**Course: Advanced Bioinformatics**

**Module title: Data Integration**

**Module no. : 214**

**Data Integration**

Dbs are great: they let us manage huge amounts of data

Data sets are often created independently

Tie together different sources, controlled by many people, under a common schema.

**DBMS**

Abstraction

Logical vs. Physical

What vs. How.

**Data Integration**

* Obtain data from a data source
* Reformat the data
* Submit the formatted data to another data source for analysis
* Parse the analyzed result
* Combine result with data obtained from another data source

**Objectives of Data integration**

* Establish automated and efficient ways to integrate large, heterogeneous biological datasets from multiple sources
* Uniform query access to a set of data sources
* Scale of sources: from tens to millions
* Heterogeneity
* Autonomy
* Semi-structure