A ResiliEnt Solution
For Data Warehousing

Jessica Bottone, Delivery Technical Specialist
John Gogel, Delivery Technical Specialist

ResiliEnt Business Solutions, LLC™

Introduction

Information is an important component in any organization today. Over the last decade, data warehousing has been incorporated into many businesses in response to the need for information.

Building a data warehouse affects the entire business. It can be an expensive, time-consuming, complex, and resource intensive undertaking that changes the way people work and think about data and managing the business. Many factors influence the ultimate success or failure of any data warehouse project.

ResiliEnt Business Solutions believes the key to a successful data warehouse project is not only in how the data warehouse is structured, but also in the planning and implementation details, several of which are unrelated to the technique or technology.

There are varying definitions for a data warehouse. To some, it's the data repository – the databases, tables, etc. To others, it's the entire business intelligence (BI) solution, of which the data repository is one part. This paper will address that difference but the primary focus will be the methodology on building the data repository. Building an effective BI solution will be addressed in a subsequent paper.

What is a Data Warehouse?

Perform an internet search on "What is a data warehouse?" and you'll get different answers. Even the two leading experts in data warehousing today, Bill Inmon and Ralph Kimball, have differing view points.

Inmon defines a data warehouse as: **"A subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process."** (Reference: *Building the Data Warehouse,* 2nd edition, by W.H. Inmon page 33)

Kimball defines a data warehouse as: **"A system that extracts, cleans, conforms, and delivers source data into a dimensional data store and then supports and implements querying and analysis for the purpose of decision making."** (Reference: *The Data Warehouse ETL Toolkit*, by Ralph Kimball and Joe Caserta, Page 23)

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

What is in common among the definitions is it's a collection of data used to make business decisions. The difference between Inmon and Kimball lies in whether it includes the tools that "extracts, cleans, conforms and delivers".

ResiliEnt Business Solutions agrees with Inmon that a data warehouse is simply a "collection of data". We believe Kimball's definition better describes a complete BI, or a data warehouse, solution. This white paper will describe the methodology for building the data warehouse and not the BI solution.

However, when considering whether or not to build a data warehouse, not taking these tools into account would be short sighted. Building the data warehouse is just one piece of an effective BI solution. Not taking all the components into account would be like building a bicycle without wheels: you won't get very far.

The general characteristics of a typical data warehouse can be further defined:

- Data is overwhelmingly sourced from transactional systems (yours or third party).
- Specifically structured for query, reporting and analysis.
- Creates a coherent picture of the business at a point in time.
- Usually based on relational database technology.
- Usually delivers information via web based technology.

Why build a Data Warehouse?

To that end, the following questions must be considered when a company commits to building a data warehouse:

- Why do some data warehouses fail?
- Why are some data warehouses, while being a technical success, viewed as a failure by key business clients?
- Why do some key business customers quickly accept some data warehouses while others are never truly accepted?
- Why do some data warehouses have a low cost of ownership while others are maintenance intensive?
- Why do some warehouses offer business value at many levels of the organization, while other data warehouses only offer value to a select few?
- Why do some data warehouses offer quick change management while the release cycle of others is very slow?

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

A ResiliEnt Solution for Data Warehousing

The Data Warehouse industry developed from the need for more and better information. Some of the obstacles that keep businesses from effective data reporting are:

- The IT systems needed to run businesses are so complex that getting and using the operational data takes a great deal of time and skill. Data in operational systems is typically stored in third normal form for optimal data storage and transaction processing. This works against effective reporting. Often, the complexity will lead to different people coming up with different answers to the same question.
- Adequately reporting the health of a business requires information stored in more than one IT system. This data is often on different platforms requiring different tools to access it.
- The IT systems needed to run businesses are designed and tuned for that purpose and not to offer a strong reporting suite.

Companies and departments invest in data warehouses because of these inherent limitations in operational systems. Data warehouses offer numerous benefits that more than pay for the cost of building and maintaining them. We agree with the benefits outlined by Alan Perkins in his white paper on data warehousing. (Reference: http://www.ies.aust.com/~visible/papers/dw.htm, *Developing a Data Warehouse, The Enterprise Engineering Approach*, By Alan Perkins, Managing Principal, copyright 1995-1996, Visible Systems Corporation)

Better Enterprise Intelligence

A good data warehouse is structured for reporting so the work to retrieve complex data is eliminated. Business users are free to concentrate on analyzing the information retrieved, possibly leading to insights and re-engineering of business processes.

More Cost Effective Decision Making

A data warehouse can reduce the staff needed to run reports, resulting in significant savings in time and resources. It also eliminates the resource drain against production systems that result from long, complex queries and reports.

Enhance Customer Service

An enterprise can maintain better customer relationships by correlating all customer data via a single data warehouse.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Business Process Re-Engineering

Unlimited analysis of enterprise information can provide insights into processes that may yield breakthrough ideas for re-engineering. Just defining the requirements for a data warehouse results in better enterprise goals and measures.

Information System Re-Engineering

A data warehouse based upon enterprise-wide data requirements provides a cost-effective means of establishing both data standardization and operational system interoperability.

Purpose of a Data Warehouse

The purpose of a data warehouse is to: **"Publish the organizations data assets to most effectively support business decision making**." (Reference: *The Data Warehouse ETL Toolkit, By Ralph Kimball and Joe Caserta, page 22)*

Every business and department has goals and key business indicators (KBIs) that are used to meet these goals. These KBIs can range from customer demographics to measuring turns on inventory. Most of these KBIs are quantitative measurements stored in operational systems. Collecting and publishing these KBIs offers insight into the business which facilitates good decision making and improves the health of the business.

To fulfill this purpose, a data warehouse must have the following functionality (Reference: *The Data Warehouse Toolkit*, By Ralph Kimball):

- Provide access to corporate and/or organizational data.
  - This access must be immediate and high performance.
  - Access to the data warehouse must be available from every applicable desktop at every conceivable location including remotely.
- Have consistent data. Every person asking the same question gets the same answer.
- Structure the data in such a way that it can be sliced and diced by every important business requirement.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Data Warehouse Critical Success Factors

Since data warehouses involve data and databases, many turn to tools and technologies as the key to a successful project. While tools and technologies are an important aspect, there are other factors that impact the success of your project. At ResiliEnt Business Solutions, we believe successful data warehousing projects have these common factors:

➢ Executive Sponsorship

ResiliEnt Business Solutions believes executive sponsorship is the most important factor to ensure a successful data warehouse project. Only an executive sponsor has the position to ensure the project team has the organizational cooperation necessary to achieve the project's objectives.

The responsibilities of the executive sponsor include promotion, budgeting, ensuring organizational cooperation, and resolving conflicts.

To effectively manage these responsibilities, the executive sponsor should form, and lead, a cross-functional BI Governance Committee. This committee will oversee the BI solution for the entire organization.

This committee should be made up of members at an equal or similar level as the executive sponsor; however, a delegate who has been given authority to act on behalf of a particular area could suffice.

As the lead of the BI Governance Committee, he:

- Guides the committee in establishing BI solution mission statements and project charters for each project.
- Establishes the priorities and scope of each project.
- Resolves issues that occur between the business organizations.
- Establishes the success criteria by which each project and the BI solution will be judged.

(Reference: www.dmreview.com/article_sub.cfm?articleId=1051114, *Intelligent Solutions: Business Intelligence Executive Involvement*, by Jonathan G. Geiger, DM Review Magazine, April 2006)

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

➢ Good Governance

In addition to having a committed executive sponsor, every member of the Governance Committee must be equally committed. The Governance Committee must be actively involved in the BI intuitive and proactive in all aspects of the BI Solution to ensure that it stays in line with the goals of the organization.

➢ Project Management

Good project management is critical to a successful data warehouse project. In fact, good project management is critical to the success of any complex project. But employing good project management practices isn't always enough. Too many organizations overlook the 'soft' skills a good project manager should possess.

Good Project Management Practices

From a purely logistical standpoint, why is project management important?

Three reasons:

(1) If you don't have a comprehensive list of everything to be accomplished, how do you know when you're done? As funny as that may sound, it is very true. Simply having a list of all the tasks to be accomplished keeps the project team in line with the business requirements.

(2) Having a list ensures all business requirements are satisfied and makes it very clear what is and is not inside the scope of the project.

(3) By associating a timeline to the list of tasks to be accomplished, you now have a project plan. A project plan makes it easy to track the progress towards the goal and allows for proactive responses to issues that could derail the project.

Good Project Manager Skills

Employing good project management practices can be for naught if the project manager does not possess good communication and leadership skills. Many believe the project manager should possess the technical and business knowledge particular to the project. While that would be the optimum situation, good communication and leadership skills are far more important.

A project manager with good leadership skills will recognize that they lack particular technical and business knowledge. They will make sure their project

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

team has staff with these skills and will rely on them to contribute to that part of the project. Their good communication and leadership skills will help them build a cohesive team that will work together to keep the project on track.

Good communication and leadership skills will also allow the project manager to effectively deal with management and the business community in matters of project scope, deadlines, budget and staffing.

➢ Iterative Incremental Development Process

ResiliEnt Business Solutions usually advocates utilizing an Iterative Incremental Development Process (IIDP) over the more traditional Waterfall approach. Whether you chose the Iterative Incremental Development Process (IIDP) or the Waterfall development process, involving the user community early and often is critical and business clients see value sooner if the project can be broken down into small phases. Then with each phase, you can apply whichever methodology you prefer.

The best definitions of the IIDP and Waterfall approaches that we have found are in the editor's notes at the bottom of an article on www.computerworld.com by Bill Walton (Reference: *Iterative vs. waterfall software development: Why don't companies get it?,* February 20, 2004):

(1) The iterative and incremental approaches involve a number of short cycles in which steps such as requirements gathering, coding, testing and deployment, are conducted to produce small parts of the final project. The software system grows incrementally, and user feedback can be used throughout the process.

(2) The waterfall philosophy is a strictly sequential approach in which a project is completed in a series of steps, such as analysis, design, coding, testing and deployment. Each step, such as requirements gathering, is undertaken only once and must be completed and verified before the next phase.

Whichever methodology you chose, showing progress quickly provides political benefits and shows justification for the data warehouse project expenditures. Additionally, it is common for business users, once they see the results and begin to consider the possibilities, to change the requirements, resulting in re-work. Given this, it makes more sense to provide smaller 'bursts' of information so the amount of re-work is smaller.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

➢ Dependable Infrastructure Architecture

The BI infrastructure includes (but is not limited to):

- o computing hardware
- o operating system
- o database(s)
- o ETL tool(s)
- o access methodologies
- o data access tool(s)
- o data access server

While this may seem like a lot of pieces, creating a dependable architecture is much easier than it used to be. The hardware and software available today is more reliable, robust and affordable. There are many acceptable solutions for building a successful BI infrastructure.

Important criteria for choosing the infrastructure components include:

- o internal expertise
- o IT infrastructure strategy
- o vendor viability
- o cost
- o performance
- o scalability of the solution to the business environment

➢ Creating a Business and IT Partnership

The benefits of creating a partnership between the business and IT should be self-evident and yet many organizations don't put a lot of effort into this. The business needs the data warehouse to make business decisions. IT is tasked with designing a data warehouse solution to support the business. The relationship between the business and IT must be seen as symbiotic – each needs the other to ensure a successful data project.

➢ Well Defined Business Problems

Each data warehouse project should include a project charter that clearly states the business problem to be solved, systems involved, timeframe, major deliverables, people involved, risks, assumptions and responsibilities. This document should be altered and reissued as the scope of the project changes.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

➢ Communicating and Managing Client Expectations

The importance of proactive communication by the data warehouse project team to the user community cannot be over emphasized. Every member of your organization will develop an opinion on the value the data warehouse will have for them. The minimum amount of information to be delivered is the "who, what, when, where, and why" of the project. Additional communication includes describing the ease of use, value of the data and importance of the project. The data warehouse project needs to be successfully marketed to internal customers the same way that new products and services are marketed to existing and target customers.

➢ Training

Every successful data warehouse project must include training. Skill assessments may be needed to determine the level of computer training required. Besides traditional classroom setting training, computer based training, training CDs and creating business user 'trainers' are other options which should be considered.

Data Warehousing Methodology

While the data warehousing industry has matured significantly in the last decade, data warehousing methodologies are still evolving. Debates continue as to whose methodology is 'widely recognized', that of Bill Inmon or Ralph Kimball.

However, these two leading experts agree on several important data warehousing construction issues that we also agree with:

- Data warehouses contain atomic level data based on the facts stored in the operational systems.

  Every data warehouse must contain the most detailed information available (atomic) from the source system(s) regardless of the defined reporting requirements. This offers the most flexible solution because the data can be sliced and diced in every way supported by the grain of the data and can be aggregated as requirements change.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

- Breaking projects into subject area sub-projects.

  Designing data warehouses to be subject based means they can be constructed and managed largely independent of the other data warehouses in your BI portfolio and can be enhanced with incremental releases independently.

- Use an iterative incremental develop process.

  This approach lowers project risk and provides value faster when larger deliverables are broken into smaller phases than the traditional waterfall approach.

In addition to this, we also believe every data warehouse should:

- Be constructed utilizing a dimensional modeling approach with conformed dimensions, standardized facts, star schema designs, bus architecture (as advocated by Kimball).

  This type of approach is more easily understood by the user community, more easily extendable and the star schema is more efficient in processing changing user requests.

- Include and present all the available source data for a subject area (as advocated by Inmon).

  Many DW practitioners follow Kimball's approach by only storing the information requested and aggregating that data to the level specified. ResiliEnt Business Solutions believes this practice to be devastatingly shortsighted because these data warehouses are inflexible, resulting in a high cost of ownership.

We advocate a hybrid approach. The **ResiliEnt Hybrid Method** is a combination of the strengths of each approach. We believe this hybrid approach offers a more flexible product with a lower total cost of ownership than either approach.

To illustrate the benefits of the **ResiliEnt Hybrid Method**, we will introduce the "traditionalist" and "ResiliEnt" data warehouse project teams and use them in a real world example of a data warehouse project.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Traditionalist Data Warehouse Project

The traditionalist data warehouse project team is commissioned to build an orders and shipments warehouse. The needs assessment phase of the project indicates the business wants to track all customer shipments and orders from every location, by SKU, on a monthly basis.

The traditionalist team builds the orders and shipments data warehouse according to the specifications, presenting aggregated orders and shipments data to the customer by SKU, from every location on a monthly basis.

One month later, the business customer says, “This is great, but we also want to track orders and shipments between our own locations as well as inter-company order and shipments.”

The traditionalist team tells the business client that the requested information was not part of the original specification and therefore, not included in the data warehouse. The traditionalists must go back to the source systems, re-analyze the data, change the ETL jobs, alter the data warehouse staging areas, change the data warehouse data model and reload the entire history of orders and shipments to accommodate this need.

After the data warehouse is updated with the inter-company orders and shipments information, the business customer says, “This is great, but we also need to track order and shipment lead times.” The traditionalist team again tells the business client that the data warehouse only contains the information that was specifically requested, and order and shipment lead time data was not requested. The traditionalist team must go through an entire development cycle again to meet this additional need.

As the traditionalist team is incorporating the changes for orders and shipments lead times, the business client makes another request. They say, “We also need to track late shipments and late deliveries, and since you are doing work getting the shipment lead time information, why not just get this information at the same time?” Since the traditionalist team was focused on the stated problem (track orders and shipment lead times), they view this change as scope creep and want more budget dollars and time to address this additional requirement.

While the business client is waiting for this list of enhancements to be completed, they call the data warehouse project team asking how they can see the cancelled

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

orders. The traditionalist team tells the business client that cancelled orders was not requested and not included in the data warehouse.

Another business client learns of the orders and shipments data warehouse, thinks it will be a great source for them but needs the information by each order item and each shipment item on a daily basis. Once again, the traditionalist team must go back to the drawing board to solve this additional need.

The business clients are very upset. They feel like IT doesn't service their needs, is slow to react as well as being reactionary and that the data warehouse was a giant waste of money.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

ResiliEnt Data Warehouse Project

The ResiliEnt data warehouse project team is commissioned to build an orders and shipments warehouse. The needs assessment phase of the project indicate the business wants to track all customer shipments and orders from every location, by SKU, on a monthly basis.

The ResiliEnt team agrees to build the orders and shipments warehouse but builds the data warehouse using the "ResiliEnt" methodology: building the data warehouse at the atomic level, incorporating all orders and shipments information available in the data sources, including cancelled orders, inter-company orders and shipments, all order dates, late shipment and delivery information amongst others.

The ResiliEnt team receives the same change requests listed above. Their data warehouse can already service these additional business needs without going through an additional build cycle because they captured all the data elements available in the source system for this subject area, and stored them in the data warehouse at the atomic level.

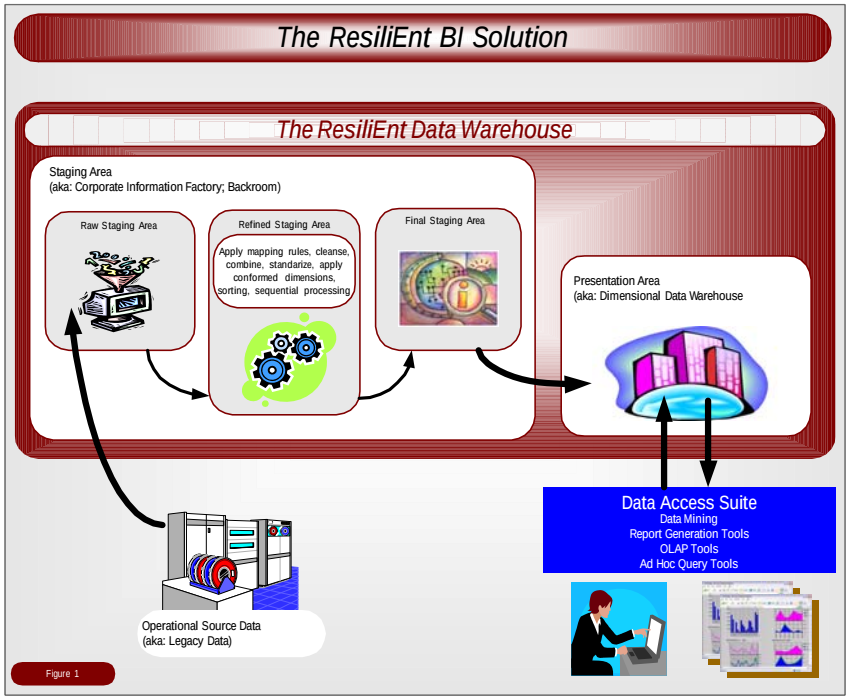
The traditionalist approach, or the "build to spec" method, has inherent flaws made obvious by the examples above. In the "real world" of data warehousing, is this example extreme? No, this example comes out of the direct data warehousing experiences of the authors. The traditionalist approach often fails in spite of a good governance process, strong project management and a great functional specification. This is because business needs change over time and the business clients often don't realize what they really need until they start analyzing their data.

The **ResiliEnt Hybrid Method** understands these business ambiguities. In a ResiliEnt data warehouse, incorporating changes beyond the original project scope is often a short project of a couple of days to a couple of weeks. This is because the data warehouse contains all the source data. The traditionalist approach often takes months to make modifications to the data warehouse.

The result of the **ResiliEnt Hybrid Method** is a more usable, lower cost of ownership data warehouse that provides better service to your business customers.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

The Data Warehouse Architecture



We like to look at a data warehouse as analogous to a restaurant (substitute the text in quotes with the text in parenthesis):

- The Operational Source Data is the pantry or refrigerator, where all the 'raw food' (raw data) is originally stored.
- The Staging Area is the kitchen. As with a restaurant, 'customers' (end-users) are not allowed in the kitchen.
- The Raw Staging Area is the counter top where you place the 'raw food' (raw data) for preparation.
- The Refined Staging Area is where the 'preparation' takes place. The 'meal' (final cleansed data) to be prepared dictates the 'amount of work' (number of refined staging tables) needed. To create your 'meal' (final cleansed data), you may simply need a bowl and a spoon; or you may need bowls, pots, pans, and an hours worth of cooking in the oven.
- The Final Staging Area is the plate where you place the 'meal' (final cleansed data).
- The Data Mart (aka: Presentation Area) is the dining room. The 'meal' (final cleansed data) is carried from the 'kitchen' (Staging Area) to the dining room where your 'customer can enjoy the meal you have prepared' (end-users can query the final cleansed data).

This analogy is similar to one you will find in *The Data Warehouse ETL Toolkit,* by Ralph Kimball and Joe Caserta, page 17.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Characteristics and Benefits

Important points about the ResiliEnt Business Solutions Data Warehouse:

- The Operational Source Data can be any source data, from operational systems to an Excel spreadsheet. Each project in the data warehouse can have many different sources, but is limited to one subject area at a time.
- The Staging Area is where all the data is brought together, cleansed, transformed, and readied for the Presentation Area. This area is off-limits to end users.
- Both the Staging Area and Presentation Area contain complete histories of the data.
  - The Staging Area contains the original source information and all transformed information. If a data requirement changes or an error in processing is identified, the data warehouse can be reprocessed completely from the staging area. Also, source systems often have a shorter archiving requirement than data warehouses, making it impossible to reprocess the entire data warehouse. Storing each source system extract in the staging area archive mitigates this risk.
  - All data in the Presentation Area is in a dimensional data warehouse structure. This data has been transformed to be in line with the conformed dimensions.
- The Presentation Area (dimensional data warehouse) is:
  - Where the organization is allowed access for their data reporting.
  - The final product of the extraction, cleansing, conforming and transformations that occurred during the development and loading steps.
  - Contains all data elements at the atomic level, eliminating the need for customers to go into the staging area.
- The ResiliEnt Business Solutions Data Warehouse uses the Kimball Dimensional data structure. The goal of the data warehouse is to publish an organization's data to facilitate decision making. To achieve this goal, the data warehouse should be constructed in the simplest structure possible. This lowers the IT knowledge required to retrieve information. Business customers will understand the data better because it's structured in the way they think about their data rather than the way IT thinks about their data. This will increase user acceptance and lower training costs.

(Reference: *The Data Warehouse ETL Toolkit*, by Ralph Kimball and Joe Caserta Part I.)

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

An Example

ABC Company decides to build an orders and shipments data warehouse.

Their business requirements are:

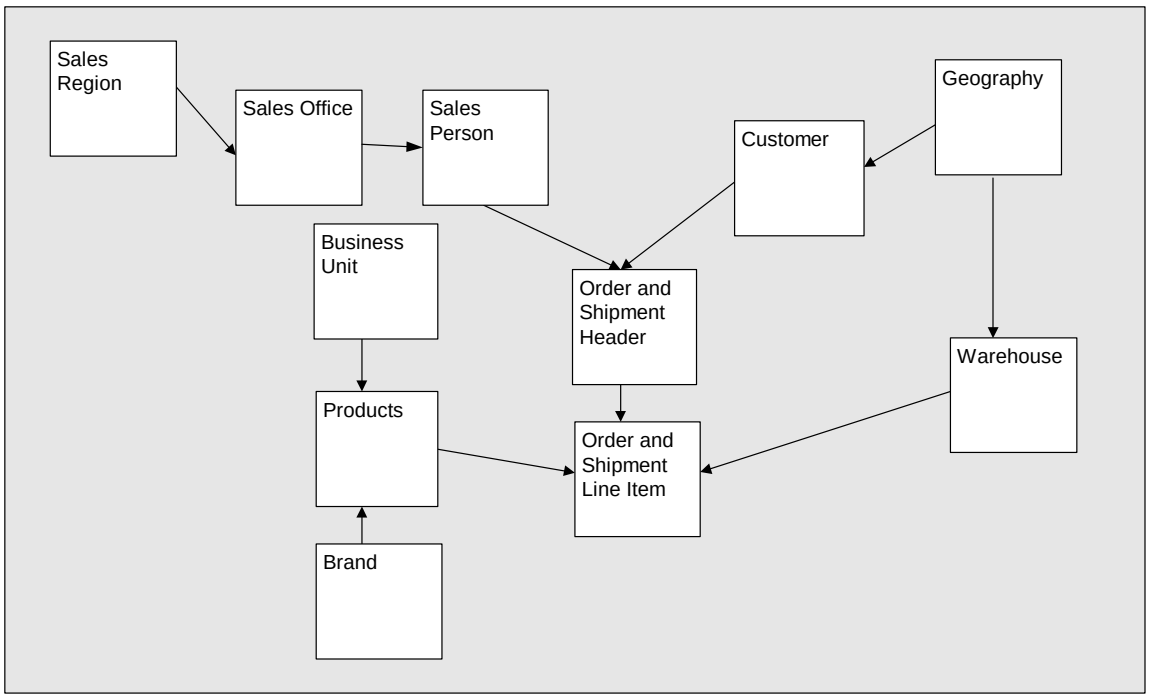
Track all customer orders and shipments by one or all of the following:

- Product
- Brand
- Business Unit
- Customer
- Order item
- Location
- Sales Representative through sales region
- Volume and net sales dollars

Also available in the data warehouse but not specifically requested or identified during the needs analysis stage or in the functional specifications:

- Shipping costs
- Gross Sales
- Rebates
- Discounts
- Cancelled Orders
- Order Lead Times
- Late Shipment Information
- Late Delivery Information

An Inmon solution might look something like the following diagram (simplified for clarity).

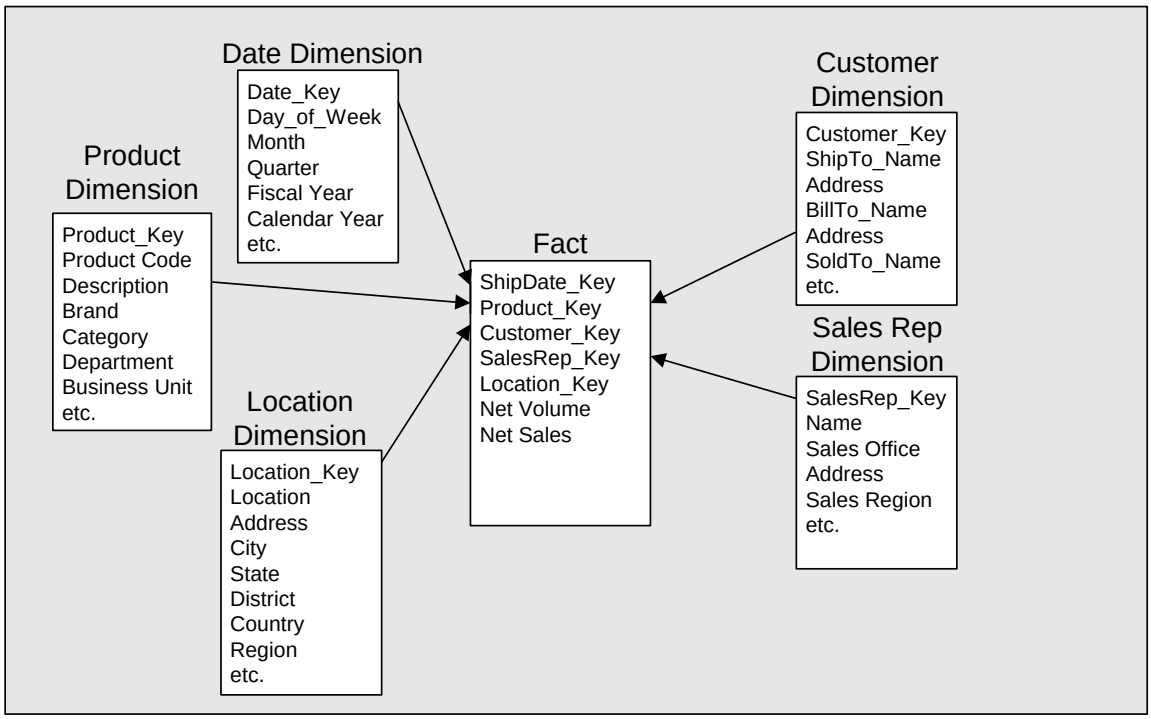


(Reference: *Building the Data Warehouse*, 2nd edition by W.H. Inmon)

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Operational Data is often stored in third-normal form. A data warehouse built following Inmon's methodology might look very similar to that kind of structure. All data would be stored, including information like address, even though the business requirements made no mention of addresses. The drawback of this approach is that any required joins would have to be done via the actual data fields, which could be many, leading to inefficient join processing and slow response time.

A Kimball build to specification traditional solution might look like this:



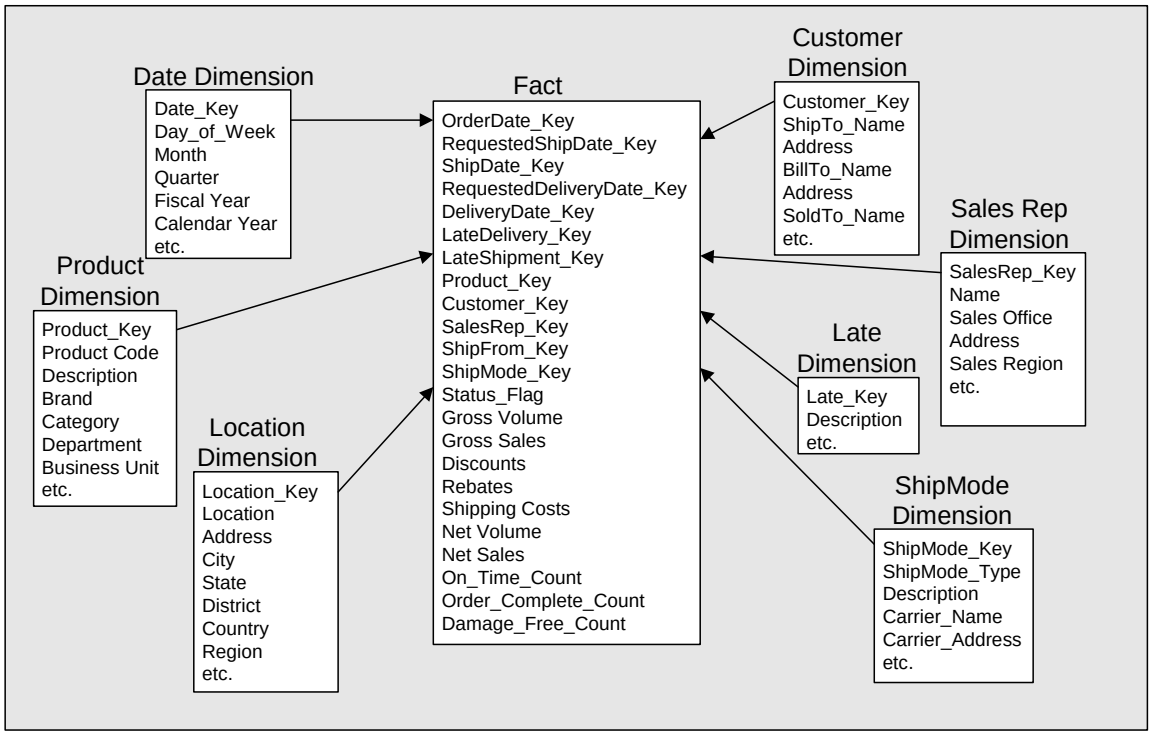
This is a star schema atomic solution built according to the Kimball methodology, but only includes the information specified in the business requirements. All other fields are 'left behind'.

While addressing the issues of being 'easy to use', it does not include all the data. This may be fine for the needs of the business today. But if future information needs require additional data that is only in the operational system, your choices are either:

(1) Do without.
(2) Read directly from the operational system again, which didn't work well the first time.
(3) Add the additional fields to your data warehouse and rebuild your history, which could prove time consuming.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

The ResiliEnt Business Solutions Data Warehouse Solution



The ResiliEnt Business Solutions Data Warehouse, employing **The ResiliEnt Hybrid Method**, looks much more like the Kimball solution, but with all data fields carried forward from the operational system.

While the business needs of today may not dictate that all data fields are incorporated into the atomic and summarized tables, the business needs of the future are more easily addressed by carrying all data fields in the data warehouse. By only having to address changes to the presentation area, and not the data warehouse as a whole, total cost of ownership is lower and changes occur quicker.

Another important note: In this example, the raw staging area of the data warehouse could have been completed while the specifications for the final staging area tables, summarization tables and subsequent reporting needs were still being compiled. Since all data fields are pulled from the operational system, once the operational data is identified and the IT staff instructed on how to extract the data, that portion of the project could begin. Working in parallel means shorter project timeline.

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Conclusion

Data warehouses are an expensive, time-consuming, complex and resource intensive undertaking that affects the entire organization. Not being able to easily get the information you need to run your business can be more expensive, time-consuming, complex and resource intensive.

While the definition of a data warehouse may be different depending on who you ask, one thing is always the same: they are an integral part of the business landscape today.

A successful data warehouse must:

- Consider the information needs of the business today as well as the future.
  - Bringing all the data fields forward provides for the future by making changes simpler and making cost of ownership low.

- Make the information easily accessible.
  - Building a dimensional data warehouse makes data retrieval simpler, providing ease of use for your business users.

- Present it consistently.
  - Using conformed dimensions provides for a common report look and feel as well as the same answer to the same question.

By examining the reasons why many data warehousing project fail, along with their years of experience, ResiliEnt Business Solutions developed a methodology to create data warehouses that are:

- viable in the shortest amount of time possible
- are ‘resilient’ as your business needs change
- have a low total cost of ownership

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

Glossary

Bus Architecture

The practice of attaching several independent data warehouses together via conformed dimensions.

Business Intelligence (BI) Solution

The set of tools used to collect and publish data designed to support business decision making.

Conformed Dimensions

Conformed dimensions are dimensions that have common fields, and the respective domains of the values in these fields are the same. That guarantees that you can perform separate queries on remote fact tables connected to these dimensions and you will be able to merge the columns into the final result. (Reference Ralph Kimball, www.kimballgroup.com, Design Tips 2002, number 34, February 28, 2002)

Data Mart

Synonymous with data warehouse.

Data Warehouse

A subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process (Reference: *Building the Data Warehouse,* 2nd edition, by W.H. Inmon page 33)

Data Warehouse Solution

Synonymous with BI solution.

Dimensional Modeling

Dimensional Modeling is a logical design technique that seeks to present the data in a standard, intuitive framework that allows for high-performance access. It is inherently dimensional, and it adheres to a discipline that uses the relational model with some important restrictions. Every dimensional model is composed of one table with a multipart key, called the fact table, and a set of smaller tables called dimension tables. Each dimension table has a single-part primary key that corresponds exactly to one of the components of the multipart key in the fact table. (Reference: Ralph Kimball, *A Dimensional Modeling Manifesto*).

Iterative Incremental Development Process (IIDP)

A development process that involves a number of short cycles in which steps such as requirements gathering, coding, testing and deployment, are conducted to produce small parts of the final project. The software system grows incrementally, and user feedback can be used throughout the process. (Reference: editor's notes at the bottom of an article on

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com

www.computerworld.com by Bill Walton. *Iterative vs. waterfall software development: Why don't companies get it?,* February 20, 2004)

Standardized Facts

A fact represents a business measure. A row in a fact table corresponds to a measurement. A measurement is a row in a fact table (Reference: Ralph Kimball, *The Data Warehouse Toolkit*). A standardized fact is one in which a common derivation is used. For example, cost of goods sold is computed in the same manner whether the data came from the east coast office or the west coast office.

Star Schema

A standard technique for designing a data warehouse. "Fact" tables join to a larger number of independent "dimension" tables. The term 'star' comes from the look the final result achieves: the fact table sits in the middle of the dimension tables, with each dimension table circling the fact like points of the star.

Waterfall Development Process

A strictly sequential approach in which a project is completed in a series of steps, such as analysis, design, coding, testing and deployment. Each step, such as requirements gathering, is undertaken only once and must be completed and verified before the next phase. (Reference: editor's notes at the bottom of an article on www.computerworld.com by Bill Walton. *Iterative vs. waterfall software development: Why don't companies get it?,* February 20, 2004)

11175 Cicero Drive, Suite 100
Alpharetta, GA 30022
(678) 242-5242
www.resilientbiz.com