

SODA:

5 Best Practices for Building Trusted Data Products



Data products, the new competitive advantage.

Is data a differentiator? There's no question that data continues to transform business practices, as more companies scramble to adopt data collection, cloud, and machine learning capabilities to drive competitive advantage. Indeed, high-performing organizations today are not simply using data to support better decision making, they are building data products to discover new insights, automate business processes, and deliver extraordinary customer experiences.

This shift to thinking about data as a product is being driven in large part by the emergence of “data-native” companies. The ride-hailing and delivery industries are great examples. Using data from across all customer touchpoints, these companies have fully automated the traditional order-to-cash process, and in doing so, disrupted how these businesses traditionally operated. Data products – in this case, accurate delivery time predictions – define competitive advantage.

As your customers demand more from every digital experience, data product management will become increasingly important. So how can we, as data teams, build reliable, trusted data products and services?

In this article, we draw on our experience with over 250 leaders in data, data management, and analytics to take a closer look at how companies can design and build data products more efficiently and effectively.

Here are five best practices to consider:

01

Good data starts with good data engineering practices.

Although data is owned by the business, good data starts with good data engineering best-practices. Therefore it's important for data engineering teams to start introducing some concepts of software engineering like documentation, testing and version control in their day-to-day. This way, issues in our data products become more transparent and reproducible. At Soda, we address this through Sodalite, an opinionated open-source package that helps data engineers measure timeliness(arrival times), stability (schema changes) and completeness (row counts).

02

Data can only be good if you have strong data processes.

Whether data is generated by a system (e.g. events) or a person (e.g. sales data), the business will eventually change, resulting in structural changes in the data. Each change can impact the reliability and quality of your data. Good operations mean that you have transparency and automated controls. When things go wrong, you can react even before there's a material impact. At Soda, we're helping data engineers and data SMEs to formalize these controls in one place, going as deep as validity (valid values, ranges, reference data, ...), missing values and business rules. Alerts, workflows and tasks allows us to manage work and trigger integrations into issue tracking tools (Jira, ServiceNow, ...).

03

Data teams should be empowered to manage their data.

To manage your data operations you need a collection of people (roles), processes and systems. Not every member of your data team will be a seasoned engineer that can produce complex SQL to understand what's going on. To create a great data culture, you need to equip and empower everyone to take ownership and responsibility of data. That's why at Soda, we focus on self-service data management. Our user experience is tailor-made for analysts, engineers and SMEs so they can collaboratively monitor their data.

04

Data producers & consumers need to be closely aligned.

Data problems often occur because producers of a dataset (internal team, external vendor, ...) are not aware of all downstream consumptions of their data, nor are the consumers really sure about how that data is created in the first place! To align producers and consumers, we need to make the implicit assumptions explicit, validate those continuously, and visualize the results in dashboards that allow teams to be proactive. At Soda, data owners and data consumers have their own dashboards so they can see results, alerts and issues over time, as well as communicate their implicit assumptions to the data owners for better alignment.

05

Data reliability issues should be transparent to everyone.

Problems in your data value chain (or flows, lineage) will happen and data issues will inevitably have a material business impact. That's an unfortunate fact of life as we're collectively increasing data management maturity.

Therefore, it's important to build a data culture that's rooted in transparency and trust. When a problem occurs, we make it explicit and transparent so consumers can trust that data owners and SMEs as they remediate the problem. At Soda, we've built an intelligent alert routing system to reduce alert fatigue and create unprecedented data transparency.



Want to learn more?

Schedule a call with our solution engineers to get a full overview of Soda's capabilities and how these would fit into your environment.

hello@soda.io

(929) 920-1414

www.soda.io

[Let's Talk](#)