RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

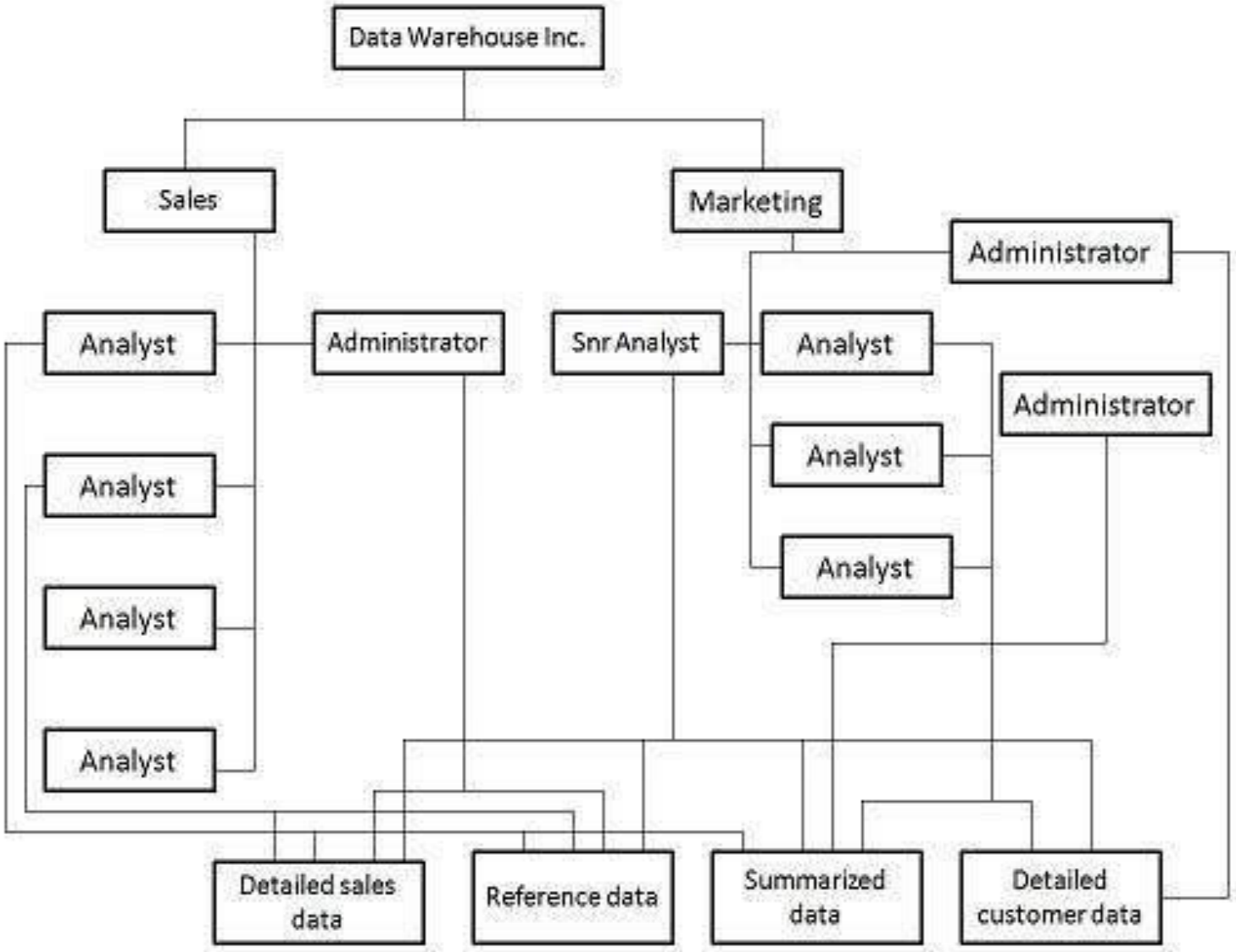
DATA MINING & WAREHOUSEING
LECTURE-14

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- USER ACCESS HIERARCHY
- BACKUP AND RECOVERY
  - TYPES OF BACKUP
  - TAPE TECHNOLOGY
  - BACKUP STRATEGIES
  - RECOVERY STRATEGIES
- MCQ
- REFERENCES

User Access Hierarchy



Backup and Recovery

- Backup is one of the most important regular operations carried out on any system.
- It is important in the DWH environment because of the volumes of data involved and the complexity of the system.

• Types of Backup

- In a complete backup, the entire database is backup at the same time. This includes all database data files, the control files and the journal files.
- **Partial backup** is any backup that is not complete.
- **Cold backup** is a backup that is taken while the database is completely shutdown. In a multi-instance environment, all instances of that database must be shut down.
- **Hot backup :** any backup that is not cold is considered to be hot.
- **Online backup** is a synonym for hot backup

Backup and Recovery

Hardware: When you are choosing a backup strategy, one of the first decisions you will have to make is which H/W to use. The choice depend on factors such as speed at which a backup or restore can be processed, H/W connection, N/W bandwidth, backup S/W used, speed of server's I/O subsystem and components.

Tape Technology

Tape media (reliability / life / cost of tape medium/drive, scalability)

Standalone tape drives (connection directly to servers, as a network-available device, remotely to another machine)

Tape stackers (a method of loading multiple tapes into a single tape drive. Only one tape can be accessed at a time, but the stacker will automatically dismount the current tape when it has finished with it and load the next tape)

Tape silos (These are large tape storage facilities, which can store and manage thousands of tapes. These are generally sealed environments with robotic arms for manipulating the tapes.)

Disk-to-disk backups (the backup is performed to disk rather than to tape)

Backup and Recovery

Software (Omniback II, ADSM, Alexandria, Epoch, Networker)

Performance- such as degree of parallelism, I/O bottlenecks

Requirements : When considering which backup package to use it is important to check the following criteria.

- ❑ What degree of parallelism is possible?
- ❑ How scalable is the product as tape drives are added?
- ❑ What platforms are supported by the package?
- ❑ What tape drives and tape media are supported by the package?
- ❑ Does the package support easy access to information about tape contents?

Backup strategies:

- ❑ Effect on database Design- such as DB partitioning strategies.
- ❑ Design Strategies -main aim should be to reduce the amount of data that has to be backed up on regular basis, e.g. Read-only tablespace, automation of backup

Backup and Recovery

Recovery Strategies- depend on kind of failure and consist of a set of failure scenarios & their resolution. Each of the following **failure scenarios** indicated below needs to be centred for recovery steps and must be documented:

1. Instance Failure
2. Media failure
3. Loss or damage of table space or data file
4. Loss or damage of a table
5. Loss or damage of control file

6. Failure during data movement The plan must be made for following data movement scenarios :

- ❑ Data load into staging tables
- ❑ Movement from staging to fact table
- ❑ Partition roll-up into larger partitions
- ❑ Creation of Aggregations.

Testing the Strategy- The backup and recovery tests need to be carried out on a regular basis, but it is advisable to avoid performing tests at busy times such as end of year and try to test to be run at low load in the business year.

Backup and Recovery

Disaster Recovery- A disaster can be defined where a major site loss has taken place or destroyed or damaged beyond immediate repair. It is advisable to decide a criteria for making that judgment before any situation occurs, as attempting to make such a decision while in the middle of crises is likely to lead problems. Deal with following requirements for disaster recovery:

- Planning Disaster recovery with minimal system required.
- Replacement / standby machine
- Sufficient tape and disk capacity
- Communication links to users.
- Communication links to data sources.
- Copies of all relevant pieces of software.
- Backup of database.
- Application-aware systems administration and operation staff.

Multiple Choice Question

1. _____ is data collected from natural systems.
a) MRI scan.
b) ODS data.
c) Statistical data.
d) Historical data.

2. _____ is an example of application development environments.
a) Visual Basic.
b) Oracle
c) Sybase
d) SQL Server.

3. The term that is not associated with data cleaning process is _____.
a) domain consistency.
b) deduplication
c) disambiguation
d) segmentation

4. _____ are some popular OLAP tools.

a) Metacube, Informix.
b) Oracle Express, Essbase.
c) HOLAP
d) MOLAP

5. Capability of data mining is to build _____ models.
a) retrospective
b) interrogative
c) predictive
d) imperative

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture
- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education
- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier
- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education
- Mallach," Data Warehousing System", TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE'97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB'96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD'97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.