U.S. Student Aid Data Warehouse Evaluation

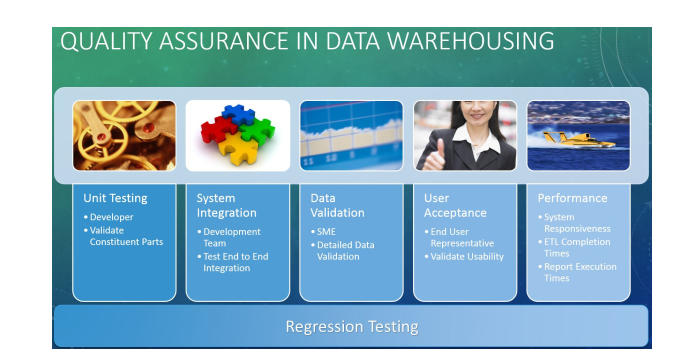
**SHANE ẼIRE BYRNE**

**DAT390**

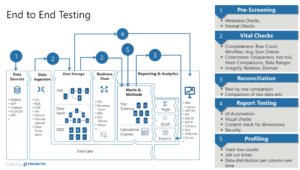
**Addressing data incompleteness in a Data Warehouse**

How does one address incompleteness of data within a Datawarehouse? As stated in Mr. Risk’s article, Overview of Data Quality Assurance in Data Warehousing**, “Nothing casts doubt on a data warehouse quicker than incorrectly reporting information.  It is critical that data warehousing projects do everything in their power to mitigate this risk.”1**

**This can be addressed through unit testing (a process of validating each of the constituent parts of a solution), System Integration Testing (sections of the overall finish product are tested as individually as they are integrated into one another), Data Validation (this can be done using an ad hoc query tool to retrieve data in a format similar to existing operational reports), User Acceptance Testing (testing to ensure the desired data is being given to the user), Performance Testing (which validates the performance of the database under real world conditions that meet the acceptable level), and Regression Testing (this being the process of resting functionality of the database to ensure that future development has not damaged any constructs that had worked prior).**



**1**

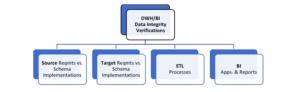
There is also Verifying Data Integrity in Schemas, ETL Processes, and BI reports as Mr. Yaddow goes over in his article, ”Ensuring Data Integrity in DWH/BI Systems with 3 Types of Testing”2. 

**2**

In order to adequately ensure integrity in Schemas the following steps must be taken:

1. Business and technical requirements for all source and target data have to be met.
2. Data integrity specification should be verified (this including DBMS, file systems, text files, and so on).
3. Data models for each implemented data schema.
4. Data mappings of source and target are loaded into their appropriate constructs.

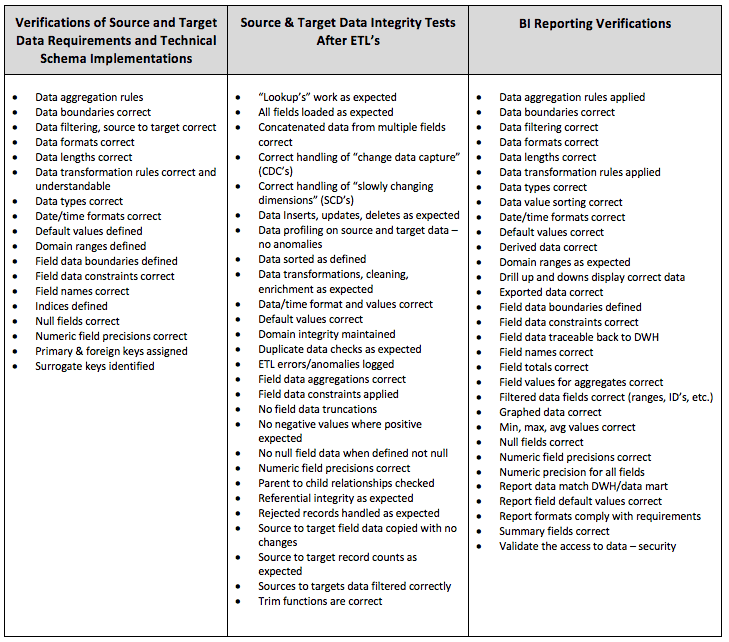
An ETL system should be able to extract data from source systems, enforce data quality and consistence standards, conform data to insure separate sources can be used congruently, and delivered data in a format allowing developers to build applications and enables end users to make decisions.



**2**

BI applications provide an interface that helps end users interact with the back-end control of a database. This involves what content users can view on information maps, what ranges are leveraged in which indicators derived from the data, and where indications exist between indicators is required to build a full suite of test cases.

**2**



**Addressing Nulls in a Datawarehouse**

It is prudent for a developer to know where a null value is coming from when it appears in a database. If it is known, it allows a developer to decide how to address it. One can leave it alone or hide it. If the origin of the null value is unknown one must address data quality.3 Ideally data should be captured correctly when placed within a database, but this is not always the case. It is possible for the sources data to be invalid for the field type or an improperly handled domain constraint. If this is not the case, one should then determine if it is a dimensional attribute. It is possible to have a null value with a Datawarehouse if one has a customer who is not a resident of a state nor a resident of an organizations country of business, for example. Ergo it would be logical to have a null value in a data column asking for the State the customer presides in. The null value could also be derived from a foreign key. Any foreign key should be assigned a value to avoid the violation of referential integrity but if this does not happen it is possible for a null value to appear. In order to limit the possibility of a foreign key being null one can take steps to ensure this.

In MySQL, NULL can be address by the use of the keywords NOT NULL at the point of instantiation of the column containing the data. As data is passed to the column from its source the values are checked to ensure they are not a NULL value but in fact contain some form of data (String, Text, Int, Double, etc.).

One can also use the keyword CASE and assign it a variable. This variable will look at the data being passed to the column it represents. If one gives it the condition (variable IS NULL) THEN “ “, the value being passed in is recorded as an empty string if it is a NULL value initially. If “ “ is replaced with 0, then any number-based characters will be substituted for null.

It is also possible to use the MySQL function LOWER() or LCASE() to convert String data to its lower case equivalent or UPPER()/UCASE() to convert String data to its upper case equivalent. This in principal helping to make data more readable to the user.

**References**

* Cohen, B., Pepi, N., & Mishra, N. (2017). Real SQL Queries: 50 Challanges (2nd ed.).
* Coronel, C., Morris, S. (20160126). Database Systems: Design, Implementation, & Management, 12th Edition [VitalSource Bookshelf version]. Retrieved from vbk://9781337509596
* Coronel, C., Morris, S. (20160126). Database Systems: Design, Implementation, & Management, 12th Edition [VitalSource Bookshelf version]. Retrieved from vbk://9781337509596
* DuBois, P. (2013). MySQL Developer's Libary (5th ed.). Upper Saddle River, NJ: Pearson Education, Inc..
* Elmasri, R., Navathe, S. B. (05/2017). Fundamentals of Database Systems, 7th Edition [VitalSource Bookshelf version]. Retrieved from vbk://9781323765883
* Krog Iverson, H. & Oates, J. (2016). How to Manage Null Values in Your Data Warehouse. Retrieved from <http://tdan.com/how-to-manage-null-values-in-your-data-warehouse/19450#>
* Murach, J. (2019). Murach's MySQL (3rd ed.). Fresno, CA: Mike Murach & Associates, Inc..
* Risk, S. (2013). OVERVIEW OF DATA QUALITY ASSURANCE IN DATA WAREHOUSING. Retrieved from <https://www.blue-granite.com/blog/overview-of-data-quality-assurance-in-data-warehousing>
* Short, T. (2016). SQL Beginner to Pro Guide . : DCW Web Trading LTD.
* SQL Quickstart Guide The Simplified Beginner's Guide to SQL . : ClydeBank Technology.
* Tahahhoghi, S. M. M., Williams, H. E., & Oram, A. (2009). *Learning MySQL* . Sebastopol, CA: O'Reilly Media, Inc..
* Tale, S. (2016). SQL The Ultimate Beginners Guide . : Steve Tale. (2016).
* Vanier, E., Shah, B., & Malepati, T. (2019). Advanced MySQL 8 . Birmingham, U.K.: Packt Publishing.
* Yaddow, W. (2018). Ensuring Data Integrity in DWH/BI Systems with 3 Types of Testing. Retrieved from <https://www.tricentis.com/blog/dwh-bi-testing-3types/>