KONKAN GYANPEETH

KARJAT COLLEGE OF ARTS, SCIENCE AND COMMERCE

# Data Warehousing

B.Sc. (IT) Sem. VI

## EVALUATION SYSTEM

| | Time | Marks |
|---|---|---|
| **Theory Exam** | 3 Hrs. | 100 |
| **TW / Tutorial** | – | 50 |

## SYLLABUS

1. **History of data warehousing :** database management system, personal computers and 4GL technology, spider web environment, evolution from business perspective, data warehouse environment, what is datawarehouse?, integrating data, volumes of data, different development approach, evolution to DW 2.0 environment, business impact of the data warehouse, components of datawarehouse environment, evolution of data warehouse from the business perspective, other notions about data warehouse, federated data warehouse, star schema, data mart.

   **Introduction :** Lifecycle of data warehouse, reasons for different sectors, metadata, Access of data, structures data/ unstructured data, Textual analysis, blather, issue of terminology, specific text, metadata - a major component, local metadata, changing business requirements, flow of data within DW 2.0, volumes, useful applications, DW 2.0 and referential integrity, reporting in DW 2.0.

   **DW components :** Interactive sector, integrated sector, Near Line sector, Archival sector.

2. **Metadata in DW :** Reusability of data analysis, Metadata, Active/ passive repository, enterprise metadata, metadata and the system record, Taxonomy, Internal and external taxonomy, metadata in archival sector, maintaining metadata, using metadata - an example, end user perspective.

   **Methodology and Approach for DW :** Spiral model methodology, seven streams approach, enterprise reference model, enterprise knowledge coordination stream, information factory development stream, Data correction stream, infrastructure stream, Total information quality management stream.

   **Statistical processing and DW :** Two types of transaction, statistical analysis, integrity of comparison, heuristic analysis, freezing data, exploration processing, frequency of analysis, exploration facility, sources for exploration processing, refreshing exploration data, project based data, Data marts and exploration facility, A backflow of data, using exploration data internally, perspective of business analyst.

3. **Data models and DW :** datamodel and business, scope of integration, making the distinction between granular and summarized data, levels of the data model, data models and interactive sector, corporate data model, transformation of models, data models and unstructured data, perspective of business user.

   **Monitoring the DW environment :** Monitoring DW environment, transaction monitor, monitoring data quality, datawarehouse monitor, transaction monitor, peak period processing, ETL data quality monitor, Dormant data.

   **DW and security :** Protecting access to data, encryption, drawbacks, firewall, moving data offline, limiting encryption, direct dump, datawarehouse monitor, sensing an attack, security for near line data.

4. **Time variant data :** All data in DW, Time relativity in the interactive sector, data relativity elsewhere in DW, Transactions in integrated sector, discrete data, continuous time span data, a sequence of records, nonoverlapping records, beginning and ending a sequence of records, continuity of data, Time-collapsed data, time variance in the archival sector.

   **Flow of data in DW :** flow of data throughout the architecture, entering the interactive sector, role of ETL, data flow into integrated sector, near line, archival sector, falling probability of data access, exception flow of data.

   **ETL processing and DW :** changing states of data, Where ETL fits, application data to corporate data, ETL in online mode and batch mode, source and target, ETL mapping, more complex transformation, ETL and throughput, ETL and metadata, ETL and an audit trail, ETL and data quality, creating ETL, code creation or parametrically driven ETL, ETL and rejects, changed data capture, ETL and rejects, Changed data capture, ELT

5. **DW and granularity manager :** granularity manager, raising the level of granularity, filtering data, functions of the granularity manager, homegrown versus third party granularity manager, parallelizing the granularity manager.

   **DW and performance :** online response time, analytical response time, flow of data, Queues, heuristic processing, analytical productivity and response time, many facets to performance, indexing, removing dormant data, end user education, monitoring the environment, capacity planning, metadata, batch parallelization, parallelization for transaction processing, workload management, data marts, exploration facilities, separation of transactions into classes, service level agreements, protecting the interactive sector, partitioning data, choosing the proper hardware, separating farmers and explorers, physically group data.

   **Migration :** Migration in perfect world, adding components incrementally, adding archival sector, creating enterprise metadata, building the metadata infrastructure, swallowing source system, ETL as shock absorber, migration to the unstructured environment.

6. **Implementation And Maintenance :** Physical design process, data warehouse deployment, growth and maintenance.

**References :**


1. Building the data warehouse, (*W. H. Inmon*), third edition, Wiley.
2. Datawarehousing, (*S. Mohanty*), TMH
3. The Data Warehouse Lifecycle toolkit, (*Ralph Kimball*), John Wilely.