

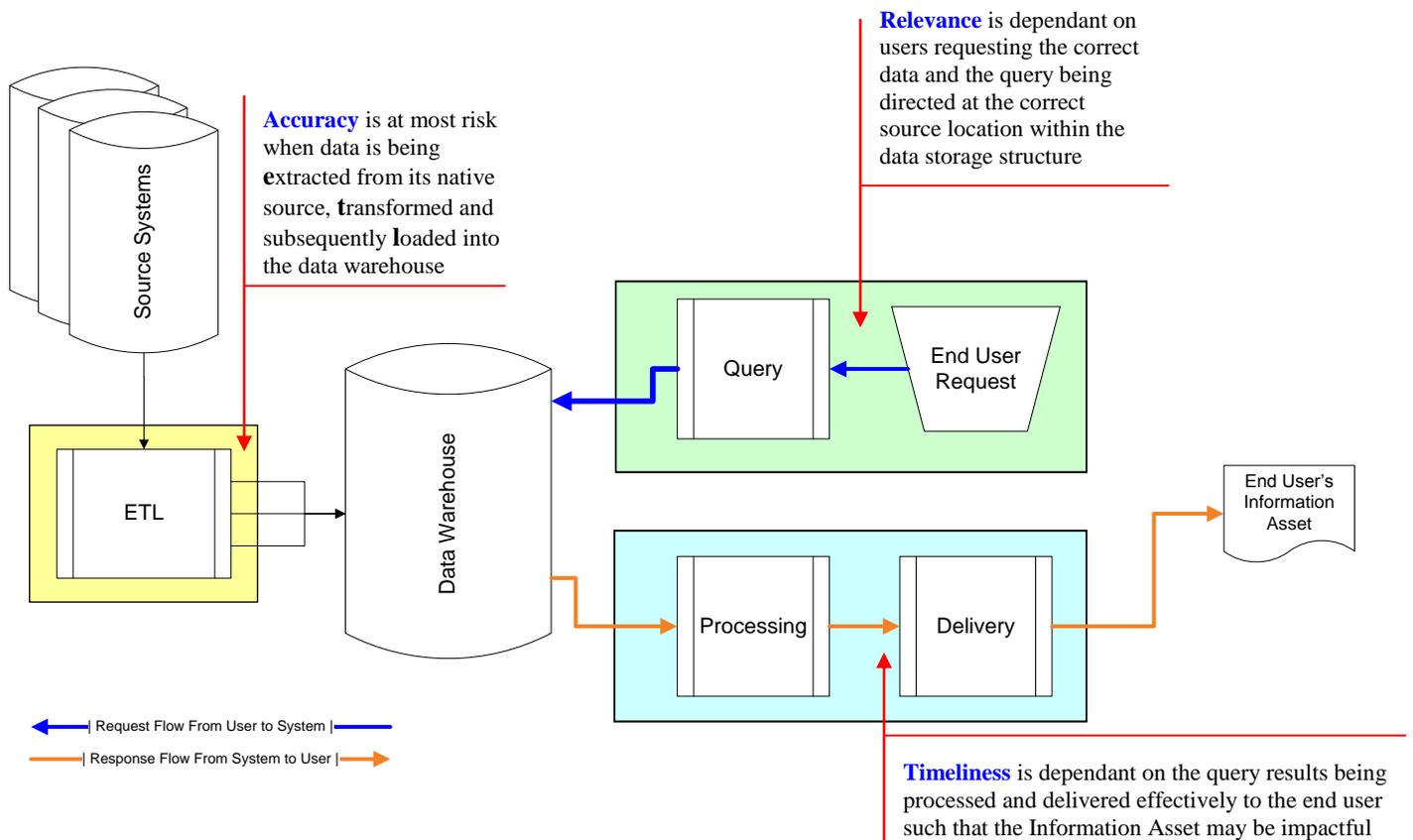
Data Quality

Data Quality is a subjective measure of the value of information based on the degree to which it possesses the following three characteristics:

- **Accuracy** – Data is accurate if it is free from error or defect. Accurate data is a precise and exact representation of events
- **Relevance** – Data is relevant if it has bearing upon or is related to the question, issue, concept, operation or strategy at hand. Relevant data is pertinent and has a material relationship to a topic of discussion, decision to be made or strategy to be defined and executed
- **Timeliness** – Data is timely if it is delivered at such point that it can be taken into consideration during the course of discussion, evaluation or execution. Timely data is delivered to the appropriate personnel such that it can be used to effect change and/or impact decision making

For data to be judged “High Quality” it must possess a high degree of all three of the aforementioned characteristics. Data that is completely lacking in any one of these qualities is considered to be of low quality. Organizations must define acceptable levels of data quality and should have fail-over procedures in place in case requisite data does not meet minimum levels of acceptability.

Given that Data is the basis and building block of all Information Assets within the Information Delivery and Business Intelligence system, Data Quality is critically important and demanding of routine evaluation and management.



Data Storage Structure & Strategy

Data Warehouses, Stores and Marts

Security, system processing speed and ease of use are all significant drivers in the configuration of any data storage design strategy.

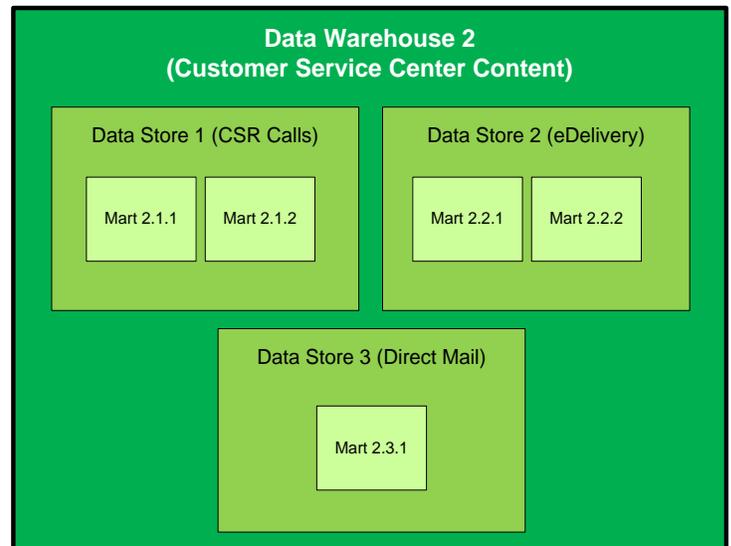
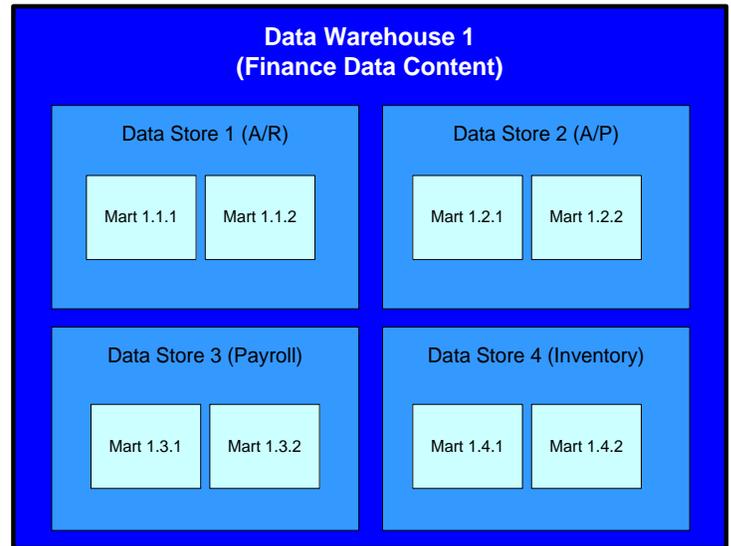
The most common approach to defining the data storage structure is to create:

- *Warehouses* that correspond to the different functions of the Enterprise
- *Stores* that correspond to the different roles within each function
- *Marts* that support the most complex and technologically demanding information assets required within each role

As the Enterprise considers the data storage structure it will use to support the Information Delivery and Business Intelligence system, it is important to keep the following in mind:

- Multiple warehouses do not necessarily mean multiple servers. It is common for entities to store multiple warehouses on a single server and use the security / user profile configuration options within the server appliance itself to keep users out of warehouses where they don't belong
- Not all delivery assets require their own Mart. For example, a simple operations report with simple calculations can be quickly pulled from the physical databases within the Warehouse. Where multiple queries, joins and algorithms are required to produce an information asset however, it is often advised to cache a mart of data for that particular deliverable.
- While warehouses are always physical technological assets, Stores and Marts may be temporary assets consisting of sub-sets of data extracted from the warehouse and cached in memory only long enough to support the user and/or deliverable for which they were intended.

The most important issues to keep in mind when defining the data storage structure and strategy is how and how often data elements will be utilized, which data elements are going to be used routinely by power users, which deliverables will be the most processor-intensive, the security requirements of the Enterprise and, ultimately, the capabilities of the hardware appliance that is to be used to host the warehouse(s).



Data Updates & Freshness

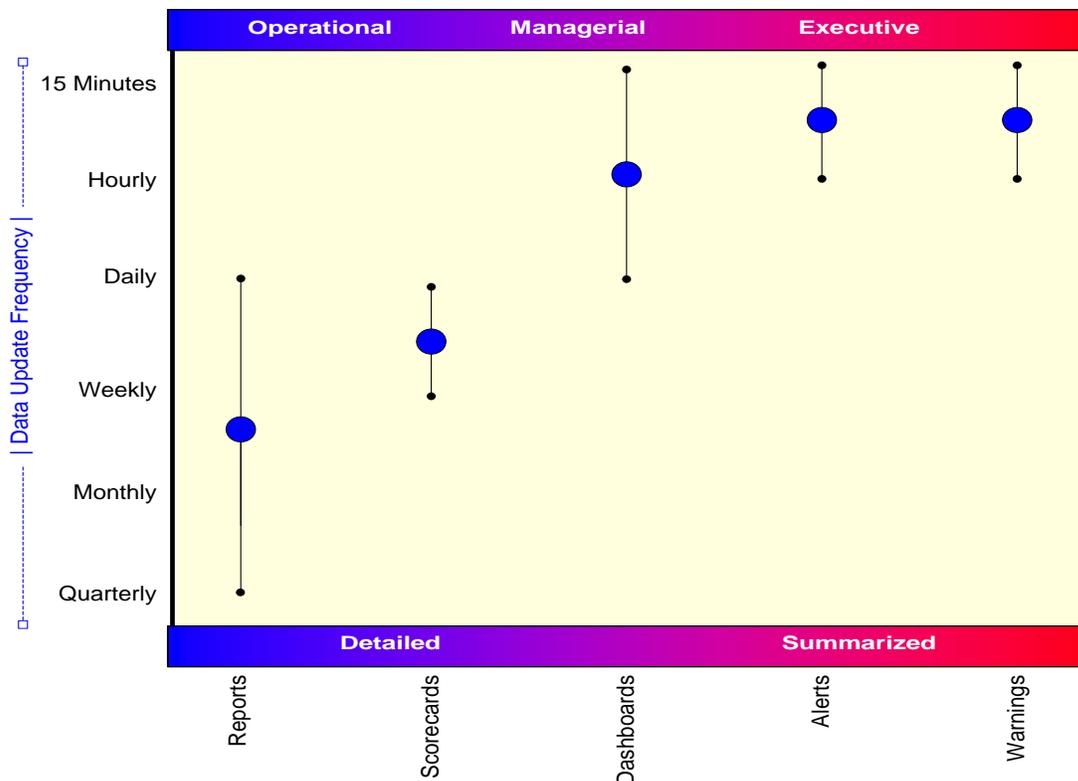
As Accuracy is a measure of data quality, the frequency with which the source of all data to the Information Delivery and Business Intelligence system is updated is a critical issue to be addressed during the design of the system.

As noted previously herein, accurate data is a precise and exact representation of events. It is the precision component that gives an Enterprise the latitude to define “acceptable” levels of accuracy. It is also important to note that not all data is held to the same standard. For example, Balance Sheet data such as the depreciated value of capital assets may be considered accurate if it has been updated within the last 90 days while Income Statement data such as gross sales to date may be considered accurate only if it has been updated within the last 24 hours. Similarly, a stock trader may require data that is updated every 3 seconds while a bank officer may not need updates any more often than once every 30 days.

Ultimately individual power users will often dictate the level of “freshness” required in order for data to be accurate enough to be judged high quality.

Data Freshness

There is an inverse relationship between the level of detail in an Information Asset and the frequency at which the underlying data supporting that content must be refreshed / updated. The table below demonstrates this relationship.



Appendix A – Glossary

Term	Definition
- A -	
Accuracy	A measure of Data Quality. An objective evaluation of the level of defect or degree of error. Precision and Accuracy are synonyms when discussing Data Quality
Agile Approach	A project management methodology based on constant interaction between the project team and user community where progress is reviewed and priorities updated constantly based on changing project forces and user demand
Alert	High-level message from Information Delivery and Business Intelligence system that predicted future state, based on current operations, may not be within acceptable thresholds
Algorithms	Collection of specifically defined rules, steps and processes, to be executed in a specified cadence in order to solve a problem
Asset	Any end-user facing output of an Information Delivery and Business Intelligence system. Assets consist of Reports, Scorecards, Dashboards, Alerts and Warnings
- B -	
Business Intelligence	“BI”. The third level of maturity in the Data Maturity Model. Business Intelligence is the comparative result of information to internal and external metrics
- C -	
Calculations	Arithmetic operations applied to numeric Data in order to convert it into Information
Concatenation	Method of textual data manipulation where multiple strings and/or sub-strings from multiple alpha-numeric data fields are combined to create a new text string. Concatenation is the opposite of Parsing.
Current State	Systems, operations, processes and work flows as defined and in force at the current time
- D -	
Dashboard	Graphic, dynamic representation of current operations updated continuously
Data	A specific characteristic and/or result of an action, work flow, process or operation
Data Architecture	The underlying structural composition of a data storage unit. Data Architecture documentation represents the “blue print” used to build and support data structures including databases, data warehouses, data stores and data marts. A Data Architecture defines the characteristics and

Term	Definition
	requirements of all data elements within a data storage unit.
Data Freshness	A measure of how recently data has been updated with regards to the most recent operations cycle. The more time that elapses between updates the less Fresh data is considered and the lower the overall data quality measure as less Fresh data is considered less accurate
Data Mart	A collection of data elements grouped according to a common job function
Data Profile	Specific details regarding the characteristics (Type, Format, Picture, Default Value) of each field within a database Architecture
Data Steward	Individual empowered to enforce and oversee the administration of a Governance Model
Data Store	A collection of data elements grouped according to a common reporting asset
Data Structure	The definition and configuration of and relationship between Data Warehouses, Data Stores and Data Marts
Data Update	A revision of existing data and addition of new data to a data storage unit based on all operations and work flows that have been executed since the last Data Update
Data Warehouse	A collection of data elements grouped according to function
Deliverable	See "Asset"
Delivery	Movement of an Asset from the source system to the end user
Downstream Flow	Flow of data from its native source system to a central storage facility (often a Data Warehouse) and the flow of assets from the central storage facility to end users. Data that goes from its native source to a warehouse is said to flow Downstream.
- E -	
Effectiveness	A measure of one's ability to complete a work stream, operation, process or function such that the end result is the anticipated / expected current state and/or Asset. And operation can be Effective irrespective of whether or not it is efficient.
Efficiency	A measure of the human and financial costs required for a work stream, operation, process or function to be Effective. An operation may be Effective without being Efficient, but it can not be Efficient without being Effective.

Term	Definition
End-User	Consumer of assets generated by and delivered from the Information Delivery and Business Intelligence System.
End State	System, operations, processes and work flows projected to be in force as of the completion of specified project or work stream.
ETL	Extraction, Transformation and Loading of data as it flows Downstream from it's native source to a centralized data storage facility
Event	A metaphysical state of being inside the Enterprise that has significance within the Information Delivery and Business Intelligence System. Also see "Trigger"
- F -	
Firewall	A system of related security measures and access devices within a technological environment in place to control and monitor access to data elements
Format	Physical presentation design of assets; Placement and presentation details of Reports, Scorecards and Dashboards. Also see "Layout".
Future State	Systems, operations, processes and work flows projected to be in force as of a point in time that has not yet occurred but is still to occur
- G -	
Gated Approach	A project management methodology based on interim interactions between the project team and user community. The occurrence of these interactions is tied to the accomplishment of certain goals or the completion of some phase of design of an end-state asset
Governance	A system of security measures, data quality measures, data management processes, user password protocols and technology standards administered and overseen by Data Stewards
- I -	
Information	Numeric, Date and Time Data elements that have been manipulated according to arithmetic and/or statistic operations. Textual Data elements that have been manipulated according to alphanumeric manipulations including Concatenation and Parsing
Information Asset	See "Asset"
- L -	
Layout	The physical configuration and design of of an Information Asset. Also see "Format"
- M -	

Term	Definition
Metadata	A component of the Information Delivery and Business Intelligence system that translates user requests for information and BI into executable queries. The Metadata component translates end-user speak into computer-speak
Metric	A unit of measure defined by an Enterprise as an indication of process, operations or work stream results and/or impact
Mining	Collecting data elements from the results of operations, procedures or work streams
- P -	
Parameters	Boundaries set to control Data, Information and Business Intelligence content presented in Reporting Assets. Parameters may be numeric, textual or date/time in format and are often provided in pairs with one upper and one lower boundary. Exceptions include single parameters provided when a standard deviation is allowed as a qualification for content inclusion
Parsing	Separating a single textual string into multiple textual strings. Parsing is the opposite of Concatenation.
Portal	A gateway through which end users can access both the Presentation Layer and certain Information Assets. Intranet Portals are contained within a private internal network while Extranet and Internet Portals are available for use to authorized users outside the Enterprise.
Pull Delivery	Method of presenting Information Assets according to a specific user request
Push Delivery	Method of presenting Information Assets automatically to end users irrespective of a specific request
- Q -	
Query	Commands and parameters passed to a data source resulting in the extraction of a sub-set of data that is valid given the limits of acceptability defined by the parameters
- R -	
Relevance	A measure of Data Quality. A subjective evaluation of the degree of impact Data, Information or Business Intelligence can have at a point in time
Report	A type of Information Asset. Detail Reports contain specified data elements for all records that fall within the defined selection range. Summary Reports contain statistically and arithmetically calculated Information that present a macro view of the content of a related Detail Report. Exception Reports display specific data elements for records that fall outside the defined selection range.
- S -	

Term	Definition
Scorecard	A type of Information Asset. A static, numeric and textual summarized representation of the results of actions, work streams, operations or processes executed during a previous time period.
Security	Protection of the Intellectual Assets that are Data, Information, Business Intelligence and the calculations, methods and processes by which they are created.
Steward	See "Data Steward"
Strategy	Collection of actions, work streams and processes designed to accomplish a defined end state
Subject Matter Expert	Individual with deep level of understanding and extensive experience in a specific area of business operations or function
Subscription Delivery	Method of presenting Information Assets to end users where the asset is automatically Pushed to the end user based on the specific delivery details included in a Subscription. A Subscription is a user request for information that is issued once by the end user and then retained within the system and repeated according to a prescribed cadence
- T -	
Timeliness	A measure of Data Quality. A subjective comparative evaluation between when an Information Asset is received by an end user and when that Asset may impact the end user's actions, decisions and/or work streams.
Trigger	An metaphysical state that, once achieved, will cause an action or series of events to occur. Please also see "Event"
- U -	
Upstream Flow	Flow of data from end users to a central data storage facility. Data that is manually entered into the system is said to flow Upstream.
- W -	
Warning	High-level message from Information Delivery and Business Intelligence system that current-state is not within acceptable thresholds