Evolving the Data Warehouse: The Next Generation for Financial Services Institutions
Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.
Executive Overview ........................................................................................................2
Introduction ..................................................................................................................2
The Crisis and its Regulatory Aftermath ........................................................................3
Competitive Drivers .......................................................................................................3
Revisiting the Data Warehouse .......................................................................................4
So, What is the Problem? ................................................................................................6
The Solution: Guiding Principles for a Next Generation Data Warehouse ...............8
  Understand the Use Cases ..........................................................................................8
  Understand the Processes that Produce or Consume Data .....................................10
  Understand the Ultimate Usage of the Analytical Outputs ....................................10
The Next Generation Data Warehouse: Architecture .................................................10
  Data Model ................................................................................................................11
  Analytical Infrastructure ............................................................................................12
  Scalable, Highly Performant Hardware/Database Foundation ................................13
Summary and Key Takeaways: The Evolving Nature of the Data Warehouse for
Financial Services Institutions .....................................................................................14
  Focus On End-to-end Data Flows to Support Key Use Cases .................................14
  Purpose-built Analytical Platform Rather than a Collection of Tools ..................15
  Deep Integration with Underlying Infrastructure ....................................................15
Conclusion ...................................................................................................................15
Executive Overview

Financial Services Institutions worldwide are operating in an unprecedented environment. Over the last 20 years, the scale and complexity of the average financial services institution, be it in banking, capital markets or insurance has increased exponentially. This Oracle Financial Services whitepaper will examine the traditional approaches to data warehousing that were in place during the recent financial crisis (many of which remain even to this day) and how many of them failed to meet the demands during this critical period. We will then examine the resulting requirements and marketplace demands that are driving this new complexity and then we will examine a next generation approach designed to eliminate these problems in the future.

Introduction

In a flurry of activity enabled by significant changes to a long-established regulatory environment (dismantling Glass-Steagall, Gramm-Leach-Bliley Act), FSI Firms entered business areas that were traditionally not considered a part of their profile (for example, structured credit, derivatives, high frequency trading to name a few) in quick succession, with varying degrees of operational preparedness, and varying degrees of success.

The defining event of the recent past, of course, was the credit crunch that occurred in the Fall of 2008 and the ensuing “Great Recession,” as many have come to call it.

The following years have made it painfully clear that critical decision-making hinges on the quality of analytical data in financial services. This is, in turn, completely dependent on the robustness of the data management processes and infrastructure within the institution.

Second, institutions no longer have the luxury of large multi-year initiatives to identify and address data management issues using a big-bang/”big data project” approach. There is an increasing need for them to respond quickly, and accurately with the right supporting analytical data, be it to regulators or their own management.

Finally, IT organizations simply cannot approach data management using a “data provider” mindset alone that is disconnected from downstream functional needs. In order to efficiently serve multiple intersecting user communities for analytical data, IT organizations need to develop a holistic mindset that spans the entire data lifecycle from operational data to
analytical systems and reporting. This represents an expansion of their role and importance in data management.

The Crisis and its Regulatory Aftermath

While there is ongoing debate about the ultimate causes of the crisis, and the set of possible remedies, one of the key lessons from the crisis is the role played by information systems, particularly analytical systems. According to the Economist, “messy IT systems are a neglected aspect of the financial crisis.” ([http://www.economist.com/node/15016132](http://www.economist.com/node/15016132))

For example, one possible driver exacerbating the crisis was that many institutions, even the largest and most complex, had little capacity to understand up-to-the-minute funding positions and institutional liquidity precisely at a time when markets needed this information most. Another example is the fact that risk management was silo’d across businesses and functions and unable to provide holistic, integrated views of risk across businesses and asset classes.

In either case, a significant underlying contributor to the problem was that analytical systems and workflows in banks and FSIs were simply not geared towards producing a unified view that may have identified and highlighted these problems in a timely manner.

The consequences are far reaching. Banks are now faced with the prospect of increasing, detailed regulatory scrutiny, like that found in Dodd Frank and Basel III, to complement the existing regulatory regimes already in place. This results in even greater stress on their operations and analytical systems to adapt and respond accordingly. Given the lack of preparedness that was highlighted in the crisis, FSIs are realizing that a fundamental rethink of data management approaches and practices is necessary.

Competitive Drivers

Another set of equally important drivers facing banks is the nature of the competitive landscape, and how technology advances have altered it.

For example, there is increasing pressure for banks to support their operations with analytical information that can be delivered in time for in-transaction decision-making. This is in stark technological contrast to traditional, “offline” analytical applications used solely to support senior management decisions.

Additionally, real-time pricing, real-time decisions in CRM, fraud detection and surveillance and algorithmic trading all represent an emerging class of analytical workflows that are needed “on-demand,” and where the deliver of so-called “business intelligence” is no longer a periodic activity within a defined time window.
Revisiting the Data Warehouse

Data warehouses have always been considered as part of the solution to the data management problem, not just in Financial Services. Ever since the discipline was founded, it has come to represent a popular set of solution patterns applicable to data management for analytical end uses.

Without belaboring the numerous definitions and related terms, the idea of a data warehouse can be summarized as “source once, distribute many times.”

That is,

- Gather operational/transactional data
- Transform it into a common representation (an Enterprise Data Model)
- Store this data (current and historical) in a central repository that is distinct from the operational systems
- Drive all data provisioning needs for analytical communities from this “golden source”

![Figure 1: First Generation Data Warehouse Design for Financial Services Institutions – Warehouse as the Central Store for All Business Data](image)

Today, Data Warehousing itself is a thriving practice that spans tools (ETL, data quality, metadata management, data modeling), infrastructure (specialized DW hardware, DB server optimizations such as caching and indexing strategies), and architectural approaches (Bill Inmon, Ralph Kimball). Historically, the Financial Services industry has been one of the earliest adopters of these data warehousing approaches and disciplines.
However, it is increasingly apparent that data warehousing has not delivered the many benefits and values promised, particularly to the Financial Services industry. The reality of data warehousing in financial services looks more like the following:

![First Generation Data Warehouse Design for Financial Services Institutions – Hard to Deliver Value](image)

This is corroborated by findings from TDWI that illustrate that almost 46% of surveyed DW owners are looking to replace their DW platforms.
This reality is not an indictment of the tools or technologies themselves – there has been no shortage of significant advances in any area from DW infrastructure, database technologies, to ETL tools, to BI technologies. At the same time, data warehousing projects and initiatives have always been considered strategic, so the problem is not a lack of sponsorship or focus.

So, What is the Problem?

In our estimation, the fundamental reason for the high failure rates for data warehousing in FSI can be attributed to a lack of understanding of the end uses of a data warehouse within a Financial Services Institution.

Analytical needs in financial services are arguably the most complex, varied and fastest evolving compared to other industries. The sheer number of existing regulatory, competitive and operational drivers as well as the computational complexity of analytical techniques and the associated data flows (eg. derivative pricing, simulation-based approaches to risk etc) has resulted in a proliferation of analytical silos in FSIs where specialized solutions are deployed, sourcing only the data relevant to them, often with little or no regard for a coherent overarching data management strategy.
Against this backdrop, ‘traditional’ data warehousing places emphasis on generic patterns and approaches to provide a data foundation for analytics without particular reference to the problems unique to FSI. This impedance mismatch is the key contributor to the questionable success in applying these techniques to FSI.

The typical symptoms of this are:

- **Too much focus/effort on data sourcing:**
  
  - The sourcing exercise is typically the most time-consuming, protracted part of a warehouse implementation.
  
  - Source systems in FSI are numerous, and have fragmented, possibly overlapping views of key data concepts like Customer, Product etc., as well as their own data model idiosyncrasies, not to mention tremendous variations in data volume and quality.
  
  - The standard data warehousing practice is to use an Enterprise Data Model, which is typically a conceptual view of an abstract bank or FSI. By definition, converting this abstract model to a concrete physical schema requires a mapping exercise whose scope is usually too large to be reasonable. Time and again, this is the biggest contributor to complexity in a data warehouse – primarily because it is tied to the effort involved in the development of data transformation processes (ETL) – often a long, protracted custom development exercise.
  
  - The critical problem here is a failure to bound the data mapping and ETL exercise by focusing on the actual end use of the data elements being sourced. As stated earlier, if there is no understanding of the use case, ETL development can and does quickly become an unbounded exercise with no demonstrable value, and an end unto itself.

- **Analytical processing seen as ‘external’ to the warehouse:**
  
  - Another ‘anti-pattern’ in data warehousing for financial services is the tendency to view analytical processing as external to the warehousing environment. As stated earlier, the owners of a data warehouse (IT organizations) view their primary and sole responsibility as being one-stop shop providers of source data to any number of analytical consumers.
  
  - The fundamental outcome of this narrow view, is a clearly observed disconnect between users and the data providers. The analysts and business managers, who lack an understanding of the source data landscape, find it difficult to communicate their needs to the IT organizations that correspondingly lack an understanding of the uses of the data.
  
  - As a result, all analytical processing, computations and calculations are viewed as external to the warehouse environment. For example, the same set of core contractual/instrument data is fed through multiple distinct flows and transformations to different risk groups as well as finance. This results in the proliferation of ad-hoc, brittle data flows to consumers from the warehouse, and in extreme cases, consumers bypassing the warehouse altogether.
  
  - This is further unsustainable in the light of the credit crisis. As another example, a coherent enterprise-wide view of the funding/liquidity position of a large institution requires stress testing the entire balance sheet in a uniform manner – which equates to providing a consistent set of data
inputs (contracts and stress parameters) to all risk groups in order to generate asset values under baseline and stress scenarios and more importantly, managing these data flows in a uniform way.

• Having an ad-hoc, fragmented landscape of data flows into different communities makes this impossible – and this situation is no longer tenable.

• Lack of unified views of results and analytical outputs:

  • Another consequence is that the results of analytical processing end up in data mart silos outside the purview of the data warehouse. Since the actual results are derived within these data marts, there is no guarantee of consistency in the reports/results across different data marts unless this has been designed from the outset.

  • This is an unacceptable situation in an age when a unified view of risk is considered necessary for the very survival of an institution.

  • Additionally, there is an emerging class of analytics such as risk-adjusted performance measurement (RAPM) that require unification of results from the traditionally disparate risk and finance realms.

  • The lack of a unified view of results and analytical outputs hence prevents financial services institutions from realizing strategic information goals – like a unified, risk-adjusted view of customer profitability. It also prevents them from being able to respond to regulatory inquiries and audits, which could result in reputational risk as well as financial impact in the form of fines.

The Solution: Guiding Principles for a Next Generation Data Warehouse

All these symptoms lead to the necessity to rethink standard data warehouse practices as they are applied to the financial services industry. Below, we lay out a set of core principles necessary to accomplish this:

Understand the Use Cases

The core underlying principle driving the data warehouse should not be ‘data for data’s sake’ but a clear, unwavering focus on the end uses supported by the data warehouse environment.

The first step is to clearly identify the key, top-level analytical solution areas in financial services:

• Risk

• Performance/Profitability

• Customer Insight

• Compliance
These represent the constituencies in any financial services institution that are the owners of analytical data flows, and hence consumers of data from operational systems that are used as inputs to these flows.

Second, it is equally critical to understand the use cases that exist at the intersection of the above areas. This is a key idea – since it involves moving beyond a silo’d view of departmental data needs to a holistic view that includes all the cross-functional use cases. Increasingly, this is a strategic imperative for financial institutions, and a new class of emerging regulatory and competitive mandates is driving this.

Any “single source of truth” data platform has to therefore recognize and account for these new emerging needs.

Examples of these include:

• Liquidity Risk
• Regulatory and Economic Capital
• Risk Adjusted Performance Measurement
• Customer Profitability

Figure 4: Oracle Financial Services Analytical Applications Representation of Typical Financial Services End Use Cases
Understand the Processes that Produce or Consume Data

The second critical and often overlooked point in rethinking data warehousing for financial services, is the need to truly appreciate the computational methods and techniques themselves that are used in financial services analytical processes.

Consider the following; Monte Carlo simulation is a statistical technique that is used in many different contexts in Market Risk (Simulation-based VaR), Credit Risk (Full-revaluation based Exposure calculations), Operational Risk (Simulation based Op-VaR) and Derivative pricing. Likewise, there are a number of similar shared statistical and computational approaches that underpin the engines and applications used in financial services analytical processing – even if the engines themselves are disparate, black-box components and provided by different vendors.

In any case, these computations place specific, clear demands on the underlying data provider (and hence the data model) – for example, the need to clearly identify specific scenarios used in a stress test, the scenarios used in developing or calibrating a computational model, scenarios used in production runs and the detailed dataset corresponding to each such scenario.

Therefore, knowledge of the computational processes that depend on the warehouse expands the warehouse from being merely an ‘external’ data provisioning platform, to a central platform that encompasses all these analytical data flows, thereby guaranteeing consistency, traceability and verifiability of both the data inputs and the generated results.

Understand the Ultimate Usage of the Analytical Outputs

Finally, the warehouse environment needs to support not just an enhanced approach to data provisioning as outlined above, but also the complete set of use cases with respect to analytical outputs/results used in reporting and business intelligence delivery.

As highlighted earlier, the proliferation of silo’d data marts is a key problem that needs to be addressed in financial services data management. To do so, the warehousing environment should provide for a single, unified data foundation for reporting and business intelligence (BI) that allows data to be gathered from all of the supported analytical processes within the institution. Minimally, this requires the data model to provide a consistent set of conformed dimensions across all relevant computed facts and measures.

This is usually considered a ‘data mart’ issue whose ownership is poorly defined. Rather than allow such issues to develop, the data warehouse should provide such a unified foundation from the outset. While individual reporting/business intelligence solutions can continue to serve individual user groups, the common results/reporting foundation ensures the ability to support reporting/business intelligence for the class of cross-functional analytical use cases highlighted earlier.

The Next Generation Data Warehouse: Architecture

With these basic principles in mind, we can lay out the architecture for a next generation data warehouse environment for Financial Services:
Based on the foregoing, there are two key, inseparable parts of this architecture:

Data Model

The data model represents the combination of all the different representations of data within the warehouse environment, for different needs. Specifically, the minimum set of capabilities of the model should include:

- A readily deployable model for a unified repository of business/operational (‘source’) shared/shareable across the four key analytical areas outlined above
- The model should clearly be mapped in detail (at the level of each attribute) to a well-understood set of analytical use cases.
- The ability to deploy this model incrementally, to deliver or support specific analytical solutions in a phased, time-boxed manner (more pointedly, to avoid ‘boiling the ocean’ with respect to source system mapping/ETL
- The ability to extend this model to support customization for individual institutions using a clear, practical conceptual model

Results/Reporting

- A readily deployable, unified data model for capturing the results/outputs across the various key analytical processes in Financial Services
• The ability to identify, and trace the detailed, granular data inputs and outputs associated with specific scenarios of analytical interest

• Conformed Dimensions to support development of cross-functional analytical reporting without needing to integrate data mart silos

Figure 6: Next Generation Data Warehouses Will Need to Effectively Provide for a Single, Unified Results Area Common to All Key Analytical Areas

Analytical Infrastructure

As highlighted earlier – the next generation warehouse needs to provide full support for end-to-end analytical processing, rather than be a passive data provider for such engines/computational capabilities.

This is a critically important idea – only by supporting an operational model focused on complete analytical data flows rather than addressing piecemeal source data requirements, owners of a data warehouse can prevent the proliferation of ad-hoc, silo’d data repositories, thereby gaining far better control and visibility over data management than they do today.

Therefore, even more critical than the model itself is the need for a set of processing components or frameworks that constitute a unified analytical infrastructure necessary for financial services. Together, these components and frameworks deliver a single data management platform supporting all classes of financial services analytics that provides one centralized point of control and operation that across different analytical user communities.

Minimally, such an infrastructure would provide the following capabilities:

• End-to-end data traceability and metadata management capabilities

• Data quality assurance developed specifically for financial services institutions
• An open data provisioning platform that can easily support existing legacy analytical applications/solutions

• A full menu of computational capabilities required to develop/deploy all classes of computational models required to support analytics, including

• Statistical libraries
  - Numerical methods
  - Pricing calculators for different instrument types

• Ability to quickly configure and deploy end-to-end solutions without extensive development effort dedicated to ‘plumbing code’

• A full set of centralized operational/administrative capabilities (security, scheduling, etc.) to operate the environment

Figure 7: Next Generation Data Warehouse Common Frameworks for Operation, Control and Management

Scalable, Highly Performant Hardware/Database Foundation

One of the defining characteristics of data warehousing has always been the focus on processing high data volumes (due to historical data storage) and extreme scalability/performance needs of an analytical environment. Given the foregoing discussion, it should be clear that the next generation warehouse architecture for financial services is no different in this regard.
It requires corresponding hardware/database server infrastructure that can handle the diverse set of analytical workloads encountered in financial services. These can range from complex single query performance to execution of a complex, interrelated set of processing rules with minimal latency over Petabyte-sized data sets, while simultaneously reducing/eliminating the need for extended batch windows.

So far, the general approach to these problems has been to position highly specialized data warehousing appliances that purport to address scalability specifically using approaches like MPP, columnar compression, column stores etc. However, these approaches have been characterized by:

- **Cost:** Specialized hardware invariably carries high costs of acquisition and ongoing maintenance. Also, it is very difficult for a CIO to size the data warehouse upfront – typically resulting in over or under-investment in infrastructure. Hence, the need is for a platform that can scale with the growth of the data warehouse.

- **Lack of integration with higher-level layers in the data warehouse:** As discussed previously, a significant portion of analytical processing in the financial services industry involves complex statistical and computational processing flows. The inability of analytical applications to leverage the power of the database platform can hence become a critical gating factor in scaling these processes. Increasingly, there is a move towards locating the computational processes as close to the data as possible – rather than require costly, network-constrained movement of large quantities of data to and from the analytical engines/processes. Hence, the scalability is not just a function of whether the hardware/database platform can handle extreme query workloads, but critically a function of how closely integrated the data warehouse platform is with the infrastructure, and how it leverages the power of the database.

**Summary and Key Takeaways: The Evolving Nature of the Data Warehouse for Financial Services Institutions**

Perhaps the key overarching theme is that the discipline of data warehousing in financial services is shifting fundamentally – from an assembly of generic components and tools towards a specialized platform approach that supports the unique analytical needs of financial services institutions worldwide.

**Focus On End-to-end Data Flows to Support Key Use Cases**

Rather than merely being a provider of operational/business data for downstream analytical consumers, the warehouse should be a single foundation that supports end-to-end analytical processing including data sourcing, calculation and aggregation processes, and results/reporting for every use case.

Minimally, this requires the IT organization and the user communities to necessarily work together to understand and agree on:

- The list of use cases for each distinct user community
• The list of use cases that straddle different user communities
• The minimum, common set of underlying data elements to support each case through its full processing lifecycle from data sourcing to results/reporting

Purpose-built Analytical Platform Rather than a Collection of Tools

Custom assembly of a data-warehousing environment has historically proven costly and prone to high failure rates partially because of ill-defined and often overreaching scope. As presented in this document, a more reasonable way to address this problem is to utilize a unified analytical platform that can support all the key requirements and usage patterns of a typical financial services institution – rather than attempt to combine general-purpose tools to achieve when facing next generation data warehousing projects.

Such a platform should include all components – a full hardware and software stack from disk to applications, as well as the maximum practical level of pre-integration to enable quick deployments. It should necessarily be open and extensible – from the ability to create and deploy new analytical processing flows, to extending the data model, to adding capacity at the infrastructure level.

Deep Integration with Underlying Infrastructure

At a technical level, the warehouse platform should be deeply integrated with the underlying infrastructure – specifically being able to leverage the power of the infrastructure to scale out in a flexible, cost-efficient manner. Adding capacity should be transparent, and require minimal disruption of the existing environment.

Conclusion

We have outlined a set of principles that can drive the next generation of data warehousing for financial services. Fundamentally, this involves a rethinking/modification of “established” data warehousing practices to meet the specific analytical needs and goals of the financial services industry. Approaching the data warehouse “problem” using these principles can provide IT organizations in the industry with a unified analytical platform that will enable them to support both current and future analytical requirements in a cost-effective, scalable and flexible manner.