ETL Testing in Informatica

Data Sheet | LumenData

PRODUCT REQUIREMENT

Informatica (Any ETL Tool)

What is ETL?

- **Extract:** This is the first step of the process. Here, data is collected from different sources such as flat files, RDBMS, etc. and stored while the other two processes get carried out.

- **Transform:** Data undergoes procedures to make sure that it is in conformity with the required use cases, and it makes sure to make the data reliable, accurate and complete before the final step.

- **Load:** This is the last step that moves the transformed data from source to target systems like data warehouses or any other downstream systems. Once the data has been loaded, the entire process is considered complete.

LUMENDATA
Masters in Data Management

What is ETL Testing in Informatica?

ETL stands for Extract, Transform and Load as discussed above. In Informatica, it involves the movement of data from the source to the target, and its testing involves testing the processes which take part in this entire procedure within an Informatica ETL (Extract, Transform, Load) system. ETL testing within Informatica ensures that the data we have at the end is loaded from source to target properly and is reliable as well as accurate and as per the requirement.

Here's how ETL testing is typically performed in Informatica:

1. Requirement Analysis:
First, we must recognize the business requirements of what we are doing, as well as the data and transformations and mappings that need to be used.

2. Test Strategy and Planning:
After that, the next thing we do would be to come up with a strategy for ETL Testing, and a test plan with test cases, expected results and actual results for the whole thing.

Test Case	Test Case Description	Step	Steps Description	Expected Result	Actual Result
Test Case Number 20	Validate Data Quality Rule: Job Family	1	Login to Customer 360 application	Login Successful	Pending
		2	Search the Person record	Person record is displayed	Pending
		3	Validate the Job Family is loaded in regards to the reference lookup value defined	Job Family is populated.	Pending
		4	Validate the Job Family is part of reference data lookup up	Record is loaded to Error Log File with error message: "For for input value is invalid"	Pending

3. Test Data Preparation:
Next, we have to design test data that shall cover all the required data types and be our guide for how we carry the testing process out.

4. Integration Testing: The interactions between different components of the ETL process will have to be validated. We should ensure data flows smoothly through the entire process and everything is applied accurately.

5. Data Quality Testing: We have to verify the completeness and accuracy of the data we have, then take care of any errors or duplicate values or such.

6. Regression Testing: After changes are made, we have to re-run tests to see if any existing features have not been affected.

7. Error Handling and Logging: During the entire process, we should record errors in the appropriate manner when they pop up for reference.

	A	B	C
	Defect No.	**Issue Description**	**Assigned to**
	Defect Number 1	The new feature has not been added.	Development Team

8. Compatibility Testing: Test if the ETL processes are compatible with different databases as well as technologies.

9. Data Reconciliation: Compare the source and target data to see its loaded accurately without any defects.

10. Validation of Business Rules: Validate that the defined business rules and transformations are applied to the data during the ETL process.

11. User Acceptance Testing (UAT): Let the business users perform testing as well to make sure the data meets their requirements.

12. Documentation and Reporting: Document the test cases, as well as errors that are found.

13. Automated Testing: Set up the tests to run on automation, so multiple test cases can be run at once.

14. Load Testing: Load testing is validating a process' ability to manage various data loads. It is an assessment of performance, scalability, and reliability. This testing analyses metrics and optimizes as needed, thus ensuring data can be handled efficiently by process and identity issues that pop up before its deployed.

ETL Testing is important as it helps to check the reliability, accuracy and completeness of the data throughout the entire process.

Validations done during the "Extract" phase:

During the "Extract" phase of ETL (Extract, Transform, Load) Testing in Informatica, the main focus is to extract data from source systems into a landing area and then to a staging area. In this phase, we verify that the extracted data is correct, and in the required format. Here are some validations required for this phase-

1. Data Completeness: First, we must verify that all of the data has been extracted from source records, so we have everything required.

2. Data Accuracy: Then, we should verify that the extracted data matches the data in the source system and is consistent with it.

3. Data Consistency: We should ensure that data relationships and hierarchies are maintained as well.

4. Data Integrity: It should be checked that primary key and any unique constraints are also maintained during this phase.

5. Data Filtering: Validate that the filter conditions are used correctly and so only the data that falls under the criteria of the filter conditions is filtered out.

6. Data Standardization: Check the use of data standardization rules to ensure they have been applied to the data correctly.- Put this in 'Extract'.

7. Data Format and Data Type: We should ensure that extracted data is in conformity with the expected format, so anything, including strings and date formats, should be checked.

S.No	first_name	last_name	email	gender	date	phone_number	Unique_ID	Data Size	Prefix	Suffix	Birthplace	Date_of_B	Marital_St	Employee ID	Account N	Contact Value
1	Vinita	Malone	vmalone0@yolasite.com	Female	5/27/2023	582-262-0721	3fdj3920	100 GB	Ms.	III	Hukou	4/7/1990	Divorced	981340911	VMaln	941592192
2	Kile	Ivchenko	kivchenko1@e-recht24.d	Male	10/3/2022	164-568-2335	tQD4YUZp	100 GB	Mr.	Sr	Amieirinh	########	Unmarriec	976512345	IVchk	941501371
3	Divyam	Singh	dsingh@gmail.com		1/4/2023	571-773-0736	4j6ySBk0	10 KB						914532789	Sgh	638633367
4	Odette	Scogin	oscogin3@microsoft.com	Female	8/13/2022	802-910-0513	wAdPr4GF	10 GB	Mrs.	III	Najin	########	Unmarriec	938916700	Scg	876548901
3	Divyam	Singh	dsingh@gmail.com		1/4/2023	571-773-0736	4j6ySBk0	10 KB						914532789	Sgh	638633367
6	Ama	Boadu	aboadu@microsoft.com	Female	2/6/2023	807-470-0378	3osD7z7u	1 KB	Mrs.	Sr	Yangcheng	########	Single	938720000	Abodu	212610041
7	Katherine	Femrite	femritek@fda.gov	Female	2/1/2023	370-169-0507	zXBLV81g	10 KB	Mrs.	IV	Ash Shary	########	Married	978901000	Femritek	190180123
8	Asmita	Satapathy	asatapath@livejournal.co	Female	10/11/2022	159-304-2380	3cnO9w1F	1 GB	Mrs.	Sr	Puerto Na	########	Divorced	967128900	Satasmita	807891456
9	Rob	Comport	rcomport8@psu.edu	Male	7-18-2022	216-894-4501	sxfqI5gb	100 GB	Dr.	II	Rudolfov	3/7/1990	Single			

8. Data Reconciliation: The extracted data should be compared to source data to make sure there are no discrepancies between them.

9. Error Handling: Record any errors that pop up in certain scenarios and report them to the development team as required.

10. Data Volume: Test the extraction process with different volumes of data to ensure it performs well in each case and there is no degradation.

11. Data Security: We should check that any data security measures, such as encryptions, are applied properly.

12. Metadata Validation: We should ensure that metadata, such as data lineage, is correctly captured too.

13. Parameterization: Test any parameters applied to ensure dynamic extraction with their use.

14. Data Profiling: Perform data profiling to analyse the extracted data, including its quality, to discover any risks or data quality issues.

15. Data Extraction Performance: We should assess the time it took for the data extraction process and measure the time for each source.

16. Full Load and Delta Load: In the case of Full Load, the entire data from source systems is extracted regardless of previous runs. In the case of Delta load, only new or modified data since the previous run is extracted after the new run. The former is useful for initial data migration and the latter for updating a target system with minimal data transfer.

Overall, the "Extract" phase of ETL Testing in Informatica is very important for making sure of the quality, completeness, reliability and accuracy of data that is extracted from all of the source systems. The validations that are performed in this phase help identity issues early and ensure data's reliability.

Validations done during the "Transform" phase:

During the "Transform" phase of ETL (Extract, Transform, Load) Testing in Informatica, the goal is to validate the transformations that have been applied to the data which is present in the staging area. Here, we make sure that the transformations applied to the extracted data are valid, and in conformity with business rules and requirements. Below are the validations done in this phase-

1. Metadata Validation: Any associated metadata, such as data lineage, should be validated in this validation.

2. Data Transformation Logic: We should check that the transformation logic mentioned in the mappings is applied correctly, while any calculations and aggregations are performed accurately. Various types of transformations come under this category-

- **Data Aggregation**: The data aggregation processes need to be validated to ensure the final data is accurate.
- **Data Joining and Merging**: The correctness of any joined or merged data should be validated, so as to ensure it is correct, and in conformity with the rules.
- **Data Splitting**: If there is data splitting, check that it has been performed accurately.
- **Data Derivation**: Data derived from existing values or calculated should also be validated,
- **Data Conversion**: If data conversions have occurred, ensure that they have happened correctly and any truncations due to this have not taken place.
- **Boundary Value Analysis:** This involves identifying boundaries of valid input ranges for the data fields, before selecting and running test cases through the data transformation logic to see the rules are correctly applied to the boundary values, and then documenting the errors which are caught by focusing on values just inside and outside the boundaries.

3. Data Formatting: Any formatting of data, examples being date formatting and string manipulations, should be implemented correctly.

4. Data Enrichment: If any new data elements have been added during transformation, it must be checked that they conform to business rules and expectations.

5. Data Cleansing: We should validate activities related to data cleansing, such as duplicate checks.

6. Data Validation: Data validation rules have to be applied to the transformed data, and any violations or discrepancies need to be checked.

7. Data Quality Rules: Test the implementation of data quality rules, such as data validation, data profiling, and data enrichment, to maintain the standards of data quality.

8. Handling Null and Default Values: Check that any null or default values are handled appropriately as per business requirements, and default values are assigned.

9. Data Transformation Performance: Evaluate how the data transformations perform to ensure they are up to the mark.

10. Data Reconciliation: The transformed data should be reconciled with the source data to ensure the transformations did not introduce any discrepancies.

11. Complex Transformations: Complex transformations using conditional logic and detailed expressions should be validated as well.

12. Error Handling: Test error handling scenarios during transformation, ensuring that error messages are logged and handled correctly.

13. Data Lineage: The information about the data lineage should be correctly recorded, including its various transformations.

14. Parameterization: Any parameters used in transformation logic should be checked as well.

Overall, these validations during the “Transform” phase help in ensuring that transformations have been used upon the data correctly, and the final data that we get after these transformations does not have any discrepancies with the original data. This also helps in checking that this data conforms to business rules and requirements and is now prepared to be loaded to the target system.

Validations done during the “Load” phase:

During the "Load" phase of ETL (Extract, Transform, Load) Testing in Informatica, our final goal is to validate the procedure of loading the extracted and transformed data into the target systems. This phase makes sure that the data is loaded correctly into the target databases and data warehouses, while being in conformity with the business rules. Here are some important validations to perform in this phase:

1. Metadata Validation: Metadata associated with loaded data should be validated, such as data lineage.

2. Data Completeness: Verify that all the transformed data has been loaded successfully into target systems, and make sure the number of records matches the expected count.

3. Data Accuracy: Check that there are no discrepancies, loss of data or truncations that have occurred during data loading.

4. Data Consistency: Ensure that data relationships and hierarchies are maintained correctly during loading. Check for any inconsistencies or missing data in the target systems.

5. Data Integrity: Check that the primary key and any unique constraints are maintained as well during the data load.

6. Data Loading Order: Test the order in which data is loaded, especially when there are dependencies between tables or entities.

7. Data Validation: Use data validation checks on loaded data to ensure its accuracy and conformity to business rules.

8. Data Reconciliation: Reconcile the loaded and transformed data to make sure there are no discrepancies between them.

9. Error Handling: If errors occur during loading, make sure to log them and report them to the development team as required.

A	B	C	D	E	F
Defect No.	Issue Description	Assigned to	Dev Comment	Status	QA Comment
Defect Number 1	The new feature has not been added.	Development Team	The feature has now been added.	Closed	Closing defect based on QA comment.

10. Data Integrity Constraints: Validate that any data integrity constraints, such as foreign key relationships, are maintained and 10. enforced during loading.

11. Data Duplication: If there is duplicate data, make sure the duplicates are removed and logged.

Morgan	Morgan
Wisedale	Wisedale
Morgan Wisedale	Morgan Wisedale
Prof.	Prof.
III	III
Physical Therapy Assistant	Physical Therapy Assistant

12. Data Aggregation and Summarization: If data has been aggregated and summarized, ensure its accuracy, as well as that of the aggregation process.

13. Data Lineage: Make sure that data lineage information is maintained during the loading process.

14. Data Loading Performance: Assess the performance of data loading processes to ensure they meet performance requirements and do not impact system performance negatively.

15. Data Distribution: Check that data has been distributed evenly and accurate across all of the target systems.

16. Data Consistency Across Loads: Make sure that loaded data is in conformity with data loaded during previous runs.

17. Parameterization: Test parameters used during the loading process to ensure dynamic behaviour.

18. Validation of Slowly Changing Dimensions: Slowly Changing Dimensions are used to capture changes to dimension data over a period of time. ETL Testing for Slowly Changing Dimensions involves validation of the data transformation and loading procedure handling these changes properly. Validations include-

(1) Type 1 - Overwrite:

a. Make sure that existing records are updated with new values as per the requirements.

b. Validate that the fields that are being updated in the dimension table reflect the most recent data from the source.

c. Check that no historical data is retained, and the dimension data only reflects the latest changes.

(2) Type 2 - Historical Tracking:

a. Make sure existing data records are not overwritten and new records are inserted for changed dimension data.

b. Validate the addition of effective date ranges to track historical changes.

c. Ensure that queries against the dimension table return the correct data for specific time periods.

(3) Type 3 - Previous Version Storage:

a. Check that a separate column is used to store any previous versions of attributes.

b. Validate that only certain attributes are updated, while others retain their original values.

c. Verify that queries return both the current and previous versions of the attributes.

(4) Type 4 - Outrigger:

a. Create separate tables to store historical attributes.

b. Ensure that the relationship between the main dimension table and the outrigger table is maintained accurately.

(5) Type 6 - Hybrid:

a. Confirm that both historical tracking and previous version storage techniques are applied as per the design.

b. Validate the behaviour of queries against both the main dimension table and the historical attributes.

During ETL testing for Slowly Changing Dimensions, the goal of these checks is to ensure that the transformation and loading processes maintain accurate and historical dimension data in the data warehouse which will help in data analysis for end users or business users.

With these validations that are done during the "Load" phase of ETL Testing in Informatica, we make sure the data that was extracted and transformed before has been loaded correctly into the target systems, and has no discrepancies, while conforming to the business rules.

ABOUT LUMENDATA:

LumenData is a leading provider of Enterprise Data Management, Cloud & Analytics solutions. We help businesses navigate their data visualization and analytics anxieties and enable them to accelerate their innovation journeys. Founded in 2008, with locations in multiple countries, LumenData is privileged to serve over 100 leading companies, including KwikTrip, Versant Health, US Food & Drug Administration, US Department of Labor, Cummins Engine, BCG, and others. LumenData is SOC2 certified and has instituted extensive controls to protect client data, including adherence to GDPR and CCPA regulations.

Get in touch to discuss how we can facilitate data-driven transformation for your organization.

MEET OUR AUTHORS

Amrit Roy
Associate Consultant

Jayachandra B
QA Lead

Durga Prasad
ETL Tester