Beginners Guide to ETL Testing

What is ELS Testing?

By http://softwareqatestings.com

These days more and more systems are moving from legacy to new technology, and ETL is one of the common methods used to help this transformation. It can consolidate the scattered data for any organization while working with different data format and sources. In this article, we will talk about the basic concept of ETL and how it has been tested.

Why we need ETL?

ETL stands for extract, transform, and load. In many organizations, the setup of the IT deferments happened long time back. So in most of the cases the way each deferment handles the data is different. For example, in a retail organization you many have different departments such as sales, marketing, logistics, etc. Each of them is handling the customer information but way they store that data could be quite different. The sales deferment store it by name where as marketing deferment has it in the number format. ETL can take all this data from different sources and transform it into a uniform presentation, such as for storing in a database or data warehouse.

Another challenge of the old IT infrastructure is the different data format used by different departments. So it might happen for HRD department somebody is using SAP where as the sales deferment you have Oracle Apps. So to take a business critical decision it is difficult for the higher management to get data from different platforms and consolidate them. ETL can do that job easily. It will take data from different sources and transform it to a uniform format and store it into DB tables. From the DB, you can generate the required reports.

How to do ETL Testing?

From the above points, clearly accuracy of the data after transformation is critical in ETL testing because if the data is not accurate, then the business decision will be wrong.

So following test types are commonly used in ETL testing:

No Data Losses: In this testing type, we first determine an N number of entities in the source system. Say the total number of employee is one such entity. If I have 2000 employees in the source system, then after ETL transformation and data store I should still get 2000 employees in the destination database. Only exception is when we have some business rule applied to the transformation. An example could be in the new system the business does not want to keep data for people who no longer works for the organization. So in that situation the destination system will discard the record of those employee and store number less than

2000.

This testing is done by mainly querying the source and designation database. Depending on the complexity of the system these SQL queries might be very complex and beyond the skill of a tester. However, running the query and checking the result is still the responsibility of the QA team.

Validation of Transformation rules: This is the second big validation point. In ETL testing, you will get the requirements in terms of transformation rules. One example could be the data format in the source was yyyy/dd/mm where as in the destination database this is mm/dd/yy. So you need to check them by taking some sample data. Here you apply all the good conventional test theories like boundary value analysis, equivalent partition, etc. All the database field level and record level data integrity must be tested here.

So you need to select the test data in such a way that at the end all the fields are touched into the destination database after transformation.

Business processes testing: Once the transformation is done these data feeds will go to many consumer systems (sales, marketing, HRD deferment applications). You need to ensure by testing from those systems that they can receive the data in the preformed format. All the critical end to end business flows must be tested from the GUI.

Article source: http://softwareqatestings.com