Cheat Sheet: A Leader's Data Quality Metrics Guide

Quality Data Leaders Understand, Quality.

We've spoken with hundreds of data leaders to determine the data quality metrics used by the top teams.

Where applicable, we've also included industry standards or benchmarks developed from three sources: our annual data quality surveys, the aggregate data from our data observability platform, and our team's considerable experience in this industry.

This cheat sheet is designed to help data leaders:

- Incentivize positive behavior across their teams
- Obtain early indications of flagging system performance
- Better understand the levers they can pull to improve data quality

It's not an easy task. Data is bigger, faster, and messier than ever.

But, we are confident as an industry we will continue to establish the high reliability levels required for us to push the envelope on the most advanced, impactful data projects.

Data Downtime

Traditional methods of measuring data quality metrics are often time and resource-intensive, spanning several variables, from accuracy (a no-brainer) and completeness, to validity and timeliness (in data, there's no such thing as being fashionably late). But the good news is there's a better way to approach data quality metrics.

Data downtime—periods of time when your data is partial, erroneous, missing, or otherwise inaccurate—is an important data quality metric for any company striving to be data-driven. By measuring data downtime, this simple data quality KPI will help you determine the reliability of your data, giving you the confidence necessary to use it or lose it.

Data downtime can be helpful to measure in the aggregate as well as by domain, data product, and even table.

Quality Metric	Definition And Calculation	Industry Standard	Note From Monte Carlo
Number of incidents (N)	An instance of incorrect, missing, or incomplete data. Total incidents (N). Incidents per 1,000 tables per month.	6 per 1,000 tables per month. (37% schema; 23% volume; 28% freshness; 12% quality)	Organizations will only know the number of incidents they are catching. Ironically, organizations with more mature data quality processes may actually find and report a higher number of incidents even as they reduce the overall rate. *How Resident reduced their data incidents by 90%.*
Time to detection (TTD)	The average time from when incidents occur to when they are detected. The time to response–the time between an alert and investigation–can also be a helpful metric.	4 hours	Issues can go undetected for weeks, months, or even a quarter. Building ML data monitors or acquiring a data observability platform can reduce this dramatically. It's painful when data consumers detect issues first as data trust is quickly lost. For this reason, we've seen some data leaders track their first-to-know (FTK) rate. *How Vimeo caught previously unknown issues.*
Time to resolution (TTR)	The average time from when the issue has been detected to when it has been given a resolved status.	9 hours	Resolved doesn't necessarily mean fixed. Resolved issue statuses can also include expected, no action needed, or false alarm. Data incidents can occur as a result of system issues (ex. Permissions in Snowflake), code issues (bad LEFT JOIN), or data issues (a third-party sent you garbage). Without the right tool, this can make quick root cause analysis tricky and time intensive. *How BlaBlaCar reduced their time to resolution by 50%.*
Data Downtime	N x (TTD + TTR) It can be helpful to calculate data downtime in the aggregate as well as by specific data products.	793 hours/month*	Reducing data downtime improves engineer efficiency and mitigates the risk of a severe data incident with devastating consequences. It also increases data trust which can further unlock data product value. *How Choozle reduced data downtime 88%.*
Data Downtime	Our calculator and eBook are the best	~26% of revenue	There is a hard labor cost for the time spent fixing data quality issues, but also a cost that occurs when

Cost	resources for this.	**OR** ~$280K labor cost ~$1 million in efficiency cost*	operations become less efficient as a result of wrong or unavailable data. The second is more difficult to quantify, but is often the greater cost. Even in our conservative calculations that assume just a 10% efficiency drop for each hour of downtime, it can total to millions of dollars for many organizations.

**Assumes $50m company with 3,000 tables, a team of 5 data engineers, and medium data dependence.*

Data Reliability Engineering

Data reliability engineering metrics measure the effectiveness of two areas: level of coverage and team performance. These metrics are usually related. The better your automatic incident coverage, the less time your team typically spends on data quality related tasks (especially if there is RCA functionality).

The most effective data quality coverage will be both wide and deep- there should be separate metrics that measure both dimensions. Data leaders will also want to assess how much time their team is spending on data quality. For many teams, it's a shocking, almost scandalous number. The industry standards here are not ideal–teams should aspire to be better.

Quality Metric	Definition And Calculation	Industry Standard	Note From Monte Carlo
Status Updates	The total number of alerts generated divided by the number of alerts that have a resolved status (fixed, no action needed, expected, false alarm).	>3%	Each data quality incident SHOULD be documented and kick off a root cause analysis investigation. This rarely occurs within organizations that don't have formalized processes or tooling in place. The ideal for status updates should be 100%. Mercari reached this milestone by creating a data reliability engineering team with clear lines of accountability and ownership.
Table Coverage	The total number of production tables divided by the total number of tables being monitored (either with rule based tests or with ML monitors) for schema changes, data volume anomalies, or data freshness anomalies.	1-10%	Table coverage matters because data flows from one to the other. Only having coverage on the most downstream tables will delay time to detection and conceal issues. Additionally, time to resolution will be extended as tables rarely contain the context necessary to determine the root cause of their own downtime. It is difficult to manually write tests for every pipeline or opt-in every table into ML coverage. Automatic monitoring is the best way to scale data quality coverage. *How Clearcover increased their coverage by 70%.*
Tests or Custom Monitors Set	The number of manually defined tests or custom monitors set per 1,000 tables.	~3 for every 1,000 tables	This number can vary considerably from team to team. In my experience, teams with a lot of tests or custom monitors are either very immature (and relying on tests to catch everything) or very mature (and use them as a preventative layer). Custom monitors and data tests do play an important role in supplementing automatic monitoring. They are best reserved for your most important tables and to uphold your most stringent SLAs. See how Checkout.com's data team reduced their reliance on manual tests with custom monitors as code.

Time Spent on Data Quality	Total hours spent on issues pertaining to data quality such as maintaining tests, conducting root cause analysis, fixing broken pipelines, etc. divided by the total amount of hours available to work.	40%	This metric covers time spent on tedious, avoidable tasks such as writing and maintaining hundreds of data tests. We know of some data teams that have written hundreds of tests on one pipeline and still not catch every instance of bad data. The time spent on data quality can be lowered to anywhere from 5 to 20% with a data observability platform. How Prefect saved 50% of their data engineering time.

Data Satisfaction And Adoption

Ultimately, data quality is contextual. The data has to be accurate and available enough to be useful for the use case required.

No one will be able to assess that better than your business stakeholders. So ask them in a survey. You can use multiple metric types and methodologies (we're partial to good ol' NPS)--the important thing is you formally assess the sentiment of your "user base." This is a relatively new practice for data teams, so we don't have industry standards to provide other than noting the average NPS for a B2B IT services company is about 40.

If your data quality is high, you should see correspondingly strong scores across trust, value, ease of use, and satisfaction. You should also see improvement in your data adoption numbers as well. If not, re-examine your data quality metric methodology.

Metric	Definition And Calculation	A Note From Monte Carlo
Trust	NPS methodology: Ask data consumers, "On a scale of 1-10 how likely would you be to describe the data at this organization as accurate and reliable?" All responses with a 9 or 10 score count as +100. All responses with a 1-6 count as -100. Total and average.	This metric can show the impact of different data quality process and tooling adoption over time. Ideally you should see trust levels rise as you invest time and resources in data quality. JetBlue's data team specifically conducts a survey of their data consumers to measure their level of trust in the data they provide.
Value	NPS methodology (as illustrated above). Ask, "On a scale of 1-10 how likely would you be to describe the data you use in your regular work as essential to the performance of a business operation or toward optimizing performance?"	Understanding the value your consumers place on the data is especially helpful to measure the effectiveness of your analytical efforts. Machine learning models and experimentation programs will have hard returns that are easy to add up, but showing how investments in data quality "pay off" across dozens or hundreds of analytical data products can be difficult without a survey. Remember, data quality is contextual. You may think data quality is strong compared to a baseline, but if it's not reliable enough for your stakeholders to generate value then it's not clearing an important bar.
Ease of Use	NPS methodology. Ask , "On a scale of 1-10 how likely would you be to say I can easily obtain the data I need?"	A more expansive view of data quality considers the usability as well. Can data consumers get the information they need? Can your data power users (data scientists, analysts) navigate the data warehouse or is it a data swamp? How well documented are your assets?
Satisfaction	Average your trust, value, and ease of use Net Promoter scores for a topline satisfaction number.	It can be helpful to roll up value, ease of use, and trust into a topline metric like "data satisfaction" to report up to higher levels of the organization.
Adoption	The overall adoption from users internal to the organization as	It's difficult to treat your data-as-a-product if you aren't closely following one of the most important metrics to

	measured by monthly active users of dashboards, ML models, and other data products. The number of employees at your organization divided by the number of monthly active users.	product teams: user adoption. This metric is the cold hard reality that tells you if you've made something useful…or not. If you aren't seeing adoption, the dimensions of data trust, ease of use, and value should provide direction on how to improve your data products. If these scores are high, but data product adoption remains low, consider how to improve your communications (marketing) within the organization.

Data Health

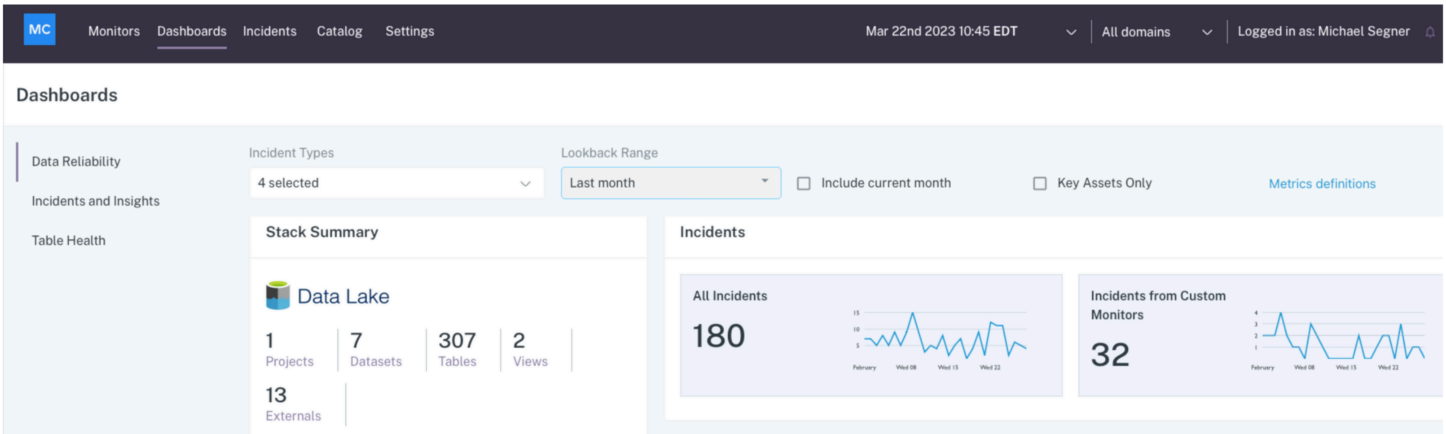
Data health metrics are those that measure the health of each layer of your data platform. Oftentimes there will be problematic "hotspots," or areas where issues pop up again and again. When you measure these failure points it can be a catalyst for doing the tough work of rebuilding a portion of your data platform rather than applying yet another band aid.

Quality Metric	Definition And Calculation	A Note From Monte Carlo
Data Consumption (BI/ML)		
SLA Adherence	Minutes the data product met SLA standards divided by the minutes it was in violation of that SLA. You won't achieve the 99.999% uptime of software applications, but the high 90s are a great level for which to aspire.	Some data teams will place their production tables or data products into three tiers with high, medium, and low SLA requirements. This categorization is often driven by use case and importance. For example, a financial analytical data product may require high precision but medium data freshness, where a machine learning model may be the opposite. Understanding your SLA adherence can help you prioritize where to invest your time and attention. See how Red Ventures leveraged data SLAs for data quality.
Dormant Dashboards	The total number of dashboards not accessed in the last 30 days. Not every dormant dashboard will be a candidate for deletion.	Unused dashboards are clutter, however many data professionals are hesitant to remove them. Our colleague Shane Murray recommends doing what he calls a "scream test." Unpublish the dashboard for a period of time and see if anyone inquires about it or makes a fuss. If not, then you have just kept your environment a bit more tidy for your data consumers.
Data Warehouse / Lake		
Table Importance	How important a table is to ongoing data operations based on its query activity and consumption. This metric can vary in sophistication, but the key is to rank tables against one another to help with prioritization.	Our data observability platform takes a more sophisticated (but automated) approach to calculating table importance which you can read more about here. The core calculation is based on: ● The number of reads ● The number of users (query executors) ● Degree of connectivity to other tables ● The update periodicity ● Age and freshness
Dormant Tables/Fields	The number of tables not	Dormant tables create clutter in your data warehouse and can introduce incidents when they are accidentally used by data

	read in 30 days. Number of fields that are not: ● Used in any downstream tables ● Used in any BI reports ● Read in a SELECT query	power users in the classic, “You’re using THAT table?!?” problem. Dr. Squatch’s VP of Data, IT & Security, Nick Johnson’s take: “The whole team is trying to build new things and I’m trying to prune them. I like being able to go into lineage and in a singular, unified view see if a table is useful.” Read the full story.
Deteriorating Queries	Number of queries displaying a consistent increase in runtime over 30 days.	This metric is helpful for containing costs by identifying queries that are not scaling well. It also identifies queries at risk of timing out, resulting in data downtime.
Documentation	Percentage of key tables with documentation covering ownership, SLAs, business logic, and incident history.	Without strong documentation, the only people that can build with, maintain, or fix data assets are those that built them. It's just not efficient. Your ability to ingest data will always be larger than the capacity of your team to document it. Prioritization is a must.
Orchestration and Transformation		
Jobs Failed	Total number of dbt model errors in the last 30 days. Total number of Airflow jobs failed in the last 30 days.	Human beings aren’t perfect and as such there will always be orchestration and transformation errors. And let’s face it, sometimes platforms just fail through no fault of our own. That doesn’t mean the data team can’t focus on lowering the number of failed jobs by leveraging best practices. If there are areas that continue to see a persistent increase in failed jobs, that could indicate a “hot spot” worthy of more time and attention. As you adopt more and increasingly sophisticated models and DAGS, there will naturally be more failed jobs. Moving from job failure totals to the rate of failures divided by total models or DAGS can be a helpful metric as well.
Documentation	Number of dbt models with documentation divided by total number of dbt models.	dbt makes it so easy to document your models that there is almost no excuse not to do it as a standard operating procedure. Documentation helps others understand key fields, metric definitions, and the why behind it all.
Ingestion		
Incidents On	Number of data incidents or	The ultimate tragedy of data engineering is that most of the time

Ingestion	anomalies from a specific third party source over 30 days.	there is very little control over the source system emitting the data. This is especially true when ingesting data from third-parties. If you are getting garbage data from the very start, no amount of transformation will turn it into high quality data. Measuring the anomalies and incidents that occur on ingestion can help organizations work with their third-party data providers to meet and maintain set data SLAs. Or if that type of relationship doesn't exist, consider alternative sources. Monte Carlo's [Ingestion Validation Monitor] can be used for exactly this purpose.

Next Steps And How To Get Started



One of the best places to start with your data quality management strategy is an inventory of your current (and ideally near future) data use cases. To understand the data quality needs and context it can be helpful to categorize them by:

- **Analytical**: Data is used primarily for decision making or evaluating the effectiveness of different business tactics via a BI dashboard.
- **Operational**: Data used directly in support of business operations in near-real time. This is typically steaming or microbatched data. Some use cases here could be accommodating customers as part of a support/service motion or an ecommerce machine learning algorithm that recommends, "other products you might like."
- **Customer facing**: Data that is surfaced within and adds value to the product offering or data that IS the product. This could be a reporting suite within a digital advertising platform for example.

Then assess the overall performance of your systems and team. At this stage you may have just begun your journey so it's unlikely you have detailed insights into every data quality metric we have covered in this cheat sheet. Do what you can and note your measurement gaps.

Categorizing your data use cases and baselining current performance will help you assess the gap between your current and desired future data quality state. Closing this gap with modern data quality solutions like a data observability platform will improve data engineering team efficiency, mitigate risk, and further unlock value from your data products.

Once you have a baseline and an informed opinion, you are ready to start building support for your data quality initiative. You will want to start by building your business case and understanding what pain is felt by different stakeholders. We have helped hundreds of data teams navigate this process and we can help you too.

Interested in learning more about how your peers are measuring and improving their data quality? Set up a time to talk with one of our experts at Monte Carlo about a complementary data quality assessment!