DATA WAREHOUSES

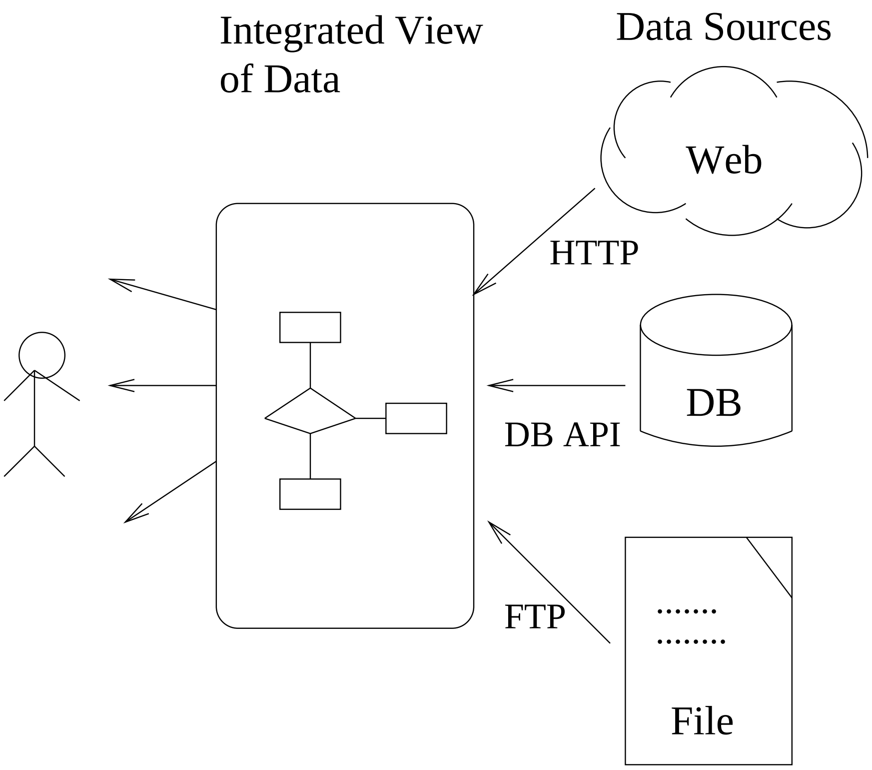
and

DATA INTEGRATION SERVICES

- Team Members:
  Christian Convey
  Olga Karpenko
  Nesime Tatbul
  Jue Yan

Data Integration

- harmonizing data from multiple sources into a single, coherent representation



Issues and Problems

- heterogeneity of the data sources
- availability of the data sources
- autonomy the of data sources
- inconsistent or missing data
- freshness of data
- performance

Approaches to Data Integration

- Federated Databases

  The sources are independent, but one source can call on others to supply information.

- Warehousing (Materialized Views)

  Copies of data from several sources, possibly in a processed form, are stored in a single database called a data warehouse.

  Important applications: OLAP (On-Line Analytic Processing) and Data Mining

- Mediation

  A software component called the mediator supports a virtual view which can be queried by the user as if the data were stored in a data warehouse.