SUMMARY

E-LT is a style of ETL (Extract, Transform and Load) that dramatically improves the responsiveness of Business Intelligence by optimizing data transformations. Many conventional ETL vendors have recently launched marketing campaigns to promote the E-LT (Extract, then Load & Transform) style of ETL – they refer to their style of E-LT as Pushdown Processing.

But not all E-LT approaches are equal. And if you care to discern between marketing hype, and solid technology that can dramatically improve performance while simultaneously lowering ownership costs, then you should learn how E-LT differs from one vendor to the next.

Only Oracle’s E-LT Approach:

- is performance optimized for your Database – whichever DB you use
- can operate without any new IT Hardware costs
- is 100% Java-based and managed within your standard Java/J2EE environment and management framework
- is easily embedded within your existing or planned SOA infrastructure without any extra hardware or proprietary Application Servers
- is not a glorified scheduler that relies on PL-SQL, or other custom-coded DB scripts to achieve maximal performance
- can entirely eliminate needless network-hops for remote data joins
- can operate with no additional energy drain in your Datacenter

No other vendor can claim any of these E-LT benefits.

But much like other E-LT vendor approaches, Oracle’s E-LT:

- uses the power of your Data Warehouse for maximum performance data transformations
- can combine engine-based operations with DB-based transformations to accomplish any level of data transformation complexity
- can scale to any multi-TB level and deploy with parallel processing configurations

To summarize, get to know all the E-LT facts before signing up for a so-called Pushdown Processing solution.
DISPELLING PUSHDOWN PROCESSING MYTHS

Oracle Data Integrator is the only data integration platform on the market to wholly implement the high performance E-LT architecture. Instead of moving all the data through an intermediate ETL transformation server, the E-LT approach leverages the power of the target RDBMS engines to perform the transformations, dramatically improving the performance at a much lower total cost of ownership. Data goes straight from sources to targets, and is transformed many times faster than other tools.

Unlike the Pushdown Processing features recently introduced by the conventional ETL providers, Oracle’s native E-LT architecture is not a mere after-thought. Most of these so-called “pushdown optimized” transformations still occur inside the ETL engines and requires the physical data to transit over the network and through their engines anyway. While pushdown optimization may indeed improve the performance of their classical ETL processes in some cases, the Oracle E-LT architecture is in a performance class of its own.

With so many years invested in the older ETL technology base, it is easy to understand why some vendors defend their entrenched monolithic architectures. But fortunately, most enterprise software buyers have choices. The first step to understanding those choices is to move beyond the hype to learn more about the E-LT architecture itself.

Persistent Pushdown Processing Myths

An E-LT architecture can be the essential key for getting the best performance from a data integration infrastructure, but all E-LT architectures are not the same. The Pushdown Processing vendors would have you believe that SQL can only accomplish the most basic of transformations, or that their engine-based ETL should be the foundation of your enterprise ETL. They would say that the E-LT approach is valuable in limited ways, that they fully support, for very targeted needs where set-based operations can be used.

Dig a little deeper though, and some major weaknesses become apparent.

Myth #1 – Pushdown Processing is an ETL Vendor Innovation

These vendors might have you believe that they went out and invented E-LT just for your benefit, the end-customer, as a way to improve on their already stellar engine-based ETL performance.

Fact #1 – The E-LT Approach is the First and Oldest Style ETL

People have been using the power of database transformations since the invention of the RDBMS. Even today, the vast majority of ETL is still performed with SQL-type scripting executed on the DB. Most conventional engine-based ETL tools are just used for job scheduling on high-end warehouses – because SQL-based transforms are still the fastest in practice. In fact, Oracle was the first to successfully commercialize the E-LT approach with Oracle Warehouse Builder and is now taking it to an even higher performance level with the recent acquisition of
the successful E-LT innovator Sunopsis, which is now the basis for Oracle's Data Integrator product family, a part of Oracle Fusion Middleware.

Myth #2 – Pushdown Processing is Best Because it is Engine-Based
Certain vendors would have you believe that “simple SQL generators” cannot handle the complexity of real-world data – that an ETL engine is required because only it can do the hard work on hard data.

Fact #2 – Oracle E-LT Can Support Highly Complex Data Transformations
Those other vendors are correct – if they’re talking about their own Pushdown Processing style. They use only simple SQL for simple data transformations, requiring bulk data processing on expensive servers to do the hard work. But Oracle Data Integrator uses the full scope of native SQL, generated uniquely for all the common databases, in order to apply the full power and performance of the database towards data transformation. Even the most complex transformations can be handled with ODI’s E-LT approach. The following table provides additional details about transformation true E-LT power:

<table>
<thead>
<tr>
<th>Complex Data Transformations</th>
<th>ODI With E-LT Processing Only</th>
<th>ODI Engine Recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneous lookup</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Lookup with data from the target system</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>If then else logic involving multiple sources</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Basic Aggregations</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Complex Aggregations</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Conversions</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Date and Time Functions</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Math functions</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Database operators (Like, between, soundex)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Strings manipulation</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Increments / counters / sequences</td>
<td>X</td>
<td>(Externally bound counters)</td>
</tr>
<tr>
<td>User Function</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Use target table as a source</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Full XML Transforms</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LOB Transformations</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Other Inline Capabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Command line access</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Emails (send / receive)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Event detection (CDC etc)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Variables management</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Loops, decision points</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Unstructured text support</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Parallel processes</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

“...Oracle Data Integrator loads and parses over 9 million source records to create 611,000 flight records in approximately 3 minutes.”

—Luc Scotts
Project Manager, OAG
Myth #3 – Pushdown Processing can Save Hardware Costs
Pushdown vendors will talk about “the intelligent use of existing computing resources,” but never fully explain where those savings are achieved – since the Pushdown Processing architecture will not actually eliminate any hardware!

Fact #3 – Only Oracle E-LT Eliminates Redundant Hardware
These Pushdown Optimization vendors routinely release benchmark studies that show how fast their ETL engines can be. In one recent benchmark, the “breakthrough” was achieved with a 64-CPU mainframe dedicated to just the ETL part of the process – and that’s something to be proud of? A true E-LT system will allow you to allocate all those additional CPUs – or no net new CPUs at all – to your Data Warehouse. Isn’t it a much better idea to invest hardware where it can be used for the greatest good and by the widest user base? Oracle thinks so.

Myth #4 – Oracle ETL is Just a PL-SQL Utility Tied to the Oracle DB
Pushdown vendors say that Oracle data integration solutions are PL-SQL centric and unnecessarily tied to the Oracle Database.

Fact #4 – Oracle Provides E-LT without Any Oracle DB Required
Somewhat ironically, most of the engine-based ETL vendors rely heavily on their customer’s use of PL-SQL for efficient Oracle DB-based ETL jobs. Ask any experienced ETL architect who has used Oracle DB as a target, oftentimes the ETL job simply drops the data on the DB machine and calls a PL-SQL job! In any case, Oracle’s Data Integrator product does not use PL-SQL by default – although it can – instead, it uses generated SQL native to any DB platform it is targeting. Popular DB’s that Oracle Data Integrator supports include Teradata, Netezza, Sybase, IBM DB2 on all platforms and, of course, all Oracle DB versions.

Myth #5 – Pushdown Processing is Best for Teradata
Pushdown vendors will say that their architecture is best for Teradata systems and other high-end warehouses, but they’re only partly right.

Fact #5 – Oracle Data Integrator is Best for Teradata…and Oracle too
While it is true that the Pushdown Processing approach is better than the old ETL approach, it is a far-cry from the deep native capabilities that a dedicated pure E-LT approach can provide for high-end data warehouses. With Oracle Data Integrator, there is no bottleneck created by a centralized ETL server. Oracle Data Integrator leverages all the native features and utilities of Teradata, such as MultiLoad, FastLoad, TPump, FastExport, parallel processing, merge joins, hash joins, nested joins, complex sort algorithms, and so on. Other high-end data warehouses like Netezza, DB2, and Sybase enjoy the same level of support. Oracle database implementations achieve even better levels of native integration.

Some E-LT Conclusions about Oracle Data Integrator
Data is one of your company’s most important assets, and data integration constitutes the backbone of your enterprise’s IT systems. Choosing the wrong
technology for data integration can cause harmful long-lasting effects that impact not only the IT budget, but also the productivity and responsiveness of critical business divisions within your enterprise.

Oracle Data Integrator (ODI), a key component of Oracle Fusion Middleware, provides a strong and reliable integration platform for your IT infrastructure. Oracle remains deeply committed to the pure E-LT approach, for lowering ownership costs, for achieving best performance, and building on the most open and modern architectures.

ORACLE’S VISION FOR DATA INTEGRATION

In summary, it should be clear that Oracle Data Integrator provides a unique data integration platform, architected for performance and productivity, which provides a high degree of flexibility and modularity. Compared to monolithic ETL or conventional Pushdown Processing approaches from our competition, the ODI solution excels in the following areas:

• **Performance** – thanks to a market leading unique approach – the high-performance E-LT architecture
• **Productivity** – with declarative design that drives reusability, accelerates development and maintenance
• **Flexibility** – a single platform that natively supports several data integration styles and latencies
• **Modularity** – with hot-pluggable knowledge modules that support application and database platform diversity

Oracle’s commitment to our customers has always been to provide the highest quality enterprise software with maximum value. Oracle Data Integrator is a cornerstone technology for providing comprehensive data integration solutions and highlights the Fusion Middleware commitment for hot-pluggability and the broadest possible support for diverse IT environments.

Regardless of the database or applications within your IT ecosystem, the ODI solution can be optimized to drive the highest-performance bulk or real-time transformations. Oracle’s vision is to combine and enable these capabilities from within a next-generation, unbreakable Service-Oriented Architecture that will continue to drive business value within your enterprise for many years to come.
Pushdown Processing Myths
September 2007

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

Copyright © 2007, Oracle. All rights reserved.
This document is provided for information purposes only and the contents hereof are subject to change without notice.
This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.
Oracle, JD Edwards, PeopleSoft, and Siebel are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.