precisely

I have trust issues with my data

How can I improve my data quality?

Paul Rasmussen | Product Management

Julie Skeen | Product Marketing

Introductions

Paul Rasmussen
Principal Product Manager

Julie Skeen
Sr. Product Marketing Manager

precisely

The leader in data integrity

Our software, data enrichment products and strategic services deliver accuracy, consistency, and context in your data, powering confident decisions.

12,000
customers

99
of the Fortune 100

100
countries

2,500
employees

Brands you trust, trust us

amazon, RBS, New York Life, Kellogg's, Westpac, Barclays, ExxonMobil, Mercedes-Benz, VISA, Travelers, blue california, Fannie Mae, T-Mobile, Truist, Merck, Cigna, Costco Wholesale, Twitter, TD Bank, L'Occitane en Provence

Data leaders partner with us

databricks, snowflake, aws, Cloudera, Google Cloud, splunk>, servicenow, IBM, Confluent, Microsoft, neo4j, TomTom, here, SAP, OpenShift

Business initiatives are driven by data

83%

of CEOs want their organization to be more data-driven

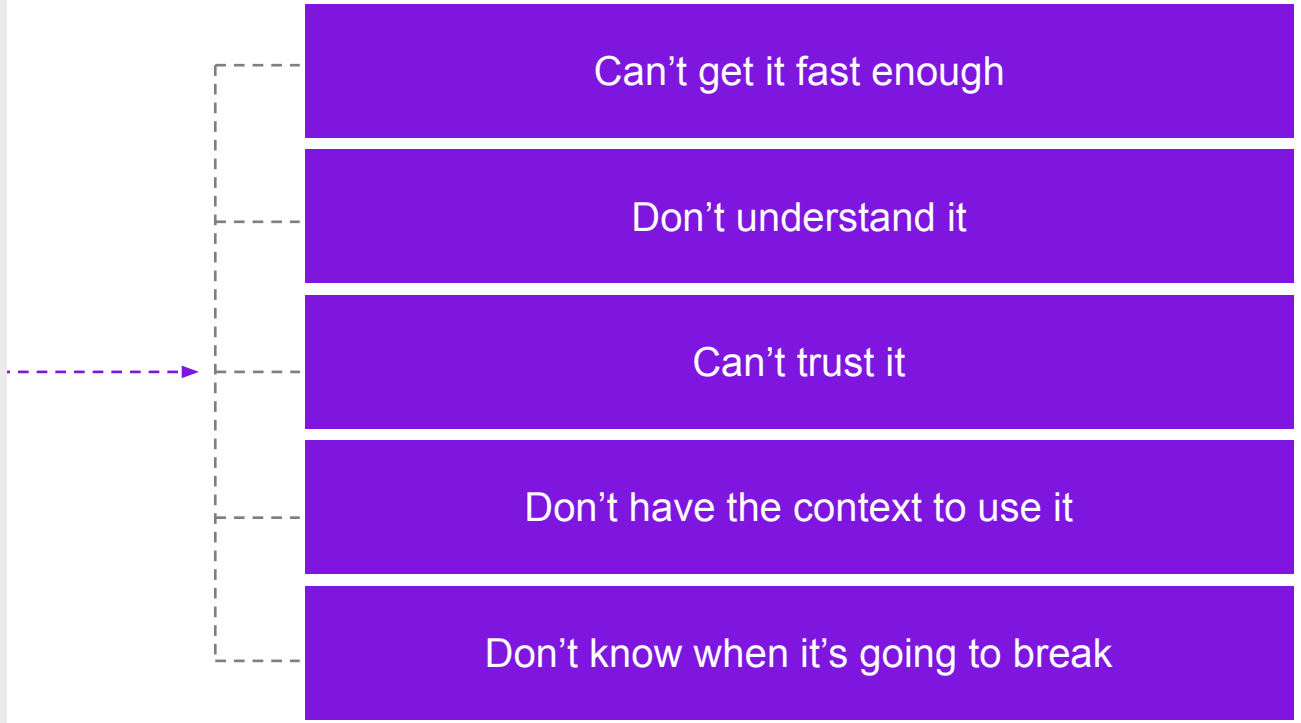
Source: IDC

But your team doesn’t trust your data

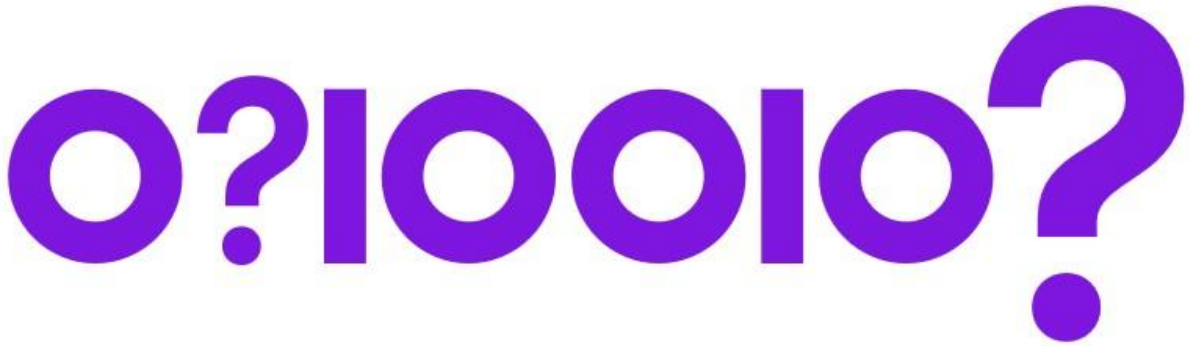
30% of data practitioners strongly believe their actions are driven by data analysis

27% of data practitioners completely trust their data

Source: IDC



The total cost of untrusted data



Common Costs

- Defects
- Bad decision(s)
- Data downtime
- Fines and Penalties
- Software/Data engineering time

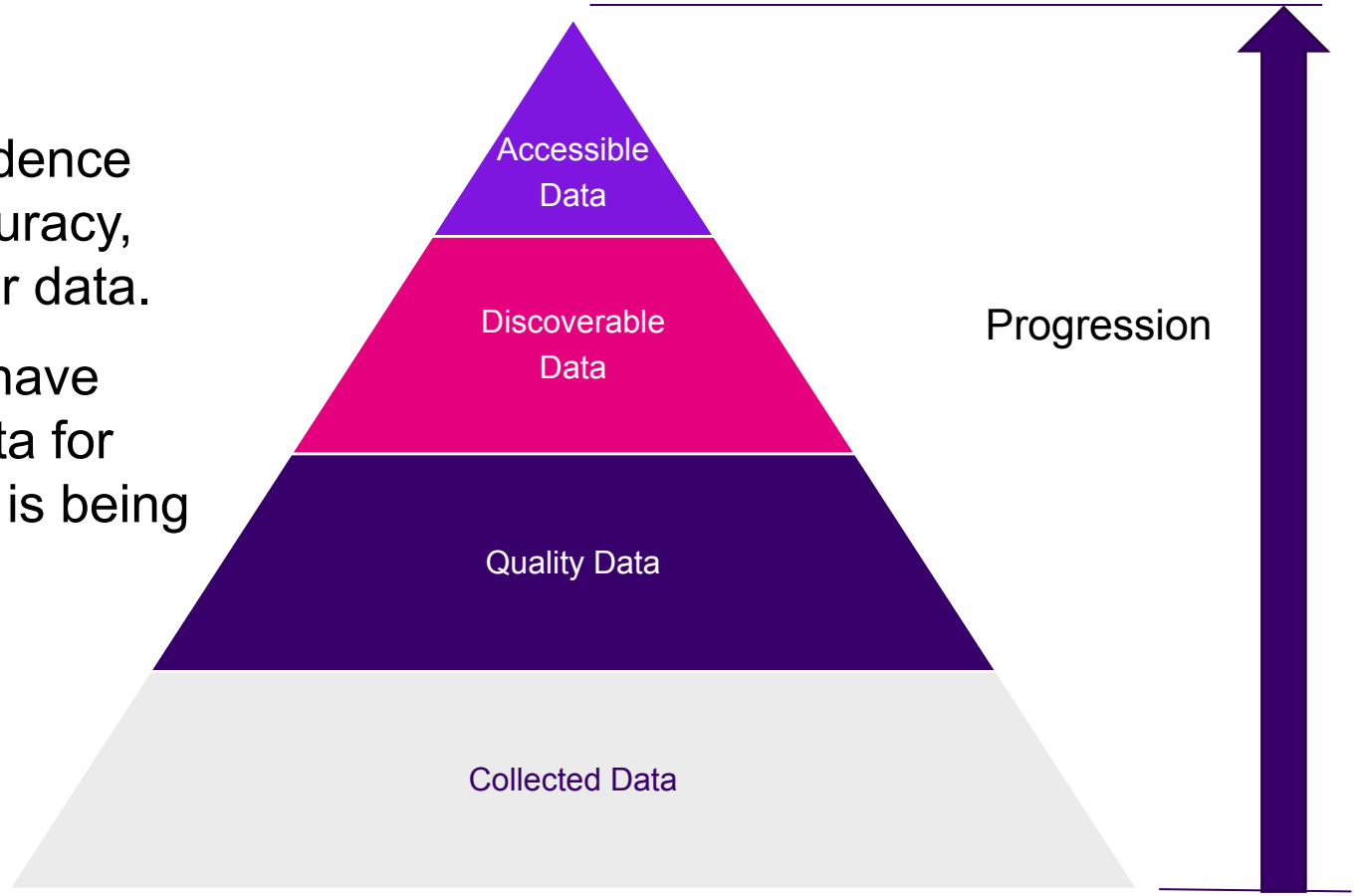
Hidden Costs

- Management time
- Decreased customer retention
- Lost sales opportunities
- Cost of Counsel

The path to trusted data

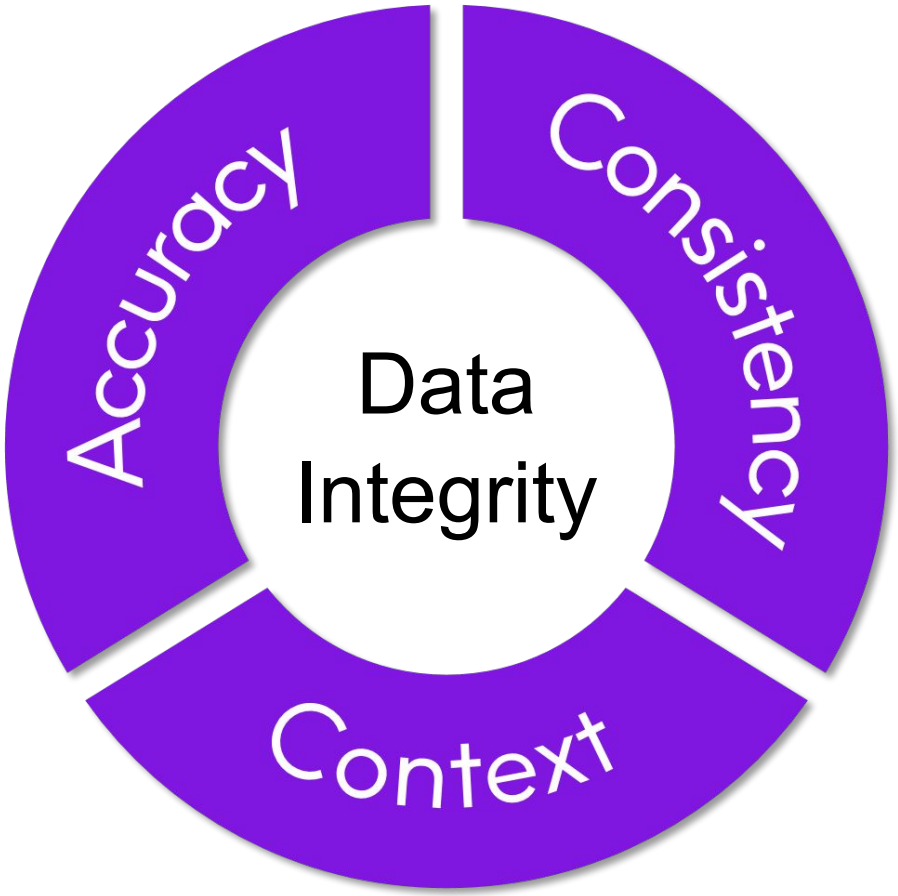
Trust in data refers to the confidence and reliability placed in the accuracy, integrity and authenticity of your data.

Trusting your data means you have confidence in the use of the data for the specific purpose to which it is being applied.



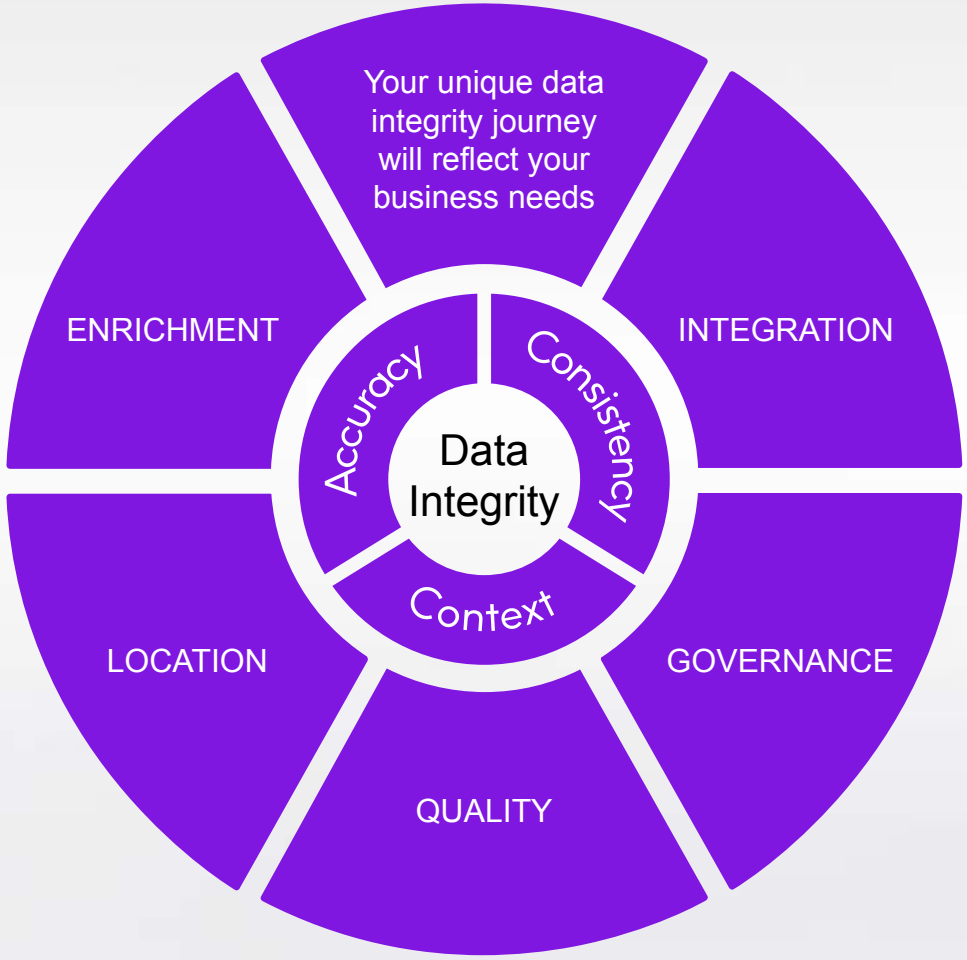
For trusted data, you need data integrity

Data integrity is data with maximum accuracy, consistency, and context for confident business decision-making



Data integrity is a journey

- Every journey to data integrity is unique and driven by business initiatives
- Market trends are accelerating the need for data integrity
- Precisely can help you at every step of your data integrity journey



Organizational needs are changing…

THEN AND NOW

Responsive

Data Mgmt. / IT teams cleaning data post-entry

Operational in use case

Focused on supporting business function efficiency and effectiveness

On-premises data stores

On-prem databases supporting operational systems and BI

Proactive

Data engineering embedding data quality to build and maintain data pipelines

Analytics-driven

Focus on analytics, artificial intelligence & machine learning, and decision-intelligence use cases

Data cloud

Companies now migrating and centralizing data in the cloud

…and so are data quality needs

THEN | AND | NOW

Manual deployment processes

Manually deploy and maintain software and data quality processes

Technical SME to manage DQ

Dedicated resources to configure and manage data quality

Data replication to validate

Replicate data within data quality tool to identify data issues

Automated deployment processes

Automated access to latest features and data quality process deployments

Intelligent data quality and usability

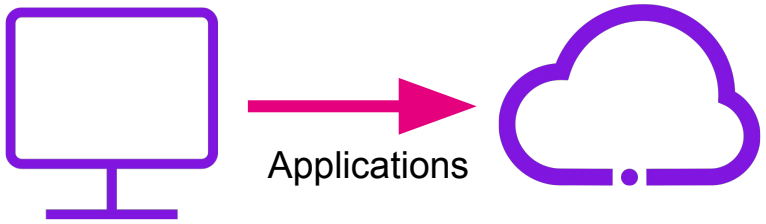
Leverage semantics, profiles, and observations in a seamless user experience to enable more users

Native data quality execution

Run data quality natively within environment data is stored

Common themes of successful data quality

- Understanding of the importance of data quality at a management and leadership level
- Institutionalized quality mindset, methodology, processes, and operating model
- Metrics and measurements
- Systems and automation
- Openness and Interoperability – Liberating the data



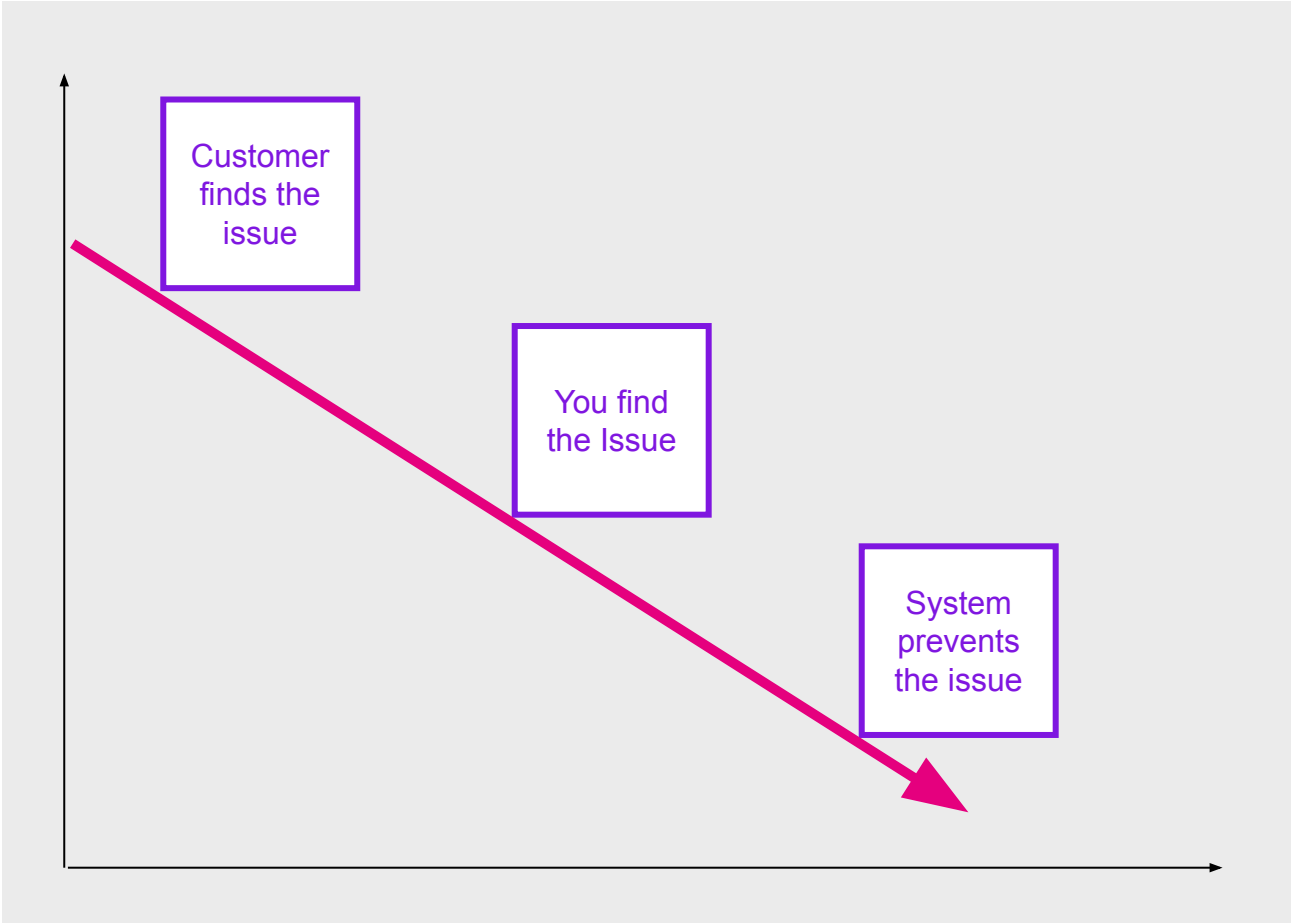
To ensure high quality data:

1. Find early and automate
2. Standardize and measure
3. Reuse
4. Educate

Find early and automate

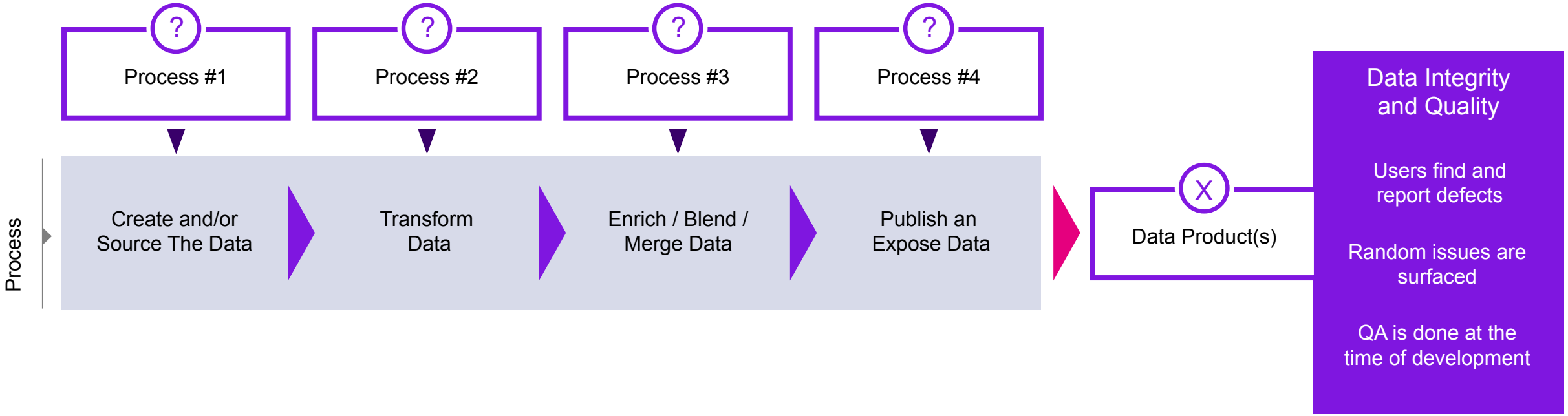
$

Cost of the Issue



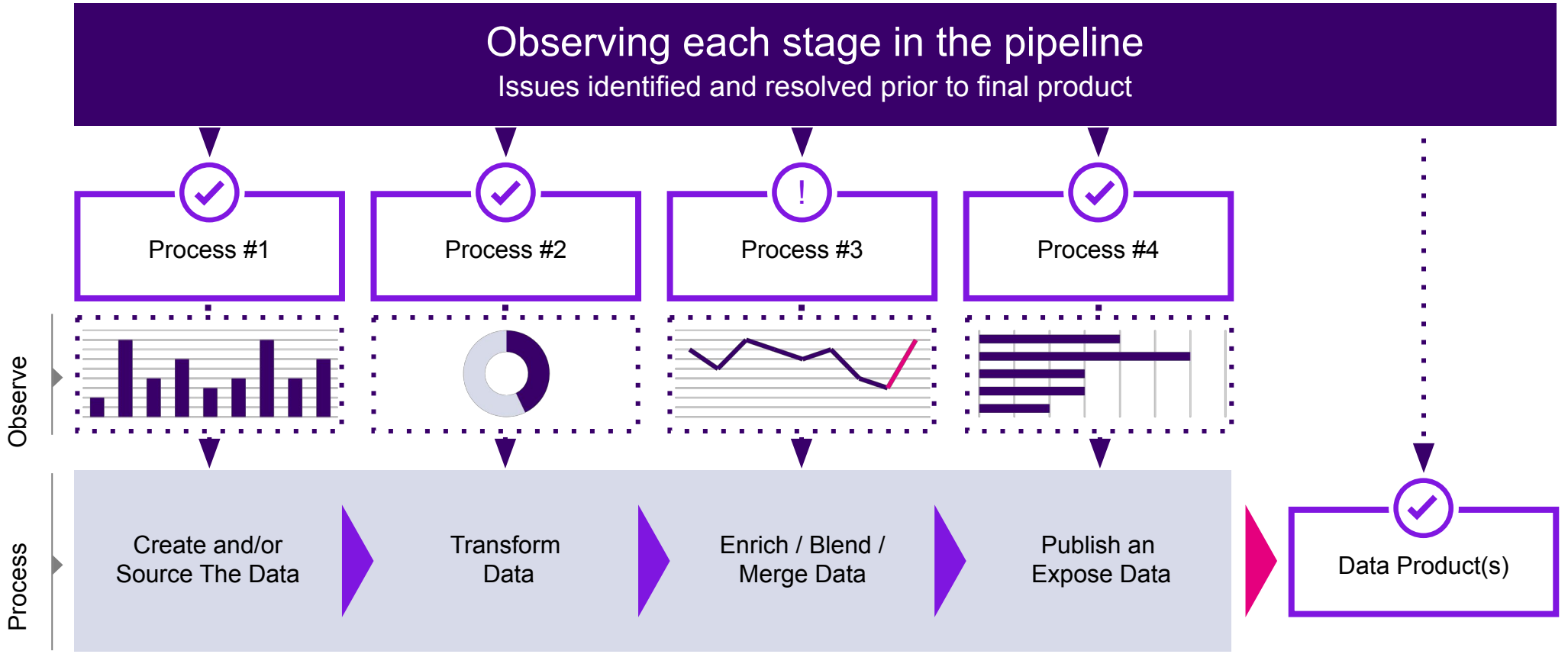
Typical data products and pipelines

Traditionally, the quality of a data product or pipeline is ensured during the development process and not throughout the operational lifecycle.



Data pipelines with data quality

Data quality tools observe the performance of data products and processes in order to detect significant variations before they result in the creation of erroneous work product in reports, analytics, insights and outcomes.



Standardize and measure

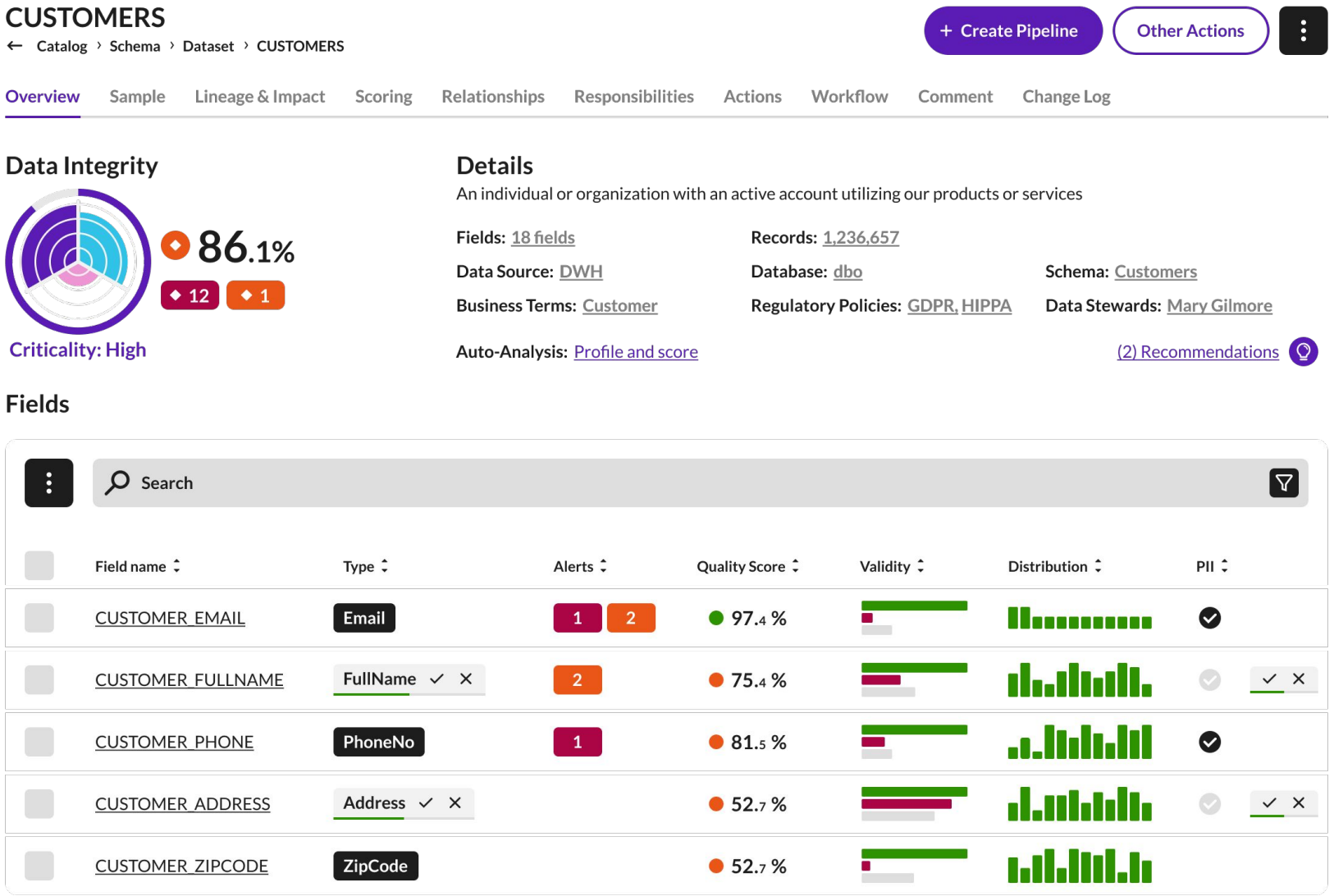
- Improvement starts with the identification of a problem
- Data quality scores are often used as a measurement system that identify correctable or incorrect conditions
- Different organizations may use other different types
- Repeatability and reusability is key

	Quality Score	Governance Score	Trust Score
Why?	How good is my data? Which data is failing the criteria?	How well am I stewarding and governing my data?	• Can I trust my data?
Examples	• Is field null? • Is field blank? • Has a semantic been identified? • Do the records have outliers? • Is the correct business logic applied?	• Is a business owner assigned? • Is a data steward assigned? • Does the asset have a description populated? • Is the status certified?	Should I trust my data if… • There are any alerts? • The average user rating is less than 5 stars? • My governance score is high, but my quality score is low?

Demonstration

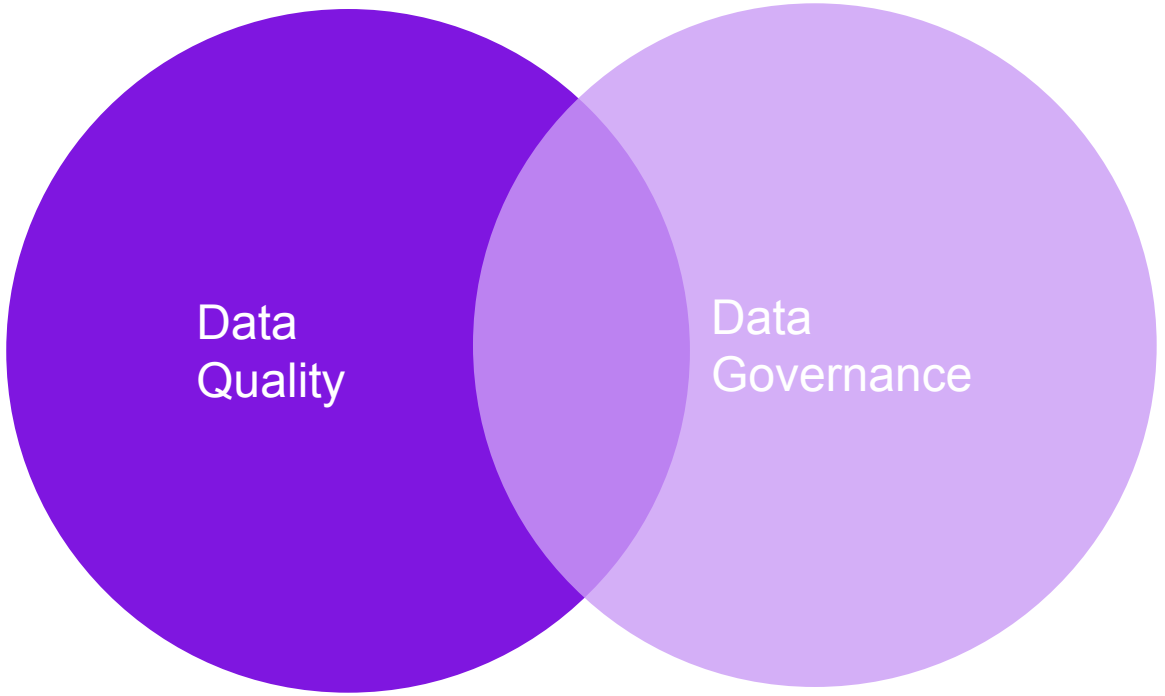
Educate

- Communication is crucial
- Demonstrate improvement over time



Data quality and governance are interdependent

- Data quality program is better with governance and catalog systems
- Improved data quality requires corrective action
- Scope to critical data and achievable improvements, not everything



Categories of data quality

Profiling, rule management and data validation

- Data validation
- Profiling
- Business rules
- Automated verification, standardization, and cleansing
- DQ scoring to support data governance

Address & contact validation

- Address validation, geocoding, and enrichment
- Validation of contact data
- Front-end application validation (CRM, ERP, etc.)

Comparison and consolidation

- Enterprise reconciliation of data in place
- Entity matching
- Single view of data

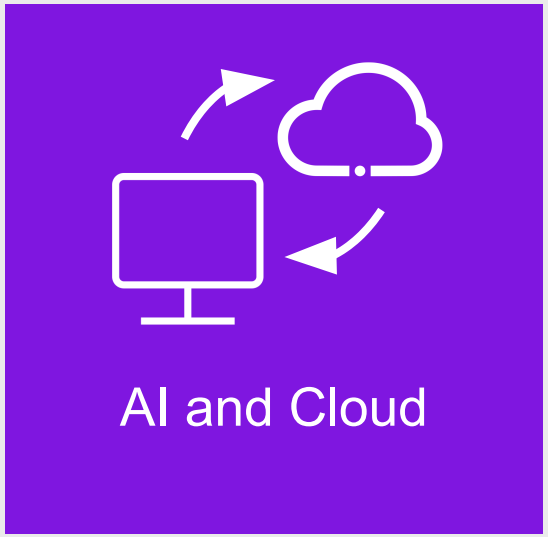
Data on the move

- System to system balancing & reconciliation
- Data observability

The effect of data quality on AI

AI's Business Impact

AI as an enterprise Initiative driving business transformation



AI and Cloud

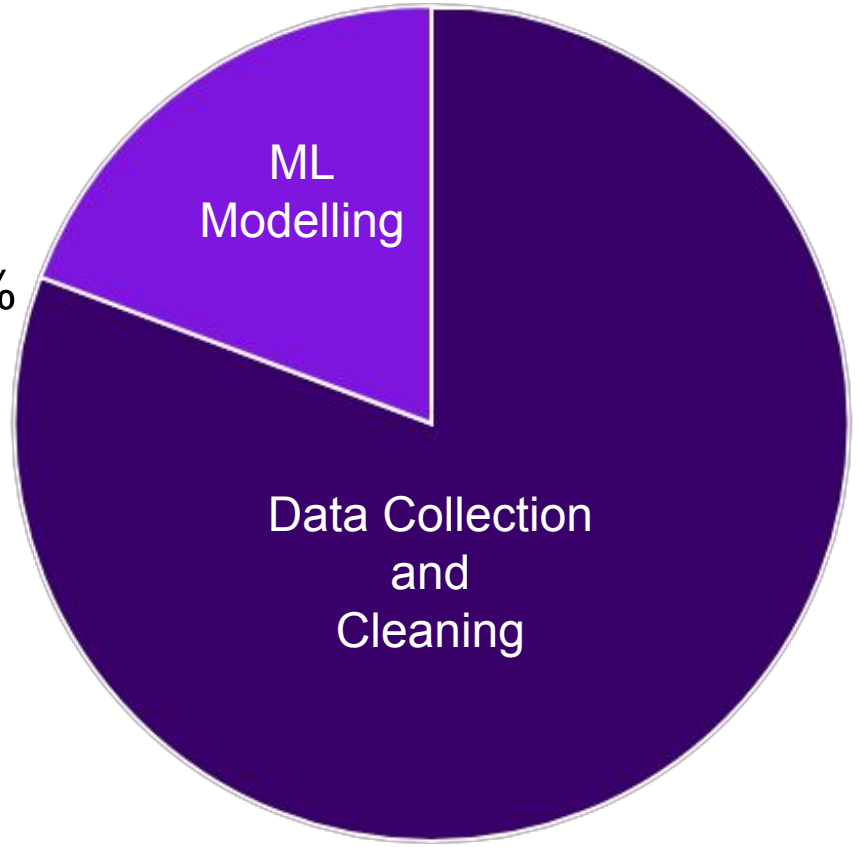
Increased pace of digital And AI initiatives is fast-tracking cloud adoption

AI and Data Integrity

Trusted AI requires data Quality and governance practices

AI example #1 Data Quality: High quality data saves data scientists time and improves ML accuracy

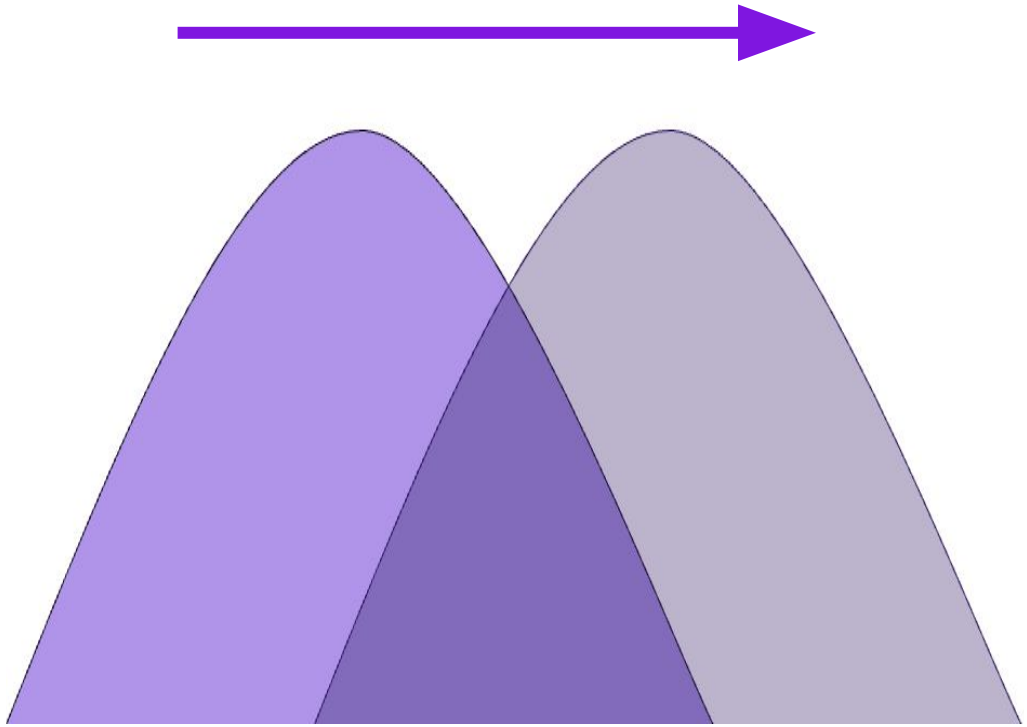
- *ML Models are an abstraction of the data they are trained on, so the quality of a model is reflected in the data it is trained on (Mohammed Taboun, PhD)*
- It is well known that data scientists can spend up to 80% of their time collecting and cleaning data.
- An estimated 54% to 90% of machine learning (ML) models don't make it into production from initial pilots.*
- A robust data quality system will reduce the amount of time that a data scientist will spend analyzing & fixing poor quality data and increase the amount of time experimenting and providing valuable insight from data.



*The power of MLOps to scale AI across the enterprise (Venturebeat)

AI example #2 MLOps: Data changes over time – so should your ML models

- There are many ways data can change over time:
  - New products may be added to a sales distribution center
  - More vehicles are being made with focus on fuel efficiency
  - Population density in certain areas change
  - New IT policies due to 3rd party AI apps in organizations
- ML models which are trained to consume data from sources which are impacted by changing data need to be continuously updated to capture changes in data (data drift).
- Data observability allows data scientists and ML engineers to know when data drift occurs, allowing them to retrain models in their continuous training processes.

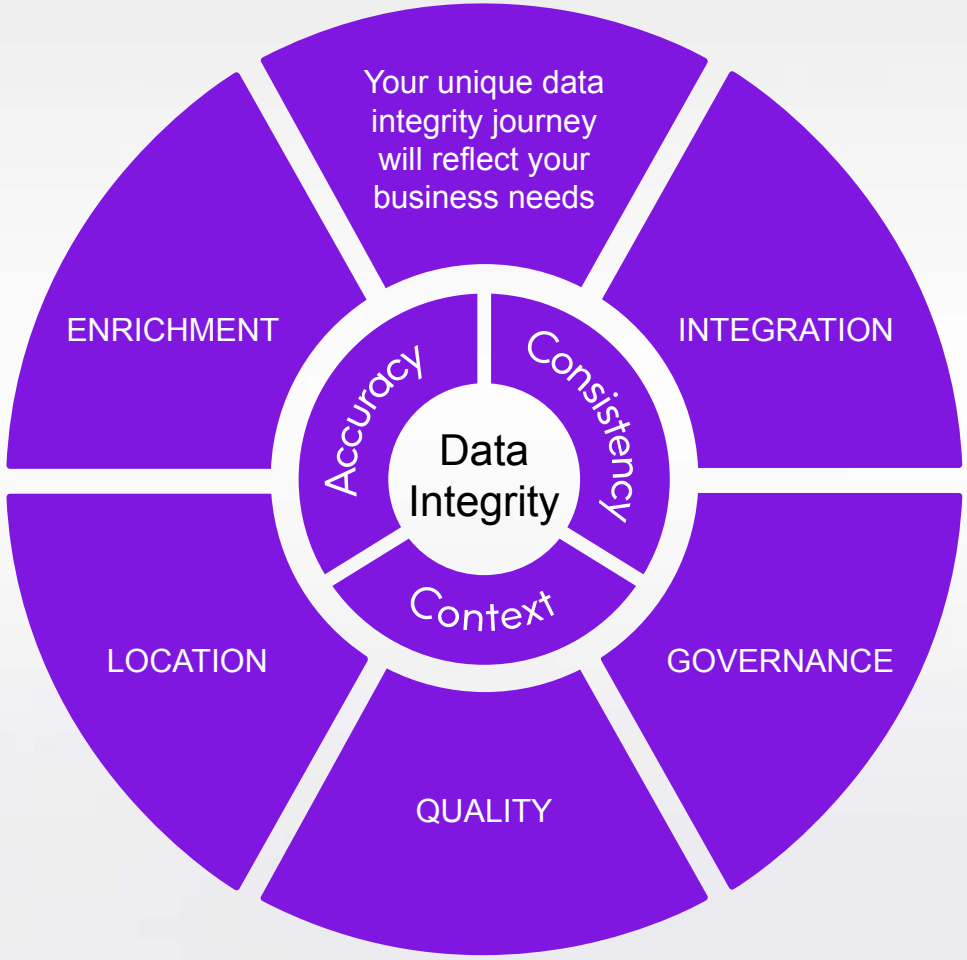


Does this sound familiar?

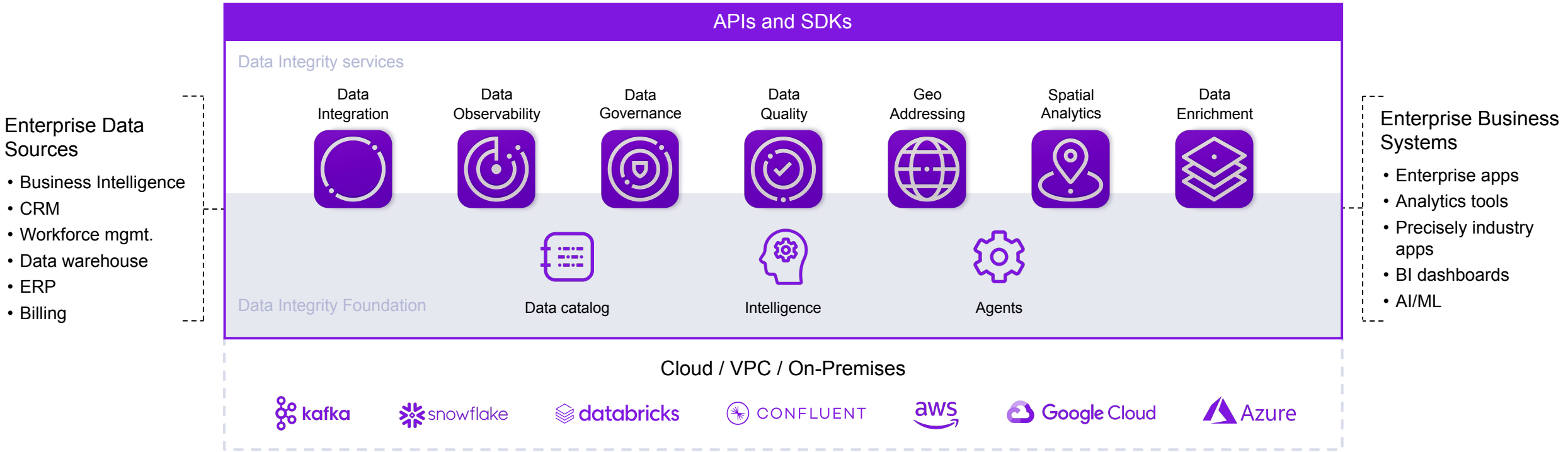
I need to drive our new marketing campaign based on lists come from outside vendors and internal sources, but how do I know If those lists are any good?

Data integrity is a journey

- Every journey to data integrity is unique and driven by business initiatives
- Market trends are accelerating the need for data integrity
- Precisely can help you at every step of your data integrity journey



Data Integrity Suite



Questions?

Thank you!

Learn More!

www.precisely.com

Sales: 1 (877) 700 0970