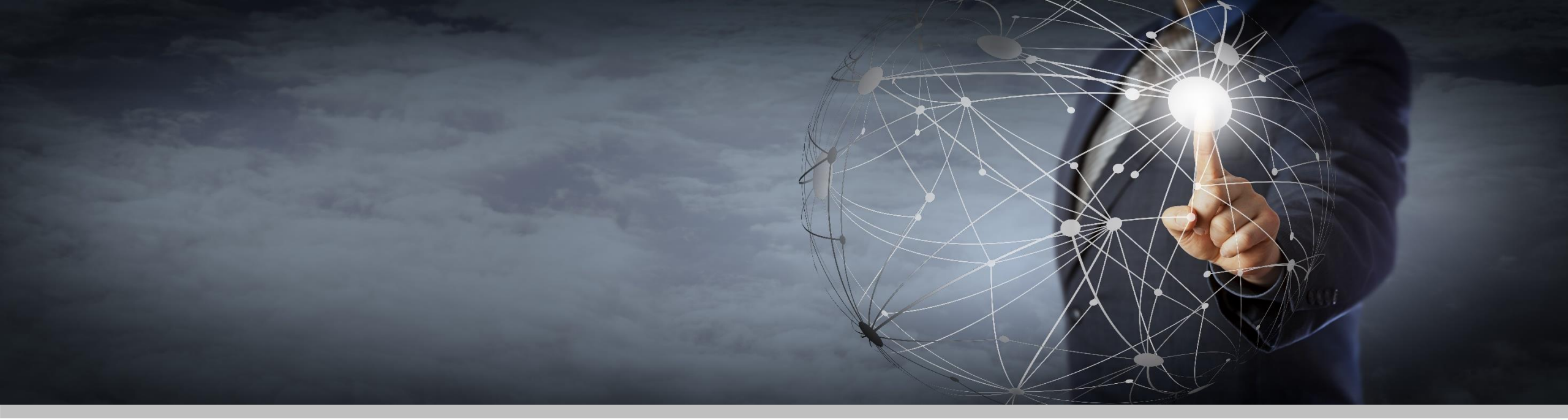


Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization

Donna Burbank
Global Data Strategy, Ltd.

July 25, 2024

About me….

- Steve Totman
- Chief Architect - Informatica Field CTO Office
- stotman@informatica.com
- Ardent, Informix, Ascential, IBM, Precisely (Syncsort), Cloudera, Privitar

Informatica

The biggest

transformation

of our lives

Forbes

Harnessing The Power Of Generative AI For Your SMB

Generative AI, when used strategically, can help many small- and medium-sized businesses improve their current practices.

Time Magazine

The AI That Could Heal a Divided Internet

For years, online spaces have struggled to foster positive conversation from Google's Jigsaw unit hopes to change that.

How I Built an AI-Powered, Self-Running Propaganda Machine for $105

I paid a website developer to create a fully automated, AI-generated 'pink-slime' news site, programmed to create false political stories. The results were impressive—and, in an election year, alarming.

Axios

AI means intelligence agencies must rethink how they work, report finds

AI's advent means the U.S. intelligence community must revamp its traditional way of doing business, according to a new report from an Eric...

The term

"AI"

is everywhere and overhyped

The New York Times

A.I. Has a Measurement Problem

Which A.I. system writes the best computer code or generates the most realistic image? Right now, there's no easy way to answer those...

USA Today

Best Buy will use AI to help customers, cuts costs with layoffs

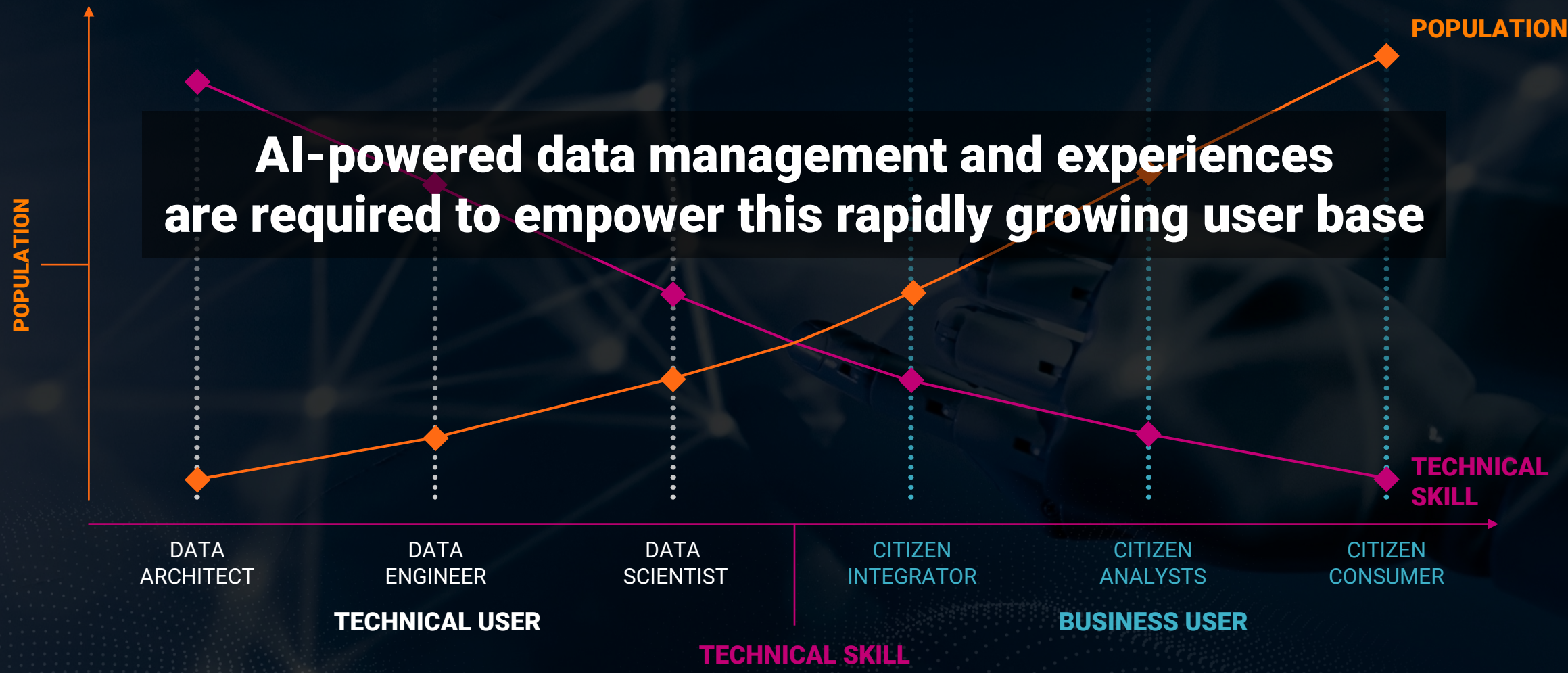
Best Buy has made significant cuts to its workforce, including the Geek Squad. On the heels of the cuts, the Minnesota-based company...

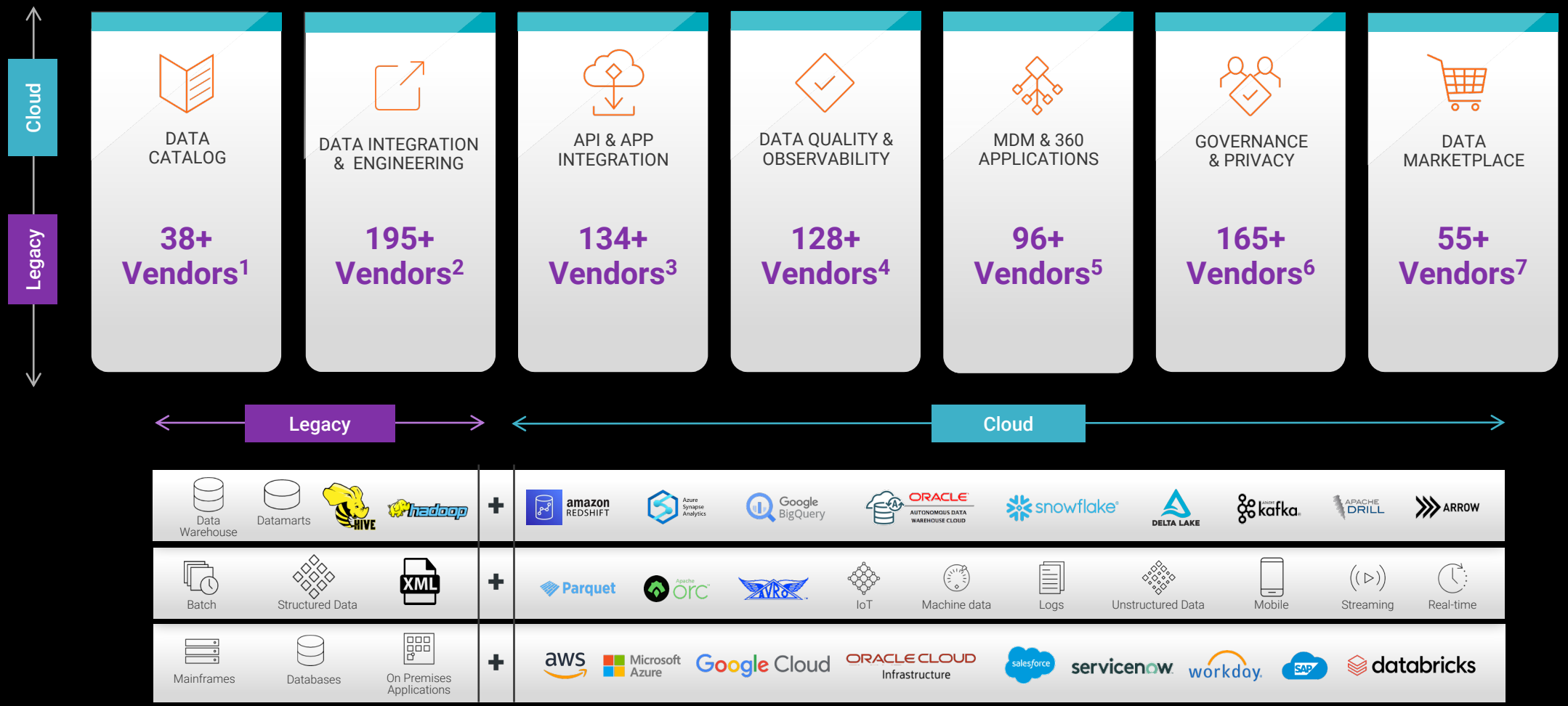
Everybody's ready for  except your data™

Data concerns are the most cited challenges related to GenAI implementations

Source: 2024 Gartner Generative AI Planning Survey

Inability to Address the Expanding Breadth of Data Users





[1]Trust Radius, Data Catalog Software, May 2023
[2]Trust Radius, Data Integration Tools Software, May 2023
[3]Trust Radius, API Management Tools, May 2023
[4]Trust Radius, Data Quality Software, May 2023
[5]Trust Radius Master Data Management (MDM) Tools, May 2023
[6]Trust Radius, Data Governance Software, May 2023
[7]Trust Radius, Data Collaboration Tools, May 2023

DATA CONSUMERS

Data Architect | Data Engineer | Citizen Integrator | Data Scientist | Data Analyst | Business Users

Intelligent Data Management Cloud™

DISCOVER & UNDERSTAND	ACCESS & INTEGRATE	CONNECT & AUTOMATE	CLEANSE & TRUST	MASTER & RELATE	GOVERN & PROTECT	SHARE & DEMOCRATIZE
DATA CATALOG	DATA INTEGRATION & ENGINEERING	API & APP INTEGRATION	DATA QUALITY & OBSERVABILITY	MDM & 360 APPLICATIONS	GOVERNANCE ACCESS & PRIVACY	DATA MARKETPLACE

CLAIRE®
AI-Powered Metadata Intelligence & Automation

Connectivity
Metadata System of Record

DATA SOURCES

SaaS Apps Sources + Mainframe | Applications | Databases | On-premises Sources + IoT | Machine Data | Logs | Real-time / Streaming Sources

Informatica

The Impact of GenAI for Data Management

Allowing the World to Speak the Language of Data

Data for All

Allow everyone in the organization to be data driven, regardless of technical ability

Create Usable Data

Enable decentralised data producers to create, document, catalog, and test new Data Products

Drive Exponential Productivity Growth for Your Data Team

Drive exponential growth in productivity of your existing data professionals with AI

Informatica

Get Ready for GenAI

Donna Burbank

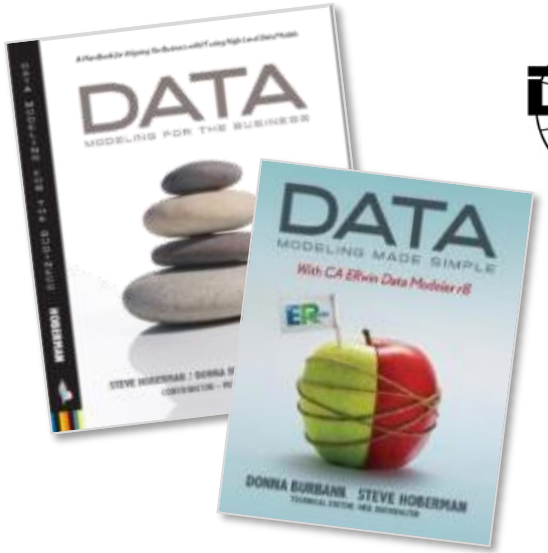
Donna is a recognized industry expert in data management with over 25 years of experience in data strategy, data governance, data modeling, metadata management, and enterprise architecture. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, and business leadership.

She is currently the Managing Director at Global Data Strategy, Ltd., an international data management consulting company that specializes in the alignment of business drivers with data-centric technology.

In past roles, she has served in key brand strategy and product management roles for several of the leading data management products in the market.

As an active contributor to the data management community, she is a long time DAMA International member, contributor to the DMBOK 2.0, Past President and Advisor to the DAMA Rocky Mountain chapter, and was awarded the Excellence in Data Management Award from DAMA International.

She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored several books and is a regular contributor to industry publications. She can be reached at donna.burbank@globaldatastrategy.com
Donna is based in Boulder, Colorado, US.



DATAVERSITY Data Architecture Strategies

This Year's Lineup

- **January** Emerging Trends in Data Architecture – What's the Next Big Thing?
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Master Data Management - Aligning Data, Process, and Governance
- **April** How do Data Governance & Data Architecture Support Each Other?
- **May** The Role of the Chief Data Officer (CDO) in Business Transformation
- **June** What Does It Mean to be a Data-Driven Organization?
- **July** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Enterprise Architecture vs. Data Architecture
- **December** The Business Benefits of Data Modeling

What We'll Cover Today

- The increasing focus on data in organizations has **increased demand for critical roles such as data architect, data engineer, and data scientist**.
- **But there is often confusion and ambiguity around what these roles entail**, and what overlap exists between them.
- This webinar will **discuss these data-centric roles and their place in the data-driven organization**.

Audience

- There exists a great deal of confusion and differing terminology in the data management industry.
- This webinar has generated a great deal of pre-interest from, at a minimum, two main audiences:

"How do I get the right mix of skills on my team?"

"How do I find someone who understands my business?"

"Where are the right people to help us build our data-driven vision?"

Those Hiring

"How do I position my skills effectively?"

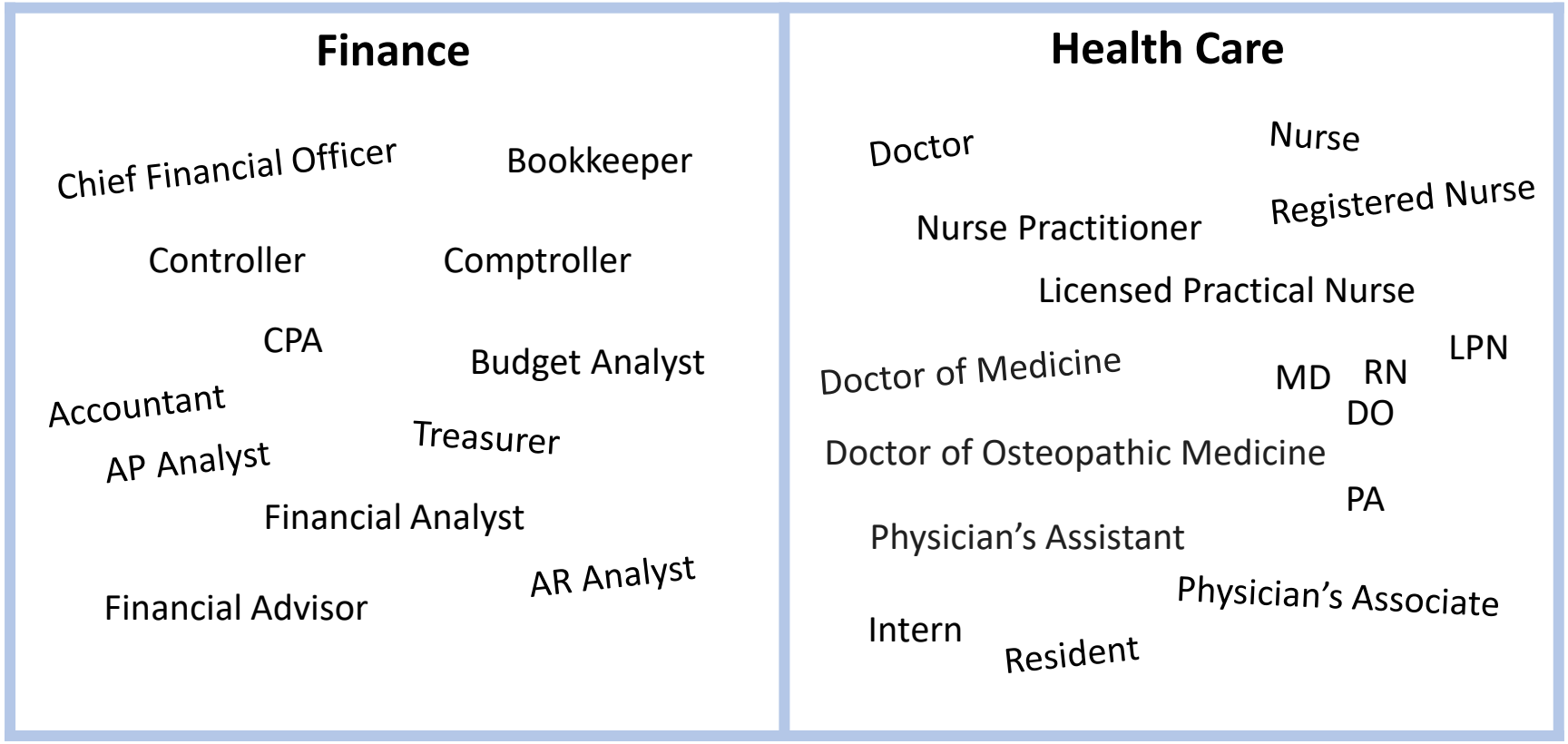
"How do I find the right role that fits my strengths and interests?"

"What's the right company to help me grow?"

Those Looking for Work

We're Not Alone

- A similar issue exists in other industries as well.
- Can we easily articulate the difference between the following? Where do their tasks & skills overlap?



What Are the Standard Roles in a Data Organization?

It Depends ….

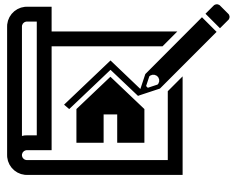
Consultant

OK – Kidding! ☺

While there are certainly differences in every organization …

… There are certain common best practices that exist.

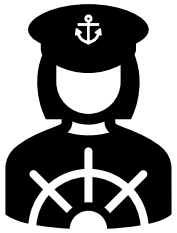
Let’s Break This Into Role “Patterns”



Design, Maintain & Build
Design, build, and maintain the data platform & ecosystem.

Govern & Orchestrate
Support the governance of and accountability for data across the ecosystem.

Analyze & Explore
Produce trends & insights from data via reporting & analytics.



Own & Use
Use data for business advantage. Accountable for the quality & protection of data.

A Sample Data Ecosystem

Data Catalog & Metadata Management – data lineage, data dictionary, business glossary, etc.

Operational Data

- Structured/Relational Data Storage
  - Registration System
  - Facilities Mgt
  - Order Mgt
  - Etc.
- Sensor Data
- Log files
- Social Media
- Video

Data Lake/ Raw Landing

- Structured Data Storage
- Unstructured Data Storage

ELT

Streaming

Structured Data Storage

- Structured format for trending over time
- Facts & Dimensions (often)

Transformation

Semantic Layer

- Business-friendly view
- Common terms & calculations
- Ability to easily "slice & dice"

Reporting

- Business Intelligence Reporting

Advanced Analytics

- AI & ML
- LLM
- Graph Relationships
- Etc…

Discovery & Analysis

Publish & Subscribe

Master Data

- Core data cleansed & governed for reuse (Customer, Product, Vendor)

Modern Data Platform (Fabric)

Data Governance – roles, organization, policies, standards, etc.

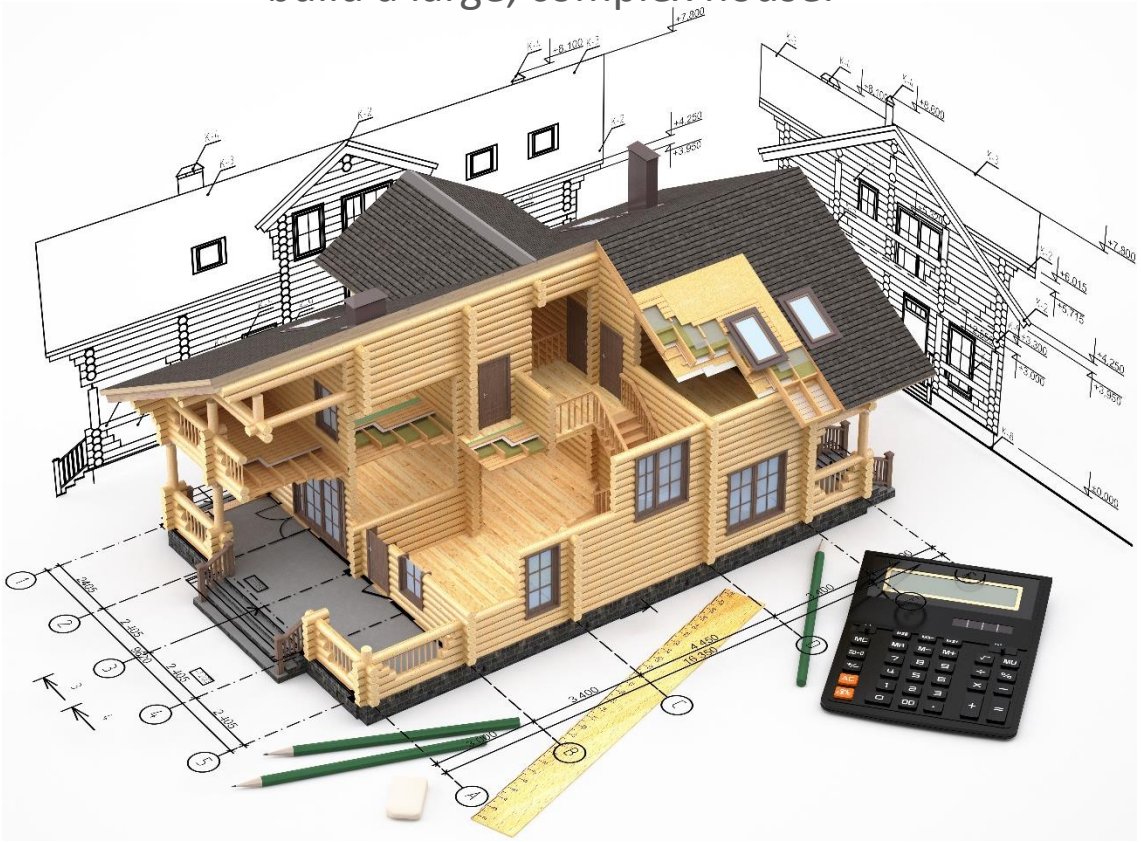
Design, Maintain & Build

Roles & Capabilities, Not Headcount

- The following descriptions describe **roles & capabilities , not headcount**
- In a small organization, they may be embodied in a single person.
- In a larger organization, there may be dozens of staff filling separate roles

Just as a single person might design and build a backyard shed

But it takes many in various roles to build a large, complex house.



A Sample Data Ecosystem – Where Do Roles Fit?

Data Catalog & Metadata Management – data lineage, data dictionary, business glossary, etc.

Operational Data

- Structured/Relational Data Storage

Registration System | Facilities Mgt | Order Mgt | Etc.

- Sensor Data
- Log files
- Social Media
- Video

ELT

Streaming

Data Lake/ Raw Landing

- Structured Data Storage
- Unstructured Data Storage

- Transformation

Structured Data Storage

- Structured format for trending over time
- Facts & Dimensions (often)

Semantic Layer

- Business-friendly view
- Common terms & calculations
- Ability to easily "slice & dice"

Reporting

- Business Intelligence Reporting

Advanced Analytics

- AI & ML
- LLM
- Graph Relationships
- Etc…

- Discovery & Analysis

- Publish & Subscribe

- Core data cleansed & governed for reuse (Customer, Product, Vendor)

Master Data

Modern Data Platform (Fabric)

Data Governance – roles, organization, policies, standards, etc.

A Sample Data Ecosystem – Where Do Roles Fit?

(Data) Solution Architect: Designs the overall data ecosystem and their interactions and document via a solution architecture diagram.

Metadata Architect/Analyst: Designs & manages the metadata classification & storage ecosystem.

Application/System Engineer (IT)

- Manage the performance of system applications to support business operations
- System customizations and maintenance

Data Architect

- Designs the data structures for data storage (e.g. Data Warehouse) to align with business requirements for reporting or other use cases.
- Creates high-level data flow diagrams to align data transformation & movement
- Builds data standards for business semantics and technical formats

Data Engineer

- Builds and maintains data structures for data storage aligned with data architecture design
- Develops data movement pipelines (ETL/ELT) for data transformation.

DQ Analyst

- Designs Data Quality Rules & Dashboards

Data Architect

- Designs master data structures & hierarchies

Reporting

- Business Intelligence Reporting

Advanced Analytics

- AI & ML
- LLM
- Graph Relationships
- Etc…

(Data) Platform Engineer: Manages the performance and reliability of the Data Platform Ecosystem.

Data Governance – roles, organization, policies, standards, etc.

Analyze & Explore

A Sample Data Ecosystem – Where Do Roles Fit?

(Data) Solution Architect: Designs the overall data ecosystem and their interactions and document via a solution architecture diagram.

Metadata Architect/Analyst: Designs & manages the metadata classification & storage ecosystem.

Application/System Engineer (IT)

- Manage the performance of system applications to support business operations
- System customizations and maintenance

Data Architect

- Designs the data structures for data storage (e.g. Data Warehouse) to align with business requirements for reporting or other use cases.
- Creates high-level data flow diagrams to align data transformation & movement
- Builds data standards for business semantics and technical formats

BI Reporting Analyst

- Business Intelligence
- Designs & Builds BI Reports & Dashboards
- Answers Business Questions via Visualizations

Data Engineer

- Builds and maintains data structures for data storage aligned with data architecture design
- Develops data movement pipelines (ETL/ELT) for data transformation.

Data Scientist

- Develops Analytical Models & Reports
- Builds Artificial Intelligence Models
- Provides Insight to Business Questions

DQ Analyst

- Designs Data Quality Rules & Dashboards

Data Architect

- Designs master data structures & hierarchies

(Data) Platform Engineer: Manages the performance and reliability of the Data Platform Ecosystem.

Data Governance – roles, organization, policies, standards, etc.

Govern & Orchestrate

A Sample Data Ecosystem – Where Do Roles Fit?

(Data) Solution Architect: Designs the overall data ecosystem and their interactions and document via a solution architecture diagram.

Metadata Architect/Analyst: Designs & manages the metadata classification & storage ecosystem.

Application/System Engineer (IT)

- Manage the performance of system applications to support business operations
- System customizations and maintenance

Data Architect

- Designs the data structures for data storage (e.g. Data Warehouse) to align with business requirements for reporting or other use cases.
- Creates high-level data flow diagrams to align data transformation & movement
- Builds data standards for business semantics and technical formats

BI Reporting Analyst

- Business Intelligence
- Designs & Builds BI Reports & Dashboards
- Answers Business Questions via Visualizations

Data Engineer

- Builds and maintains data structures for data storage aligned with data architecture design
- Develops data movement pipelines (ETL/ELT) for data transformation.

Data Scientist

- Develops Analytical Models & Reports
- Builds Artificial Intelligence Models
- Provides Insight to Business Questions

DQ Analyst

- Designs Data Quality Rules & Dashboards

Data Architect

- Designs master data structures & hierarchies

(Data) Platform Engineer: Manages the performance and reliability of the Data Platform Ecosystem.

Data Governance Lead: Orchestrates the Data Governance Organization

A Sample Data Ecosystem – Where Do Roles Fit?

(Data) Solution Architect: Designs the overall data ecosystem and their interactions and document via a solution architecture diagram.

Metadata Architect/Analyst: Designs & manages the metadata classification & storage ecosystem.

Application/System Engineer (IT)

- Manage the performance of system applications to support business operations
- System customizations and maintenance

Data Architect

- Designs the data structures for data storage (e.g. Data Warehouse) to align with business requirements for reporting or other use cases.
- Creates high-level data flow diagrams to align data transformation & movement
- Builds data standards for business semantics and technical formats

BI Reporting Analyst

- Business Intelligence
- Designs & Builds BI Reports & Dashboards
- Answers Business Questions via Visualizations

Data Engineer

- Builds and maintains data structures for data storage aligned with data architecture design
- Develops data movement pipelines (ETL/ELT) for data transformation.

Data Scientist

- Develops Analytical Models & Reports
- Builds Artificial Intelligence Models
- Provides Insight to Business Questions

DQ Analyst

- Designs Data Quality Rules & Dashboards

Data Architect

- Designs master data structures & hierarchies

(Data) Platform Engineer: Manages the performance and reliability of the Data Platform Ecosystem.

Data Governance Lead: Orchestrates the Data Governance Organization

A Sample Data Ecosystem – Where Do Roles Fit?

(Data) Solution Architect: Designs the overall data ecosystem and their interactions and document via a solution architecture diagram.

Metadata Architect/Analyst: Designs & manages the metadata classification & storage ecosystem.

Application/System Engineer (IT)

- Manage the performance of system applications to support business operations
- System customizations and maintenance

Data Architect

- Designs the data structures for data storage (e.g. Data Warehouse) to align with business requirements for reporting or other use cases.
- Creates high-level data flow diagrams to align data transformation & movement
- Builds data standards for business semantics and technical formats

BI Reporting Analyst

- Business Intelligence
- Designs & Builds BI Reports & Dashboards
- Answers Business Questions via Visualizations

Data Engineer

- Builds and maintains data structures for data storage aligned with data architecture design
- Develops data movement pipelines (ETL/ELT) for data transformation.

Data Scientist

- Develops Analytical Models & Reports
- Builds Artificial Intelligence Models
- Provides Insight to Business Questions

DQ Analyst

- Designs Data Quality Rules & Dashboards

Data Architect

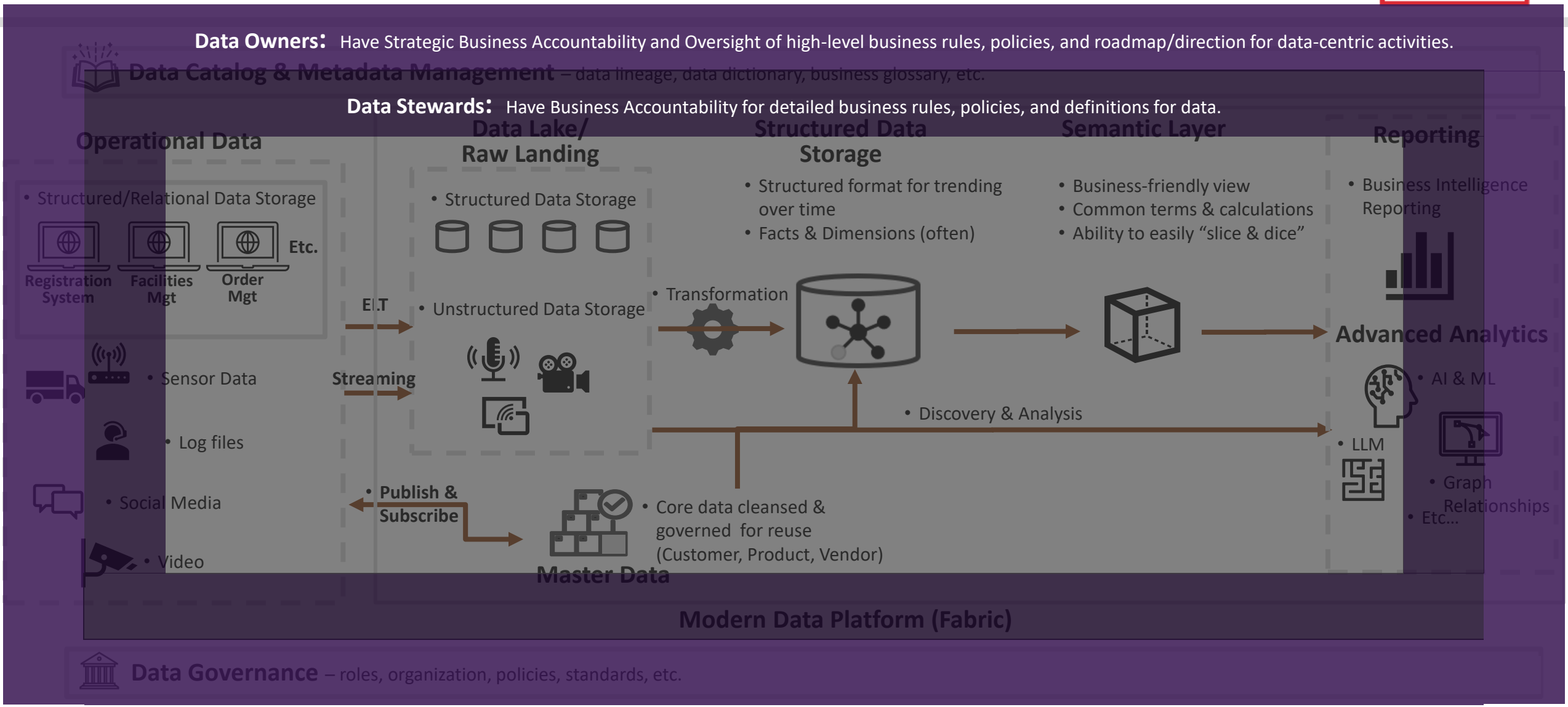
- Designs master data structures & hierarchies

(Data) Platform Engineer: Manages the performance and reliability of the Data Platform Ecosystem.

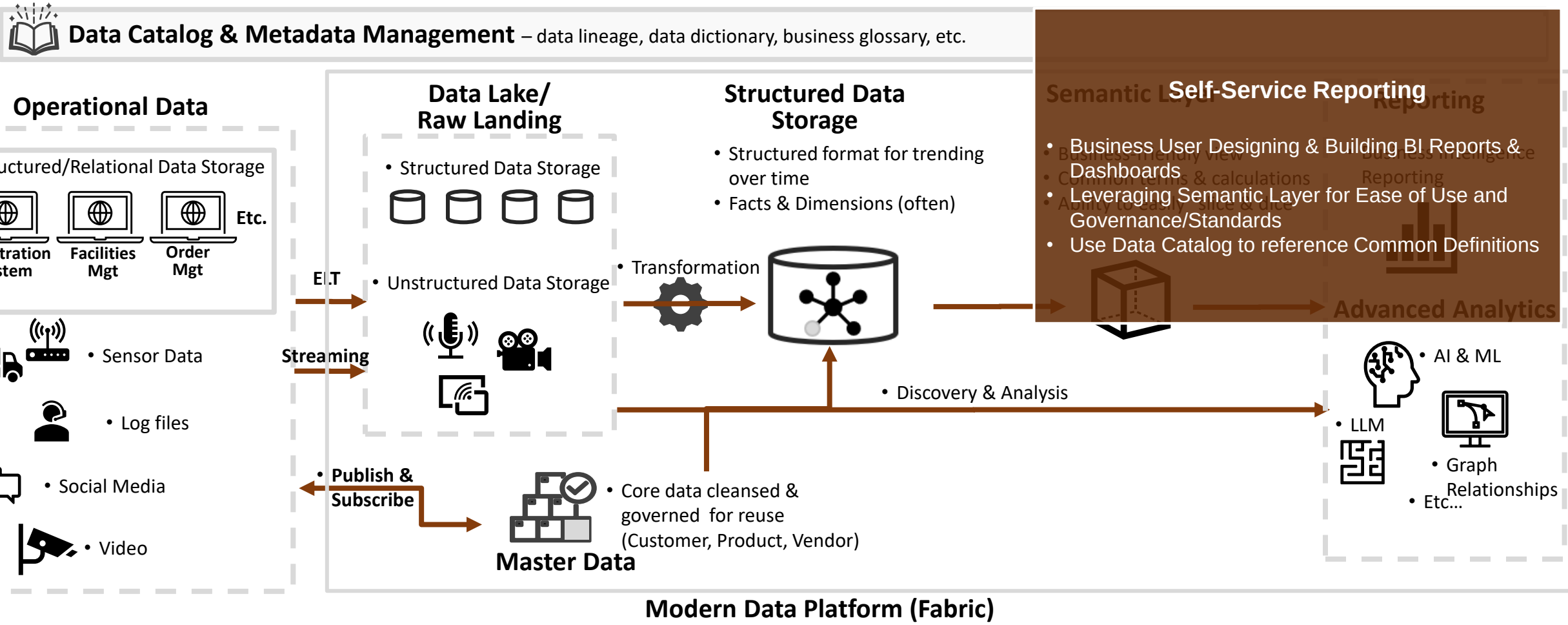
Data Governance Lead: Orchestrates the Data Governance Organization

Own & Use

A Sample Data Ecosystem – Where Do Roles Fit?

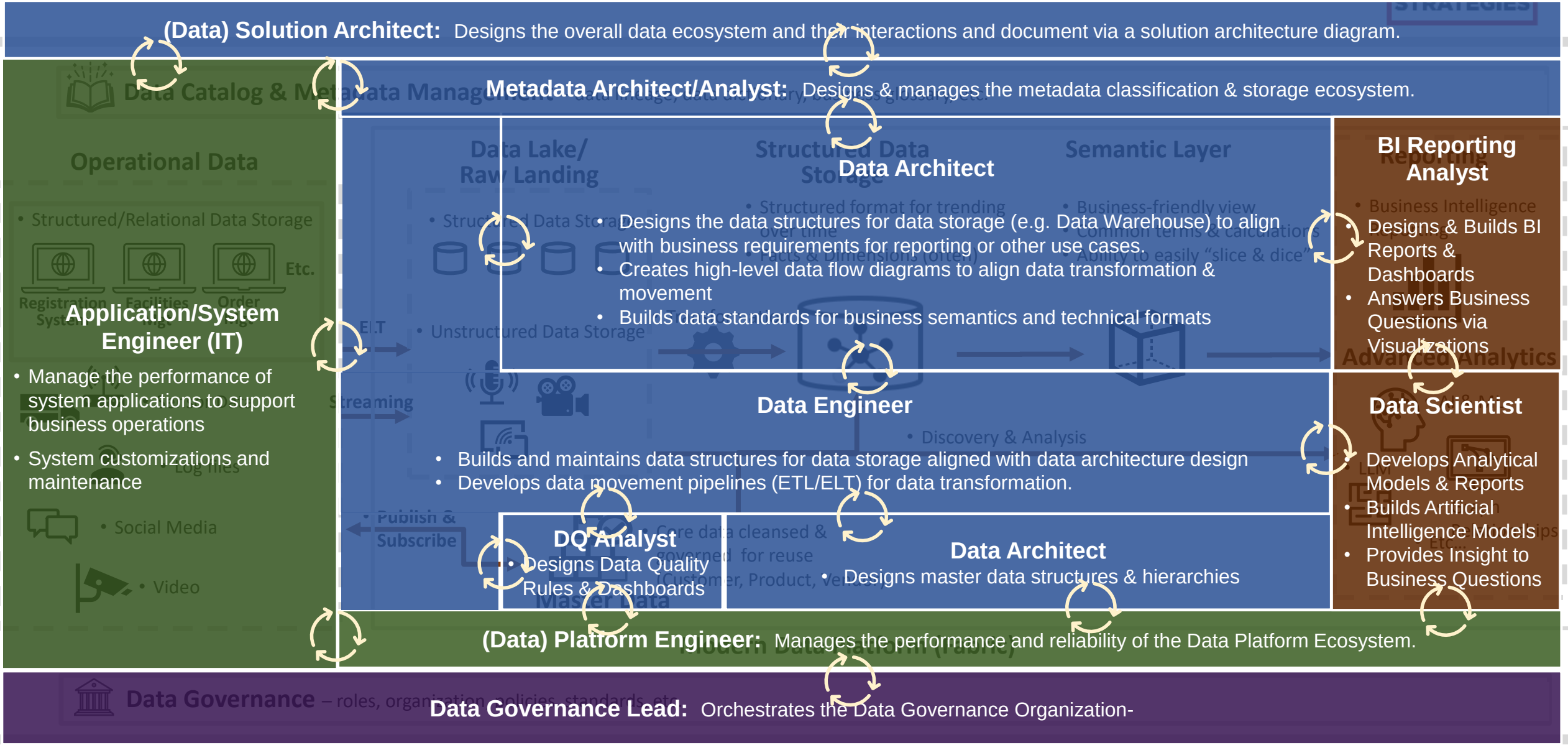


A Sample Data Ecosystem – Where Do Roles Fit?



Execution / Implementation

Roles Work Together through Touchpoints/Deliverables/SLAs



Summary

- While there can be confusion around Data Roles, there are common patterns for responsibilities
- Responsibilities Designing, Building, Analyzing, Governing, and Using Data should be coordinated through clear, documented SLAs
- Both business and technical stakeholders have a role in creating a successful data ecosystem
- There is no “one size fits all” approach – while there are patterns, each organization is unique

DATAVERSITY Data Architecture Strategies

This Year's Lineup

- **January** Emerging Trends in Data Architecture – What's the Next Big Thing?
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Master Data Management - Aligning Data, Process, and Governance
- **April** How do Data Governance & Data Architecture Support Each Other?
- **May** The Role of the Chief Data Officer (CDO) in Business Transformation
- **June** What Does It Mean to be a Data-Driven Organization?
- **July** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Enterprise Architecture vs. Data Architecture
- **December** The Business Benefits of Data Modeling

Who We Are: Business-Focused Data Strategy

Maximize the Organizational Value of Your Data Investment

In today's business environment, showing **rapid time to value** for any technical investment is critical.

But technology and data can be complex. At Global Data Strategy, **we help demystify technical complexity** to help you:

- Demonstrate the ROI and **business value of data** to your management
- Build a data strategy **at your pace to match your unique culture** and organizational style.
- Create an **actionable roadmap for "quick wins"**, which building towards a long-term scalable architecture.

Global Data Strategy shares experience from some of the largest international organizations scaled to the pace of your unique team.

Global Data Strategy has worked with organizations globally in the following industries:

Finance · Retail · Social Services · Health Care · Education · Manufacturing · Government · Public Utilities · Construction · Media & Entertainment · Insurance …. and more

Thoughts? Ideas?

Questions?