Unstructured Data Profiling Service.

A - Archive First, B – Backup Less and C – Control Storage Costs.

Background.

Data Information Management is more critical than ever in today's dynamic business climate. Few organisations have the resources and time to manually manage and maintain the "Information Lifecycle" of data within their organisation, while also planning for future data growth. Merely adding more storage capacity every time capacity is exceeded only adds and compounds the infrastructure management burdens, creates potential points of failure, and consumes precious IT resources and budgets.

The seamless, smooth and continuous operation of many IT business services often depends on the availability, performance, capacity and health of the storage infrastructure. As the business and IT environment grow, it becomes all too easy for the storage of data to become disorganised, generally unhealthy and unmanaged.

It has become generally reported and publicised by many leading IT analysts that on average 40% of the organisation's unstructured data can be classified as Redundant, Obsolete or Trivial (ROT).

The Unstructured Data Profiling service provides a detailed analysis and report, based upon metadata collected from a reprehensive data sample (Maximum 5TB). The findings from this analysis provide invaluable insights into the level of file duplication; individual data file size/ ageing characteristics and the immediate identification of data files with zero business value.

Service Deliverables.

Firstly, a general purpose Unstructured Data "ROT" analysis. This analysis immediately identifies areas of cost saving, by the identification of:

- Redundant - Data Files that contain identical content (duplicates).

- Obsolete - Data files whose "Business Value" has reduced to a point where the data file contents have not been "Accessed" for multiple years (>5yrs).

- Trivial – Data files which are classified as not company related.

Secondly, it delivers invaluable insights into how data backup time could be significantly reduced and optimised (via the future deployment of intelligent archiving and stubbing technology).

Thirdly, provides a snapshot insight into the rate at which the business is consuming storage. By utilising this information, an organisation can begin to formulate and implement an Information Data Management Strategy.

Moreover, **fourthly**, understanding the rate at which an organisation is creating new content, allows future infrastructure investments to be focused and aligned to the business.
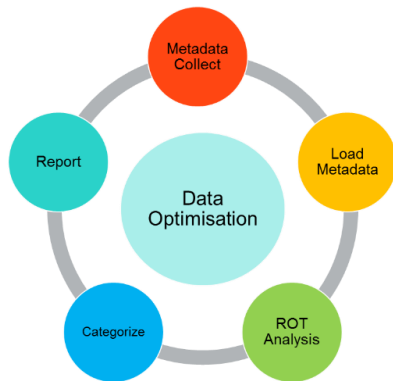
Service Activities.

Figure 1. Data Optimisation Methodology.

Figure 2. Data ROT Analysis.

This service comprises the following tasks and deliverables:

1. Pre-Installation conference call. A general technical exchange/data gathering conference call. Access to a secure site where the Remote Analysis Agent (RAA) software kit can be downloaded.

2. The Remote Analysis Agent collects Metadata information from the customer designated MicrosoftWindows file shares.

3. Upon completion of the file scan activity, the collected metadata Index files (IDX) should be transferred to StoreXltd for Analysis.

4. Ingestion and Analysis of the transferred IDX files, for example:-

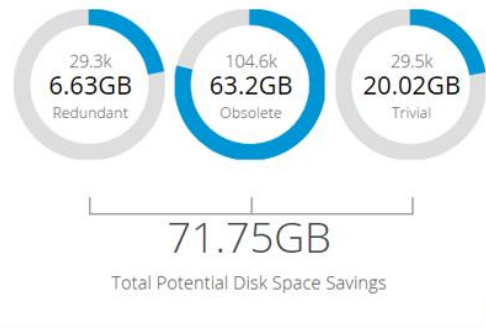   4.1. Perform initial high-level data analysis (ROT, data type, growth…).

    4.2. High-level files type classification.
    4.3. Determine the level of "structured" vs "unstructured" data.
    4.4. Identify the amount of file duplication.
    4.5. Determine the ageing charactertics of the ingested data.
    4.6. Detailed file analysis by file extension.
    4.7. Determine data volume addition rate per year.
    4.8. Data file analysis by User.
    4.9. Publish "Findings & Recommendations" report.

5. Present and discuss report findings.

Service Exclusions.

It is important to note that this service excludes explicitly the following:

- Any remedial activity. However, if assistance is required to address some or all of the findings, it is suggested you request details of additional packages or consultancy services

- Configuration changes.

- Capacity Planning or detailed monitoring of the system.

- A hardware/patch audit of the system.

Delivery Pre-Requisites.

To ensure this product is delivered efficiently and effectively it is critical that the customer.

- Ensures that technical assistance is available as and when requested by the technical specialist performing this study.

- The customer identifies the file share (UNC path) to be indexed using the Remote Analysis Agent.

- The Metadata files generated by the Remote Analysis tool are copied to the appropriate media for off-site analysis by StoreXltd.

Customer's server Pre-requisites.

Software Pre-Requisites.

The server where the Remote Analysis Agent will be installed must have Windows 3.5 and 4.5 (or higher) ".Net Framework" installed.

Hardware Pre-Requisites

None.