ORACLE
FUSION MIDDLEWARE
DATA INTEGRATOR

An Oracle White Paper
May 2011

Maximize Success on Data Integration Projects with Oracle Data Integrator

Executive Summary

An increasing number of companies recognize that collecting, integrating and analyzing data is a critical lifeline of their business. Without it, enterprises have an incomplete picture of business performance, trends and lack the visibility into business patterns over time. Without the right information, companies risk losing market share and being surpassed by the competition. Yet data, often never makes it to the mission critical systems that need and require it. Data integration initiatives are seen as a critical enabler to improve the accessibility, timeliness, and quality of mission-critical data. In order to make enterprise data available to the systems that need it, you need the right tools in place to ensure data makes it to where it needs to be and data can be represented, changed or transformed as the business requires it.

This whitepaper reviews the 5 important success factors that drive data integration projects to success and how Oracle's flagship product for ETL - Oracle Data Integrator Enterprise Edition (also referred to as ODI), is a core component for any data integration project. In addition, this paper will discuss what the key challenges of data integration projects are and why ODI is suited for any data integration initiative, big or small.

The State of Enterprise Data

Many enterprises are facing the challenge of how to handle so much data, fragmented across the organization. This data is often found in applications, various data stores, master data management and enterprise resource planning systems, as well as software as a service applications. Since bringing together disparate systems of information is critical for decision making, the need for data integration has grown significantly over the last few years, and is now viewed as a critical enabling technology for any information driven decision making enterprise.

At the same time, the amount and type of data stored in these systems has changed dramatically. In recent years, there has been a deluge of information capture within the enterprise, and IT organizations are struggling to cope with the increased volume of information captured on a daily basis. Therefore, IT organizations are seeking ways to:

1. Optimize business insight with corporate information to help take proactive action and increase market share and value
2. Maximize existing investments in their core systems so that they can avoid additional IT purchases and stretch their dollar
3. Further leverage the various treasure troves of information in their enterprise in order to provide timely, actionable business insight.

Successful data integration projects enable organizations to extract insight, deduce trends and make predictions about their business, and data warehouses are at the core of these projects. Data warehouses provide historical data across a wide area, reveal trends visible in such data to help management make better predictions, and provide greater visibility into past and future trends. Looking at the information in a data warehouse across time can reveal patterns of behavior that aren't otherwise visible. For example, a

rental car company might learn that in Western Europe the highest number of convertibles is rented on Fridays by German men between 50 and 60 years old. The company can then create a special promotion aimed solely at this group.

Data Integration projects: Failure is not an Option

The Data Warehousing Institute reports that only 11% of data warehouses represent a consolidated view of data. In most cases data warehouses do not reflect a successful data integration strategy and often contain only partial information. This is not due to a lack of desire on part of the integration team, but rather due to the fact the data integration projects can be extremely challenging. The reasons for this are many, but here are some of the key ones:

Multiple, fragmented siloed systems. Data volumes increase through any number of sources. Data sources are no longer only databases, but include anything from a mainframe to an Excel spreadsheet and add to this packaged applications that help make customization convenient. The proliferation of packaged applications where users can easily customize to suit their needs also adds to the volumes of data and data sources across a company. As organizations add to their IT repertoire, over time what one may find is that there are hundreds of systems, running on multiple platforms, older versions and in geographically unique locations. Fragmented data resides in silos and adds a whole new dimension of difficulty because of different data models and structures. Furthermore fragmented systems have the huge additional burden of carrying multiple versions of the truth in these different information silos.

Complexity of implementation. As businesses turn to packaged applications to satisfy business needs, they are often customizing them and indirectly adding complexity to their applications. Business needs often dictate and drive custom coding efforts to get applications "to do what they need". This results in hundreds of custom modules that are developed, not only to achieve business-driven functionality but also to move data between multiple data sources – resulting in a maintenance and upgrade nightmare. Ultimately, this translates into higher cost to continue to use, maintain and upgrade these customized systems. With numerous intertwined systems, business users find it very difficult to obtain custom views of data to view trends, patterns and make predictions in business performance. As a result, organizations have to "make do" with partial information giving partial business insight which means they can ultimately lose market share, customer loyalty and fall behind the competition.

Lack of consolidated, unified approach to data integration. Businesses need to consolidate their disparate data and move it from place to place so that it can be used by the appropriate applications. As a result, the need to transform and move data has been the leading reason for adopting data integration tools. Yet, once the decision is made to embark on a data integration project, many companies fall short on successfully implementing and leveraging the technology in such a way that maximizes benefit to the business. A lot of data integration is handled by point-to-point custom scripts that grow organically to the point where they are un-maintainable. Furthermore these scripts also encode business logic, leading to proliferation of such logic across multiple scripts, languages, and frameworks for integration. This means that changes to such business logic are extremely hard to implement and consume significant amounts of time and effort -- furthermore they result in poor quality of information integration. Data must be synchronized in an automated and reliable manner across all platforms for companies to have one version

of the truth. Errors caused by inconsistent, data and manual data entry can prove very costly for companies and disrupt business activities.

Criteria for Success

Success on a data integration project varies depending on who you ask. Business users primarily view data integration success as getting "a single version of the truth" and not having to access multiple applications to answer simple business questions like – how many service level agreements were in compliance with this month or how many sales prospects were converted over to customers, etc. For the line of business, seamless, consolidated information upon which business action can be is extremely valuable. However, IT staff view a successful data integration initiative as one that is not time-intensive, adeptly handles complex, heterogeneous systems and offers high performance and handles exceptions easily and cleanly. Here are a few success factors or criteria for data integration projects:

- **Integrate data from diverse data sources.** This is a key factor that determines success because data integration initiatives are riddled with complex, cumbersome data sources. Often applications overlay multiple databases, data may reside in data marts or "slices of databases" and being able to seamlessly connect to these data sources, extract and move bulk data is critical. Business users demand consolidation and for many enterprises, bringing together master data from CRM, ERP, and BI systems involves unifying data from dozens, and possibly hundreds, of diverse data sources. In a CDI scenario, for example, you may need to bring together CRM data from a relational database management system (RDBMS), sales data from mainframe file extracts, customer segmentation data from Microsoft Excel, and third party demographic data via Web services. Assembling customer data from these diverse data sources requires an integration solution that can successfully access and interpret their distinct interfaces, structures.
- **Accomplish high performance information processing.** Corporate IT systems like data warehouses, various repositories, etc have terabytes of valuable information. Over time the amount of data collected has grown either through additional IT implementations, mergers and acquisitions or third party data sources. As a result of this information explosion in enterprises, there is a need for fast, agile development cycles. Often more times than not, there are pre-built maps for millions of records, as well as new maps that are being built which require fast bulk/batch performance while leveraging existing data warehouse or IT investments. With a rise in the numbers of transactions, organizations seek solutions to help them move only the necessary data, in the easiest and fastest way possible without having to spend more money on net new IT systems.
- **Increases in productivity.** Data integration projects are often stalled because of the amount of time it takes to build maps between known systems that are not integrated or because there is a huge amount of custom coding and scripts for which transformations need to be built. Take this coupled with the fact that often there are multiple runtime environments, several developer tools where transformations are created and need to be reconciled- this results in extended project time lines and delayed implementations. Development productivity gains play an important role in getting to the finish line of these projects faster. These gains are seen through declarative

design tooling, built-in templates and modules that help jump start connections to critical applications.

- **Results in high quality enterprise data.** At the end of the data integration project, one of the most visible improvements that any user can see is the quality of data. If the data quality has improved and has fewer inaccuracies – it results in real, trustworthy business insight and ultimately analytics and reporting you can base business decisions on. All enterprise software projects are at risk of failure from bad, "dirty" data. Here's why – poor data quality has a cascading effect - It doesn't just remain at the source. Data warehouses, CRM, BI or ERP systems incorporate this data and reports are executed on it and shared with other systems through shared interfaces and complex data relationships. Now, what you have is bad data quality rampant throughout the enterprise. Unfortunately, "dirty data" also costs money it results in expensive corrections for marketing campaigns that target incorrect groups, returned mailings, etc. Creating standardized and consistent information ensures that business users are more confident about business information and in a better position to grow the business and remain competitive.

Oracle Data Integrator: Benefits and Advantages

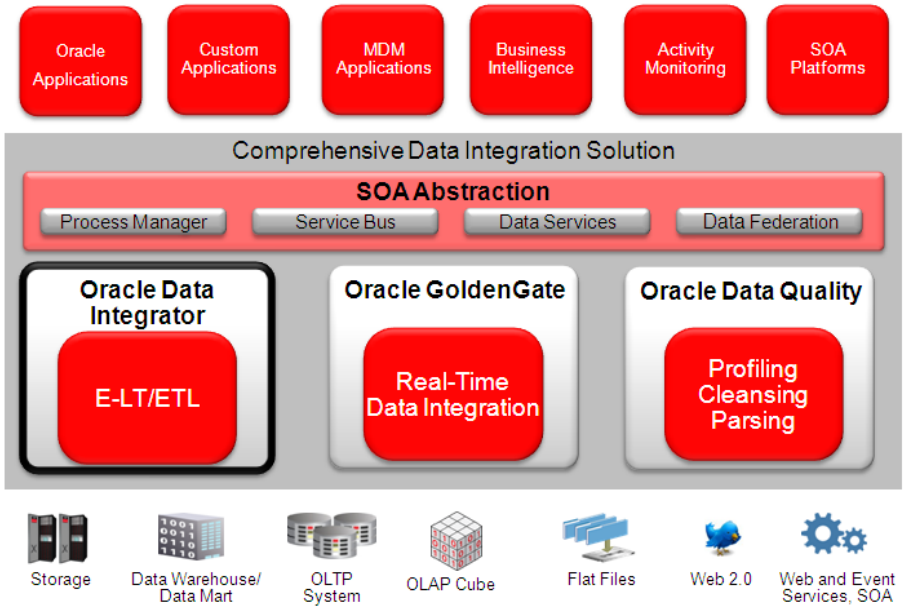
Oracle's data integration product line, including Oracle Data Integrator Enterprise Edition, Oracle GoldenGate, and Oracle Data Quality, offers the best-of-breed solution for providing timely, accurate and continuous access to your data. Easy-to-use, and standards-based, Oracle's data integration solutions dramatically improve productivity, provide unparalleled efficiency, and lower the cost of ownership. Oracle's data integration solution is open to work with heterogeneous systems, and fully certified with and optimized for Oracle Exadata and Oracle Business Intelligence Solutions.

Oracle Data Integrator offers organizations the power of high performance, increased productivity, coupled with data movement and transformation across different platforms. It is the tool of choice for data integration projects. ODI makes it possible to connect to all major databases, data warehouses, service-oriented architectures, BI systems and helps lower the total cost of ownership among information-centric architectures in the enterprise. A large component of data integration projects, ODI specializes in bulk data movement without compromising performance and maintaining existing data relationships and transformations. Here are a couple of key areas that set ODI apart from other data integration solutions:

- **Reduce costs and faster time to value.** Oracle Data Integrator Enterprise Edition's declarative design environment enables developers to focus on architecting interfaces so that they can focus on what to do and not how to do it. This in turn leads to quick implementation times, without the need for intermediaries and additional solutions. Based on next-generation ELT architecture, ODI doesn't require a separate standalone server be purchased – so no extra ETL server. Additionally, ODI enables technical teams to extend their existing investments in IT systems and leverage resources more effectively to partition workloads more efficiently. Keeping track of data integration activities is easy with integrated management and monitoring modules and connections to Oracle

Enterprise Manager. ODI eliminates the need for multiple development tools and IDEs and relies on Developer, so organizations don't have to invest in multiple development tools that add time to project cycles and increase project costs.

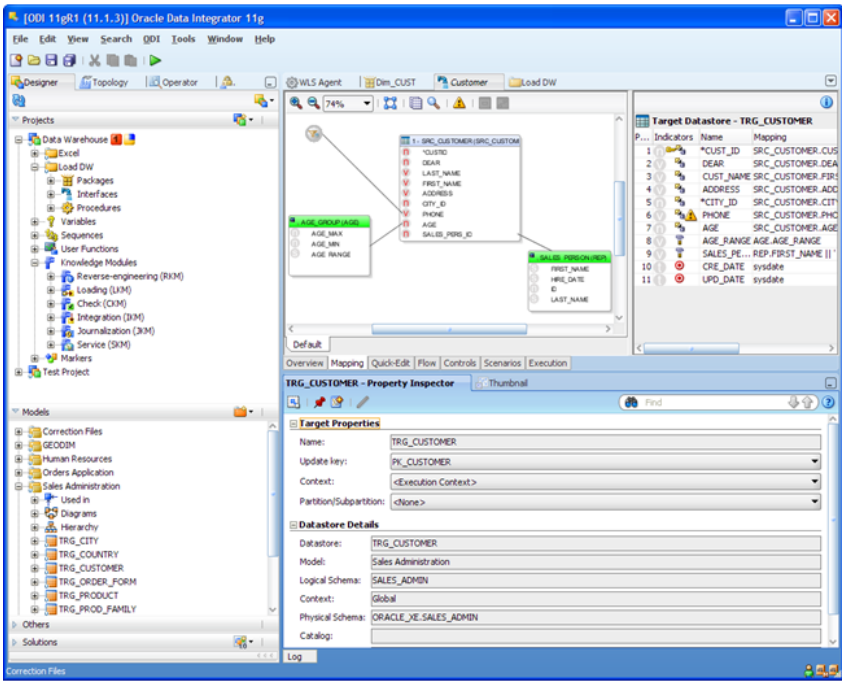
- **High performance.** Conventional ETL causes additional hardware dollars to be required when data sets grow or there is a need to scale and increase performance. ODI delivers top performance and sprints on Oracle Exadata. Oracle Data Integrator Enterprise Edition's (ODI can run natively on Oracle Exadata to offer the most efficient data integration platform. Unlike other ETL solutions, ODI allows every transformation to occur on Oracle Exadata, without adding any additional hardware. With this architecture, data never leaves the Exadata server, and ODI scales along with the Exadata server.
- **Heterogeneity.** This is a key feature of ODI and offers deployments the flexibility to implement on any database – Oracle or non-Oracle. ODI is optimized to scale linearly as data sets grow and processes only execute during integration runs. This means that server resources are not unnecessarily wasted during interim periods. On benchmark studies, ODI has shown that up to 7TB per hour of real world data loading and complex transformations are easily handled when used with Exadata. On engineered systems like Exadata, ODI is optimized to take full advantage of the enhanced server's power, management facilities and modularity to deliver the best bulk data loading solution with the simplest set up and the fastest performance.



Title: Oracle Data Integration product suite – Oracle Data Integrator provides bulk data movement, loading and data transformations to help extend exiting IT investments and drive data integration projects to completion on-time.

- **Boosts project delivery.** Establishing connections and access for bulk data movement to and from multiple types of applications is very efficient with ODI. ODI has over a 100extensible integrations known as Knowledge Modules to help leverage best practices and jump start data integration

projects. The advantage of out-of-the box modules is that developers have a clear methodology to relay on and can use these extensions to help develop robust code quickly versus 'just making it work'. The Knowledge Modules contain best practices Developers can leverage in order to quickly and efficiently move data to and from applications along with sample code and scripts. These out-of-the box integrations are available for GoldenGate, BI Apps for SAP, SOA and many others. Standardized code via Knowledge modules enables faster implementations and easier maintenance going forward. ODI enables development productivity gains of up to 40% and makes environment setup less difficult by removing the gotchas that often delay projects. This translates into successful data integration projects that are on time and under budget. Hot pluggable, modular, flexible and extensible capabilities are what make ODI a clear choice for successful data integration projects.



Title: Oracle Data Integrator Knowledge Modules

- **Strong connections with BI, data warehousing and SOA technologies.** More and more customers are seeing the value of implementing both ODI and SOA technologies for uses case such as decoupling business applications, orchestrating services and transforming large files of data. A major retail customer adopted ODI with SOA and saw a 75% improvement on performance over running alone on SOA. This stems from ODI boosting performance of business process executions and taking a standards-based approach versus an inflexible approach. Many of today's BI solutions are simply batch transfers of operational data into data marts or "data slices" that are served up into dashboards and batch-generated reports. BI systems are only as good as the data they represent. Oracle's customers report to us that over 60% them realize that data integration enhances the value of BI applications by increasing the quality and consistency of data. ODI's extensibility for BI applications, Hyperion Essbase, Financial Management and Planning helps

eliminate the symptoms of most BI implementations: inaccuracy, obsolete data or out-of-sync data warehouses. ODI is a critical enabler of business intelligence and provides users with accurate, high quality information that they can confidently take business action by making actionable BI a reality. Through assured high performance bulk-data movement, data warehouses are receiving the information they need to make sure BI applications can present the most accurate and complete information to the business user.

- **Real-time data warehousing.** Oracle Data Integrator's tight integration with Oracle GoldenGate provides customers with continuous feeds from operational systems via real-time change data capture as well enables a thin middle tier transformation on the database platform (target). Through a log-based CDC approach, organizations can source data from OLTP systems without impacting performance and feed the ETL/E-LT system with a continuous stream of real-time data. This method not only decreases data latency for BI systems, but it also eliminates the reliance on batch-processing windows, while allowing continuous operations for business-critical systems, ultimately resulting in a true real-time integrated solution.

- **Data quality for Trusted Business Insight.** Knowing the quality of your data and being able to analyze it for patterns and historical information is vital to implementing business improvements as well as planning for future business action. ODI seamlessly works with Oracle Data Quality and Oracle Data Profiling to provide business users with the ability to assess the quality of their data through metrics, to discover or infer rules based on this data, and to monitor historical metrics about data quality. ODI coupled with Data Quality brings critical data quality in the heart of any enterprise data integration strategy.

Oracle Data Integrator Enterprise Edition: Making it Real

Oracle Data Integrator has enabled numerous customers realize success, quick time-to-value and seamless integration without interruption to existing systems. Let's specifically take a look at two customers and the challenges their businesses faced and how ODI helped them achieve better business insight, cut costs and improve efficiency.

Raiffeisen International Bank-Holding

As one of the leading financial institutions in Central and Eastern Europe, with over 14 million customers and over 3,000 locations globally, Raiffeisen first-hand sees the value of successful data integration. They faced several challenges which included replacing the existing ETL tool which failed to provide the scalability to populate data warehouses in 12 countries, 2) reduce the number of hand-coded ETL jobs to minimize complexity and reduce high maintenance costs, 3) lastly they wanted to reduce the costs associated with system migrations and upgrades. ODI not only addressed these challenges, but also provided significant productivity gains and eliminated a large amount re-development of existing ETL jobs. Raiffeisen chose Oracle Data Integrator because of its clear benefits, flexible ELT architecture and develop tools to expedite data integration projects. The benefits an ODI solution provided to Raiffeisen include:

- Reduced manual effort for migration of 14.000 ETL jobs through accurate conversion
- Scaled as data volumes increased due to massive business growth
- Improved data warehouse processing performance by 40%
- Reduced development and maintenance efforts and cost from a template based approach

NYK Line

Another customer that succeeded in its data integration efforts while reducing costs and improving efficiencies is Japan-based, **NYK Line**, a 120 year old shipping company. This company coordinates 700 ships and handles tens of thousands of cargo containers from around the world. NYK Line needed to implement a new global shipping solution and meet new US Customer reporting requirements to notify Customs within 24 hours before a container was loaded on a ship in a foreign port. To accomplish this they need tighter integration between key business systems such as: e-commerce, corporate information systems, ERP and legacy integration with regional systems.

In the short term, NYK Line was able to meet the U.S. Customs Service requirements for notification. "In less than a month, we were able to implant new business processes using Oracle Data Integrator that report on container movements and provide that information to U.S. Customs," says Loganathan. "That's when our business managers realized how good the time to market is with Oracle Data Integrator-based solutions. Now when they need a solution, they ask that we do it in Oracle Data Integrator." NYK Line uses Oracle Data Integrator to provide the data integration services for its container-tracking application used by customers around the world, which needs to be accessible 24 hours a day.

Conclusion

Data integration project success depends on a variety of factors: business and IT joint ownership of the initiative, a consistent, proactive approach to improving the delivery and quality of data and last but not least a foundational solution to bring all the technical elements together. Oracle Data Integration's comprehensive offerings encompass bulk data movement, loading and data transformation, data quality for trusted business insight, data services and Oracle GoldenGate for real-time, log-based change data capture, routing and delivery between heterogeneous systems.

Oracle Data Integrator is a necessity for any successful data integration initiative. ODI's solid data integration engine offers high performance along with tools to improve productivity, cut costs and fit into any IT infrastructure. ODI empowers companies with trusted business insight by providing reliable, timely and trusted data which is essential for any agile enterprise to compete and win in today's competitive landscape.

Maximize Success on Data Integration Projects with Oracle Data Integrator

May 2011
Author: Mala Narasimharajan]

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together