DATA CATEGORIZATION

Data categories are groupings of data with common characteristics or features. Data classification is the process of organizing data into categories for its most effective and efficient use. A well-planned data classification system makes essential data easy to find and retrieve. This can be of particular importance for risk management, legal discovery, and compliance.

Data categories are groupings of data with common characteristics or features. Data classification is the process of organizing data into categories for its most effective and efficient use. A well-planned data classification system makes essential data easy to find and retrieve. This can be of particular importance for risk management, legal discovery, and compliance.

1. How to start process of data categorization:

First step is to evaluate and divide the various applications and data into their respective category as follows:

1. Relational or Tabular data (around 15% of non audio/video data)
    1. Generally describes proprietary data which can be accessible only through application or application programming interfaces (API)
    2. Applications that produce structured data are usually database applications.
    3. This type of data usually brings complex procedures of data evaluation and migration between the storage tiers.
    4. To ensure adequate quality standards, the classification process has to be monitored by subject matter experts.

2. Semi-structured or Poly-structured data (all other non audio/video data that does not confirm to a system or platform defined Relational or Tabular form).

1. Generally describes data files that have a dynamic or non-relational semantic structure (e.g. documents , XML ,JSON ,Device or System Log output , Sensor Output).
2. Relatively simple process of data classification is criteria assignment.
3. Simple process of data migration between assigned segments of predefined storage tiers.

2. Types of data classification:

1. Geographical: i.e. according to area (supposing the rice production of a state or country etc.)
2. Chronological: i.e. according to time (sale of last 3 months)
3. Qualitative: i.e. according to distinct categories. (E.g.: population on the basis of poor and rich)
4. Quantitative: i.e. according to magnitude(a) discrete and b)continuous.

3. Benefits of data classification:

Benefits of effective implementation of appropriate data classification can significantly improve ILM process and save data centre storage resources. If implemented systemically it can generate improvements in data centre performance and utilization. Data classification can also reduce costs and administration overhead. "Good enough" data classification can produce these results:

1. Data compliance and easier risk management. Data are located where expected on predefined storage tier and "point in time"

2. Simplification of data encryption because all data need not be encrypted. This saves valuable processor cycles and all related consecutiveness.

3. Data indexing to improve user access times

4. Data protection is redefined where RTO (Recovery Time Objective) is improved.

Related Posts:

1. Information Life Cycle Management (ILM)
2. Storage infrastructure
3. Integrated VS Modular Array
4. Data proliferation
5. Component architecture of intelligent disk subsystem
6. Intelligent disk subsystems overview
7. Mapping n operations
8. Storage system architecture
9. RAID
10. Hot spare
11. SAN security
12. JBOD
13. Elements of DAS,NAS,CAS,SAS
14. Limitations of DAS
15. Cloud vocabulary
16. NAS security

17. Management of DAS,NAS,CAS,SAN
18. FC Connectivity
19. Memory virtualization
20. Data center concepts & requirements
21. Network virtualization
22. Server information storage and management
23. ISM Architectural Framework