Steps for developing a data transformation plan

So far, you've learned about the challenges of data transformation, and the importance of a data transformation plan. This plan should be designed to help you manage your resources and ensure data integrity, so that you can provide critical insights to your data team and your stakeholders. In your career as a data analyst, you'll likely design data transformation plans for your organization. In this reading, you'll explore the key components of a cloud data transformation plan to optimize data for analysis, and learn how to mitigate challenges through an example.

Identify a business need

The first step of designing your data transformation plan is to understand the business need you want to address. Starting with the end in mind will ensure that the plan you're designing is actually relevant to your organization's needs. It also ensures that your data is usable through transformation.

For example, a fast-food restaurant chain is considering changing its menu items in a specific region due to low sales in the last quarter. But, they want to be sure not to make a decision based solely on data from the largest city in the area, and still consider smaller cities in the region. So, the data analyst team needs to find a way to analyze regional data to determine the popularity of different menu items.

Explore the data

Once you've identified the business need your plan is addressing, you'll need to understand how the available data could be used to achieve your goals. To do this, spend time exploring your organization's data, and identifying what data you need that isn't available.

Returning to the fast-food example, think about the kind of data these restaurant locations might be generating. They might be generating a lot of daily sales data that includes what items are being purchased. But, the data is probably not separated out yet. As the team explores the data more, they find that it often has entry-errors since it's being entered as transaction data. This means they might have the data they need, but it isn't clean enough to be useful.

Define transformations

At this stage, you've identified the business need and determined what data you need. But, this data isn't ready to use yet because it still needs to be transformed. Now, you can define how the data should be when it is fully transformed and usable. This will help you understand what transformations need to happen along the way to get it there.

The data analyst team for the restaurant chain has identified that the data they need from the transaction systems needs to be cleaned. So, this will be part of the transformations they define. Also, the data doesn't currently contain one of the metrics they'd like to analyze: the popularity of different menu items. In order to make analysis easier, they also want to create calculations that can showcase how often customers are ordering particular items.

Choose tools and techniques

As a data cloud professional, you know that selecting the right tools and techniques for the job is important. For cloud systems in particular, making sure you're considering storage, cost, and available resources must be part of your transformation plan. This is how you will actually execute the plan. The restaurant chain analyst team is working with the transaction systems from the individual restaurants. This means they need to ingest the data they want from those systems, into the storage system they're using, to store and stage their data. From there, they need to decide how to use their organization's internal tools effectively.

Test the plan

Now that your plan is fully developed, you'll want to test it to ensure you're getting the results you need. Sometimes, it's difficult to know how something will work until you execute it. This is an opportunity to check for any potential errors, and ensure that the transformed data is usable for the business need you identified in the beginning. That way, you can review and rework your data transformation plan as needed, so that it delivers the data you need for your project.

Once the team sets up their pipeline and transformation systems, they do a test batch to find out if it's working. They discover that during the transformation, the data type of the sales data wasn't maintained from the transactional system. In order to ensure that error doesn't happen, they revisit their pipeline to address this issue before rolling out the plan.

Iterate

The restaurant chain is able to use the data coming from regional locations to monitor menu popularity to make decisions about menu changes. The data analyst team continues to monitor the transformation plan's performance. During the seasonal menu change, the chain's

stakeholders decide they also want to be able to analyze the performance of limited edition items. So, the data team makes some adjustments to the transformation plan to account for these metrics.

Key takeaways

A data transformation plan enables you to effectively manage your resources and ensure data integrity. A plan can also optimize your data, improve data quality, create better data governance, and make data more accessible. And understanding the components of a data transformation plan will help you design and implement your own plans.