

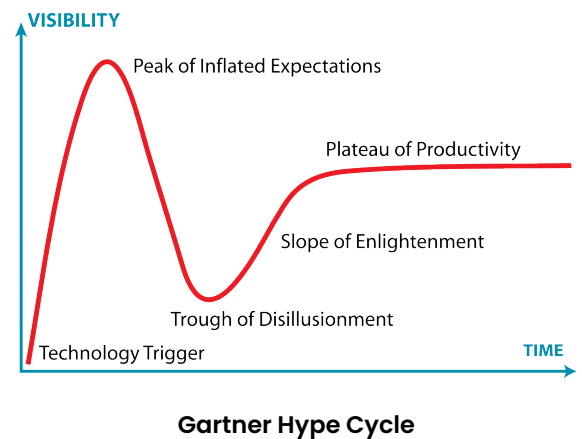
The Cost of Data Migration to the cloud

We've decided on a new Cloud based Data Warehouse. Now what?

We've recently seen an uptick in data validation inquiries with Cloud Data Migration projects. Obviously, this is driven by the undeniable success and cost savings associated with moving to the Cloud. We are well past the Hype Cycle of the Trough of Disillusionment as described by Gartner. We believe that we are well down the path of the Plateau of Productivity. Some believe we never went through the Trough of Disillusionment.

When you pass through the Slope of Enlightenment, the challenges are surfaced more clearly. As we are a data and BI validation company, our clients that are migrating to the cloud ask three things.

- How do I validate that the data is migrated correctly?
- How will this affect my ETL processes?
- How do I check that my analytics tools function correctly in the new environment?



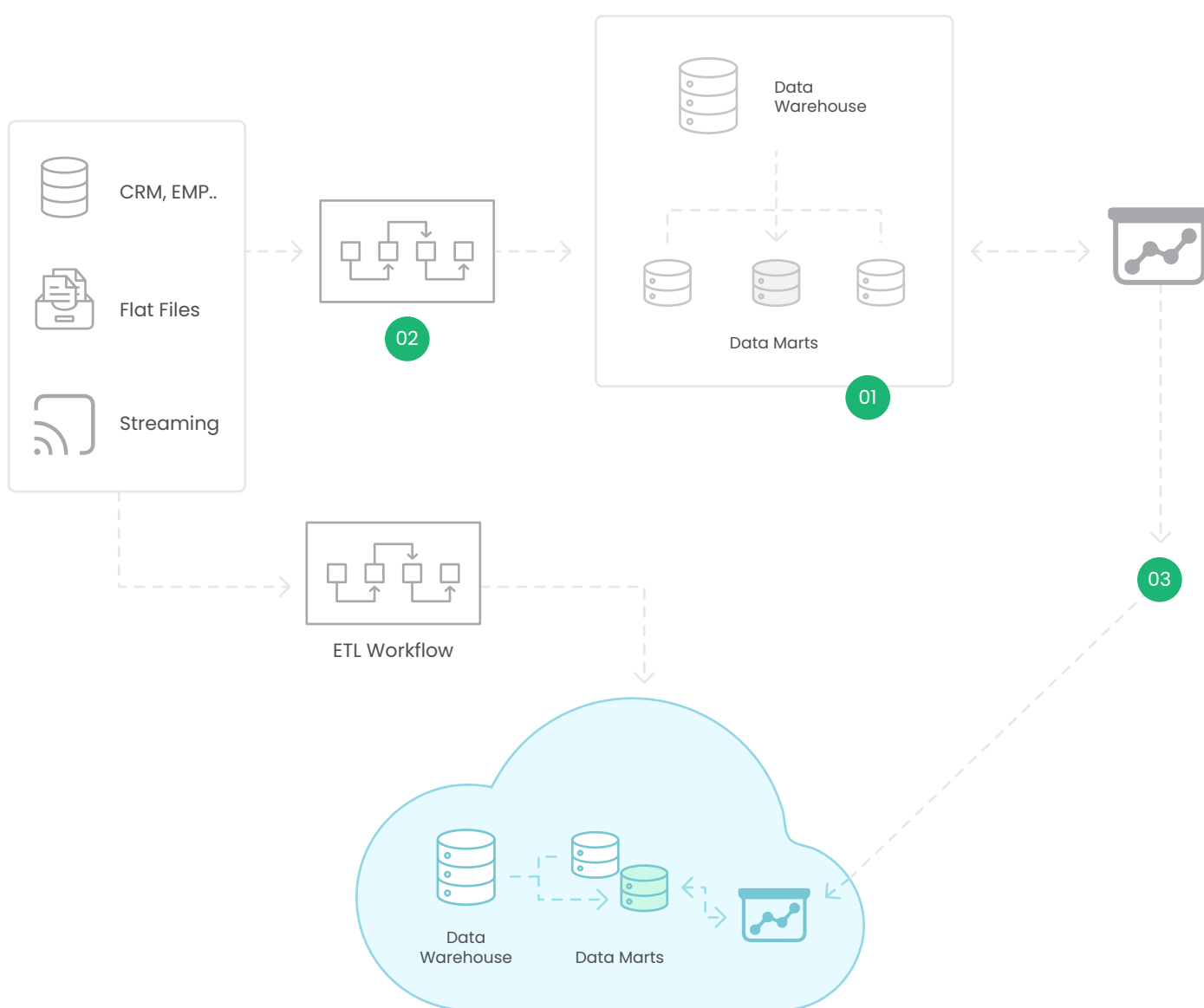
Challenges in Migrating to the modern Data Stack

- 01 Implementing the Modern Data Warehouse:** Moving to the Modern Data Warehouse typically takes the shape of fork lifting the current data in the on-premise Data Warehouse and migrating that data with minimal changes into the modern data stack. Often this movement of the data results in mistakes due to a variety of reasons, not the least of which is the data formats in the new warehouse aren't completely compatible with the old. The new Data Warehouse and the old need to be validated by comparing 100% of the data elements in each. Not doing so can lead to mistrust and loss of credibility.
- 02 Incremental ETL Testing:** The ETL processes that once feed the on-premise Data Warehouse incrementally now have to feed the cloud data warehouse generally in a different data store such as Snowflake/Databricks or Redshift. The change in datastore typically, drives a shift in the ETL tool and/or the incremental ETL processes. So now, there are two ETL processes.



One is the old ETL process, and the other is what feeds the new Cloud Data Warehouse. Often, a change is made to swap out the ETL tool to take advantage of some of the differences in the new modern data stacks. Sometimes, this takes the shape of the new process as an ETL process to transform the data once it is loaded. Again, these results need to be validated together to determine if the new techniques replicate all of the data transformations required to support the data analytics to follow.

03 Impact on Data Analytics Finally, you need to point your analytics tools to the new Cloud Data Warehouse platform. Often multiple data analytics platforms are involved as well. In some cases, companies take this opportunity to consolidate analytics platforms, and most don't due to the enormous amount of additional development and testing required.



In each of these steps, data must be extensively validated.



Automating the data migration validation with Datagaps

- 01 **Bulk Migration Testing:** With the bulk migration of the data to the Cloud comes a variety of means for issues to be introduced. Truncations, precision, referential integrity are just a few of the fundamental problems that creep into the process. We provide a data migration test case builder that automatically creates migration source and target comparisons between the newly formed Cloud data structure and the older on-premise structure. This shaves months off the data migration project.
- 02 **Incremental ETL Testing:** In conjunction with getting the data moved successfully, most organizations run a parallel ETL replacement project to prepare the ETL processes to do the incremental loads in the new data warehouse. Whether the ETL tool is being replaced with a newer technology or not, data validation testing is still an essential requirement. The old and new ETL processes run in parallel for a period of time and need to be continuously tested to ensure the processes are working correctly. Much like Datagaps provides data migration testing for the Bulk migration test, we can produce these test cases in an automated fashion without the need to do extensive transformations, which save time and cost significantly.
- 03 **Analytics Testing:** When the data analytics platforms are redirected to the new cloud implementation, several tests need to be performed in parallel.
 - a. Regression testing between the new and the old platform.
 - b. Metadata testing to make sure the schemas haven't been changed inadvertently and to validate the ones that need to be changed have been.
 - c. Data validation testing is needed between the old and the Cloud to ensure the data is correct in both instances.
 - d. Report to database query in both systems are needed to verify the expected results.
 - e. Finally, old pixel to new cloud app pixel and old dashboard data to new dashboard data comparisons at the GUI interface level should be performed to check the technologies are performing the same in each environment.
 - f. Performance and stress testing in the new environment.

Benefits of automating the Data Migration Testing

Reduction
in migration
testing time by

45–60%

Reduction
in data quality
testing time by

25–35%

Reduction in
QA Cost by

20–30%

Overall
reduction of
TCO by

20%