**DATA WAREHOUSE AUTOMATION**

**Presented By:**

Sushma Rani Reddy Aleti

Leela Krishna Kosaraju

Blake Picklo

**ABSTRACT**

Data warehouse automation (DWA) refers to the process of accelerating and automating the data warehouse development cycles while assuring quality and consistency. DWA works on the principle on design patterns. It comprises on central repository of design patterns which encapsulates architectural standards as well as best practices for data design, data management, data integration and data usage. DWA primarily focuses on automation of each and every step involved in the life cycle of a data warehouse and reduces efforts required in managing. DWA provides automation of entire life cycle of data warehouse from source system analyst to testing the documentation.

1. **INTRODUCTION**
   1. **What is Data Warehouse?**

Data warehouse is a large repository where huge amount of data is stored gathered from more than one source. Data warehouses are constantly evolving to support new technologies and business requirements. Data warehouse is a subject-oriented, integrated, time-variant, non-updatable collection of data used in support of management decision-making processes. For example, in business world, a report on current inventory information can include more than 12 joined conditions. This can quickly slow down the response time of the query and report. A data warehouse provides a new design which can help to reduce the response time and helps to enhance the performance of queries for reports and analytics.

* 1. **Overview of Data Warehouse development**

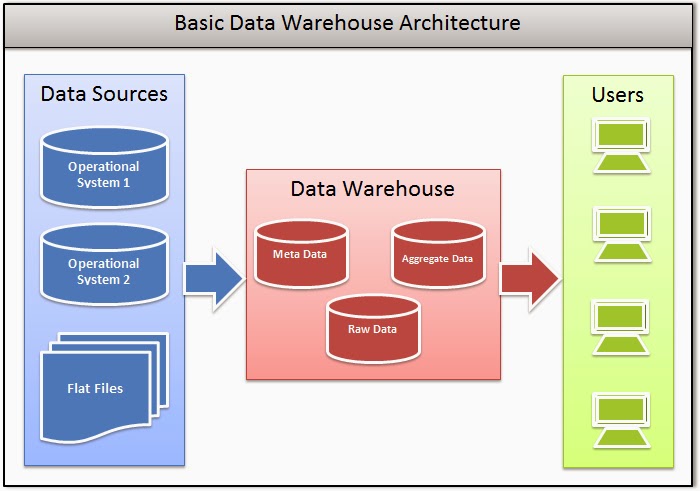


FIG 1.2.1 Architecture of Data Warehouse

From the Fig 1.2.1, data from different sources is extracted, and stored into temporary location. Here data is transformed and loaded into the data warehouse which can be accessible by the users all around the world. This process can be called as ETL/ELT (Extract, Load and Transform).

1. **DATA WAREHOUSE DEVELOPMENT**
   1. **problems with traditional data warehouse development**

Problems caused by traditional development includes:

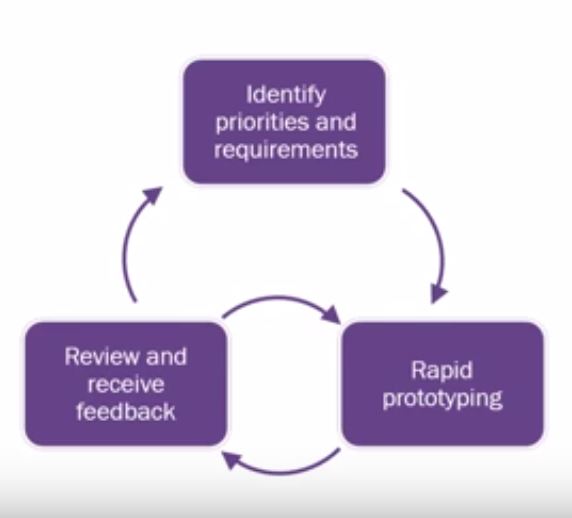
* Lot of manual coding.
* Time consuming.
* Due to manual coding, many mistakes are made, very difficult to find and fix errors.
* Very costly due to development resources and time taken to complete projects.

These problems can be overcomed by:

1. Using Evolutionary Analytics Development (EAD) and
2. Data Warehouse Automation (DWA)

Now let us discuss each of the above methods.

* 1. **Evolutionary Analytics Development**

****

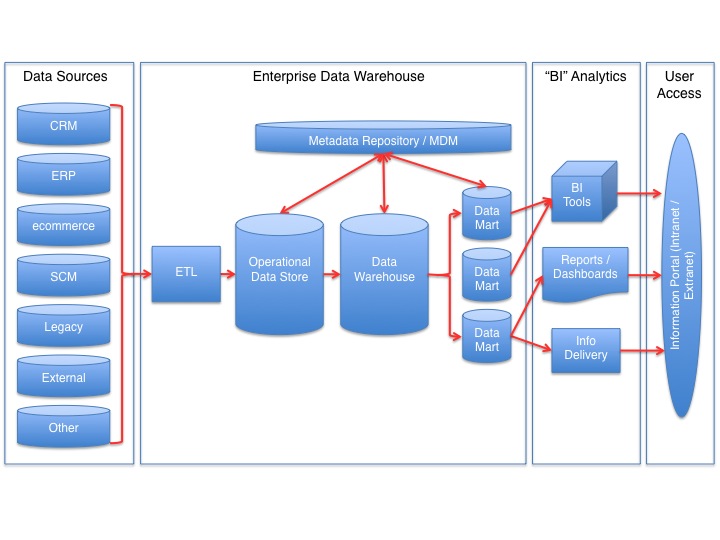
**Fig 2.2.1** EAD lifecycle.

In Fig 2.2.1, the life cycle of EAD is explained. First, we gather priorities and requirements of a product for developing the warehouse. Then, we set rules for developing for developing a product and implement them which is called as Rapid prototyping. After dispatching the developments, we gather review and feedback from clients and think about how to modify the previous version based on feedbacks. EAD aims to deliver functionality incrementally in short cycles. EAD systems delivers values in short time frames and have a tighter fit with organization’s decision making requirements. The problem with the EAD is that it is extremely difficult with traditional data warehouse development methods because manual coding takes long time.

So, we use Data Warehouse Automation to solve this problem.

* 1. **DATA WAREHOUSE AUTOMATION**

In DWA, Modelling is based on metadata. In this code development is not needed and no SQL knowledge is required. The code generated in DWA is 100% open for further modification. DWA improves productivity, reduces cost, and improves overall quality of Business Intelligence projects.



**FIG 2.3.1 DWA Lifecycle**

1. **Detailed Discussion**

To improve effectiveness in data warehousing process and gain effiency using technology, we use Data warehouse Automation. DWA is simply the automation of ETL development. The life cycle of data warehouse automation is planning, analysis, designing through development, operations, maintenance and management.

Data warehouse Automation has many benefits same as manufacturing which are:

* Increase in speed and quality of products.
* Less manual effort.
* Better quality and consistency.
* Improved controls and process optimization opportunities.

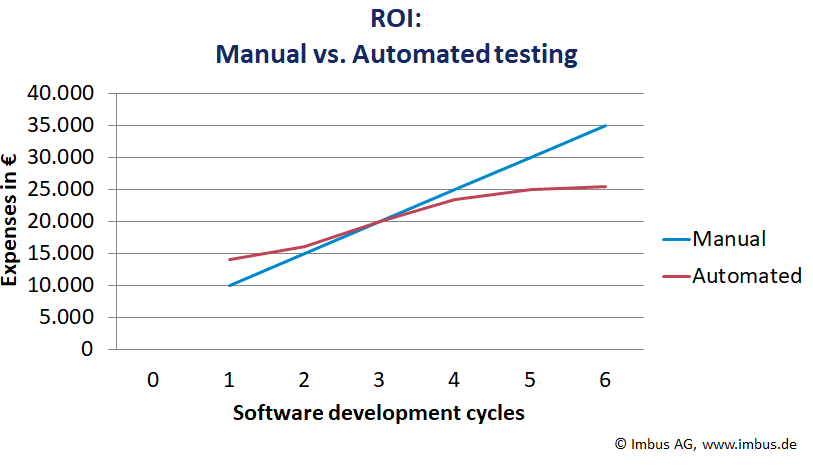
**ADVANTAGES OF AUTOMATION**:

* Warehouse data models.
* ETL generation
* Data management
* Manage deployment
* Scheduling
* Change in path analysis and easier maintenance
* Modifications of data warehouse.

To solve problems of Traditional data warehouse development and Evolutionary Analytics Development, we use DW Automation for following reasons

1. Rapid data warehouse delivery with agile / incremental changes.
2. Maintenance of database i.e. pull data from any source and deploy new functionalities rapidly.
3. Reduce development cost (meta-data driven).
4. Minimal coding (consistent high quality code).
5. Automated documentation.
6. Choice of data warehouse platform (eg:- Wherescape)
7. Export data to analytical tools
8. Significant ROI (Return On Investment).

The difference between manual and automated testing in ROI graph is as shown below



**Fig 3.1** [**ROI graph**](https://www.google.com/search?q=roi+full+form+in+automation+testing&rlz=1C1GCEB_enUS831US831&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjHj7yjv8bhAhUJKa0KHSemB9IQ_AUIECgD&biw=1366&bih=576#imgrc=2auEQILkle5L4M:)