### Introduction

Many of Bigeye’s customers are cutting-edge tech companies that are building with the modern data stack. A core tenet of the modern data stack is using data warehouses not only for data storage, but also data transformation. While this has numerous advantages, such as making it possible for data analysts to write transformations in SQL, it also means that data is arriving in the data warehouse without having undergone cleanup. In this blog post, we explain why having a data observability solution is even more important when you are using ELT rather than ETL.

### ETL vs. ELT: What's the Difference?

In the world of data warehousing, ETL and ELT are two common approaches for managing data pipelines. ETL (extract, transform, load) involves pulling data from various sources, transforming it into a standardized format, and then loading it into a target data warehouse. On the other hand, ELT (extract, load, transform) flips the order of operations, loading raw data into the target warehouse first and then performing transformations as needed.

### Bigeye monitoring with ELT and ETL setups

Data observability solutions like Bigeye can work with both ETL and ELT setups.

**ETL setups**

With ETL setups, the transformation step happens first, so Bigeye will likely be monitoring the tables (post-transformation) as they land in the data warehouse. You can use Bigeye’s [Metadata Metrics](https://www.bigeye.com/blog/introducing-metadata-metrics-instant-data-observability-for-your-entire-data-warehouse) to track things like:

* Whether the data has landed
* Volume of data that has landed
* Freshness of data that has landed

Alternatively, Bigeye might be used to monitor a more upstream OLTP database like Postgres or MySQL, before the transformation step.

**ELT setups**

With ELT setups, on the other hand, since transformations are occurring in the data warehouse, Bigeye can help monitor at every step of the process.

With Bigeye’s [Metadata Metrics](https://www.bigeye.com/blog/introducing-metadata-metrics-instant-data-observability-for-your-entire-data-warehouse), you get out-of-the-box monitoring on all your tables for freshness and volume. This is a good way to monitor raw tables and ensure that they have actually landed.

Alternatively, you can use [Bigeye Deltas](https://docs.bigeye.com/docs/deltas) to create a Delta between the application DB and the snapshotted raw table in the data warehouse. Deltas tracks two tables and checks that the two are the same. This ensures that nothing in the “load” step of the data pipeline has gone wrong.

Finally, with Bigeye’s deeper columnar checks, you can monitor your most important tables, either raw or processed, in a more detailed fashion.

### Why it’s more important to have a data observability tool with ELT setups

With ELT setups, data transformation happens *inside* the data warehouse. This means that there is no cleanup of things like duplicate records and incorrect formatting in the data before it lands in the data warehouse.

In such a scenario, a data observability solution that monitors the raw bronze layer can go a long ways towards alerting you to pipeline breakages before they filter downstream, and reducing warehouse costs.

With modern data warehouses like Snowflake or Redshift, you are paying not only for storage, but also for compute - i.e. you are charged for every query and every transformation that you run. **By using a data observability solution to track raw tables as they land in the data warehouse, you reduce the number of transformations necessary, and thus your ELT expenditures.**

Finally, ideally, over time, data observability on your raw bronze layer of data should help inform your transformation models. If you are perennially getting lots of low-quality data with nulls, duplicates, strings lengths spanning outside what they should be, or UUIDs that aren’t what they should be, you may want to modify your transformation SQL to handle them.