# Glossary

| | |
|---|---|
| **Access** | The operation of seeking, reading, or writing data on a storage unit. |
| **Algorithm** | A set of statements organised to solve a problem in a finite number of steps. |
| **Application** | A group of algorithms and data interlinked to support an organisational requirement. |
| **Archival Data** | A collection of data of a historical nature. As a rule, archival data cannot be updated. Each unit of archival data is relevant to a moment in time. |
| **Attribute** | A property that can assume values for entities or relationships. Entities can be assigned several attributes. Some systems also allow relationships to have attributes as well. |
| **Audit/Log trails** | Data that is available to trace activity, usually update activity. |
| **Authentication** | The act of verifying the identity of a user and the user's eligibility to access computerised information (usually by means of password verification). Designed to protect against fraudulent logon activity. |

**Availability**

Component of internal control risk. Availability relates to information being available when required by the business process now and in the future. It also concerns the safeguarding of necessary resources and associated capabilities.

**Breadbox Analysis**

A phase of the data warehouse development. An analysis which determines the volume of data which will be retained within the data warehouse environment.

**Capacity Planning**

An assessment used to determine whether the data warehouse and associated interfaces will be able to manage an increase in the volume of data transferred to the data warehouse.

**Central Processing Unit**

Computer hardware which houses the electronic circuits that control/direct all operations in the computer system.

**Closed-loop Business Performance Management**

Future functionality which may provide data warehouse information on a real-time basis.

**Compliance**

Component of internal control risk. Compliance deals with complying with those laws regulations and contractual arrangements to which the business process is subject, i.e. externally imposed business criteria.

**Confidentiality**

Component of internal control risk. Confidentiality concerns the protection of sensitive information from unauthorised disclosure.

| | |
|---|---|
| **Continuity Plan** | A formalised procedure developed by management to assist in addressing loss of operations due to controllable and/or uncontrollable disasters. |
| **Conversion Plan** | The conversion plan details the mapping of data from the operational environment to the data warehouse environment. It also identifies the best route to migrate source data to the data warehouse. |
| **Data Administrator** | The individual responsible for the specification, acquisition, and maintenance of data management software and the design, validation, and security of files or databases. The data model and the data dictionary are classically the charge of the data administrator. |
| **Database** | A collection of interrelated data stored (often with controlled, limited redundancy) according to a schema. A database can serve single or multiple applications. |
| **Database Administrator** | The organisational function charged with the day-to-day monitoring and care of the databases. The database administrator is more closely associated with physical database design than the data administrator is. |
| **Database Management System** | A computer-based software system used to establish and manage data. |
| **Data Element** | (1) An attribute of an entity; (2) A uniquely named and well-defined category of data that consist of data |

items and that is included in a record of an activity.

**Dependant Data Mart**

A database, or collection of databases, designed to help managers make strategic decisions about their business. Whereas a data warehouse combines databases across an entire organisation, data marts are usually smaller and focus on a particular subject or department.

**Data Mining**

The automated analysis of detailed operational customer transaction data for the purposes of discovering hidden, unidentified or underlying patterns.

**Data Model**

(1) The logical data structures, including operations and constraints provided by a database management system for effective database processing; (2) The system used for the representation of data.

**Data Model Structure**

A logical relationship among data elements that is designed to support specific data manipulation functions.

**Data Warehouse**

A collection of integrated, subject-orientated databases designed to support the decision support function, where each unit of data is relevant to some moment in time. The data warehouse contains the lowest level data and lightly summarised data.

**Decision Support Systems**

A system used to support managerial decisions. Usually they involve the analysis of many units of data in a heuristic fashion. As a rule, their processing does not involve the update of data.

| | |
|---|---|
| **Distributed Data Warehouse** | A data warehouse environment dispersed over a large geographical area to address the unique needs of such a structured organisation. |
| **Effectiveness** | Component of internal control risk. Effectiveness deals with information being relevant and pertinent to the business process as well as being delivered in a timely, correct, consistent and usable manner. |
| **Efficiency** | Component of internal control risk. Efficiency concerns the provision of information through the optimal (most productive and economical) use of resources. |
| **Encryption** | A technique used to protect the plain text by coding the data with suitable algorithms such that it is unintelligible to the reader. |
| **Extract** | The process of selecting data from one environment and transporting it to another environment. |
| **Firewall** | Software used to protect against denial of services and any unauthorised access to internet resources. The system should control any application and infrastructure management flows in both communication directions, i.e. data sent and received by the organisation. |
| **Granularity** | The level of detail contained in a unit of data. The more detail there is, the lower the level of granularity. The less detail there is, the higher the level of granularity. |

| | |
|---|---|
| **Heuristic** | The mode of analysis in which the next step is determined by the results of the current step of analysis. |
| **Index** | The portion of the storage structure maintained to provide efficient access to a record when its index key item is known. |
| **Information** | Data human beings assimilate and evaluate to solve problems or make decisions. |
| **Information Technology Department/Personnel** | Personnel and/or department appointed as custodians of computer resources. |
| **Interface** | Program logic responsible for transferring data from the operational environment to the data warehouse environment. |
| **Integrity** | Component of internal control risk. Integrity relates to the accuracy and completeness of information as well as to the validity in accordance with business values and expectations. |
| **Internal Control Objective** | The specific goal of an audit or review. These often center around substantiating the existence of internal controls to minimise internal control risk. |
| **Internal Control Risk** | The risk that management's plans, organisation and associated procedures will not provide reasonable assurance that the organisation's goals and objectives will be achieved. See **Effectiveness, Efficiency, Confidentiality, Integrity, Availability,** |

| | |
|---|---|
| | **Compliance** and **Reliability** regarding components of internal control risk. |
| **Key** | A data item or combination of data items used to identify or locate a record instance (or other similar data groupings). |
| **Key, Primary** | A unique attribute used to identify a single record in a database. |
| **Key, Secondary** | A non-unique attribute used to identify a class of Records in a database. |
| **Knowledge Discovery in Databases** | A framework applied in decision support systems. The Knowledge Discovery in Databases process is split into five stages, viz. Selection, pre-processing, transformation, information access layer systems and, interpretation and evaluation. |
| **Load** | To insert data values into a database that was previously empty. |
| **Metadata** | (1) Data about data; (2) The description of the structure, content, keys, indexes, etc., of the data. |
| **Migration** | The process by which frequently used items of data are moved to more readily accessible areas of storage and infrequently used items of data are moved to less readily accessible areas of storage. |
| **Network** | A system of interconnected computers and the communications equipment used to connect them. |

| | |
|---|---|
| **Operational Data** | Data used to support the daily processing a company does. |
| **Populate** | To place occurrences of data values in a previously empty database. See **Load**. |
| **Primitive Data** | Data elements whose existence depends on a single occurrence of a major subject area of the enterprise. Primitive data is distinguished from secondary data thereby ensuring that duplicate data elements are avoided and that the most accurate data element is selected if duplicates are found. See **Secondary Data**. |
| **Project Team** | The data warehouse project team consists of a project leader, business analysts, data administrators, database administrators, systems support, computer programmers and end users. These personnel will be responsible for the overall project administration, design of the warehouse structures; analyse source data; identify how data is to be linked and, and if applicable, integrate external sources. |
| **Quality Assurance Review** | The aims of the quality assurance review are to determine whether the data warehouse development is being administered according to established standards and that project deficiencies are identified timeously and corrected with minimal resources. |
| **Query Optimiser** | A sub-module of the query processor used to govern and expedite the processing and data transmission required for responding to queries. Its aim is to ensure that either the total cost or the total response |

time for a query is minimised.

**Record**

An aggregation of values of data organised by their relation to a common key.

**Redundancy**

The practice of storing data more than one occurrence of data. In the case where data can be updated, redundancy poses serious problems. In the case where data is not updated, redundancy is often a valuable and necessary design technique.

**Referential Integrity**

The facility of a database management system to ensure the validity of predefined relationships.

**Reliability**

Component of internal control risk. Reliability of information relates to the provision of appropriate information for management to operate the entity and for management to exercise its financial and compliance reporting responsibilities.

**Repository**

See **Database**.

**Return on Investment**

A tool used for weighting expected benefits against the costs of a specific project. Return on Investment calculations include cash flow analysis, net present value calculations, return on investment and payback calculations.

**Security Plan**

A plan which identifies the organisation's standards regarding system security and the user's responsibilities while utilising computer equipment and/or system software.

| | |
|---|---|
| **Secondary Data** | Data elements whose existence depends on two or more occurrences of a major subject. See **Primitive Data**. |
| **SEMMA Methodology** | A data mining methodology. It consists of 5 stages, viz. sampling, exploring, modifying, modeling and assessing. |
| **Snapshot** | A database dump or the archiving of data out of a database as of some moment in time. |
| **Source Systems Analysis** | A phase of the data warehouse development. An analysis which defines the source data elements, evaluates the accuracy of data before migration and considers how to manage the volume of data elements. |
| **Stability Analysis** | An analysis which considers the grouping together of data elements which will change based on similar conditions. |
| **Strategic Planning** | A phase of the data warehouse development. Strategic planning identifies whether the organisation is in need of a data warehouse environment, and if so, what would the extent of such a development be. |
| **Subject Area Analysis** | A phase of the data warehouse development. Identification of suitable population data from existing applications which should be introduced into the data warehouse environment. |
| **System Development Life Cycle** | The framework applied in the development of an application. The framework is split into distinct phases to improve controllability of the development |

process.

| | |
|---|---|
| **Table** | A relation that consists of a set of columns with a heading and a set of rows. |
| **Technical Environmental Preparation** | A phase of the data warehouse development. The primary aim of this phase will be to determine how the organisation's network capabilities will be affected by the increased traffic created by the data warehouse. |
| **Time Stamping** | The practice of tagging each record with some moment in time, usually when the record was created or when the record was passed from one environment to another. |
| **Technical Assessment** | A phase of the data warehouse development. It focuses on determining the architectural configuration needed to house the data warehouse. |
| **User** | A person or process issuing commands and messages to the information system. |
| **Year 2000 compliance** | Identified as the risk that hardware and/or software will be unable to process transactions with 4 digit year date fields (i.e. DD/MM/**YYYY**). |