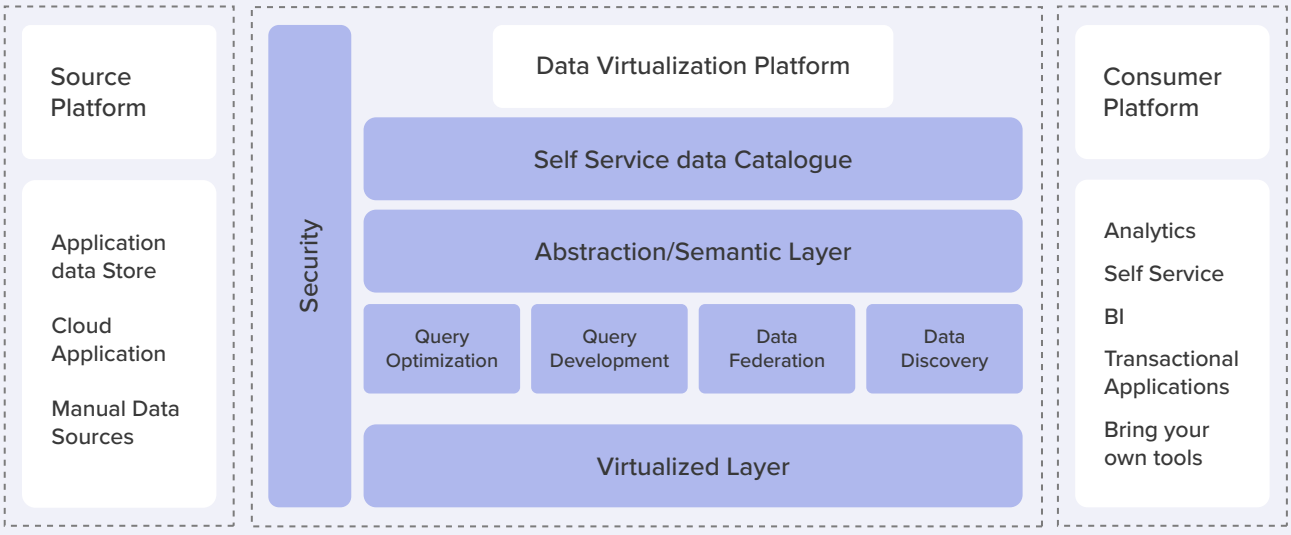
- The data virtualization platform can run natively on each cloud platform and scale depending on the type of workload.
- As each virtualized platform works with localized data in each cloud platform, data movement is reduced, and cloud egress costs are minimized.

Single centralized virtualized layer across multiple cloud platforms

- Different localized virtualized platforms are connected to each other.
- Single comprehensive view of data with active monitoring.
- Containerization and Kubernetes accelerates overall management.

Enhanced query federation and self-service data catalogue

- Every Data virtualization cluster is connected to all different data sources and a data catalogue is maintained.
- Each virtualization cluster has visibility of the datasets available across all other clusters and can request this data from its peer clusters based on requirements.



7. A smarter way to manage performance

Multiple factors impact performance like - data sources, network latency, complexity of query and processing, consumer ingestion rates, etc.

Some principles to improve performance are:

- Minimize data moved through the network.
- Maximize 'local' data processing.
- Move processing to data.

The various components of data virtualization

- Query Rewriting and Delegation
    - Push processing to where the data lives, minimize the amount of data going through the network.
    - Query is analyzed to determine views – and data sources – being queried.
    - Rule-based optimizer rewrites the SQL query to maximize delegation and performance.
    - Cost-based optimizer uses statistics, indices, etc. gathered from data sources to 'cost' different execution plans.
    - The best plan is selected for execution. It orchestrates execution of sub-queries, collation of results, any processing within the data virtualization Platform and returns results to the user.
- MPP Query Acceleration
    - Offload processing to co-located cluster.
    - Data copied to cluster and operation is delegated for processing, and data is copied to a Parquet file.
- Caching
    - Caching data from slow sources and caching costly data combinations/ transformations.
    - Client queries are deflected to the cache system instead.
- Smart Query Acceleration (Aggregate-aware Caching)
    - Detects common query patterns and proposes pre-computing summaries to accelerate execution of analytical queries.

8. The Denodo Advantage: Logical Data Management

The Denodo Platform is a powerful solution that can ease the challenges of multi-cloud data processing. As more organizations adopt multi-cloud strategies to distribute workloads across different cloud providers, they often encounter complexities related to data integration, management, and accessibility. The Denodo Platform, a logical data management solution built on data virtualization, addresses these challenges by providing a unified and streamlined approach to accessing and processing data across multiple cloud environments. Here's how the Denodo Platform can ease multi-cloud data processing:

- ✓ Connecting to data instead of collecting data with sophisticated two-way communicating data wrappers.
- ✓ Central data model with flexibility to democratize and delegate management control.
- ✓ Data abstraction reducing complexity of data access and improved security.
- ✓ Decoupling sources and consumers to make the data integration environment agile.
- ✓ Central security and governance model for ease of deploying security policies.
- ✓ Low code and no code platform, user friendly graphical user interface for developers
- ✓ Improved query performance with a state-of-the-art query engine.
- ✓ Query rewriting and delegation: Push processing to where the data lives, minimizing the amount of data going through the network. This feature is automatic but can be controlled.
- ✓ MPP Query Acceleration: Offload processing to co-located MPP cluster, with embedded MPP based on Presto.
- ✓ Caching: Caching data in a local cache for performance improvement, smart query acceleration (aggregate-aware Caching). Based on recurrent queries, cache summary tables to accelerate queries
- ✓ Granular cell level security with data masking to table, row, column, and cell level.
- ✓ State of the art data catalogue with built in AI, which is intelligent enough to recommend data sets based on user login and role. Shows lineage and provides option for sample querying.
- ✓ DevOps and auto cloud scaling management
- ✓ Few click data services (Rest, SOAP etc)
- ✓ Integration with Kafka /Live stream
- ✓ Certified on major cloud platforms and OEMs.
- ✓ Optimize your cloud data spend new FinOps dashboard in Denodo with real-time insights into cloud data costs.
- ✓ Democratize Data with Big Data Analytics & improve Scale of data operations.
- ✓ Democratize data with Gen AI - OOTB Integration with OpenAI and ChatGPT
- ✓ For Accelerating enterprise-wide privacy compliance
- ✓ Location-based compliance: manage data sovereignty easily with Denodo.
- ✓ Empower domains to build data products to help in data discovery and DIY data models and reports.

9. How Accenture helped its customers in achieving Multi Cloud Virtual Environment

LARGE MULTINALTIONAL MINING COMPANY

Customer overview

This company has multinational mining, metals, and petroleum operations, and is headquartered in Melbourne, Australia. It spans over ninety locations worldwide, with an employee strength of 80,000 and revenue of above US$44billion. It has four primary operational units Coal, Copper, Iron ore & Petroleum.

It ranked as the world's largest mining company, based on market capitalization.

Problem

In the world driven by data today, data management poses challenges to companies. Many of the common problems were seen and Data Virtualization was a solution to this.

Consider a multinational mining, metals, and petroleum company with its operations spread globally. Your business operates like a puzzle with various pieces of information scattered around various places such as departments and systems, representing data about your operations that range from production figures to market trends.

However, there has been a problem in trying to assemble this puzzle. Across ninety locations on Earth your data is kept in many formats, databases, and software systems. Consequently, looking for the crucial information you need wants to find a needle in a haystack each time.

This leads to slow and cumbersome decision making. You cannot see what is happening throughout the whole corporation at any given moment. And when different teams are dealing with various versions of data there is room for mistakes or miscommunications.

Solution

Now pictures have a tool that brings together all these pieces of dispersed information right away no matter where they are located. That is exactly what the Denodo Platform does for your company.

The Denodo Platform enables you to create an abstract layer that connects all your data sources whether it is coal fields in Australia, copper mines in Chile or oil sites.

Denodo offers a comprehensive data virtualization platform that addresses these challenges effectively:

- **Unified Data Access:** The Denodo Platform enables organizations to access and integrate data from disparate sources – databases, cloud services, data lakes, APIs, etc. – seamlessly, eliminating data silos.
- **Abstraction Layer:** It provides a logical abstraction layer over the underlying data sources, allowing users to query and consume data in a unified manner without concerning themselves with the complexities of underlying data structures. The Denodo

Platform's data access layer is user friendly; any database team can simply plug in and start playing with data with Security.

- **Real-time Data Integration:** The Denodo Platform facilitates real-time data integration by providing capabilities like caching, query optimization, and parallel processing, ensuring timely access to the most up-to-date information.
- **Scalability and Performance:** The Denodo Plaform's architecture is designed for scalability, enabling organizations to manage large volumes of data efficiently. It employs techniques such as distributed caching and query optimization to deliver high performance.
- **Data Security and Governance:** The Denodo Platform incorporates robust security features, including data masking, encryption, and role-based access control, to ensure data security and compliance with regulatory standards. It also provides auditing and lineage capabilities for governance.

10. Conclusion

Virtualized layer enabling mesh architecture and helping us to build governed Data Product journey. This approach of virtual first in data management future-proofs an organization's data architecture and makes it agile. This layer provides an **abstraction layer** between consumers and data migration phases.

- The Denodo **Hybrid / Multi-Cloud Deployment** model enables the clients to **optimize for execution** and **adhere to the regional/corporate policies.**

- The Denodo Solution Manager (Operations Console) provides for Automated AWS/Azure Deployment, and the Denodo Platform can be deployed in other environments such as GCP, Containers, etc. **Denodo FinOps Dashboard** gives users critical visibility into data adoption, consumption, and possible associated costs with the data storage and compute which is critical for cloud initiatives.

The Denodo Platform's data virtualization technology makes data integration quicker and improves the ease of manageability. It also brings huge business benefits and cost savings mentioned below.

Data Integration Cost reduction: **60-80%** savings

Traditional Call Centers, Portals: **30-70%** savings

BI and Reporting: **40-60%** savings

ETL and Data Warehousing

- Project timelines of 6-12 months reduced to **3-6 months.**
- **Up to 85%** reduction in time

The Denodo Platform, with all of the above technical and business benefits, enables organizations to effortlessly implement data fabric and data mesh architectures.

11. References

- www.denodo.com
- https://community.denodo.com
- https://denodo.zinfi.net/concierge/crm/beta/#/Home
- https://www.denodo.com/en/document/analyst-report/forrester-wave-enterprise-data-fabric-q2-2022