**Course: Advance Bio Informatics**

**Module Title: Data Integration Techniques**

**Module No: 82**

Single unified interface of biological databases.

It has three techniques

1. **Data warehousing**: Uses a single physical database for all biological data. Data is downloaded from different databases, converted into unified format and stored.
2. **Federated Data Integration:** Data is stored into multiple databases. User query is divided into sub-queries and each query is forwarded to relevant database. Result is combined and presented to user. Query Manager handles query division and result integration.
3. **Semantic Data Integration:** Similar to data warehouse technique in way that data is stored into single location. Instead of using relational databases, data is stored in the form of ontologies.

**Some Existing Databanks**

|  |  |  |
| --- | --- | --- |
| **Databank Name** | **Integration Technique** | **Data Sources** |
| BioWarehouse | Data warehousing | ENZYME, KEG, BioCuyc, UniProt, GenBank, NCBI Taxonomy, CMR databases, Gene Ontology. |
| VINEdb | Data warehousing | KEGG, OMIM, IntAct, GO,  UniProt |
| BioDWH | Data warehousing | OMIM, KEGG, UniProt, Brenda, GO. |
| Columba | Data warehousing | KEGG, PDB, Swiss-Prot, SCOP, Gene Ontology, Enzyme. |
| Atlas | Data warehousing | KEGG, OMIM, IntAct, GO,  UniProt |
| BioDWH | Data warehousing | GenBank,  RefSeq, Uniprot, Human Protein Reference Database (HPRD) etc. |
| YeastMed | Federated Based | SGD,  YEASTRACT, MIPS-CYGD,  Bio-GRID, PhosphoGRID. |
| Semantic web services and Multi-Agent System (SEMMAS) | Semantic Web | GenBank, EMBL, DDBJ, SWISS-PROT, UniProt, PDB. |
| FungalWeb | Semantic Web | NCBI, NEWT, BRENDA, SwissProt. |