

CHAPTER 1 INTRODUCTION

With the current trend in the industries to learn from one's mistakes and those of others, setups from small retailers to large corporations are looking towards the "Knowledge Discovery" paradigm to get a more comprehensive overview of their own data. This has been further made possible due to the increased performances in the data warehousing and data mining algorithms, falling costs of storage units and increased processing speeds. The sections that follow illustrate techniques used to store and mine data, in order to derive information therefrom, that had never been available or thought of heretofore.

1.1 Data Warehousing

With the excessively large customer transactions happening every second, in large super markets, internet sites, bank, insurance or phone companies, there is a need to come up with alternative methods to store the historical data, so that there is no loss of information. Also there could be a need, at a later date, to draw out the *hidden knowledge* from the data without a tangible loss of information. The study of data warehousing comprises everything from the machine architecture and the data store, that could be best compatible to store the specific form of data, to the algorithms used to handle and store data.

A data warehouse can be thought of as a collection of different sources of data, put together, that have been cleaned and checked for inconsistencies prior to merging. These data sources could be from different locations of a chain of super markets, like Wal-Mart, or data recorded at the same location over time. The