Magnitude Dynamic Information Warehouse
Performance on Teradata

The Magnitude/Teradata partnership offers greater agility & flexibility in data warehousing.

Introduction

Teradata builds fast data warehouses, and Magnitude Dynamic Information Warehouse (DIW) builds data warehouses fast. Warehouses built using the automation Magnitude offers can go live in 90 days or less. Magnitude DIW on Teradata was specifically engineered to take advantage of the Teradata architecture. As a result, customers get:

- Greater agility for capturing new warehouse requirements through Magnitude DIW's business-model-driven approach to building the data warehouse
- Much faster delivery of proof-to-production, leading to faster time to value
- Management of ALL your data on Teradata, fully leveraging your investment in the platform
- Reduced TCO via automation capabilities for developing and maintaining the warehouse

Magnitude DIW (formerly Kalido) is known for rapid deployment of data warehouses. But how does that resulting warehouse perform on Teradata? To confirm how Magnitude DIW on Teradata scales and performs, Magnitude and Teradata jointly engaged in an extended testing exercise in the Teradata porting lab. The Teradata benchmark system was a two-node, 36 AMP Enterprise Data Warehouse appliance with 2.6 terabytes of storage. Each node consisted of four cores (2.66 GHz) and 8 gigabytes of memory. The tests were conducted at Teradata's Rancho Bernardo R&D Center.

This paper details the success of the benchmark tests.

Highlights include:
- Loaded all US prescriber reference data for a top-10 pharmaceutical company in 2.5 minutes
- Magnitude DIW's high-performance transaction loader delivered all the typical ETL processes required to onboard data, and ran between 12 and 60 times faster than equivalent operational ETL processes
- The Magnitude DIW transaction staging loader performed consistently on loads up to 1 billion transactions per load at a rate of 455 million transactions per hour
- Both transaction data and large dimension loads performed consistently over time
- Loaded 24 months of prescription activity for a mid-sized pharmaceutical company in 16 minutes
- After tuning, generating a result set was reduced from 22 minutes to 3 minutes, demonstrating substantially improved query performance

Common Questions about a Magnitude Warehouse

There are two types of performance that matter in data warehousing: load performance and query performance. Once the Magnitude warehouse is built, you will experience the same query performance as on Teradata. This document focuses primarily on load performance. Before we explore load performance benchmarks, it makes sense to share some common questions we receive (and their answers) regarding Magnitude DIW and how well it performs on Teradata given its highly flexible architecture.

It's my understanding that Magnitude data is stored in a "single table" – is that true and will that make it difficult for me to tune the warehouse?

It's not true that all data is stored in a single table. A very small percentage of the total data is persisted in a small number of tables, and even this data is fully replicated in a traditional relational schema for access. Magnitude DIW employs a highly generic storage structure for all metadata and low-volume reference data, though this is purely to facilitate data reorganization to meet evolving data extension requirements. Magnitude DIW also provides many features to enable the warehouse administrator to optimize the physical design that is driven directly from the Magnitude business information model (BIM). Additionally, the Magnitude solution enables the use of advanced tuning features as well as provides mechanisms for physical modelers to fine-tune the design of the warehouse schema without manually migrating the data or changing the inbound interfaces. As a result, the Magnitude DIW can be sufficiently tuned to support the data warehouse needs of many large companies, including leading pharmaceutical and insurance companies, financial institutions, the world's largest consumer goods companies and major oil and gas firms.

As the Magnitude Data Warehouse is created and managed via a "business information model," do I still need to use ETL to load the data?

The Magnitude solution performs many more operations (in addition to what is done in ETL) required to design, build and operate a data warehouse. All load operations necessary for populating a warehouse stored on Teradata are configured and executed by the Magnitude solution using a native ELT process. This is in addition to automating many of the most common functions required to build and operate a data warehouse, such as those identified in Figure 1 below.
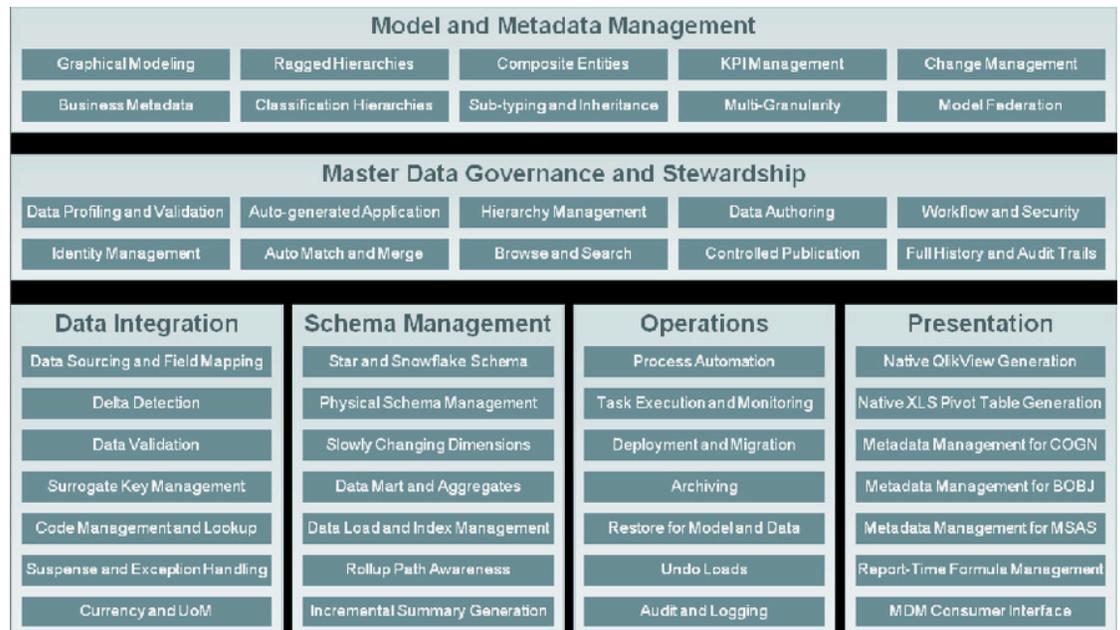

Model and Metadata Management

| Graphical Modeling | Ragged Hierarchies | Composite Entities | KPI Management | Change Management |
| Business Metadata | Classification Hierarchies | Sub-typing and Inheritance | Multi-Granularity | Model Federation |

Master Data Governance and Stewardship

| Data Profiling and Validation | Auto-generated Application | Hierarchy Management | Data Authoring | Workflow and Security |
| Identity Management | Auto Match and Merge | Browse and Search | Controlled Publication | Full History and Audit Trails |

Data Integration	Schema Management	Operations	Presentation
Data Sourcing and Field Mapping	Star and Snowflake Schema	Process Automation	Native QlikView Generation
Delta Detection	Physical Schema Management	Task Execution and Monitoring	Native XLS Pivot Table Generation
Data Validation	Slowly Changing Dimensions	Deployment and Migration	Metadata Management for COGN
Surrogate Key Management	Data Mart and Aggregates	Archiving	Metadata Management for BOBJ
Code Management and Lookup	Data Load and Index Management	Restore for Model and Data	Metadata Management for MSAS
Suspense and Exception Handling	Rollup Path Awareness	Undo Loads	Report-Time Formula Management
Currency and UoM	Incremental Summary Generation	Audit and Logging	MDM Consumer Interface

FIGURE 1: Sample Data warehouse operations automated in Magnitude Dynamic Information Warehouse (DIW)

How does Magnitude DIW perform compared to ETL?

Since its inception, the Magnitude solution has progressively advanced the use of database architecture for the most processing-intensive operations through the use of ELT (Extract, Load and Transform) architecture. This is a similar architecture to "Push down ETL" employed by the leading ETL vendors; the only difference is that due to Magnitude's comprehensive understanding of the business information model and data, we are able to fully automate and manage things such as the staging layer, in-process temp tables, dynamically driven statistics collection and the optimization of physical structures. This enables a Magnitude warehouse on Teradata to deliver massively scalable, high-performance operations without significant processing by an application server. In a customer benchmark, Magnitude's high-performance transaction loader delivered all the typical ETL processes (see Data Loading section, below) required to onboard data, running 12-60 times FASTER than equivalent operational ETL processes.

Magnitude Data Warehouse Performance

To characterize the performance of a Magnitude Data Warehouse, the following key facets of performance were considered:

- Model complexity
- Industry relevance
- Tuning opportunity
- Leverage of Teradata features

To evaluate how a Magnitude warehouse performs, a set of benchmarks was executed on a modest, two-node, 36 AMP Teradata system. For this exercise, the complete Magnitude Pharmaceutical Sales and Marketing Analytics Solution business information model was selected for complexity and industry relevance. Load and query tests focused on the key dimension and fact data related to prescription activity or Retail Sales data, typical to that of third-party data supplied to most pharmaceutical companies for prescription products.
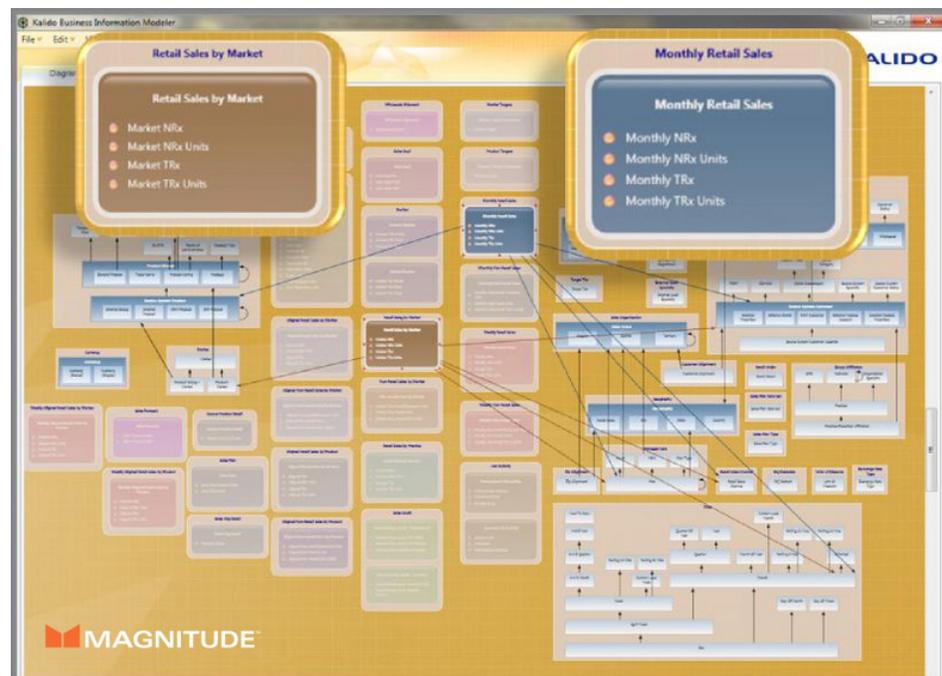

FIGURE 2: Magnitude Pharmaceutical Sales and Marketing Analytics business information model

Data Loading

What we mean by "Loading"

It is important to qualify the function of the Magnitude loaders. These loaders are responsible for executing all the typical functions required to populate and integrate data into the warehouse. In the Magnitude engine, load mappings are defined to the business model rather than physical table structures. As a result, these mappings automatically inherit the business rule logic defined in the business information model. And more importantly, they also inherit any changes to the business logic captured in subsequent changes to the business information model. Therefore, the loaders are not simply moving data from table to table or from a file to a table, etc., but rather are performing essential tasks such as:

- Surrogate key mappings
- Referential integrity enforcement (and suspense management should these fail)
- Duplicate validation
- Lookups
- Key integrity management
- Data partitioning for parallel execution
- Contra processing
- Exchange rate conversion
- Plus many more commonly required data integration tasks

Types of Loaders

In a Magnitude data warehouse, we have a main loader, Universal Load Controller (ULC), which transparently enables loading of three different types of data: transaction data, large dimension data, and generic reference data.

It is important to understand the inverse relationship between the typical data volumes and the number of elements of these three data types. The information in Table 1 below is taken from a customer implementation of the pharma sales and marketing model and clearly illustrates this inverse relationship:

Type	% of Objects in the DW	Rows	% of Data in the DW
Transaction Data	17%	2,963,345,136	98.43%
Large Dimension Data	6%	47,006,826	1.56%
Generic Reference Data	77%	382,110	0.01%

TABLE 1: Data Volumes vs. Model Scope

Generic Reference Data typically has by far the highest percentage of objects in the data warehouse, though represents a relatively minor percentage of the volume of overall data in the system (0.01% in this case). By contrast, the Transaction Data may not represent the majority of the scope of the physical model, but it typically represents the highest percentage of total data volume in the system.

These ratios will vary significantly across implementations, especially the percentage of objects (or tables) in the warehouse and the ratio of Large Dimension to Fact Data. Due to the Large Dimension feature in the Magnitude DIW, there are few if any remaining implementations where Generic Reference Data volumes are significantly close to the Transaction and Large Dimension volumes.

For this reason, our benchmarks focus heavily on the performance characteristics of these loads in the system. Of note: both types of loaders leverage a full ELT approach, while the Generic Reference Data Loaders leverage a partial push-down ELT approach.

Load Benchmarks

During extended testing in Teradata's labs, the Magnitude transaction staging loader performed consistently on loads of up to 1 billion transactions per load. Load rates increased with total load volumes up to 120 million total records, then stabilized across the higher volumes.
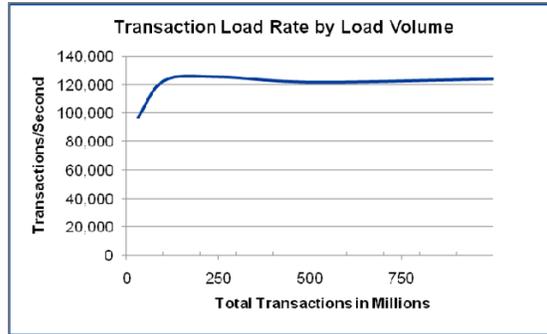

FIGURE 3: Load Rate remains constant as total transactions grow

The staging loader performed consistently at 124,000 records per second during loads up to and including 1 billion records – the top end of the test plan. That translates to **445 million transactions per hour**.

Plotting total run time against total volume, note in Figure 4 below that the Magnitude transaction staging loader performs consistently fast as total volume increases.
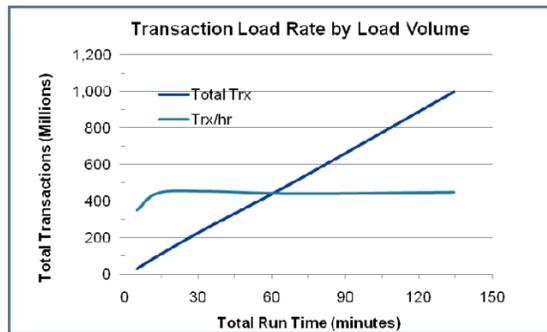

FIGURE 4: Consistently fast performance of
Magnitude Transaction Loader

Large Dimension Loader

The large dimension loader also performed consistently during tests of up to 100 million reference data or business entities (BEs). Load rates averaged 16,000 records per second for volumes from 10 million up to 100 million which was the top end of the test plan. **This translates to 57 million records per hour.**
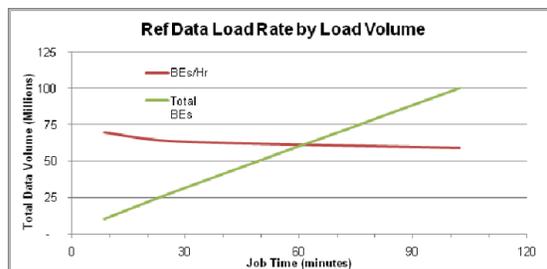

FIGURE 5: Reference data loads for large dimensions demonstrate
consistently fast performance

Query Performance

Because the Magnitude solution generates a traditional star schema or 3NF for BI access, query performance can easily be tuned and optimized on Teradata. In the graph below, Teradata tuning opportunities were identified and optimization techniques applied in subsequent iterations, including the creation of an aggregate join index. **Time to generate a result set was reduced from 22 minutes to 3 minutes.**
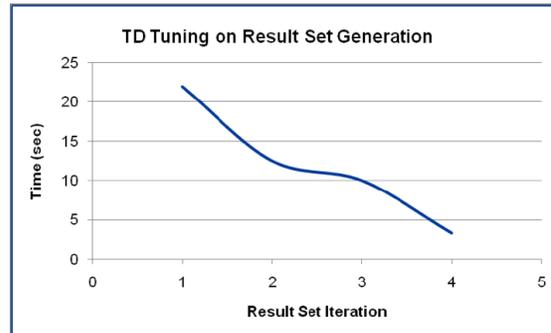

FIGURE 6: Query performance improvement as a result of tuning

Magnitude DIW: Tuned for Teradata

The Magnitude solution has been specifically tuned to ensure that the built-in functionality for data integration, schema management, load operations and data delivery and presentation support a high-performance, scalable warehouse solution. To this end Magnitude DIW has incorporated Teradata features such as TPT, automatic statistics gathering, integrated PI definition and an overall architecture that enables fine tuning of the physical structures to ensure optimized query performance.

In addition, Teradata experts can continue to tune the design of the physical Teradata schemas and take advantage of a host of performance optimization techniques, widely employed in custom built Teradata EDWs today.

Tuning Benchmarks

In Table 2 below, standard Teradata tuning techniques (statistics management and PI selection) were applied to the transaction load process. Teradata performance tools such as PMON were used to identify tuning opportunities outside the Magnitude solution. Customization within Magnitude DIW affords the solution architect the ability to configure Teradata features such as PI selection and Partitioning. As you can see from the table, the same load operation was repeated multiple times with different tuning techniques applied. Once tuned, the system continued to deliver consistently high load speeds as volume increased.

Total Volume	Tuning options	Tran/sec	Tran/hr
100M	No external tuning	94K	338M
100M	Statistics/ PI selection	123K	443M
250M	Statistics/ Column Sizing	126K	453M

TABLE 2: Transaction load performance after tuning

Leveraging Teradata Linear Scalability

The high-volume transaction data loader performance described in this paper was run on a two-node, 36 AMP Teradata machine. In a prior test, we loaded the same transaction data volumes for the same business information model through the Magnitude transaction loader on a 3 AMP machine. On the 3 AMP machine, the transaction data loader maxed out at approximately 10,000 records per second, while on the 36 AMP machine, the transaction data loader maxed out at over 120,000 records per second – over an order of magnitude faster. These initial results suggest that with Magnitude's ELT architecture on Teradata, Magnitude DIW's high-volume transaction data loading effectively leverages Teradata linear scalability, as it performed at a correspondingly higher rate on a more powerful Teradata system.

Industry Comparisons

PHARMACEUTICAL

In the medical field, there are approximately 2 million prescribers in the United States. A benchmark for one of the most familiar brand names in this space, a top-10 global pharmaceutical company, collects information on approximately 2 billion prescription sales records from the previous 24 months. A typical mid-size pharmaceutical company ($4 billion annual sales) may collect 120 million prescription activity records for the same time period from one million prescribers. Typically, pharmaceutical companies buy prescription sales data from third-party vendors. This data is often purchased at monthly or weekly cadences, and is used to support sales and marketing activities such as customer alignment, sales force goals, targeting and compensation

The following benchmarks are based on monthly data provision and an estimated 10% monthly change in prescriber information and conformance with the industry practice of restating at least the prior 6 months of transaction history during each monthly load.

Initial warehouse population (one time)
- Top-10 Pharmaceutical Company
  - 24 months of prescription activity: 4 hours 24 minutes
  - All US prescriber reference data: 2.5 minutes
- Mid-size Pharmaceutical Company
  - 24 months of prescription activity: 16 minutes
  - US Prescriber Reference Data: less than 2 minutes

Monthly warehouse refresh
- Top-10 Pharmaceutical Company
  - 6 Months restated prescription activity: 1 hour 6 minutes
  - 10% Change to US Prescriber Reference Data: less than 1 minute
- Mid-size Pharmaceutical Company
  - 6 Months restated prescription activity: 4 minutes
  - 10% Change to US Prescriber Reference Data: 30 seconds

By leveraging a fully push-down ELT process for the highest volume data in a data warehouse, Magnitude DIW on Teradata is able to deliver outstanding load performance, even on modest Teradata environments.

Summary

Magnitude DIW on Teradata delivers the key performance benefits of a Teradata data warehouse, combined with the speed of delivery and overall flexibility of the Magnitude solution. Most notably, it is still possible to achieve fast load performance with a combined solution of Magnitude DIW on Teradata. Tests suggest that Magnitude DIW leverages Teradata linear scalability as the size and power of the machine increases.

By leveraging a fully push-down ELT process for the highest volume data in a data warehouse, Magnitude DIW on Teradata is able to deliver outstanding load performance, even on modest Teradata environments.

MAGNITUDE™

For timely corporate performance insights based on your business needs, turn to Magnitude.

Dive deeper into this topic or contact us today at info@magnitude.com or **1.866.466.3849**. Visit our website: **www.magnitude.com**

About Magnitude

Magnitude Software leads the way in delivering corporate performance insights with solutions for corporate performance management (CPM), master data management (MDM) and data connectivity in heterogeneous environments. The Company's software and solutions drive real business results for hundreds of thousands of users around the globe. Magnitude's family of award-winning applications includes business intelligence, data warehousing, master data management, connectivity, analytics and reporting.