Using Data Intelligently:
Unification and Pipelining Patterns in the Digital Economy

An IDC InfoBrief, sponsored by TIBCO
January 2020

IDC ANALYZE THE FUTURE

US45653219BROI

TIBCO®

The Data Management Challenge

Facing a rising tide of data

33ZB

of new data was created in 2018, and by 2023, IDC estimates **103 ZB of new data** will be created. (1 ZB = 1 trillion Gigabytes).

IDC, Worldwide Global DataSphere Forecast, 2019–2023

14%

of the 33 ZB was original data; **86% of the new volume** was generated by replication and distribution, resulting in significant data liabilities.

IDC, Worldwide Global DataSphere Forecast, 2019–2023

25%

of useful data created in 2018 was tagged and **only 13%** of the tagged data was analyzed.

IDC, Worldwide Global DataSphere Forecast, 2019–2023

60%

of organizations are being challenged by data **quality** and **complexity**, complicating data integration, master data management, and governance.

IDC, Data Integration and Integrity End User Survey, June 2019

The Global DataSphere is a measurement of data created, not data stored.

Data is the Lifeblood of Digital Transformation (DX)

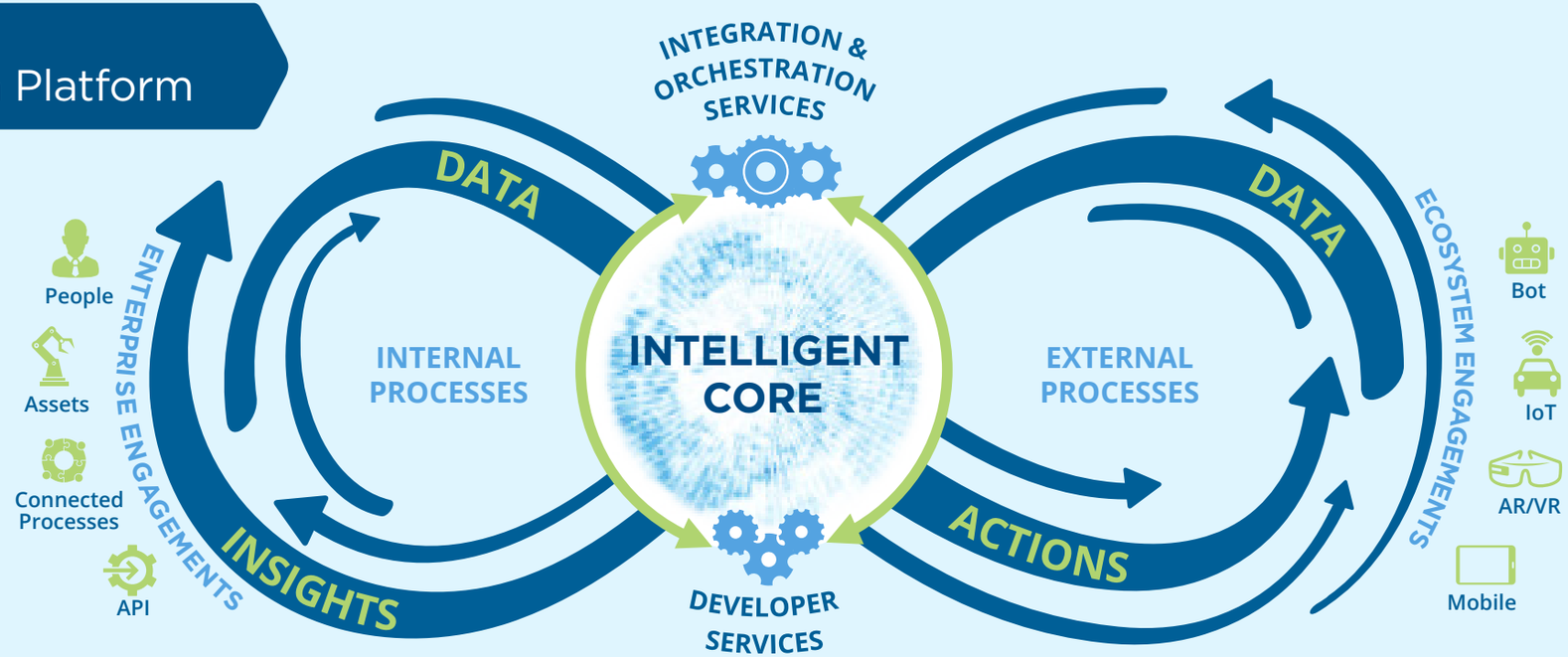
46%
of organizations are
Digitally Distraught

Organizational culture, strategy, financials, and platform
are separating the digitally distraught from the digitally determined.

54%
of organizations are
Digitally Determined

Source: IDC, Global DX Leaders Survey, June 2018. Worldwide sample, n = 1,987

IDC's Digital Transformation Platform



Source: IDC, 2018

Highly Distributed and Diverse Data Environments Are Common in the Digital Economy

Multiple data management technologies are in use across on-prem, hybrid, and cloud:

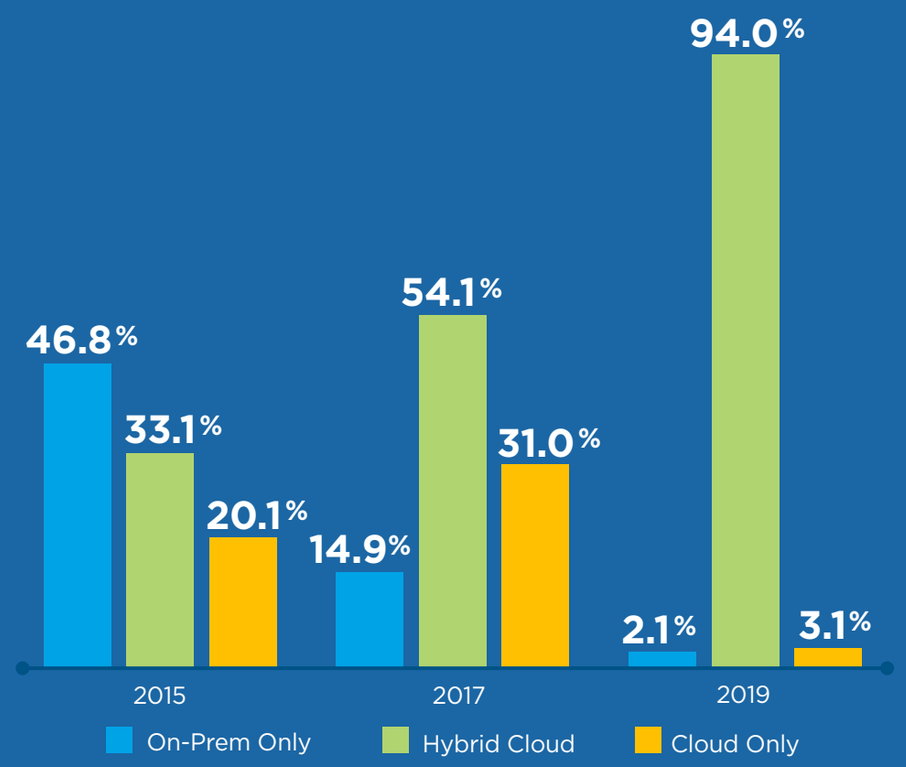
- Mainframe
- Relational databases
- Analytical databases
- Data lakes
- NoSQL
- In-Memory
- Streaming

Managing multiple types of data, such as:

- Flat files
- Master data
- Transactional data
- Spatial data
- Object data
- IoT data
- Interaction data
- Social data

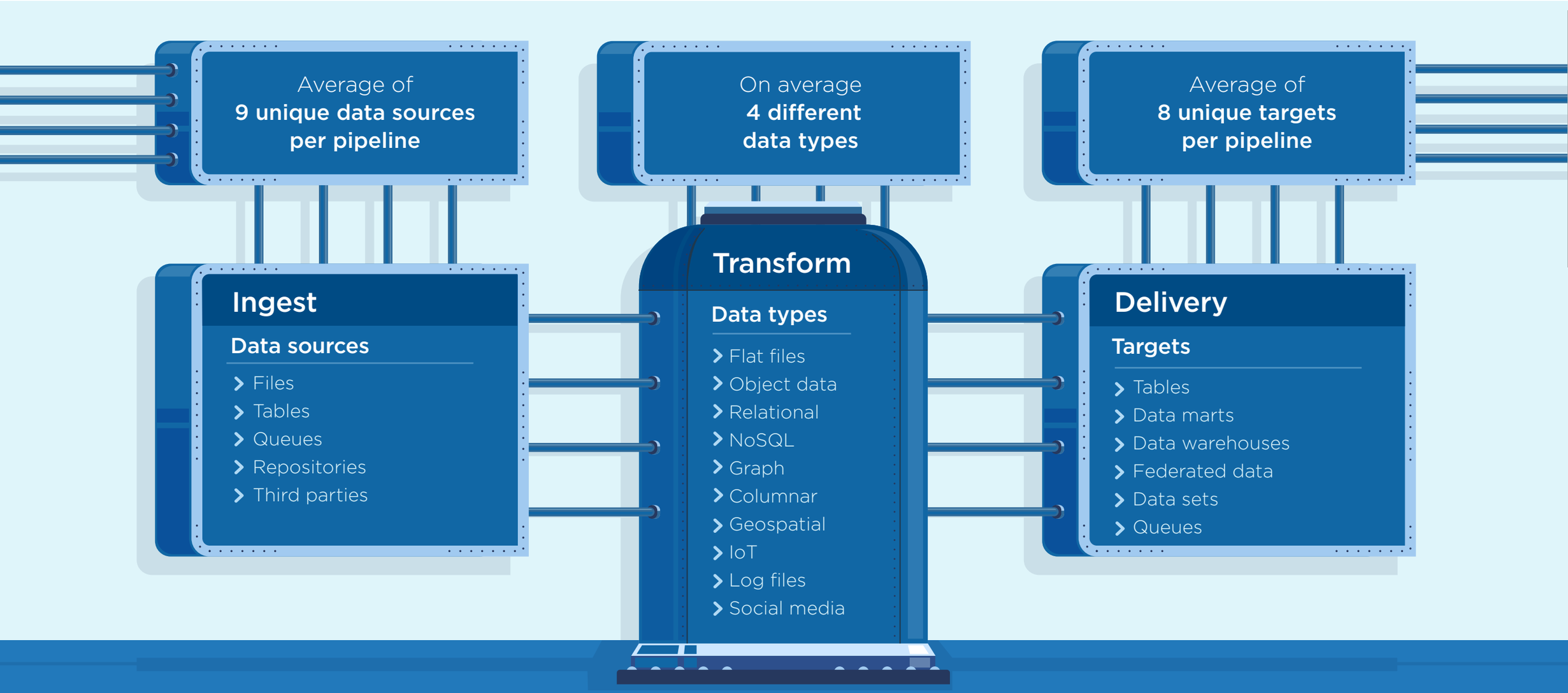
Legacy data management technologies and data types continue to be part of digital environments.

Data Environments for Data Integration Solutions



Source: IDC, Data Integration and Integrity End User Survey 2019, n=300

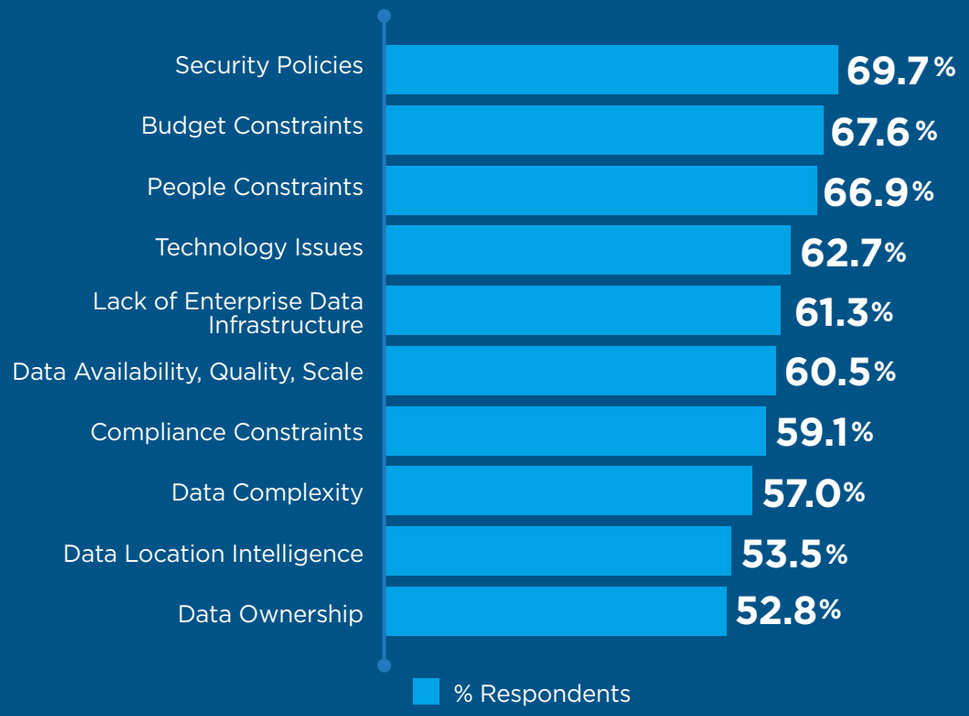
Data Distribution and Diversity is Complicating Data Integration Pipelines



Source: IDC, Data Integration and Integrity End User Survey 2019, n=300

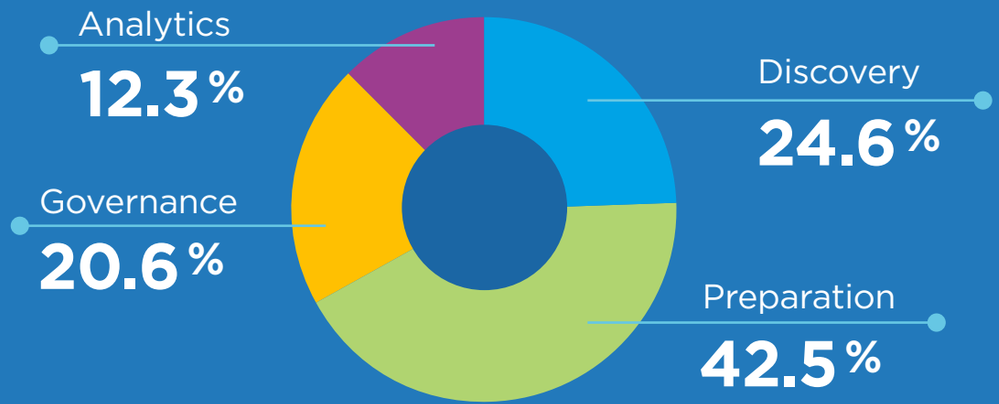
Organizations Report Data Quality, Integration, Mastering, and Governance Challenges that are Impacting Employee Efficiencies and Effectiveness

Data Management Challenges



Source: IDC, Data Integration and Integrity End User Survey 2019, n=300

% of Weekly Time Spent Working with Data



Data worker efficiencies are decreasing as workers spend more time looking for, preparing, and governing data rather than performing analytics.

Source: IDC, Data Integration and Integrity End User Survey 2019, n=300

Taming Data Complexity with Patterns

Patterns offer a framework for:

Intelligent data utilization patterns:

360-for-anything

Data governance

Data-as-a-Service

Data pipelines

360-for-Anything

Understanding the problem

- Customer-driven blended view of master and transactional data across disparate data environments
- Valuable for customer experience, operational excellence, and organizational enablement

Selecting a solution

1. Master data management with virtual federation of transaction, interaction, and relationship data
2. Master data management with physical federation of transaction, interaction, and relationship data

Implications

1. Federation with data virtualization may require caching to improve performance
2. Federation with data replication increases compliance and security risks

Data Governance

Understanding the problem

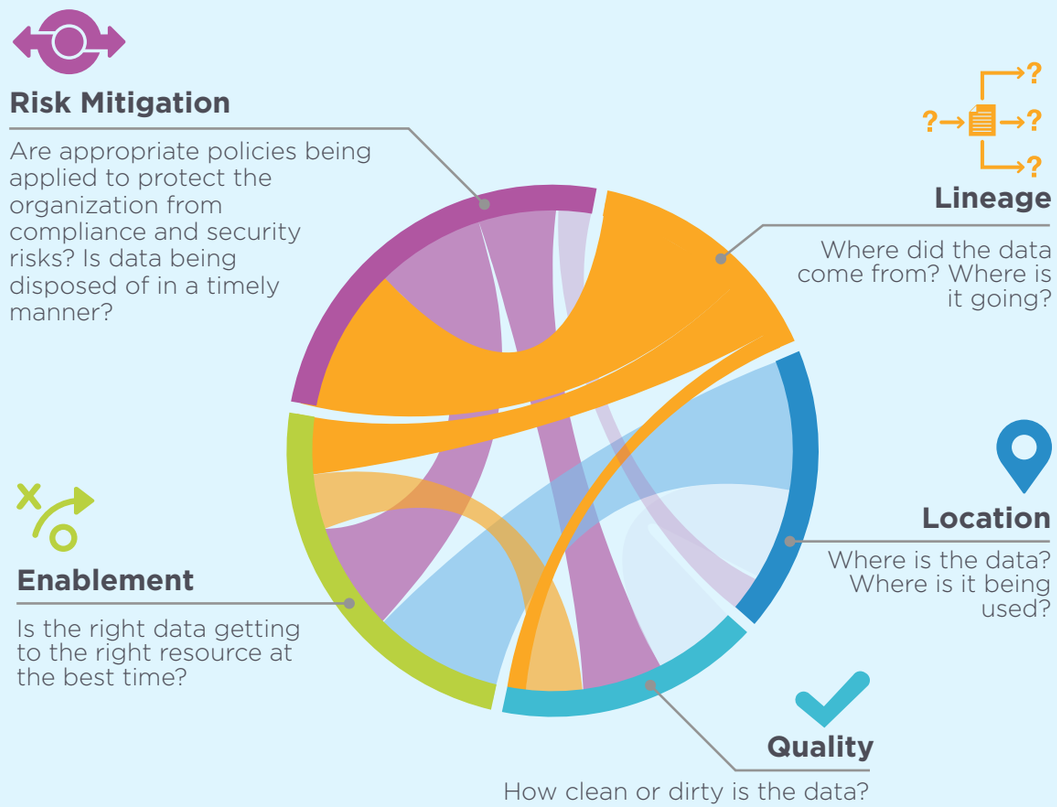
- Reduce security and compliance risks associated with data persistence, use, and life-cycle management
- Improve the levels of organizational data trust and transparency

Selecting a solution

1. Formally organize for data governance, implementing solutions to reduce risk and enable the organization with trusted data
2. Focus on compliance and security risk, assuming data quality and trust will be addressed by individuals

Implications

1. Gathering intelligence at the scale of volume and diversity of data in the digital economy requires automation
2. Applying policy focused on risk alone may not elevate data to become a trusted, valuable asset and could restrict data innovation



Data-as-a-Service

Understanding the problem

- Disparate data needs to be delivered to multiple consumers for a variety of use cases, inside and outside the organization
- Multiple views of the same source data are required, driven by the use case

Selecting a solution

1. Virtually federated fit-for-purpose view of data exposed via standard APIs
2. Physically federated data, using data marts to deliver fit-for-purpose data

Implications

1. Virtual data transformation will be required to support transactional and analytical workloads, implying caching
2. Latency of data movement and the effort to create data marts can result in lengthy timelines, and rigidity will impact ability to change, and increase liability of data duplication

Data Pipelines

Understanding the problem

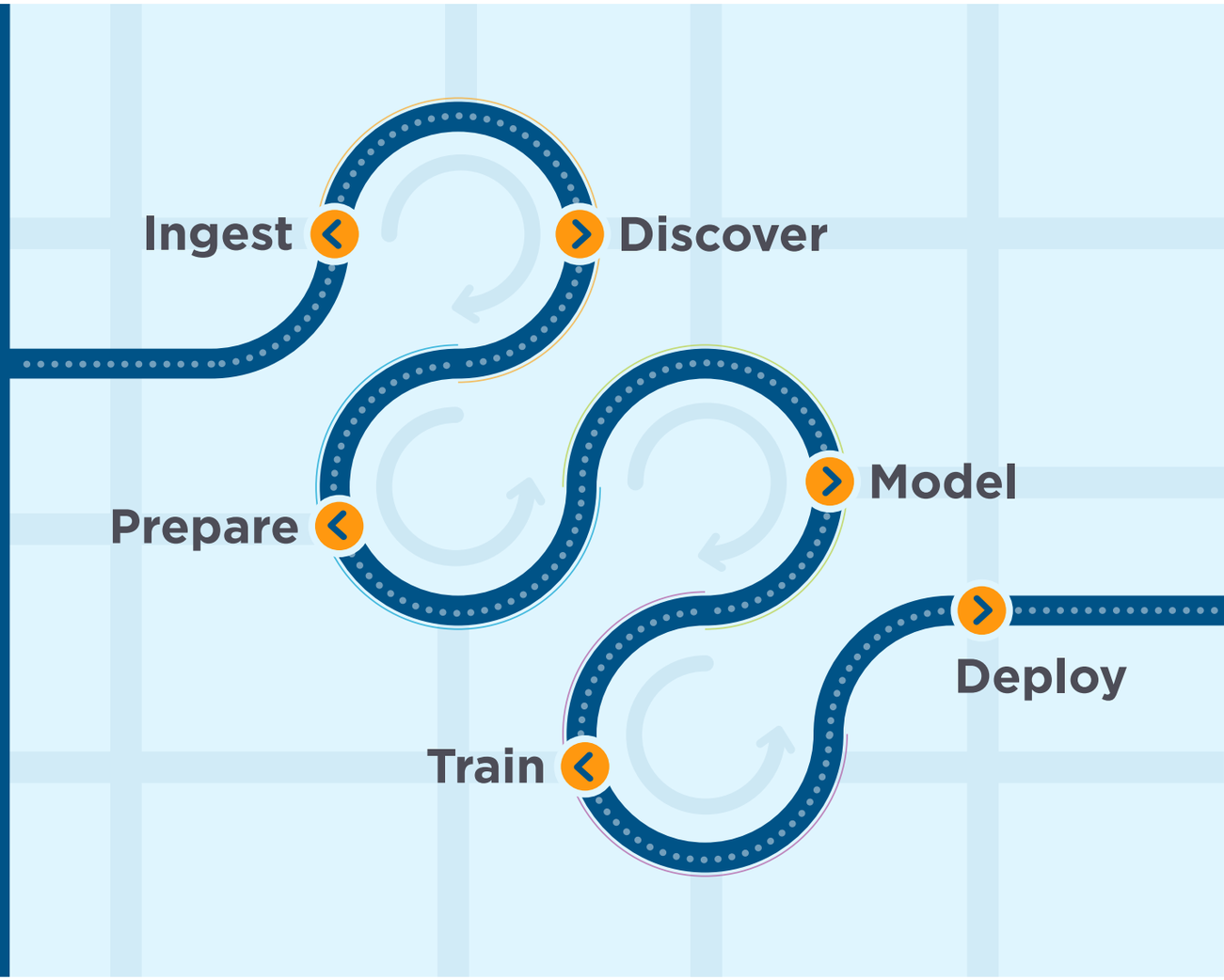
- Ingest from multiple disparate and diverse data sources, internal and external, with minimal latency
- Near-real-time data profiling, cleansing, normalization, mastering, and protection
- Deliver to analytical infrastructure for model training and deployment

Selecting a solution

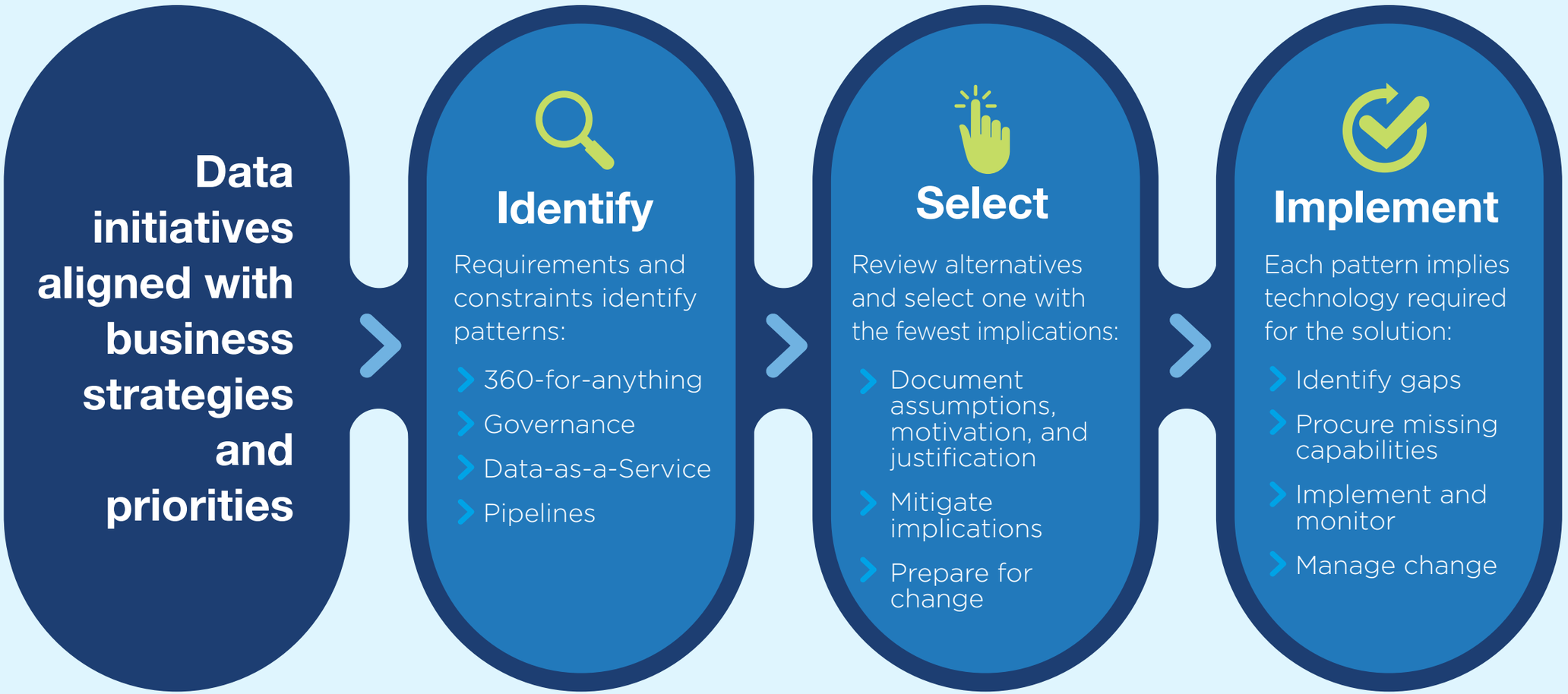
1. Data replication and movement into a persistent data store
2. Real-time virtual and in-memory data federation across disparate sources

Implications

1. Replicating data increases data latency, security, compliance, and quality risk
2. Frequent refresh of federated data will be required to mimic streaming



Applying Patterns



Closing Gaps, Moving Forward

Buy or build?

- Complex data environments require complex solutions: 96% of organizations prefer commercial or open source over custom code.
- Buyers are looking for agile solutions with lower data latency, elasticity, traceability, and lineage.
- Majority of respondents that have implemented commercial data-quality software have seen improvements in quality metrics.

Data replication or virtual federation?

- Top three metrics organizations with data virtualization have seen positive improvements in:
  1. Compliance **2.** Correctness **3.** Return on investment
- Most organizations with data virtualization are seeing benefits within a year of implementation.
- Data virtualization combined with master data management can deliver a 360-degree view in near real time without replication.

Source: IDC, Data Integration and Integrity End User Survey 2019, *n=300

96% of organizations prefer commercial or open source over custom code.

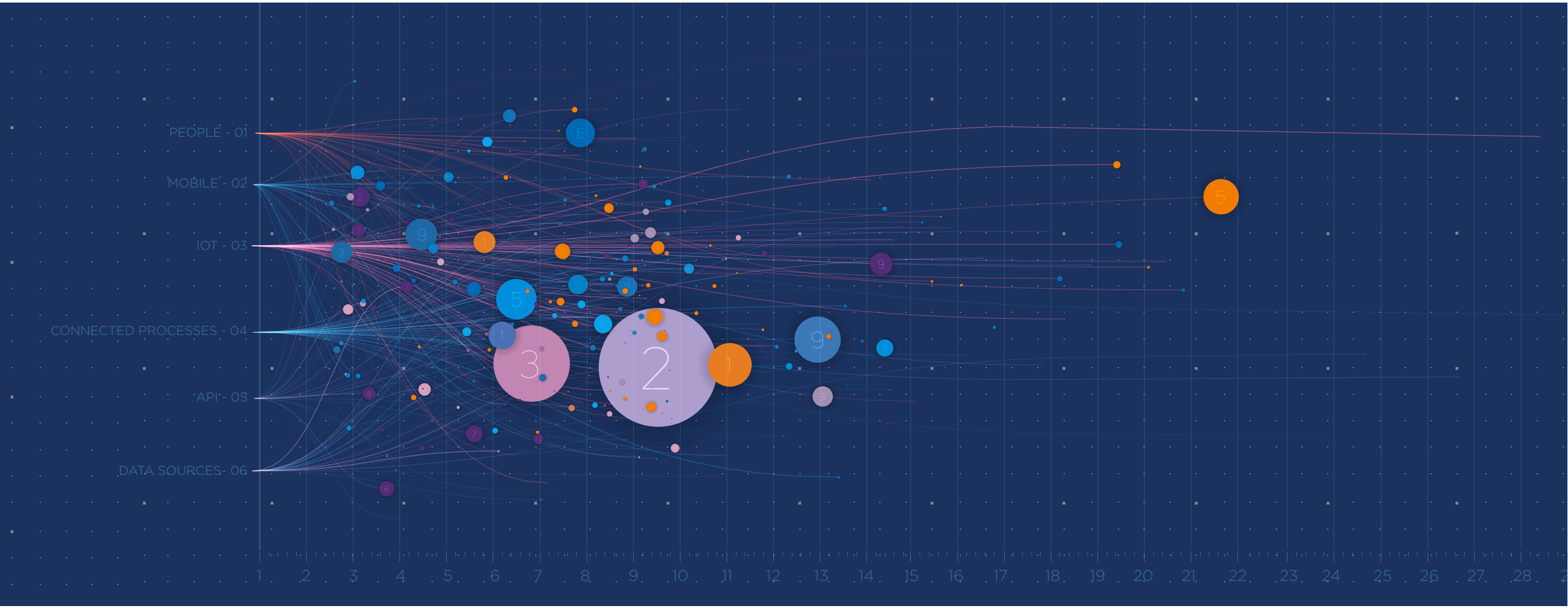
IDC Analyst Profile

Stewart Bond,
Research Director, Data Integration and Data Intelligence Software

Stewart Bond is Research Director of IDC's Data Integration and Intelligence Software service. Mr. Bond's core research coverage includes **watching emerging trends** that are shaping and changing data movement, ingestion, transformation, mastering, cleansing, and consumption in the era of digital transformation.

The content in this paper was adapted from existing IDC research published on **www.idc.com**

Message from the Sponsor



To learn more about how TIBCO can help with your initiatives, visit:
tibco.com/products
tibco.com/contact-us