Data acquisition – Getting the necessary data

Data Integration – Integrating the different dataset to a unified data set conveying relevant business concepts

Data Wrangling – dealing with messy data

Self-service Data Preparation – users given freedom to perform data preparation.

It is a challenge since data management becomes difficult for instance data protection laws can be infringed, the data can be altered.

Data prep follows ETL process

Data preparation tools

Hand coded -

Designer tools –

Data prep tool capabilities

General

1. Connect to diff data sources/dest.
2. Parse different data formats
3. Graphical/visual designer
4. Provide common control flow and data transformation tasks
5. Job automation and scheduling
6. Operation monitoring and error reporting

Advanced capabilities

1.

2.