

## **ABSTRACT**

### ***ANALYSIS AND DESIGN OF DATA CLEANSING USING OPENSOURCE TOOLS***

**By**

**SITI AULIA NOOR**

**1202151364**

*The rapid development of technology will produce diverse data or information so that it can influence the decision making process. This causes many organizations to compete to process the data that is owned by each organization. Besides that, the quality of data is an important supporting factor for processing data so that it can produce a valid information that can be useful for the company. However, in reality there are still many organizations or companies that do not have good data quality, this is caused by the many errors in the data input process and the different data formats for each application. To overcome these problems there is a data cleansing process that is useful for finding errors and cleaning data that is not in the format. This method can be done with a variety of paid or free applications, but because of the high cost of paid applications, many organizations use open source applications so that comparisons are made to determine open source based applications that suit the needs of the organization. In this study, the researcher conducted an analysis and design of a data cleansing architecture. It is expected that the results of the algorithms created will be implemented using the open source tool and carried out by comparison with other studies to see the differences in the results of the algorithms that have been made.*

*Keywords : data cleansing, open source, data quality, algorithm, data pattern, data null*