

When data is processed through multiple layers in an ETL pipeline, it is important to make sure no record is lost in the processing and if there is a data discrepancy between the layers, it is critical to be automatically notified about the discrepancy so that the required / appropriate action can be taken after analysis. Here, we use DynamoDB streams / Lambda to achieve this automatic notification.

In my earlier post, we have seen how to build a robust ETL pipeline to make sure every record sent to us by the source system is taken up for processing at the earliest, using event-driven processing. This post is an extension of this event-driven processing.