IT Project Guidance

Glossary of IT Specific Terms:   
Data Storage

Version:

0.2

## Description

A Glossary of common IT Terms related to data storage and manipulation, to establish a common understanding, while reducing duplication of effort in downstream documents.

## Synopsis

Included are the meanings of acronyms and industry terms used to describe aspects of data storage.

## Contents

[Description 1](#_Toc146637002)

[Contents 2](#_Toc146637003)

[Introduction 3](#_Toc146637004)

[Objective 3](#_Toc146637005)

[Terms & Acronyms 4](#_Toc146637006)

[System Data Storage Terms & Acronyms 4](#_Toc146637007)

[Data Hub 4](#_Toc146637008)

[Indexed 4](#_Toc146637009)

[Key 4](#_Toc146637010)

[Longitudinal Data 4](#_Toc146637011)

[Master Data 4](#_Toc146637012)

[Natural Key 4](#_Toc146637013)

[Natural Primary Key 4](#_Toc146637014)

[Normalised 4](#_Toc146637015)

[Primary Key 4](#_Toc146637016)

[Reference Data 4](#_Toc146637017)

[Relational Database 4](#_Toc146637018)

[Schema 5](#_Toc146637019)

[Secure keystore 5](#_Toc146637020)

[System of Record 5](#_Toc146637021)

[Transactional Data 5](#_Toc146637022)

[Unoptimised 5](#_Toc146637023)

[Optimised 5](#_Toc146637024)

[Appendices 6](#_Toc146637025)

[Appendix A - Document Information 6](#_Toc146637026)

[Images 6](#_Toc146637027)

[Tables 6](#_Toc146637028)

[References 6](#_Toc146637029)

[Review Distribution 6](#_Toc146637030)

[Audience 6](#_Toc146637031)

[Diagrams 6](#_Toc146637032)

## Introduction

## Objective

To develop a common understanding of terms used to deliver services with an IT component.

# Terms & Acronyms

## System Data Storage Terms & Acronyms

#### Backup

: the physical duplication of data to support restoration and thereby protect data against inadvertent logical state change (e.g.: *deletion*) or complete *erasure*.

#### Data Hub

* : A data platform architecture that provides management of integration, transformation, storage, and subsequent availability.

#### Datastore

* : any form of data storage. The primary types are classified as:
* Relational (traditional databases)
* Non-Relational (no-SQL, blob, key-value, etc.)

#### Deletion

: disambiguated from Erasure, in its strictest meaning, deletion is the unlinking to a record or changing its state to Deleted or Removed, without Erasing it. With sufficient permissions, authorised users can retrieve the record.

#### Codeset

* : a shared list of codes that is used in place of longer names or explanations. See *Reference Data*.

#### Erasure

: the physical removal of a record so as to be unretrievable.

#### Indexed

* : a table that is indexed upon insertion of new records, to speed up subsequent finding of records.

#### Key

: an attribute (column) or a set of attributes that help to uniquely identify a tuple (or row) in a table within a datastore or table. See *Primary Key*.

#### Longitudinal Data

* : the collection of repeated observations of individuals over a duration of time.   
  In an educational context, the duration is generally measured in years, even decades, during which a Lifelong Learner will transition between multiple Education Providers.

#### Master Data

: data describing the entities needed to do business. Customers, accounts, etc. See *Transactional Data.* Categorised and classified with *Reference Data*.

#### Natural Key

: a key derived from the data itself (e.g.: a national person identifier), as opposed to an attribute generated by the database (e.g.: row number or a guid).

#### Natural Primary Key

: a *natural key* used as the table’s *primary key*.   
*Note: Use thereof is a design error, as performance is severely impacted negatively.*

#### Normalised

: the process of structuring *relational databases* in accordance with a series of so-called *normal forms* in order to reduce data-redundancy and improve data-integrity.

#### Primary Key

: the *key* in a relational database table that's distinctive for each record. See *key*.

#### Reference Data

: a special subset of master data that is used for classification and categorisation. Whereas Reference data is often internal to a system or organisation, they may also be keyed to cross organisation codesets. See *Codeset.*

#### Relational Database

: a relational database is a collection of information that organizes data in predefined relationships where data is stored in one or more tables (or "relations") of columns and rows, making it easy to see and understand how different data structures relate to each other. The organisation of these tables is referred to as a *schema*.

#### Restoration

: the process of overwriting production data with data that was previously backed up, thereby protecting systems against inadvertent state change (e.g. *deletion*) and *erasure*.

#### Schema

* : the definition of how information is persisted in a datastore (generally a relational database).

#### Secure keystore

: a non-relational key-value datastore for confidential information, usually integration information.   
Note: a relatively common high value example use case is to only permit a deployment pipeline to be a member of the secure keystore, to retrieve confidential integration credentials which it then injects into deployments, removing the risk of humans accessing, knowing and disclosing these credentials.

#### System of Record

* : a data management term for an information service that is the authoritative data source.

#### Transactional Data

: data collected about interactions and events, referencing *Master Data* and *Reference Data*. See *Master Data*, *Reference Data*.

#### Unoptimised

: a database who’s performance is impacting system performance. The causes are often one or more of the following:

* Tables are not effectively *normalised,*
* tables miss useful indexes,
* tables have too unnecessary non-valuable indexes and omission.

#### Optimised

: a table or database on which work has been done to remove causes of for it being unoptimized.

Appendices

Appendix A - Document Information

### Authors & Collaborators

Author: Sky Sigal, Solution Architect

### Versions

* 1. Initial Release
  2. Minor Changes

### Images

**No table of figures entries found.**

### Tables

**No table of figures entries found.**

### References

**There are no sources in the current document.**

### Review Distribution

The document was distributed for review as below:

|  |  |
| --- | --- |
| Identity | Notes |
| Sandy Britain, Enterprise Architect |  |
| Amy Orr, Data Architect |  |
| Roger Govind, Security Architect |  |
| Archana Sahani, Business Analyst |  |
| Dijana Sneath, Business Analyst |  |
| Vincent Weirdsma, Lead Developer |  |

### Audience

The document is technical in nature, but parts are expected to be read and/or validated by a non-technical audience.

### Diagrams

Diagrams are developed for a wide audience. Unless specifically for a technical audience, where the use of industry standard diagram types (ArchiMate, UML, C4), is appropriate, diagrams are developed as simple “box & line” monochrome diagrams.