

Business Intelligence Data Warehousing

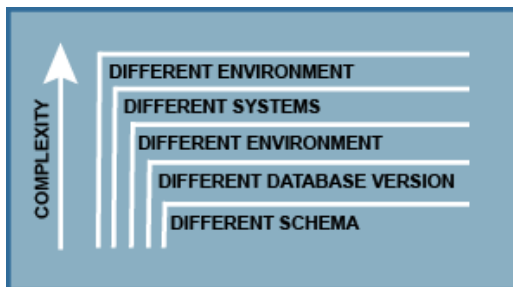
Business Intelligence solutions are always more successful (and in many cases **only** successful) when the underlying data is properly structured. Data warehousing can be a critical component of your project success, so TriCore Solutions encourages all customers to seriously consider the structure of their data prior to embarking upon a major BI deployment.

Data Source

One of the main challenges of a business intelligence system is working with data that originates from more than one source. Managing the extraction of this data and consolidating into a singular and predictable dataset requires deep business knowledge and data expertise.

TriCore's strategy is to engage early on to align with key business leaders and identify the full spectrum of data elements and their subsequent source systems. Working with business and technology stakeholders, TriCore will identify all operational data stores and design an ETL strategy for consuming this source data. This approach will guarantee that all data source complexities are properly considered, and that any risks in complexity are properly managed early on in the project.

This figure shows that as interactions between potential data elements in the data source increase, the complexity of the system also increases.



Some other considerations include:

- Different database systems
- Different schemas, database formats, and naming conventions
- Geographically separated source locations
- Source system ownership, volatility, and volume

TriCore Provides:

- Data Warehouse Assessments and Planning
- On-going warehouse support and maintenance services
- 24x7x365 Support
- Focus on building warehouses for BI projects
- Certified DW architects and ETL developers



141 Longwater Drive
Suite 100
Norwell, MA 02061
Phone: 617-774-5200

www.tricoresolutions.com



Never lose sight of this fact... Your data architecture choices directly affect the User Experience.

- Establish a strategy that fit your needs
- One Strategy does not fit all
- Stick to it!

Data Integration (ETL – Extract, Transform, Load)

Data integration is a critical component of BI because it is responsible for extracting, transforming, and loading the data into the data warehouse. This is a complex task that requires a disciplined approach. TriCore will use the following guidelines to drive the ETL development process.

Data Profiling

- Column Analysis – evaluates the distribution range, completeness, uniqueness, format, type, size, and frequency of the data
- Referential Analysis – evaluates relationships, integrity, and business rule dependencies
- Redundancy Analysis – technique for identifying duplicate and orphaned data

Data Extraction

- Source adapter connection and configuration
- Extraction Frequency
- Transactional record tracking, i.e. how to flag delta changes

Data Staging

- Store aggregates in staging tables for checks and balances

Data Transformation

- Source, split, and/or merge data from source systems.

Data Cleansing

- Properly handle inconsistent values such as absent values, empty strings, duplicate primary keys, etc.

Data Loading

- Load historical and transactional data into the data warehouse.

Data Storage

TriCore will work with key leaders to identify the logical objects for the data warehouse. The data stored in a warehouse is typically loaded by the ETL process. These requirements will then be translated and used to create schema tables and views within the Warehouse Database.

In addition, TriCore will also leverage the following where needed:

Dimensional modeling: Strategies for effectively organizing data are critical to implementing BI/DW systems. The technique used in modeling the logical data warehouse is commonly referred to as dimensional modeling. The guidelines for designing data warehouse solutions are different from transaction systems. As a designer you have to choose appropriate schema types, such as star or snowflake schema, design fact tables with measures relevant for the business and at the appropriate level of granularity, and address attributes that change over time by picking appropriate types of slow changing dimensions (SCD).

Partitions: Warehouse databases typically contain millions of rows in tables. Dividing large tables and their indexes into multiple segments and assigning them to file groups is called partitioning. As a designer you have to create a scalable partition that enables the best possible performance, you have to choose an appropriate partitioning strategy, choose optimal partitioning functions, appropriately place partitions in file groups, use index alignment, and plan for data management, such as moving new data inside a partition and removing aging data out of the partition.

Indexes: Designing an appropriate indexing strategy taking into account various factors such as usage patterns, schema design, column types, and storage needs are important for efficient operation.

Big Data Solutions: Data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. TriCore helps customer define and deploy Big Data solutions when the data Volume, Velocity, or Variety grow too large for standard modeling techniques.