

# Drive Faster Insights from Your Data



David Stodder Senior Director of TDWI Research for Business Intelligence

**GUEST SPEAKER** Thomas Ridings, Senior Director of Product Management, Matillion



## Webinar Sponsor





# Drive Faster Insights from Your Data

**TDWI Research Perspectives** on Trends in Data Engineering Productivity and Collaboration with Best Practices and Al Innovation



David Stodder Senior Director of TDWI Research for Business Intelligence



## Demand for Faster Data Insights - and Concerns

## Maximizing the value of data

- Democratized users need contextual data to drive operational decisions, customer engagement, risk management, regulatory adherence, and more
- Data-rich applications and enriched transactions
- Automated decisioning in apps

## Analytics and Al

- Data for model development and testing
- Data for LLMs & generative Al
- Augmentation of Bl and reporting

# Moderate success currently: Concerns about responding to future challenges

Overall, how successful is your organization in using data systems to enable users to gain faster data insights? How confident is your organization that it can succeed as data and workload volumes, speed, and/or complexity increase?



Research data source: TDWI Q1 2024 Best Practices Report. See https://bit.ly/43j2ajp

## Data Engineering: The "Engine" of Faster Insights

# Data engineering: Central to provisioning data for BI, applications, analytics, and AI

- Discover, collect, and prepare data
- Design and populate data warehouses and data lakes
- Develop and code data pipelines, connectivity, and transformations (ETL/ELT), often with significant technical expertise
- Validate data; manage, modify, and maintain data integration through multiple iterations
- Big data and data engineering services market size to be worth \$169.9 billion by 2029, up from \$75.55 billion in 2024 (CAGR of 17.6%) [Source: Market Data Forecast, www.marketdataforecast.com]

# Data demands put pressure on data engineering

- To be efficient and effective as data demands escalate
- 50% of organizations surveyed say project teams spend more than 61% of their time on data integration, pipeline development, and preparation, with 16% spending over 81%
- Pressure to provision data for advancing analytics and AI
- Pressure to partner effectively in teams with data scientists, analysts, developers, and low- and no-code users

## Data Pipelines, Connectivity, and ETL/ELT Challenges

## Data governance and security

• Difficulty monitoring exposure risks; interest in automation

# Identifying and eliminating unnecessary workloads

- Problems with overall visibility to reduce costs and resource usage
- 63% say managing and reducing costs is challenging

Improving performance, scalability, and orchestration

- Research shows challenges
- Challenges understanding dependencies

How challenging is it to address the following issues with your organization's data pipelines, data connectivity, and ETL or ELT processes?

Based on answers from 183 respondents. Ordered by combined "very challenging" and "somewhat challenging" responses.



#### Governing and securing data as it is moved or copied

0	0	•		
29	%	10%	20%	6% <mark>5%</mark> 6%
Identifying an	d eliminating workloads that a	re no longer necess	sary	
21%	48%		19%	<mark>4%</mark> 8%
Improving per	formance and scalability			
22%	42%		23%	6% 7%
Reducing time	required to load high data vol	umes into target pl	atforms	
25%	38%		21%	7% 9%
Managing and	reducing cost of data pipeline	s and ETL		
20%	43%		18%	<b>7%</b> 12%
Understanding	g dependencies between system	15		
24%	37%		18%	<b>9%</b> 12%
Understanding	dependencies within jobs			
19%	39%		24%	<b>8%</b> 10%
Managing and	orchestrating numerous pipel	ines and ETL jobs		
18%	40%		25%	<b>7%</b> 10%
Implementing	ELT as an alternative to tradit	ional ETL		
19%	37%	17%	8%	19%
Performing in	cremental data loading and CD	C		
13%	42%	19%	12%	14%
Managing data	egress costs when moving clo	ud data		
15%	33%	20%	8%	24%

## **Top Issues Hindering Timely** Data Insights

What are the biggest issues hindering your	Data quality and completeness are inadequate	50%
enabling users to gain timely data insights	Data is too fragmented to gain single views of all relevant data	47%
questions and solving problems?	Data governance and regulatory rules increase latency	33%
Rased on answers from 245	Data pipelines are difficult to build, launch, and iterate	32%
respondents, who were asked to select at least their top five concerns.	Too much reliance on spreadsheets	30%
	Analytics and AI/ML workloads are difficult to operationalize	29%
	Not enough people skilled in real-time data solutions; learning curve is steep	29%
	Dependence on slow batch processing for updates to most BI reports, dashboards, and analytics	27%
	Data transformations and pre-processing are slow and difficult to scale	25%
	Data is old and not updated frequently enough for business purposes	22%
	Investment is perceived as too costly to effectively reduce data latency	20%
	Users cannot easily analyze real-time and historical data together	12%
Research data source: TDWI Q1 2024 Best Practices Report. See https://bit.ly/43j2ajp	Users cannot interact with data through natural language, search, and/ or generative AI	12%

## Major Challenges Hindering Data Engineering Goals

### Disconnected and siloed data integration

- Data accessibility barriers, especially for Al
- Too many tools and jobs to track and manage, including from a data governance perspective
- Technical expertise barriers for low- and nocode users
- Data lineage and documentation is incomplete and inconsistent
- Difficulty managing and reducing costs

## Scalability and orchestration challenges

- Difficulty managing and locating all the pipelines, connectors, and ETL/ELT jobs
- Manual maintenance of manual coding: Time and complexity fixing errors and making updates

#### Support for Al growth, including generative Al

- More users want analytics and Al for datadriven decisions
- Feature engineering and vector embedding
- Selecting and validating data for LLMs and generative Al
- Data lineage and documentation for Al explainability

#### Performance and agility issues

- Difficulty performing data engineering responsibilities quickly to accelerate business use
- Need for more automation to reduce manual coding
- Accelerating costs create resource limitations

# **Poll Question**

To enable faster data insights, which of the following currently your biggest data engineering challenge?

Technical challenges leading to delays in building and managing data pipelines, connectors, and ETL jobs

Having the talent and expertise to keep pace with accelerating data demands, including AI requirements

Data and tool silos make it difficult to centrally manage and orchestrate data pipelines, connectivity, and ETL, and to manage code

Team collaboration is lacking among data engineers and no-code and low-code users

Cost issues; we cannot manage and reduce costs as data demands escalate

## Solving Challenges: Trends and Priorities

## Streamlining data insights with tech modernization

- Al-driven automation; automation of processes for putting data pipelines into production
- Potential of generative AI for improving productivity
- Improving reuse by setting up pre-built connector libraries
- Easier interfaces for less technical users

### Centralizing data integration management

- Orchestration of jobs to manage dependencies, address bottlenecks, and improve scalability
- Enabling rationalization of data pipelines, ETL, and connectors to ensure business value
- Data lineage and documentation for data and Al governance
- Reducing silos to enable faster and more complete tech modernization (rather than piecemeal)

## TDWI research: Importance of AI-driven automation

- <u>Data quality, validation, and enrichment</u>: 40% very important, 27% important but not a current priority
- <u>Data management (including for data</u> <u>warehouse and data lake)</u>: 36% very important, 26% important but not a current priority
- <u>Data governance and regulatory adherence</u>: 33% very important, 31% important but not a current priority
- <u>Data ingestion, pipelines, preparation, and</u> <u>transformation</u>: 30% very important, 29% important but not a current priority
- <u>DevOps, programming, code generation,</u> <u>and container orchestration</u>: 21% very important, 32% important but not a current priority

## Optimizing Productivity with DevOps and DataOps

# Collaborative data culture for more effective teams

- <u>DevOps and DataOps</u> for data engineering and code management
- <u>DataOps</u>: Building on agile and DevOps to provide a structure similar to software development life cycles but for data processes
- <u>Data observability</u>: Important to move beyond limited, siloed monitoring; visibility to see how problems impact business
- <u>Leadership</u>: Key for encouraging crossfunctional collaboration

Is your organization currently using or planning to use DataOps practices to optimize productivity and support growth in data pipelines and analytics and AI/ML workloads? Have you had experience in using DataOps or related methods?

Based on answers from 163 respondents.



Research data source: TDWI Q3 2023 Best Practices Report



# Thank You



## David Stodder

Senior Director of TDWI Research for Business Intelligence dstodder@tdwi.org @dbstodder





## **Thomas Ridings**

Senior Product Director Matillion





# Matillion is the **Data Productivity Cloud**

Empowering teams to make their data business-ready faster



# The landscape is more complicated

More data, from ever increasing sources, with different data formats, structures and approaches

The infrastructure (scaling, integration, security)



#### Budgets aren't keeping up

Overall demand shortage for data skills is increasing the cost of recruitment and staffing

Infrastructure and software costs are rising.

# There's growing demand

As businesses are demanding more from their data, the demand increases.

50% increase in demand yearon-year.

# Teams and organizations need to be equipped to be **productive** with data

## #1 Design for discovery and learning

Documentation is key to discovery, make your pipelines readable

Multiple tools and systems increase the learning time, keep it simple

Column level lineage can accelerate learning

→ → → → → → → → → → → → → → → → → → →			
Baw Products from Filter on va	lid Bemove duplicates	Solit Product Name	Bewrite Table
PIM product co	des	opilet roduce nume	
	Validate singling (dense singling) outra		
	Validate Dideline (demo-dideline)SHIFT + CTRL		
	· · · · · · · · · · · · · · · · · · ·	+ RETURN	
	Run pipeline (demo-pipeline) CTRL + ENTER	+ RETURN	
	Run pipeline (demo-pipeline) CTRL + ENTER	+ HETURN.	
	Run pipeline (demo-pipeline) CTRL + ENTER	+ HE I UHN	
	Run pipeline (demo-pipeline) CTRL + ENTER	* HEIURN.	
	Run pipeline (demo-pipeline) CTRL + ENTER Add note	+ HE LUHN.	
	Run pipeline (demo-pipeline) CTRL + ENTER Add note Add note using AI	+ HE LURIN.	
	Run pipeline (demo-pipeline) CTRL + ENTER Add note Add note using AI	+ HE LUHN.	
	Run pipeline (demo-pipeline) CTRL + ENTER Add note Add note using AI	+ HE LUHN.	
	Run pipeline (demo-pipeline) OTRL + ENTER   Add note Add note using AI   Undo OTRL + Z	+ HE LUHN.	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Im   Undo OTRL + Z   Redo OTRL + Y	+ HE LURIN.	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Im   Undo CTRL + Z   Redo CTRL + Y	+ HE LUNN	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Im   Undo CTRL + Z   Redo CTRL + Y	+ HE LURIN	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Im   Undo CTRL + Z   Redo CTRL + Y   Copy CTRL + C	* HE LURIN	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Im   Undo CTRL + Z   Redo CTRL + Y   Copy CTRL + Q   Paste CTRL + V	+ HE LUHN	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Image: Copy CTRL + Z   Copy CTRL + Q CTRL + Q   Paste CTRL + Q	+ HE LUNN	
	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Imponent   Undo CTRL + Z   Redo CTRL + Y   Copy CTRL + Q   Paste CTRL + V   mponent DELETER	+ HE LURIN.	
Al: Auto-documentatio	Run pipeline (demo-pipeline) CTRL + ENTER   Add note Add note   Add note using AI Image: Comparison of the second seco	+ HE LURIN	



## #2 Make your development time productive

Build what's unique to your business, leverage off-the-shelf or open-source where possible

Low-Code and High-Code development can co-exist

Reduce barriers so others can contribute but ensure there are guardrails





# #3 Optimise for management

# Ensure you have strongly versioned system

Automate your tests, mandate them before deployment

Work to identify and remove what slows you down from delivering value

#### Commit changes

Create a snapshot of your changes and push to the remote branch.

#### Branch

ይ <Branch name>

#### Changes (3)



#### **Commit message**



Cancel

Commit



Al: Auto-generating commit messages

# Fireside Chat Discussion



Thomas Ridings Senior Director of Product Management Matillion



David Stodder Senior Director of Research Business Intelligence TDWI





# Audience Q&A with Speakers





# Speaker Contact Information



Thomas Ridings Senior Director of Product Management Matillion Thomas.ridings@matillion.com



David Stodder Senior Director of Research Business Intelligence TDWI dstodder@tdwi.org



## Thank You to Our Webinar Sponsor



# Follow Us on Social Media



tdwi.org/linkedin











tdwi.org/youtube



# Thank You for attending!

