

Strategy Governance Architecture Data Mgmt Quality Modeling Operations Metadata MDM  
BI Applications Big Data

Subscribe to The Data Administration Newsletter Today - it's FREE!

REGISTER HERE!

# Current State Analysis of Your Data – Part 2 – Data Freshness

Published: February 2, 2022 | 12:11 am · Vanessa Lam



This article is the second in a series taking a deep dive on how to do a Current State Analysis on your data. ([see first article here](#)) This article focusses on Data Freshness: what it is, why it's important, and what questions to ask to determine its current state.

The questions are organized by stakeholder group to facilitate usability; hopefully you can use this as a template to start your Current State Analysis journey. A few definitions before we begin – note that these groups are not mutually exclusive:

*People who Input Data:* These are people who collect and/or input data into the system. For example, sales people inputting their sales

numbers, or survey creators.

*People who Manipulate and Analyze Data:* These are people who organize the data and create analyses. This includes Data Engineers, Business Intelligence Professionals, and Data Analysts.

*People who Make Decisions based on Data:* These are the people who use the data to make decisions. This may be a sales manager deciding where to invest resources, a product manager understanding product use demographics, or an executive trying to cut costs.

## What is Data Freshness?

Data Freshness refers to whether data is available when you need it to be. This includes the cadence of data input and data refreshes as well as the reliability of processes and tools. Data Freshness issues may be

introduced at any point of the process: during data collection, at any point in the ETL or data pipeline process, or even at the display stage. If a report relies on salespeople to input their time data weekly and some people have not yet filled out their timesheets, this would be an example of data collection delays. Data pipeline or ETL delays can be introduced in many ways, including buggy code, tool or server failures, or other (often technical) issues. The display stage, such as BI tools or Excel worksheets, can cause issues if the underlying data has changed and the dashboard doesn't update, or occasionally because there are manual steps that need to be performed before data can be displayed.

Fixing Data Freshness is usually about ensuring the process is being correctly followed, both by people and in the technical implementation. It is often helpful to write out the entire data workflow, from all the data input sources to each tool and update process the data undergoes before it is displayed. This can help find bugs and make quality checking your data easier because you can check each stage and isolate where the data differs from expectation.

## Why is Data Freshness Important?

Data Freshness is one of the core elements for data trust. If people are unsure of the last time the data was updated, they will feel hesitant to use it and may not be interested in making decisions based on it or sharing it with their managers.

However, it is nearly impossible to always keep the data fresh. Data refreshes are expensive and require a lot of computing power, so "real-time" data is often unachievable. It is important to set realistic cadences for which the data will be updated. This means taking the technical limitations in to account while also understanding the business needs. Many organizations have a daily refresh where data is set to update overnight when the servers are not as busy. However, talking to the business about how often they need data is a good practice – you may find that daily updates are unnecessary and that they only use this report once per quarter when doing company updates. If that is the case, it is possible to reduce the number of updates and save money and computing power.

Even if you set your schedule optimally, it is not uncommon for something to go wrong and for data to be stale. In this case, it is important that there is some way to alert people that the data is not up to date. This could be another dashboard, an email, or an automatic alert. Alerting data users about data freshness is important because they will be frustrated if they make reports or decisions from stale data. Although it might seem like bad PR to message out data failures, it will actually build trust in the data team because stakeholders will feel like they are being kept in the loop and they will not unknowingly make decisions based on stale data.

## Questions to Determine Current State of Data Freshness

### *To Those Who Input Data*

These questions are designed to understand the process of data collection at the organization. It is important to understand how manual data is being input because if the data isn't making it into the system, no amount of technical savvy can recover it on the back end. Manual data collection is the backbone of many data organizations, so it is essential that data collection is designed to be as easy as possible, removing any roadblocks or redundancies. Additionally, it is important to understand if there is any data being kept or exchanged outside the system and why that may be happening. In an ideal world, all data is kept centrally, where everyone can access it, so if there are external data exchanges there may be a process or tool that requires reassessment.

- How often are you inputting data?
- Are you keeping any data in manual trackers, outside the typical data input process?
- Are you ever providing manual data to data users, bypassing the typical data input process?
- What is the hardest part about inputting data? What prevents you from inputting data regularly?

### ***To Those Who Manipulate and Analyze Data***

Technical Data Freshness issues will usually be centered around people in this group. The people who manipulate data are usually the ones who are acutely aware of stale data, so they may be able to more quickly pinpoint where the issues are arising. We are able to get a lot of knowledge about processes, tools, and where there may be large data freshness vulnerabilities.

- Is your data being input or refreshed in a timely manner?
- What is your ideal refresh schedule for the data? What is the current refresh schedule?
- What are the blockers for faster refreshes or more timely data?
- What processes and tools are currently in place to ETL data and create data pipelines?
  - Are they reliable?
  - Are they brittle? What input changes could occur that could break these processes?
- What tools (such as BI tools) are you using to deliver analyses?
  - What changes can occur so that the BI tools do not display the correct values (even if the data is correct)?

### ***To Those Who Make Decisions based on Data***

Data decision makers (or their analysts) are ultimately the people who will have specific data freshness requirements, so it is important to understand when they need data so that it can be refreshed on time. They may also feel the impact of stale data as a stakeholder, so they may be in a good position to explain its impact.

- How often do you need to make decisions with this data?
- Are there specific times (eg., End of Quarter, before a Board meeting) where it is essential to have the data up to date?

- Do you find yourself seeing the data and overriding a data-backed decision because you think the data is out of date?

## Conclusion

Data Freshness is essential for building trust and reliance on data. If people feel that it is unreliable and frequently out-of-date, they will stop using it to make decisions. While it often isn't possible for data to be real time, there are other ways to meet stakeholder expectations. Understand when they need updated data and understand which datasets need to be updated often and which ones can be on a more spaced-out schedule. Additionally, understand the data input schedule and assess to see if there are ways to streamline the process and increase frequency. If things go wrong, be sure to communicate so that stakeholders do not end up using stale data to do analyses or make decisions. Although data freshness is usually seen as a problem for technical people, it requires input and communication from everyone in the workflow to optimize.

This article is the second in a series discussing the important considerations when assessing your Current State of Data. Follow along for the next article about Data Culture – the way that people in the organization interact with data!

---

### ALSO ON TDAN

#### The Data-Centric Revolution: Avoiding ...

2 years ago · 1 comment

Gartner has put "Knowledge Graphs" at the peak of inflated expectations. If ...

#### Celebrating 25 Years of TDAN.com!

10 months ago · 1 comment

Twenty-five years ago today, I published the first issue of The Data Administration ...

#### All in the Data: The Same Difference of ...

9 months ago · 1 comment

What does your head do when someone uses the expression "same ...

#### All in Gove

2 years ago

I last published in a TI