Strategic Document to Guide in the Process of the Design of SADTU's Historical Research Archive

1. Introductory overview

This strategic document is to guide in the process of the establishment of SADTU's Historical Research Archive as outlined in SADTU's five year Strategic Plan which identified **a programme to establish a SADTU knowledge institute (physical or virtual).** Its success indicator is to conduct comparative research on education policy modules within policy parameters of SADTU and to post onto the SADTU website all knowledge related to being a member so as to pursue the mission of the organisation which is to provide an administrative support to the leadership in order to achieve the mandate of the organisation thereby realising the 7th National Congress mandate theme. This will also contribute towards SADTU's strategic goal of ensuring that the membership of the union are serviced through being able to access research and historical data of the organisation in their membership and career development in a quest to create a learning nation.

Archiving is a process of moving interactive data to a storage location that can be located and accessed when necessary. Data archiving is an excellent solution for the problems that are associated with ever-increasing data. Therefore the purpose of archiving will be to store historical information about the establishment and the contributory role that SADTU played in the struggle for national transformation of our country. Through archiving our membership will be able to access such information so as to develop a concrete historical evidence of how our organisation was born, what it stood for, its values and ethics so as to preserve such information and knowledge for posterity. Through archiving our membership will also be able to access latest research information both physically and digitally that will develop their human capital and be used for the purpose of professional development in different areas of research they are engaged in and to promote a culture of life-long learning in our organisation and therefore enhance Quality of Learning and Teaching Campaign (QLTC) in their different schools.

This strategic document will focus on the need for archiving, benefits of archiving, the importance of designing an achieving strategy, mistakes that should be avoided when designing an archiving strategy, and recommendations for designing a successful archiving strategy.

2. The need for archiving

Several factors present the need for data archiving. First, the volume of databases continues to grow. Second, much of this data is inactive but must be maintained for organisational memory and business reasons. Third, it is important to maintain data as inexpensively and effectively as possible. Data centres are concerned with ever-increasing volumes of data. Databases are growing at an unprecedented rate. Some inactive data resides everywhere because of historical transactions, the use of data warehouses, and other reasons. Inactive data

is not unusable; however, it is less likely to be accessed. Several reasons why we might be maintaining inactive data include; we are following government requirements; anticipate the need for building future trend analysis; need to maintain a complete history of the organisation and membership. As the size of a database grows, the percentage of inactive data often grows as well, which can result in performance problems and extensive processing that impacts users, manageability problems because large objects become difficult to reorganize and size limits are reached and extensive hardware and storage costs. An excellent solution to these problems is data archiving.

3. Benefits of archiving

Archiving data reclaims disk space, lowers storage costs, reduces object maintenance, and improves the performance of active data. The majority of benefits of archiving are interrelated. By reclaiming disk space that was consumed by inactive data, we can effectively lower our storage costs and reduce object maintenance. As a result accesses to our active data become more efficient.

3.1. Lowering costs

Lowering costs is the most important benefit of archiving our inactive data. We can store archives on less expensive storage media than holding our active data. Assume that our active data is stored on a high-performance, expensive data storage medium. When we establish an archiving strategy, we might want to incorporate a less expensive (and probably lower performance) storage medium on which to place our archive tables and files. An example of this type of medium is magnetic tape. In addition, applications might experience improved performance and use less memory resources, which results in further cost savings.

3.2. Less object maintenance

Reducing object maintenance is one of the major benefits to archiving our inactive data. Archiving DB2 data reduces the overall number of DB2 objects that require maintenance. For example, when less data exists, utility processing requires less time, reducing batch processing times.

4. The importance of designing an archiving strategy

An archiving strategy is a plan that ensures that we are archiving the correct data in the correct way. Archiving without an archiving strategy might result in poor overall system performance. The most successful archiving implementations are those that are well-planned and take several important aspects into account. When creating an archiving strategy, we should answer the following questions:

- What data do we need to archive?
- When do we need to archive this data?
- How long do we need to maintain the archived data?
- Do we need SQL access to the archived data?
- Should the archived data be stored in DB2 table archives or in flat-file archives?

- Under which circumstances and how often will we need to retrieve archived data from either table archives or file archives?
- Do we require archived data that is stored in table archives to be moved to flat file archives (in other words, do we require a multitiered archiving strategy)?

After we have answered these questions, we can begin to use DB2 Data Archive Expert to build our archives.

5. Mistakes to avoid when designing an archiving strategy

Poorly designed and inefficient archiving schemes are the result of some common pitfalls. When we designing an archiving strategy, we should be sure to avoid the following mistakes:

- We should not design an archiving strategy in a vacuum.
- We should include key departments and units for input, especially those who will be using the archives for their information storage purposes.
- We should not separate high-level application design from the archiving strategy for those applications.

The result is that archiving is often an after-thought for those applications, losing the most efficient methods of archiving data for those applications. In most cases, changing applications is prohibitively expensive, and a less-than optimal scheme must be used. We should not archive at random or on an ad-hoc basis. Successful archiving strategies are designed to be proactive rather than reactive, well-integrated with application processing. We should not assume that the schema for the active data will remain the same forever. This is simply a poor assumption in most cases.

6. Recommendation for designing a successful archiving strategy

The success of our archiving strategy depends on a number of factors. To design a successful archiving strategy:

- We assign an application team to maintain control over the archiving for our installation.
- Understand the schemas, the data, and how they are accessed.
- Understand our archive retrieval requirements.
- Assess the retention requirements for archieved data, that is, determine how long we need to keep archieved data. This factor strongly influences the design of our strategy. For example, if we require some archieved data to be accessed periodically, and other archieved data to be accessed very infrequently, we might determine that a multitiered strategy is required in which very old archives can be deleted after a certain period of time.
- Identify truly active data (versus inactive data) early in the design process. If archived data later requires updating, it really becomes active data again. The assumption is that archived data will not be updated because once it is modified, it no longer represent accurate snapshot of that data.

- Establish a consistent scheduling process for our archiving. The best strategy is to archive data on a regular basis to maintain storage goals and to help in retrieval consistency. We can still run single archive operations on as-needed basis.
- Because many approaches to archiving exist, we should choose the one that best suits our needs.