April 15, 2024

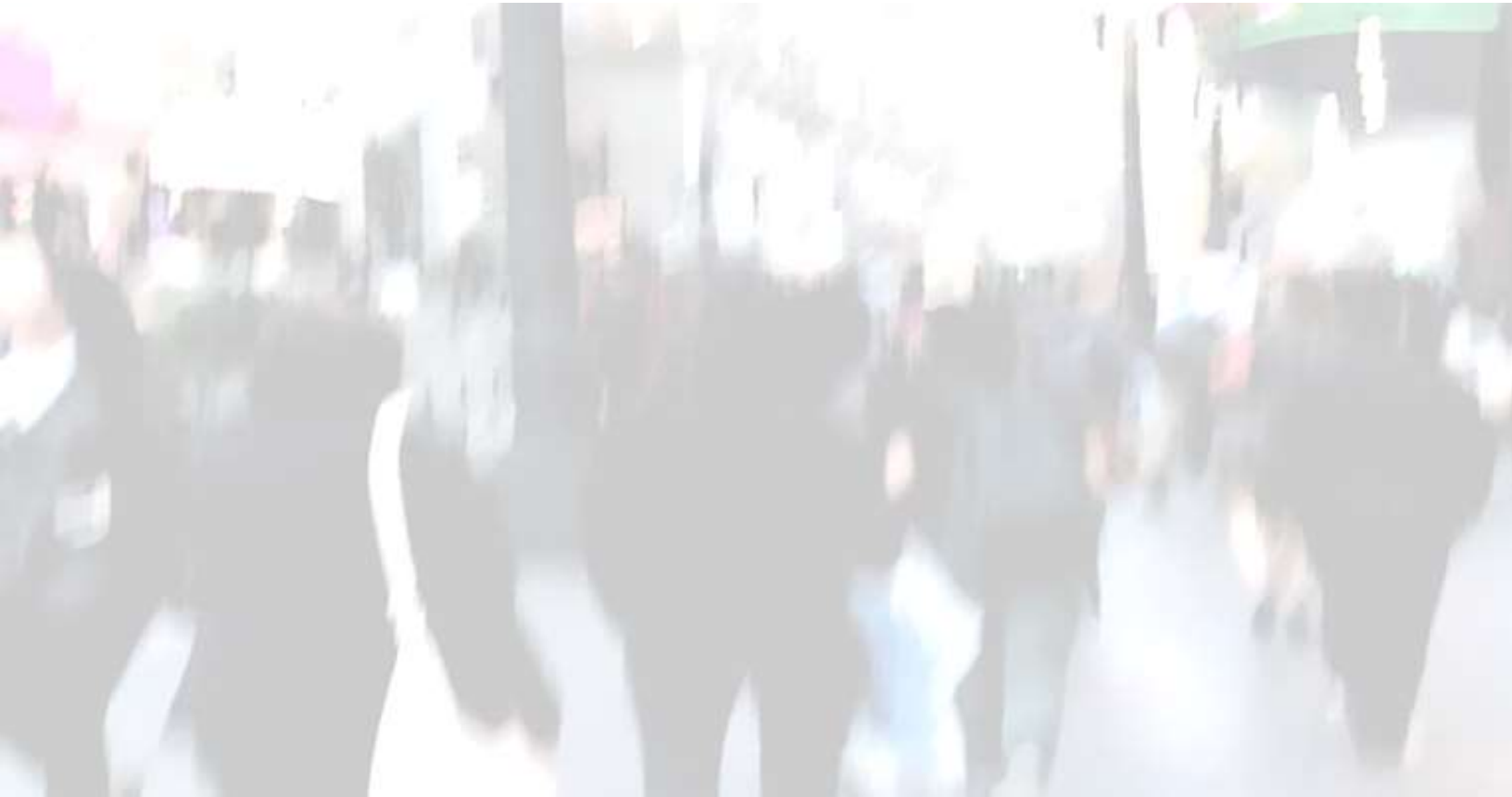
Dresner Advisory Services, LLC

2024 Edition

Active Data Architecture™ Report

Wisdom of Crowds® Series

Licensed to Denodo



Disclaimer:

This report is for informational purposes only. You should make vendor and product selections based on multiple information sources, face-to-face meetings, customer reference checking, product demonstrations, and proof-of-concept applications.

The information contained in this Wisdom of Crowds® Market Study report is a summary of the opinions expressed in the online responses of individuals that chose to respond to our online questionnaire and does not represent a scientific sampling of any kind. Dresner Advisory Services, LLC shall not be liable for the content of this report, the study results, or for any damages incurred or alleged to be incurred by any of the companies included in the report as a result of the report's content.

Definitions

Business Intelligence Defined

Business Intelligence (BI) is “knowledge gained through the access and analysis of business information.”

Business Intelligence tools and technologies include query and reporting, OLAP (online analytical processing), data mining and advanced analytics, end-user tools for ad hoc query and analysis, and dashboards for performance monitoring.

Definition source: Howard Dresner, The Performance Management Revolution: Business Results Through Insight and Action (John Wiley & Sons, 2007)

Active Data Architecture Defined

Active Data Architecture™ supports a platform-independent layer that sits between physical data stores and points of data consumption. It is comprised of various data management capabilities including virtualized and distributed data access, data governance, and security. At its core, Active Data Architecture is an abstraction layer translating business and physical structures. It is an architecture dynamically optimized for performance, scalability, and cost management.

A fundamental use case of active data architecture capabilities is the idea of data products (or data as a product) —elevating the notion of data inherent value for achieving specific business outcomes. An active data architecture helps to elevate the status and importance of data to the level of a “product” by separating the management, governance, and use of data from the specific technical systems in which it may be housed. In essence, an active data architecture provides (among other things) a layer of abstraction enabling data to be managed and applied in an application-independent manner.

Being an architecture, not a single product or piece of technology, an organization can build an active data architecture based on componentry coming from various existing markets that are well established—for example, data integration, data engineering, data governance, metadata management, operational and analytical data infrastructure, and more. This can be done with a very narrow and focused subset of these capabilities (for narrowly defined use cases and requirements) or in a much broader way to support wide-ranging needs. Different buyers will emphasize different capabilities when seeking technology to build out their active data architectures.

The terms “data mesh” and “data fabric” are often associated with the concept of active data architecture. Data mesh and data fabric typically refer to strategies for specifically

enabling data access across distributed, diverse, and connected data sources. Many practitioners use the terms interchangeably and often perceive them as specifically associated with data virtualization (enabling an abstract layer for simplifying distributed data access and query). There are many conflicting definitions of these terms. For the most part, fabric and mesh are about linking and accessing data—they do not address the broader requirements of active data architecture, such as modeling, metadata management, and data governance.

Introduction

As we mark the 17th anniversary of Dresner Advisory Services in 2024, we are pleased to present the first edition of this report.

We extend our sincere appreciation to our valued clients and partners for your consistent support and motivation. Since our inception in 2007, our focus has been on setting and surpassing high standards, driving innovation, and leading the market in providing increasing value each year.

This is the first edition Active Data Architecture Report, in which we explore market requirements and priorities for data orchestration, integration, and transformations, including advanced analytics in the active data architecture pipeline workflow.

We hope that this report serves as a valuable resource for you and your organization, offering fresh insights and perspectives on the evolving landscape of data architecture.

Thank you for your continued support, as we look forward to embarking on this new chapter together.

Best,

Chief Research Officer
Dresner Advisory Services

Contents

Benefits of the Study

The 2024 Dresner Advisory Services Active Data Architecture™ Report provides a wealth of information and analysis, offering value to both consumers and producers of business intelligence technology and services.

Consumer Guide

As an objective source of industry research, consumers use the Dresner Advisory Active Data Architecture Report to understand how their peers leverage and invest in collaborative BI and related technologies.

Using our unique vendor performance measurement system, users glean key insights into BI software supplier performance, which enables:

- Comparisons of current vendor performance to industry norms
- Identification and selection of new vendors

Supplier Tool

Vendor licensees use the Dresner Advisory Services Active Data Architecture Report in several important ways:

External Awareness

- Build awareness for business intelligence markets and supplier brands, citing Dresner Advisory Services Active Data Architecture Report trends and vendor performance.
- Gain lead and demand generation for supplier offerings through association with Dresner Advisory Services Active Data Architecture Report brand, findings, webinars, etc.

Internal Planning

- Refine internal product plans and align with market priorities and realities, as identified in the Dresner Advisory Services Active Data Architecture Report
- Better understand customer priorities, concerns, and issues
- Identify competitive pressures and opportunities

About Howard Dresner and Dresner Advisory Services

The Dresner Advisory Services Active Data Architecture Report was conceived, designed, and executed by Dresner Advisory Services, LLC—an independent advisory firm—and Howard Dresner, its President, Founder and Chief Research Officer.

Howard Dresner is one of the foremost thought leaders in business intelligence and performance management, having coined the term "Business Intelligence" in 1989. He has published two books on the subject, *The Performance Management Revolution – Business Results through Insight and Action* (John Wiley & Sons, Nov. 2007) and *Profiles in Performance – Business Intelligence Journeys and the Roadmap for Change* (John Wiley & Sons, Nov. 2009). He lectures at forums around the world and is often cited by the business and trade press.

Prior to Dresner Advisory Services, Howard served as chief strategy officer at Hyperion Solutions and was a research fellow at Gartner, where he led its business intelligence research practice for 13 years.

Howard has conducted and directed numerous in-depth primary research studies over the past three decades and is an expert in analyzing these markets.

Through the Wisdom of Crowds® Business Intelligence market research reports, we engage with a global community to redefine how research is created and shared. Other research reports include:

- Wisdom of Crowds® Flagship BI Market Study
- AI, Data Science and Machine Learning
- Analytical Data Infrastructure (ADI) Flagship
- Analytical Platforms
- Cloud Computing and Business Intelligence
- Data Catalog
- Data Engineering
- Data Governance
- Embedded BI
- Master Data Management (MDM)
- ModelOps
- Self-Service BI

You can find more information about Dresner Advisory Services at www.dresneradvisory.com.

The Dresner Team

About Elizabeth Espinoza
Elizabeth is Research Director at Dresner Advisory and is responsible for the data preparation, analysis, and creation of charts for Dresner Advisory reports.

About Kathleen Goolsby
Kathleen is Senior Editor at Dresner Advisory ensuring the quality and consistency of all research publications.

About Danielle Guinebertiere
Danielle is the Director of Client Services at Dresner Advisory. She supports the ongoing research process through her work with executives at companies included in Dresner market reports.

About Michelle Whitson-Lorenzi
Michelle is Client Services Manager and is responsible for managing software company survey activity and our internal market research data.

Survey Method and Data Collection
As with all our Wisdom of Crowds® Market Studies, we constructed a survey instrument to collect data and used social media and crowdsourcing techniques to recruit participants.

We include our research community of over 6,000 organizations as well as crowdsourcing and vendors' customer communities.

Data Quality
We carefully scrutinized and verified all respondent entries to ensure that only qualified participants are included in the study.

Executive Summary

Executive Summary

- Current data architectures and the supporting technology capabilities (i.e., analytic data infrastructure and data engineering and catalog products) are under both technology and business pressures to meet the requirements of ever more complex, diverse, and distributed BI and analytics use cases and applications. Data architectures that lack flexibility, adaptability, and scalability lead to challenges as organizations struggle to capture benefits and achieve positive ROI from their BI investments.
- Active data architecture is about correcting these shortcomings. It is about building supporting capabilities for BI that are flexible, more scalable, and dynamically optimized—ideally, they can automatically change their behaviors in response to current conditions (workloads, data source availability, infrastructure capacity, governance constraints, and so forth).
- Organizations know they need to modernize their data architectures to better support BI / analytics needs, and they know that data fabric/mesh concepts are critical. But the market is confused on what those concepts are, how they relate, and how to act on them.
- Large North American organizations express the greatest interest in active data architecture concepts, with the business services, financial services, and healthcare industries rating the topic as critical or very important with the highest frequency.
- Being an architecture, not a single product or piece of technology, an organization can build an active data architecture based on componentry coming from various existing markets that are well established—for example, data integration, data catalog, data engineering, data governance, metadata management, operational and analytical data infrastructure, and more. This can be done with a very narrow and focused subset of these capabilities (for narrowly defined use cases and requirements) or in a much broader way to support wide-ranging needs. Different buyers will emphasize different capabilities when seeking technology to build out their active data architectures
- A fundamental use case of active data architecture capabilities is the idea of data products (or data as a product) —elevating the notion of data inherent value for achieving specific business outcomes. An active data architecture helps to elevate the status and importance of data to the level of a “product” by separating the management, governance, and use of data from the specific technical systems in which it may be housed. In essence, an active data architecture provides (among other things) a layer of abstraction enabling data to be managed and applied in an application-independent manner.

- The recognition of the need to modernize data architectures is correlated with degrees of BI success—organizations indicating they are actively working toward the goal of an active data architecture also report higher levels of success with their BI investments.
- Across geographies, industries, and business functions, there is no clear winner in terms of definitions or priorities between data fabric and data mesh. This creates significant challenges for technology providers seeking to tap this emerging area of demand. Organizations must apply significant effort to educating buyers on adaptive data architecture concepts, the roles of data fabric and data mesh, and the benefits that can be captured by moving in this direction.
- Data catalog (and broader metadata management) and data virtualization are viewed as the most important enabling technologies for deploying an active data architecture. Survey data show vendors are making near-term (next 12 months) significant investments in these areas.
- The center of gravity for data architecture clearly shifted to the cloud—cloud/SaaS deployments are the clear preference for organizations working toward active data architecture (including private cloud and hybrid / multi-cloud deployments).
- A majority of the data integration workloads supporting current data architectures remains bulk/batch (ETL) in nature. This indicates organizations face substantial efforts to modernize and diversify (into real time and virtualization) the integration capabilities in support of active data architecture.
- It is critical for organizations to realize that creating an active data architecture is not a “rip and replace” exercise. Rather, it is an evolution of existing analytic data infrastructure (data integration capabilities, data warehouses, etc.).
- There is a growing recognition of the importance of governance capabilities in a modern data architecture. Specifically, organizations view security, privacy, and quality controls as the highest-priority governance needs for active data architecture.
- Vendors of data integration technology are the most frequent choices when buyers seek to source technology in support of their active data architecture work. Following closely in preference are vendors of BI and analytics technology. However, organizations basing their active data architecture work on data integration vendors’ technology tend to be much more successful in their BI implementations (presumably due to greater flexibility to adapt data delivery to changing needs, support multiple BI and analytics tools, and deal with a broader range of metadata sources).

Study Demographics

Study Demographics

Study participants provide a cross-section of responses across geographies, functions, organization sizes, and vertical industries. We constructed cross-tab analyses using the data to identify and illustrate important industry preference, priorities, and trends.

Geography

North America, which includes the United States, Canada, and Puerto Rico, represents the largest group with over 48 percent of all respondents, followed by EMEA at over 33 percent. Asia Pacific and Latin America account for the balance (over 13 percent and 4 percent respectively) of respondents (fig.1)

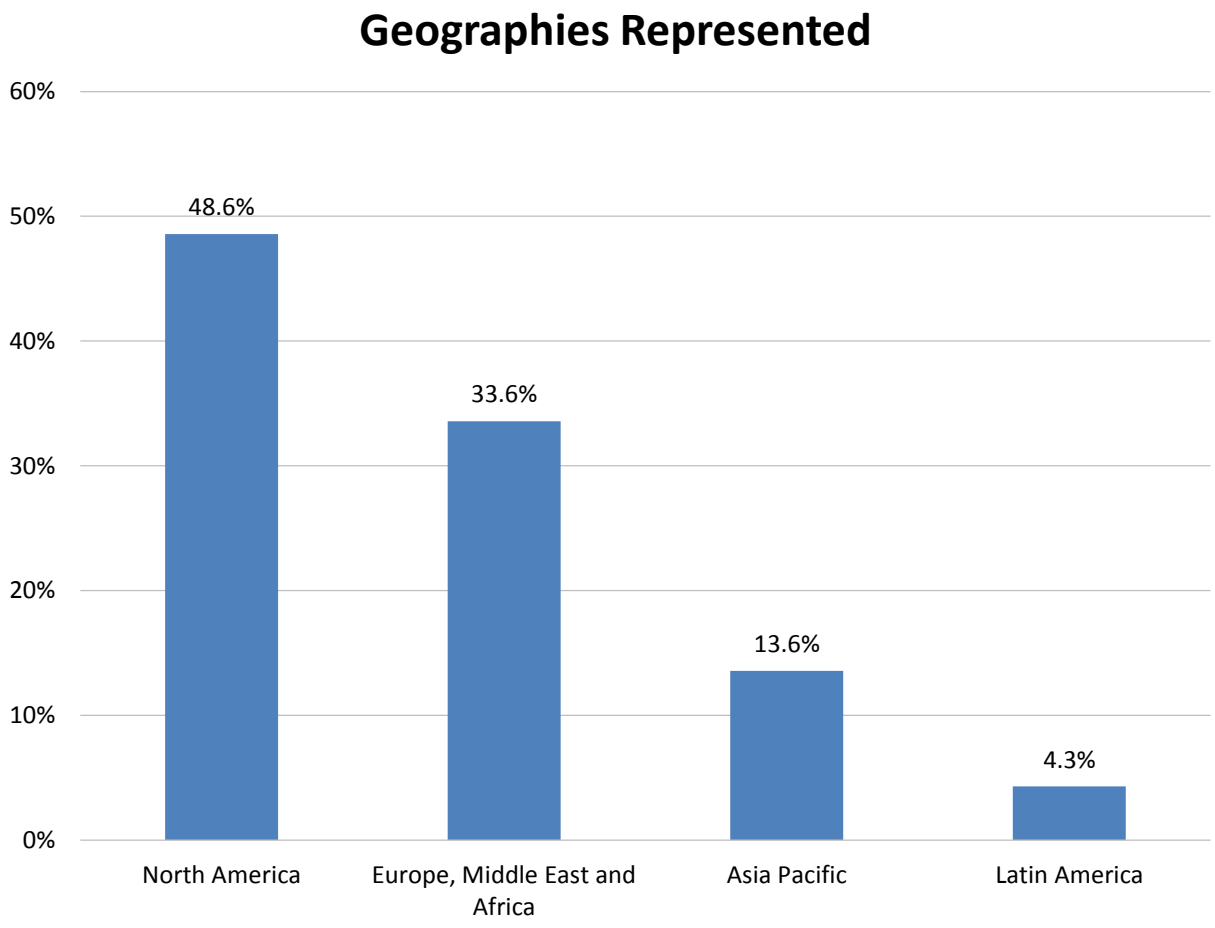


Figure 1 – Geographies represented

Functions

In 2024, respondents from IT functions account for 33 percent of our sample, followed by the Business Intelligence Competency Center (BICC) at 22 percent of the respondents (fig. 2). Participants from several business functions also responded, including data science functions, R&D, finance, operations, and executive management.

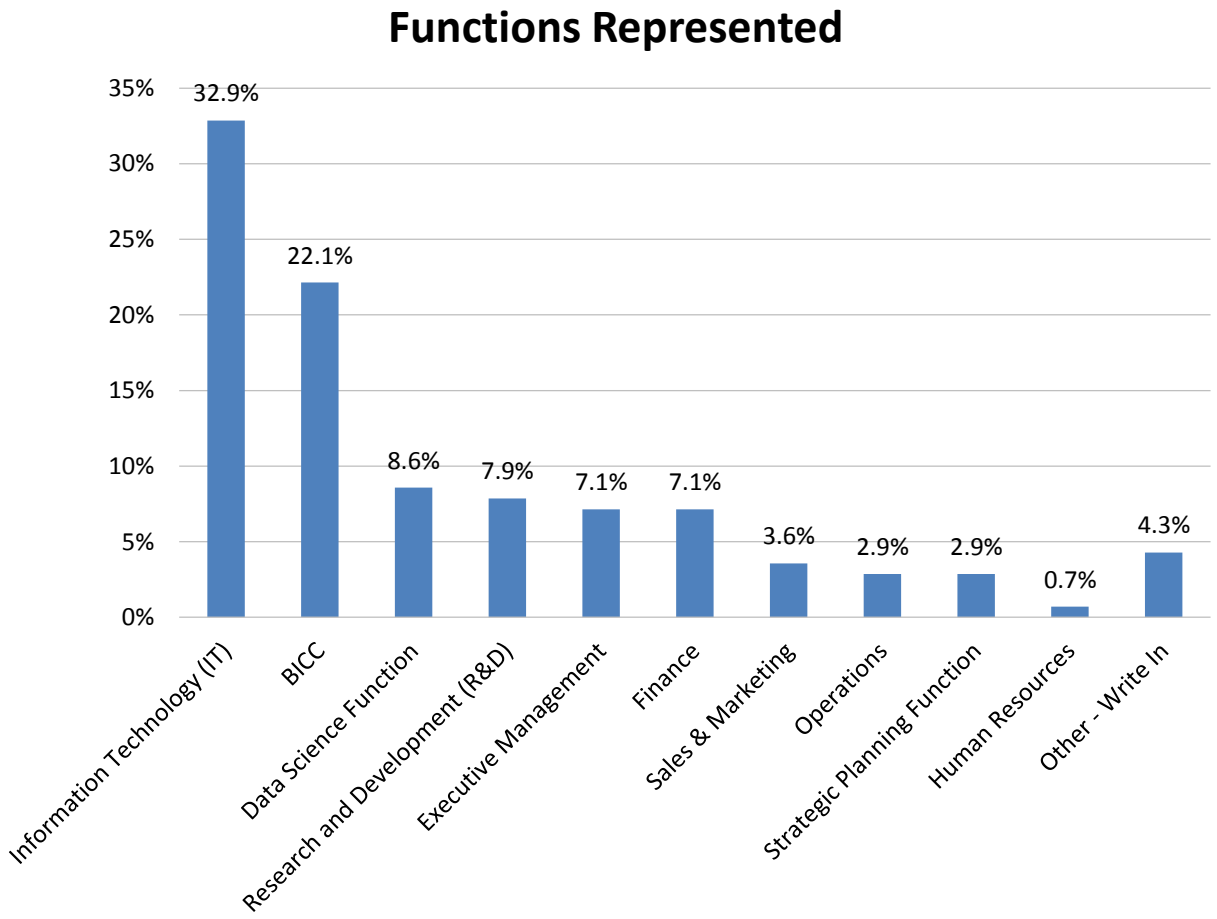


Figure 2 – Functions represented

Vertical Industries

In 2024, financial services (which includes banking and insurance) represents the largest industry segment of the sample, at 22 percent, followed by business services (which includes consulting, telecommunications, and transportation respondents) at 21 percent of respondents (fig. 3). Other key industry segments include technology (16 percent), manufacturing (14 percent), and healthcare (10 percent). The remaining industries represent less than 10 percent each of the sample.

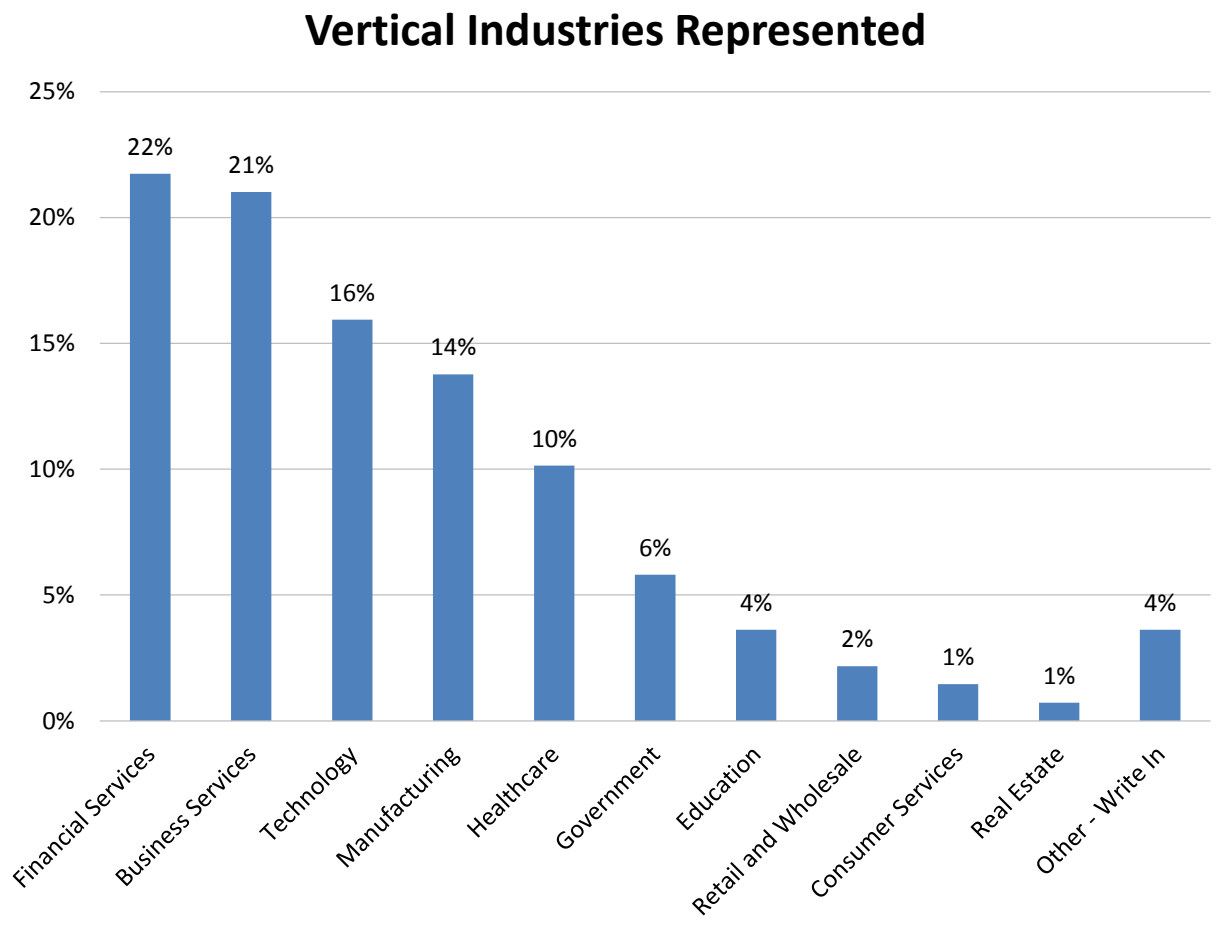


Figure 3 – Vertical industries represented

Organization Size

In 2024, our survey includes small organizations (1-100 employees), midsize organizations (101-1,000 employees), large organizations (1,001-10,000 employees), and very large organizations (>10,000 employees) (fig. 4). This year, small and midsize organizations each account for nearly 22 percent of our sample, while large organizations represent 34 percent, and 23 percent of respondents are from very large organizations.

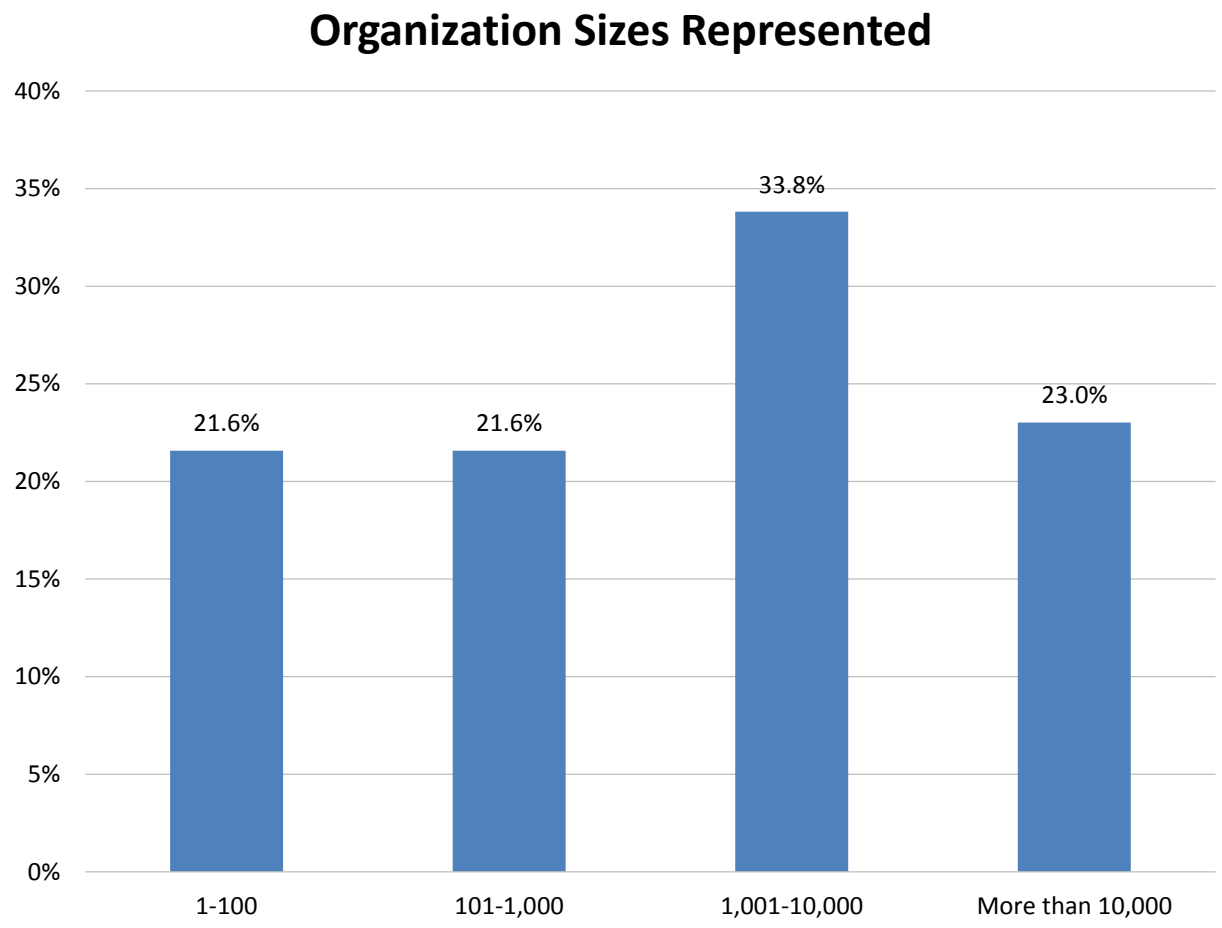


Figure 4 – Organization sizes represented

Analysis & Trends

Analysis and Trends

Importance of Active Data Architecture

We asked survey respondents about the importance of active data architecture concepts in their plans and activities. In our 2024 survey, 26 percent of respondents indicate active data architecture is of *critical* importance (fig. 5), while 39 percent indicate it is *very important*. Slightly less than 7 percent indicate that active data architecture is not important.

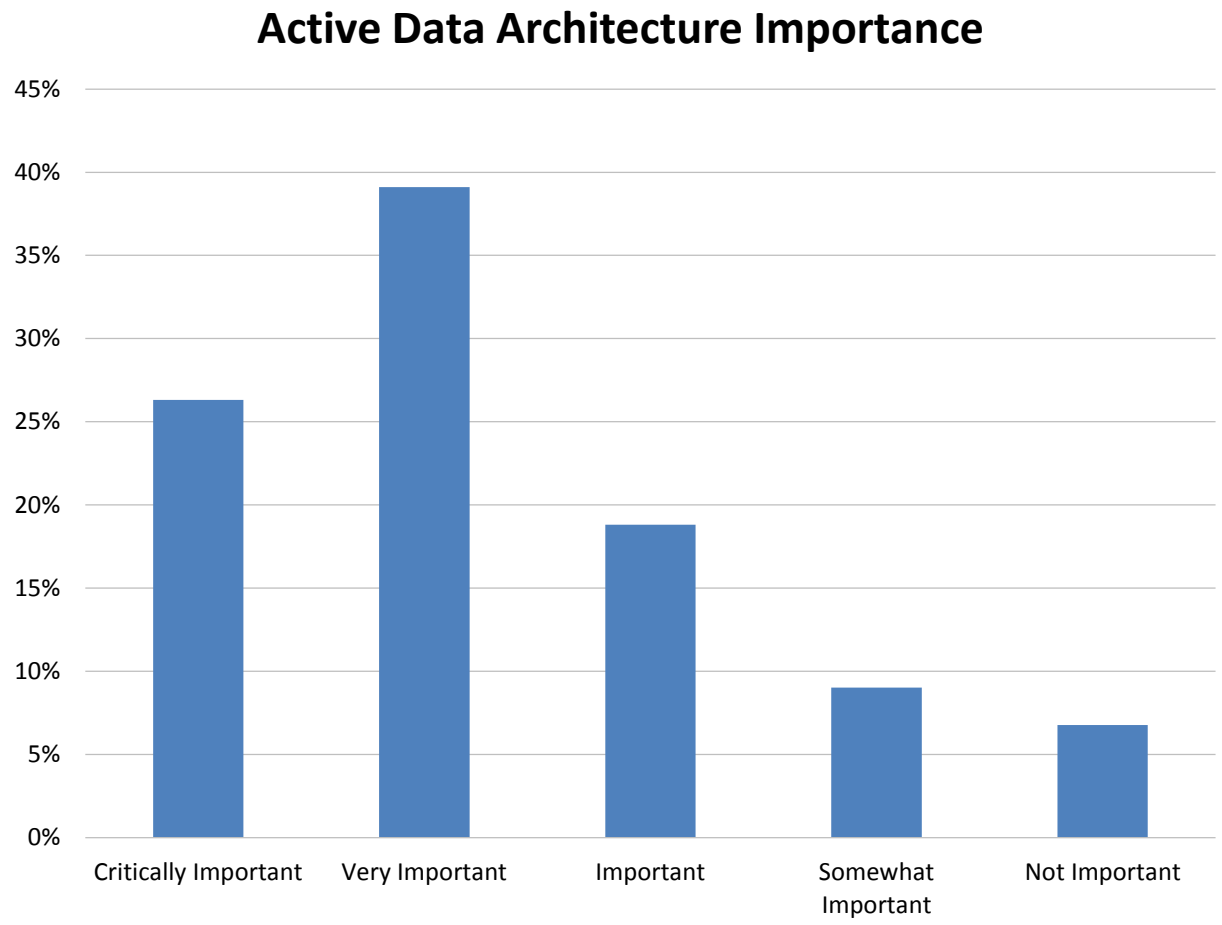


Figure 5 – Active data architecture importance

The degree of importance of active data architecture varies significantly across regions of the world. Organizations in North America view the concept as much more important than other regions, with nearly 72 percent of organizations viewing it as either critically or *very important* (fig. 6). In comparison, only 66 percent of organizations place the same degree of importance on active data architecture in Asia Pacific and 57 percent in EMEA. Latin America expresses the lowest level of importance, with only 40 percent of respondents in the region indicating the topic is critically or very important.

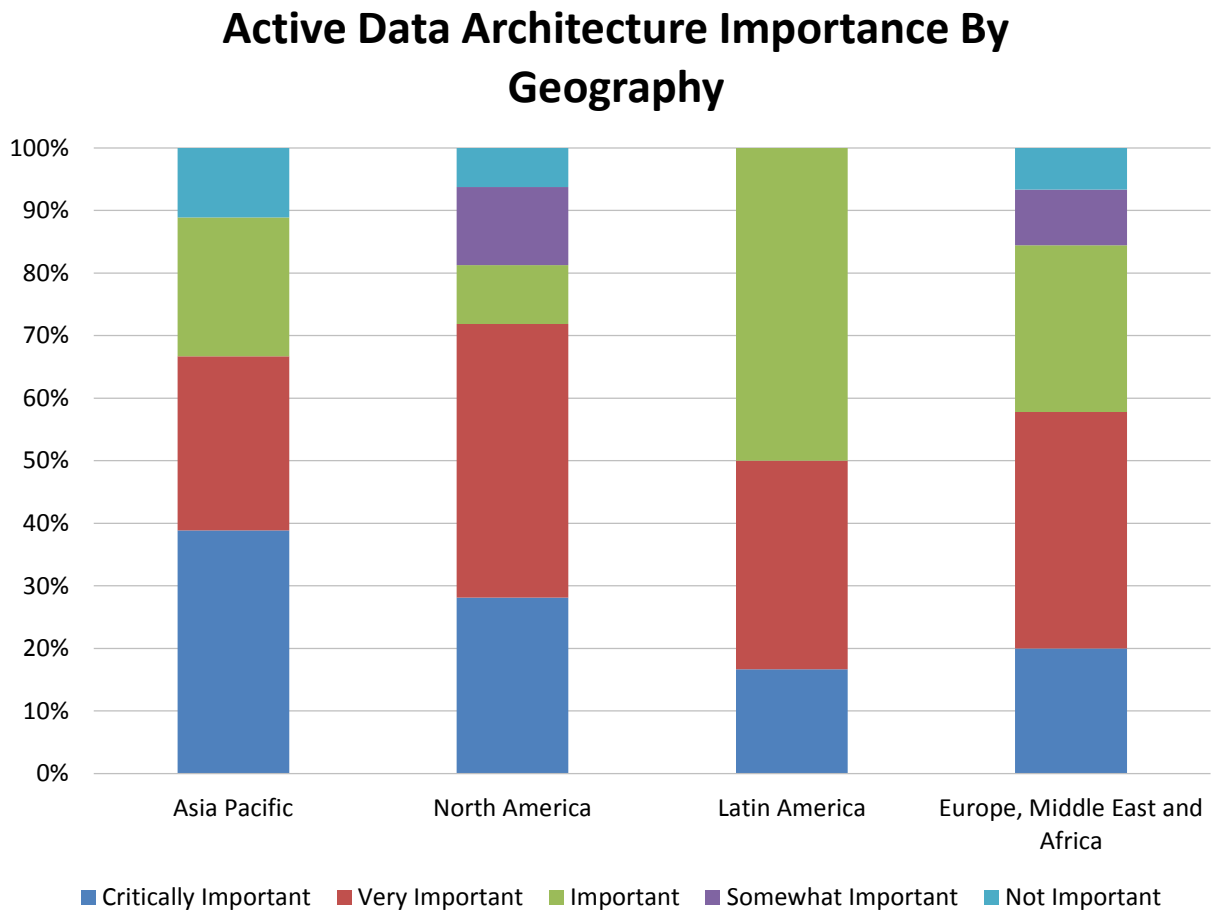


Figure 6 – Active data architecture importance by geography

Respondents from the BICC and data science functions place the highest overall importance on active data architecture. Over 53 percent of survey respondents from BICC functions indicate active data architecture is of critical importance (fig. 7). Thirty-three percent of respondents from data science functions deem it critically important. IT functions see active data architecture as very relevant, with a total of 66 percent rating it as either critically or very important. Among the purer business functions, R&D and executive management view active data architecture as most important.

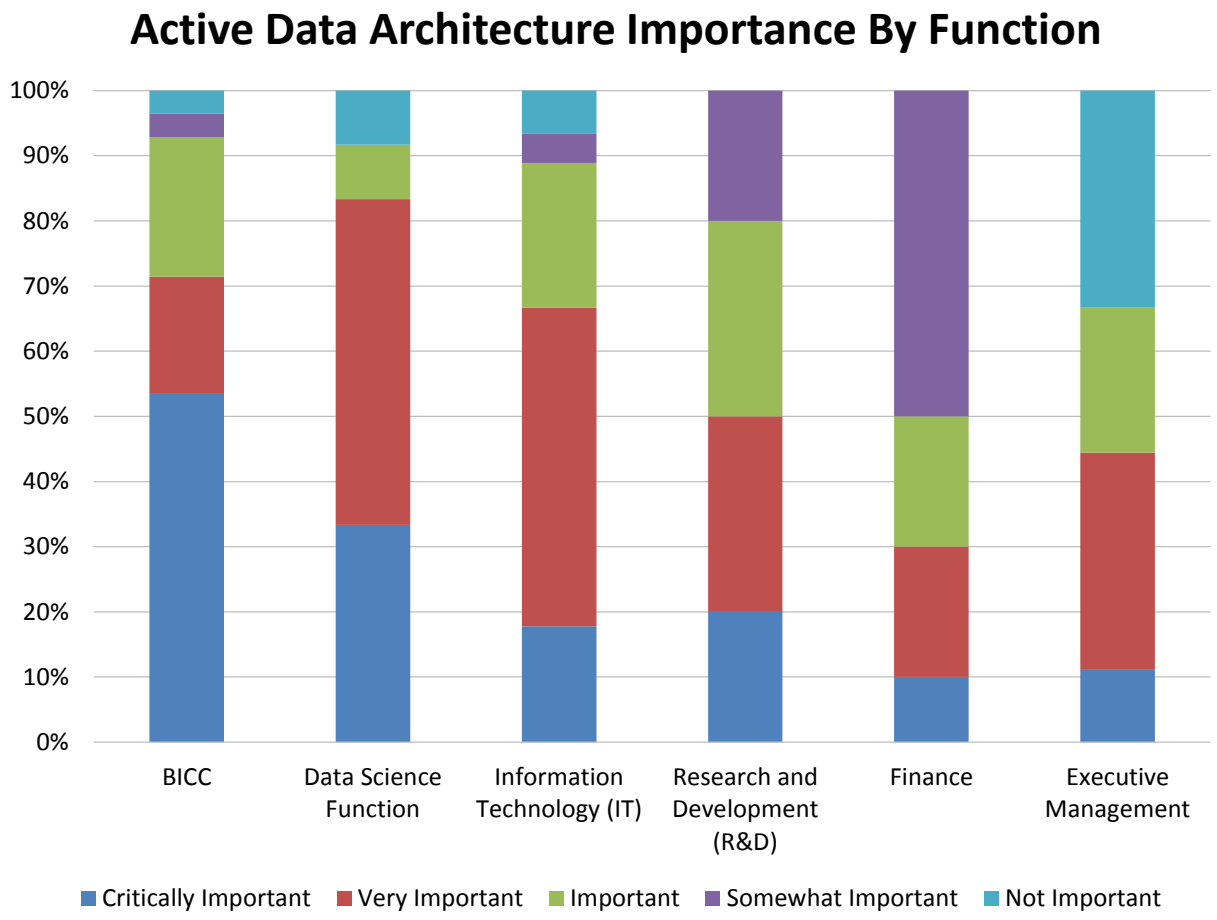


Figure 7 – Active data architecture importance by function

Active data architecture is, for the most part, an increasing priority as organization size increases beyond 1,000 employees, which is likely due to increasing complexity of data-related systems and the desire to focus more heavily on treating data as a product. Seventy-six percent of respondents in the 101-1,000 employee size range rate active data architecture as important, very important, or critically important (fig. 8). This increases to nearly 87 percent with organizations of 1,001-10,000 employees and even further to over 96 percent for organizations with more than 10,000 employees.

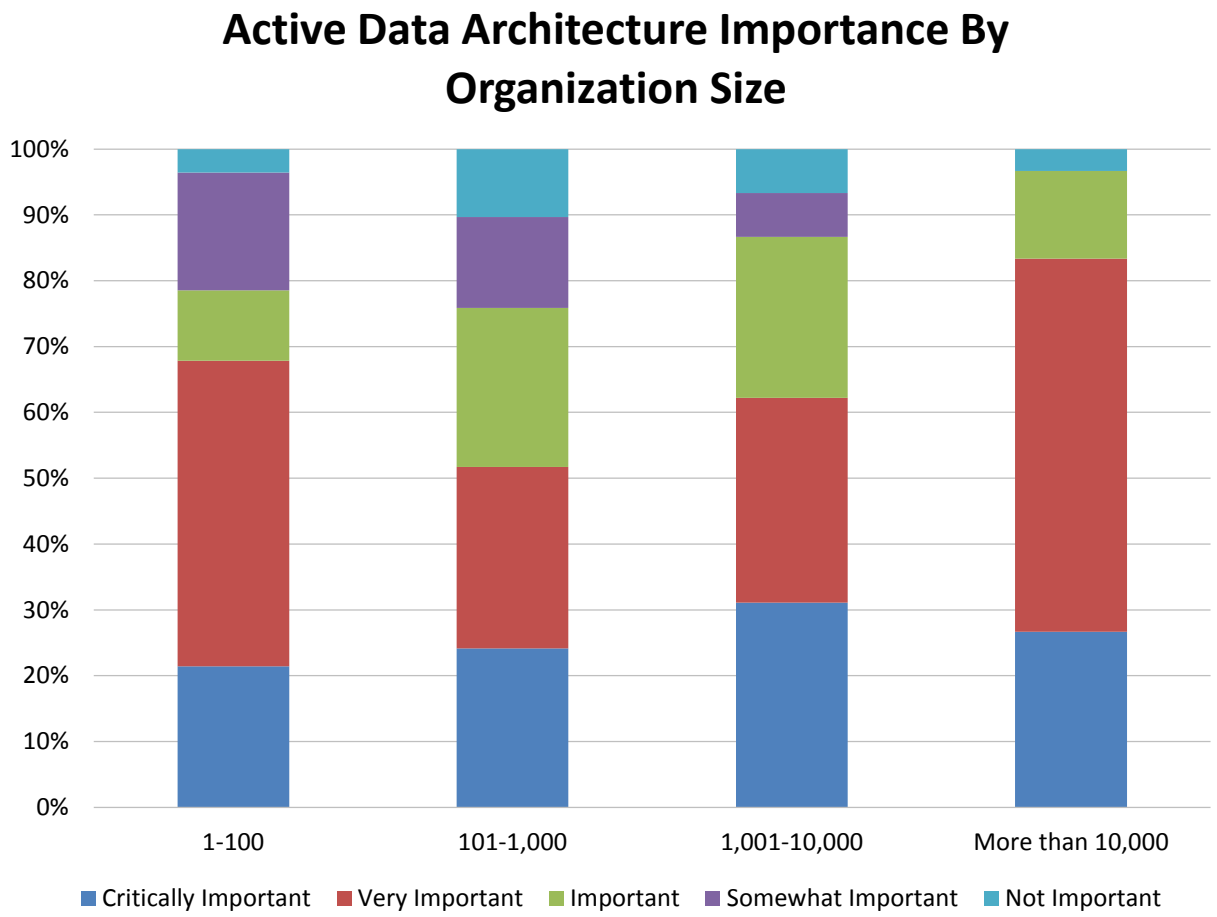


Figure 8 – Active data architecture importance by organization size

Although all industries place a relatively high degree of importance on active data architecture concepts, respondents from industries that are more data intensive (such as business services, financial services, and technology) give it the highest importance ratings (fig. 9). Manufacturing and government respondents give active data architecture lower importance scores.

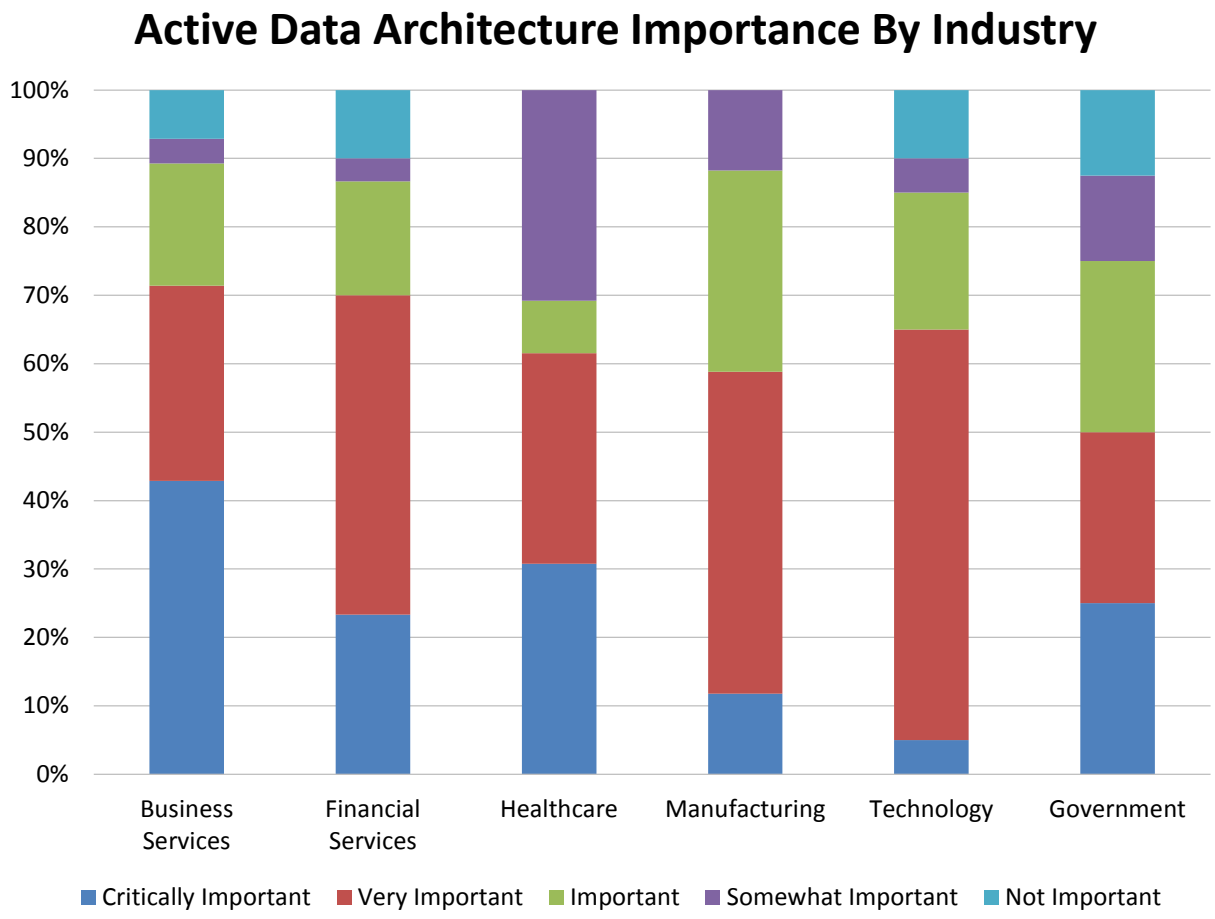


Figure 9 – Active data architecture importance by industry

The buildout of an active data architecture approach to accessing, combining, and preparing data speaks to a degree of maturity and sophistication in leveraging data as a strategic asset (fig. 10). It is not surprising, then, that respondents that rate their BI initiatives as a success place a much higher relative importance on active data architecture concepts, compared to those organizations that are less successful. Sixty-seven percent of organizations rating their BI efforts as extremely successful view active data architecture as critically important. In contrast, only 22 percent of organizations identifying as very successful and moderately successful with BI consider active data architecture at the same level of importance.

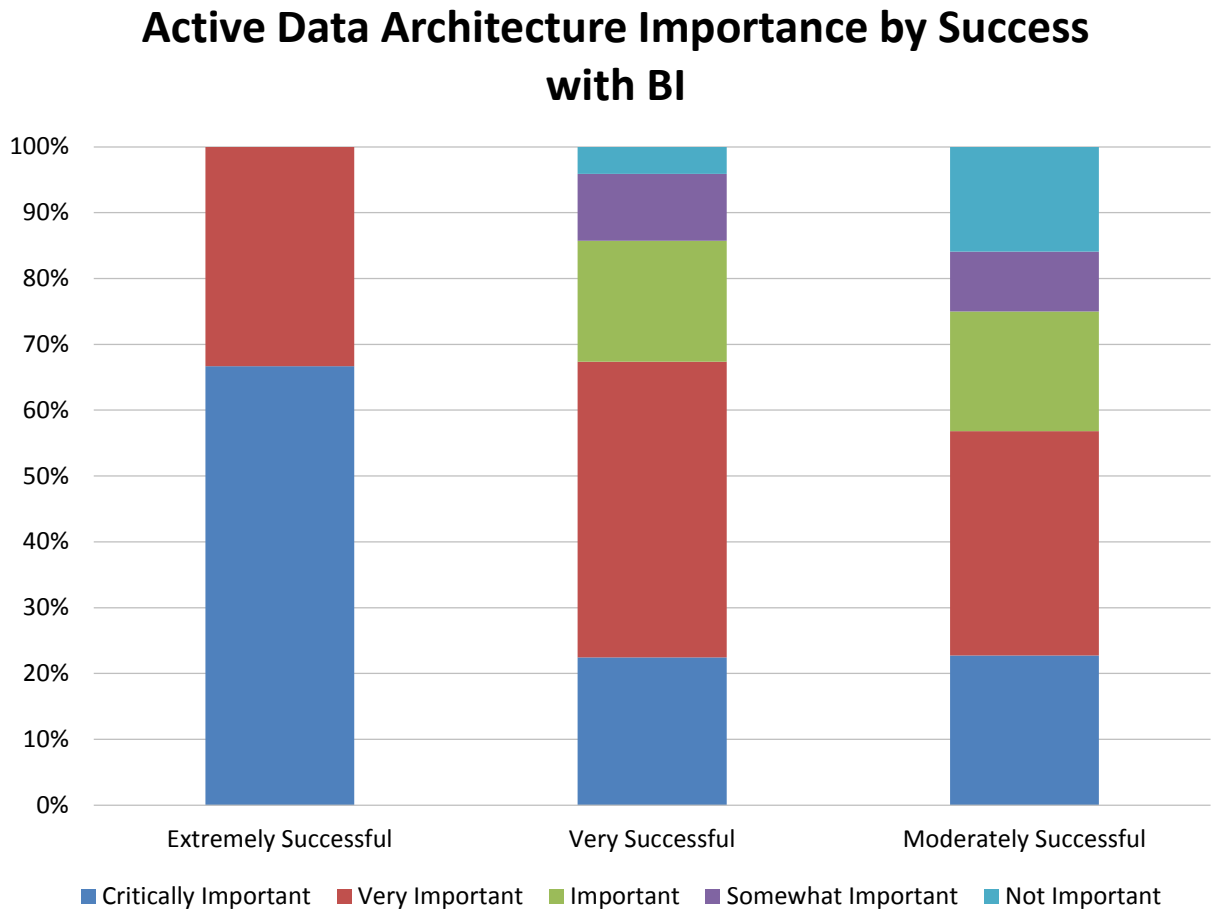


Figure 10 – Active data architecture importance by success with BI

Perception and Understanding of Active Data Architecture's Role

The concepts of data mesh and data fabric are sometimes associated with an active data architecture. They both involve managing distributed data, enabling consolidated data views, and provisioning data to various process and application points of consumption. Data mesh links together distributed data sources and enables these capabilities in a pre-programmed, practitioner-managed, and manually optimized fashion. Data fabric builds on these same capabilities and adds elements of automation to help make the active data architecture truly dynamic, self-organizing, and continually optimized.

Survey respondents clearly express widely ranging views on what these terms mean and how they apply. Nearly 28 percent indicate they have no knowledge of these concepts and how they apply to active data architecture (fig. 11). Another 24 percent see data mesh and data fabric as the same thing and use the terms interchangeably. Twelve percent of surveyed organizations consider data mesh to be their focus and data fabric as irrelevant or redundant; and 8 percent express the opposite, focusing on data fabric without need for data mesh. And 24 percent recognize the need for both ideas.

These data clearly show that a substantial amount of education in the market is required, and technology providers must help customers see the value and benefits of active data architecture and the required capabilities regardless of the terminology.

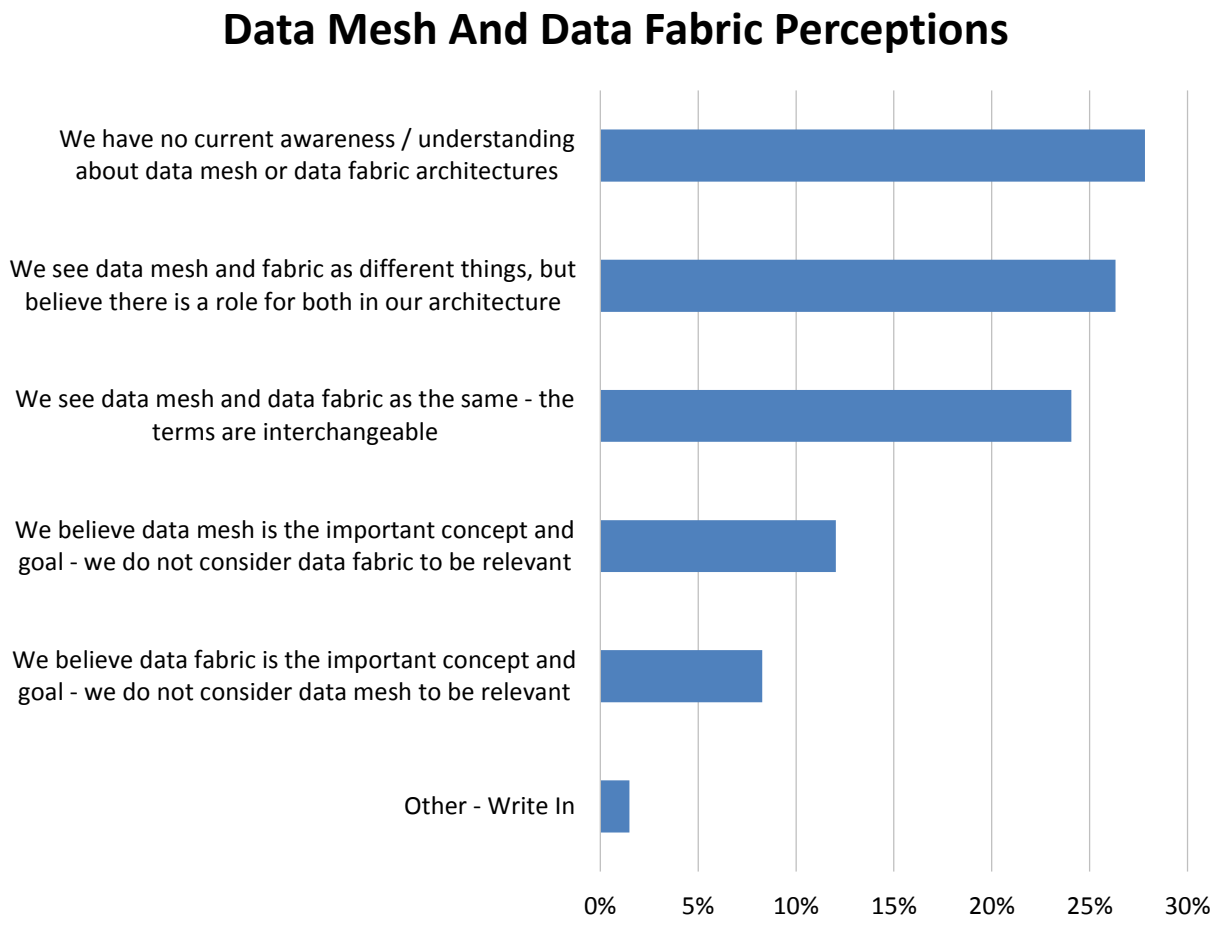


Figure 11 – Data mesh and data fabric perceptions

On a regional basis, we observe the same trends. While awareness about data mesh and data fabric concept is highest in North America (only 23 percent of respondents indicating no knowledge), lack of clarity on the importance and relationship between the concepts is clear (fig. 12). We see similar patterns in EMEA and Latin America. Asia Pacific presents different perceptions, with significantly more organizations (50 percent) expressing no awareness or understanding. For those organizations discerning a difference between data mesh and data fabric, data mesh garners more focus than data fabric in all regions except for Latin America.

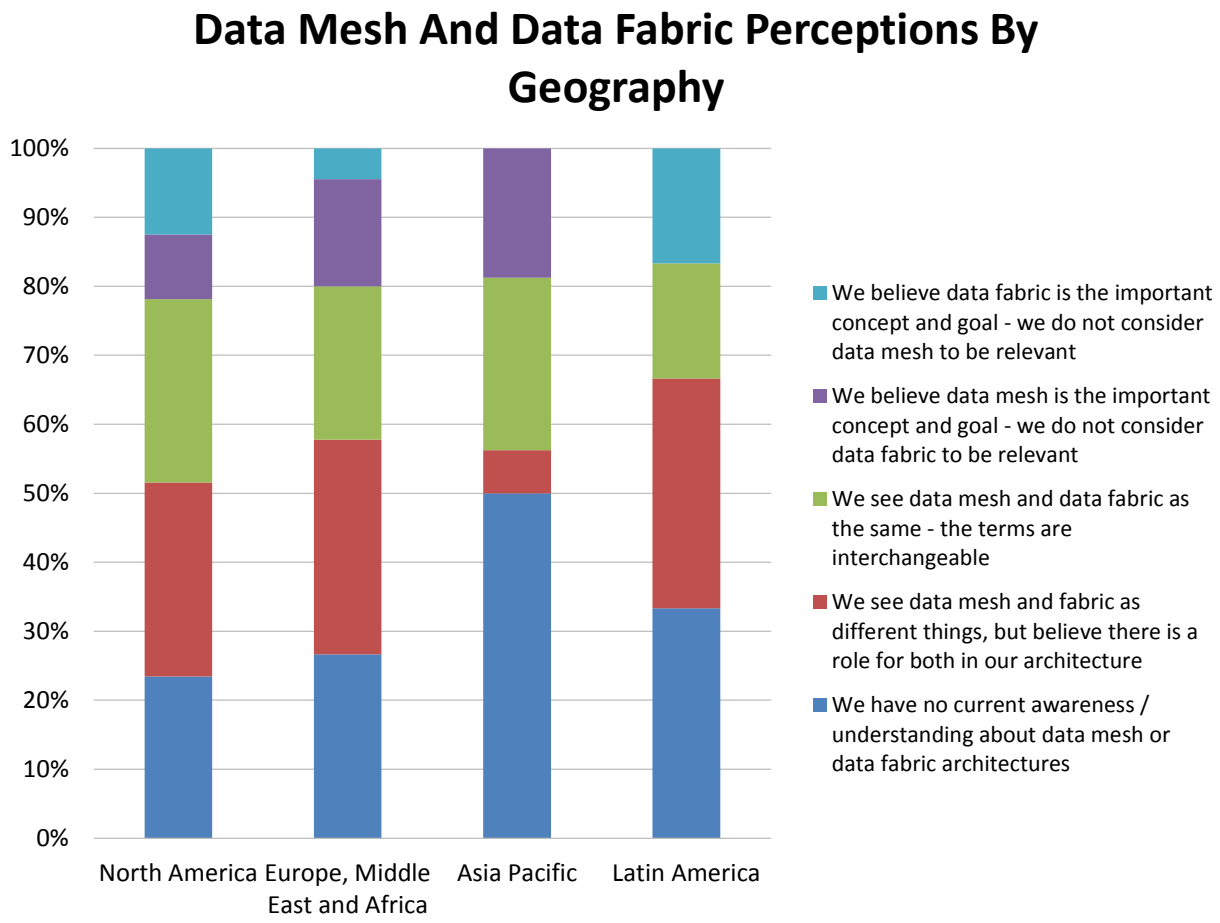


Figure 12 – Data mesh and data fabric perceptions by geography

From an industry perspective, there is wide variation in perceptions of these concepts across industries. Notable is the technology sector, where only less than 5 percent of organizations express no awareness or understanding of data mesh and data fabric concepts, with nearly 29 percent indicating they see a need for both (fig. 13). This sector likely has the highest acceptance of active data architecture principles. In contrast, most other industry segments reflect a diverse mixture of lack of awareness, confusion, or conflict about the terms and no clear direction.

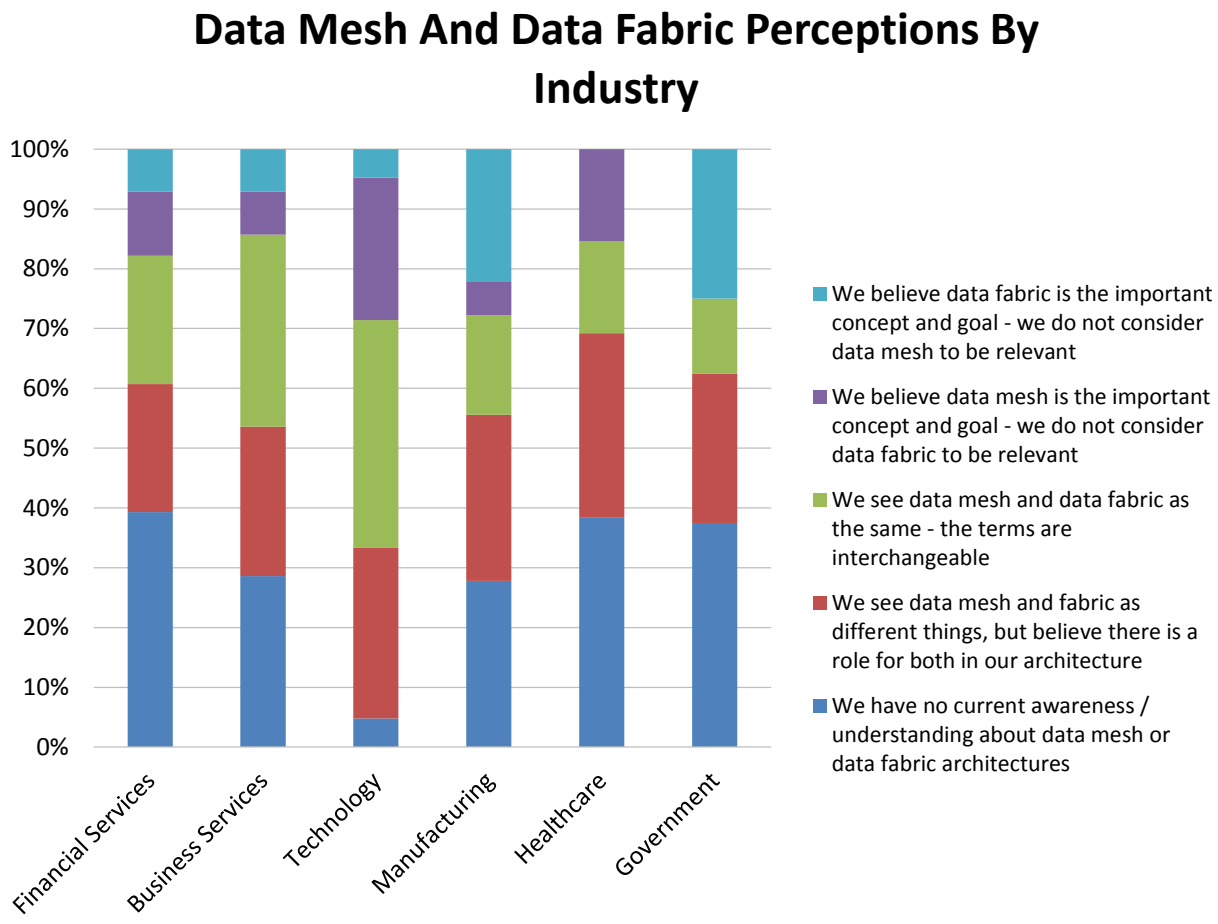


Figure 13 – Data mesh and data fabric perceptions by industry

An even wider variation in perceptions is visible at the business function level. IT and the BICC show the greatest awareness and interest in data mesh and data fabric concepts, with only 27 percent and 13 percent (respectively) of respondents in those functions indicating no awareness or understanding (fig. 14). However, the remainder of respondents in those functions are split in their views about which is more important and whether they are the same or not. Surprisingly, the data science function exhibits a much greater degree of lack of awareness (45 percent). Education here could be key to unlocking a lot of value for a function that has a strong need to gather and connect data across a diverse and shifting landscape.

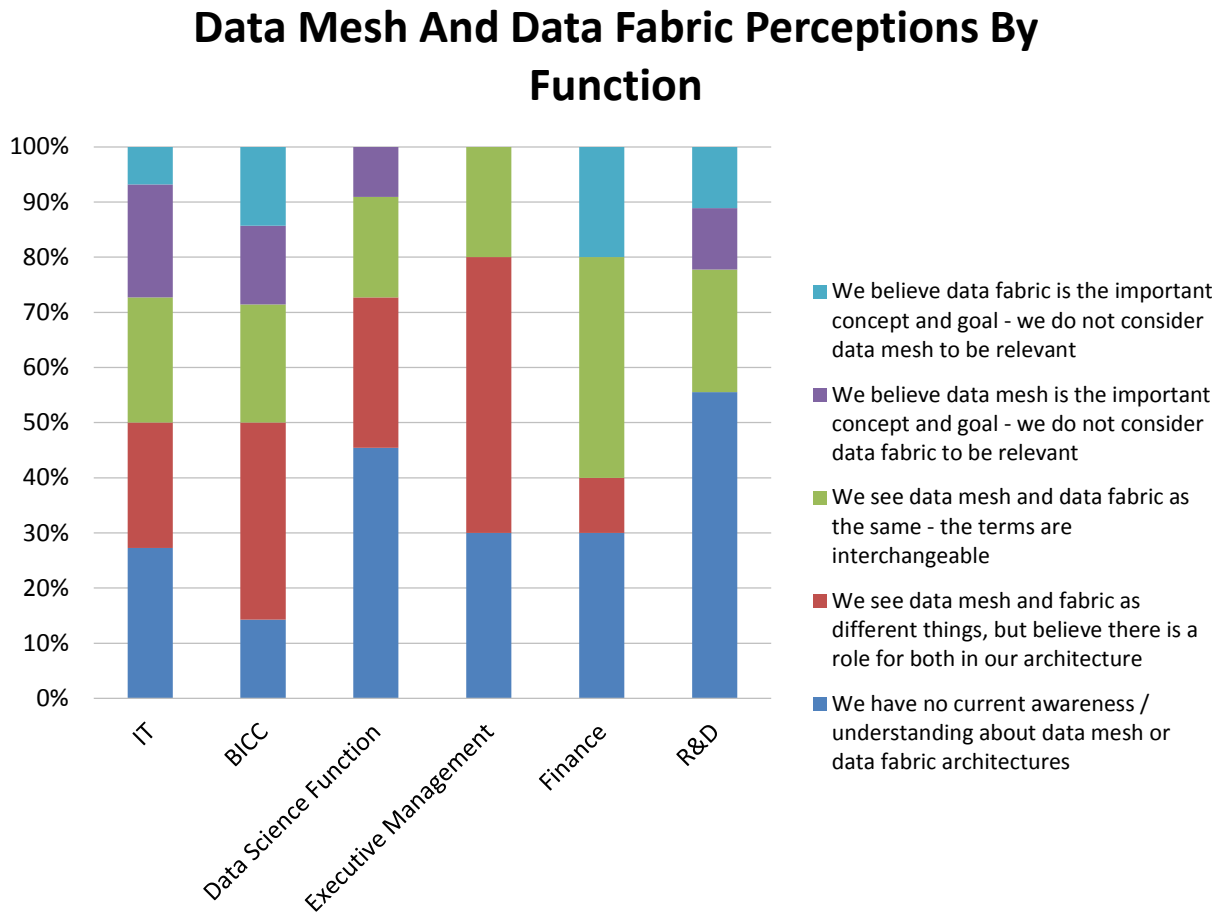


Figure 14 – Data mesh and data fabric perceptions by function

Very large organizations tend to express a greater awareness and understanding of data mesh and data fabric concepts, with only 16 percent of very large (more than 10,000 employees) respondents indicating no awareness and understanding (fig. 15). This is compared to nearly 28 percent in large (1,001-10,000 employees) and 40 percent in midsize (101-1,000) organizations. But aside from general awareness, organization size does not generally indicate consistency of understanding and direction. All organization size categories reflect a mix of perceptions that data mesh and data fabric are interchangeable, different but related, or that one is relevant and the other is not.

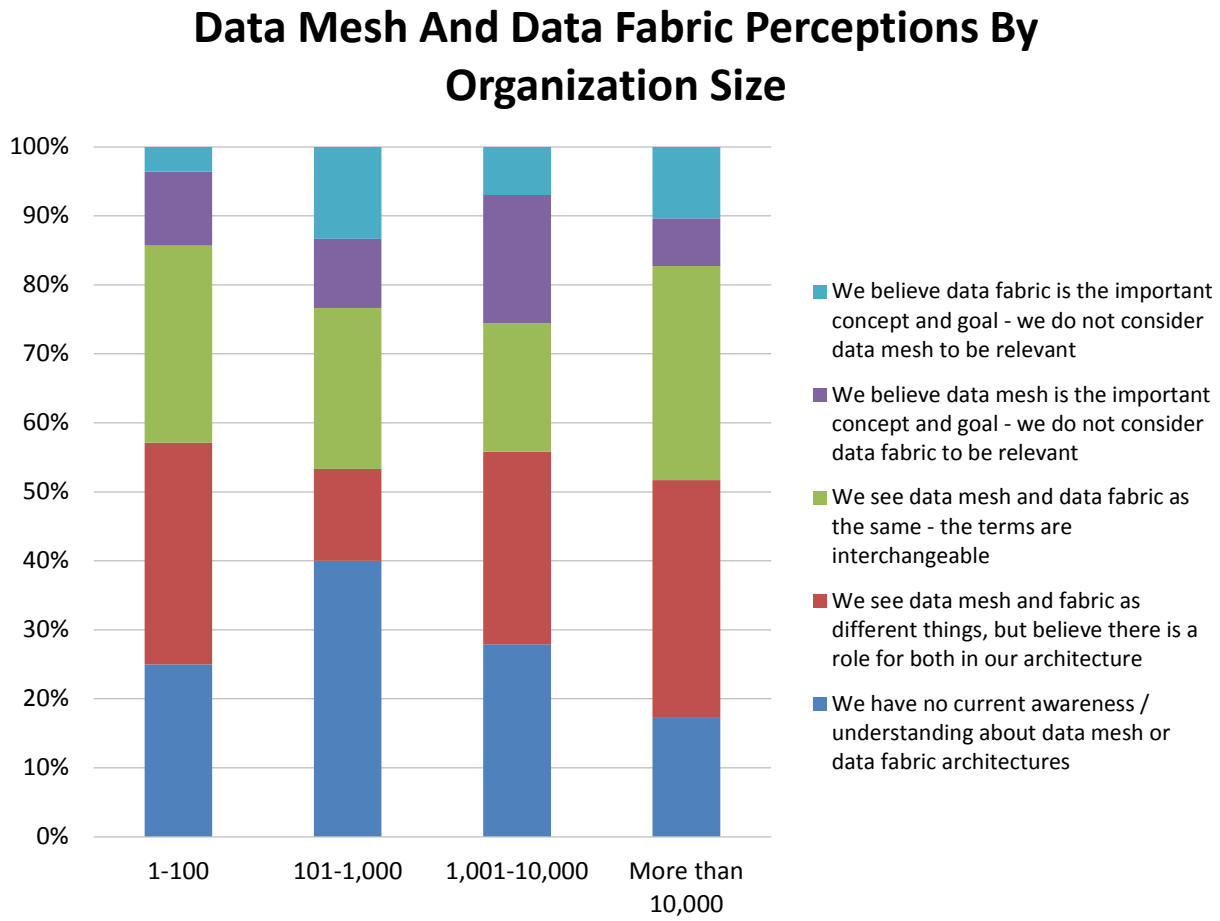


Figure 15 – Data mesh and data fabric perceptions by organization size

Relating perception of data mesh and data fabric concepts to degrees of BI success provides minimal clear correlation. While these ideas, and the benefits of active data architecture, would be expected to contribute substantially to BI success, we observe that a hefty 33 percent of organizations identifying as extremely successful with BI also claim no awareness or understanding (fig. 16). In contrast, only 21 percent of very successful organizations state they have no awareness or understanding. Notably, those organizations that are only moderately successful with BI exhibit both the greatest lack of awareness (over 41 percent) and the greatest diversity of viewpoint about these topics and how they relate.

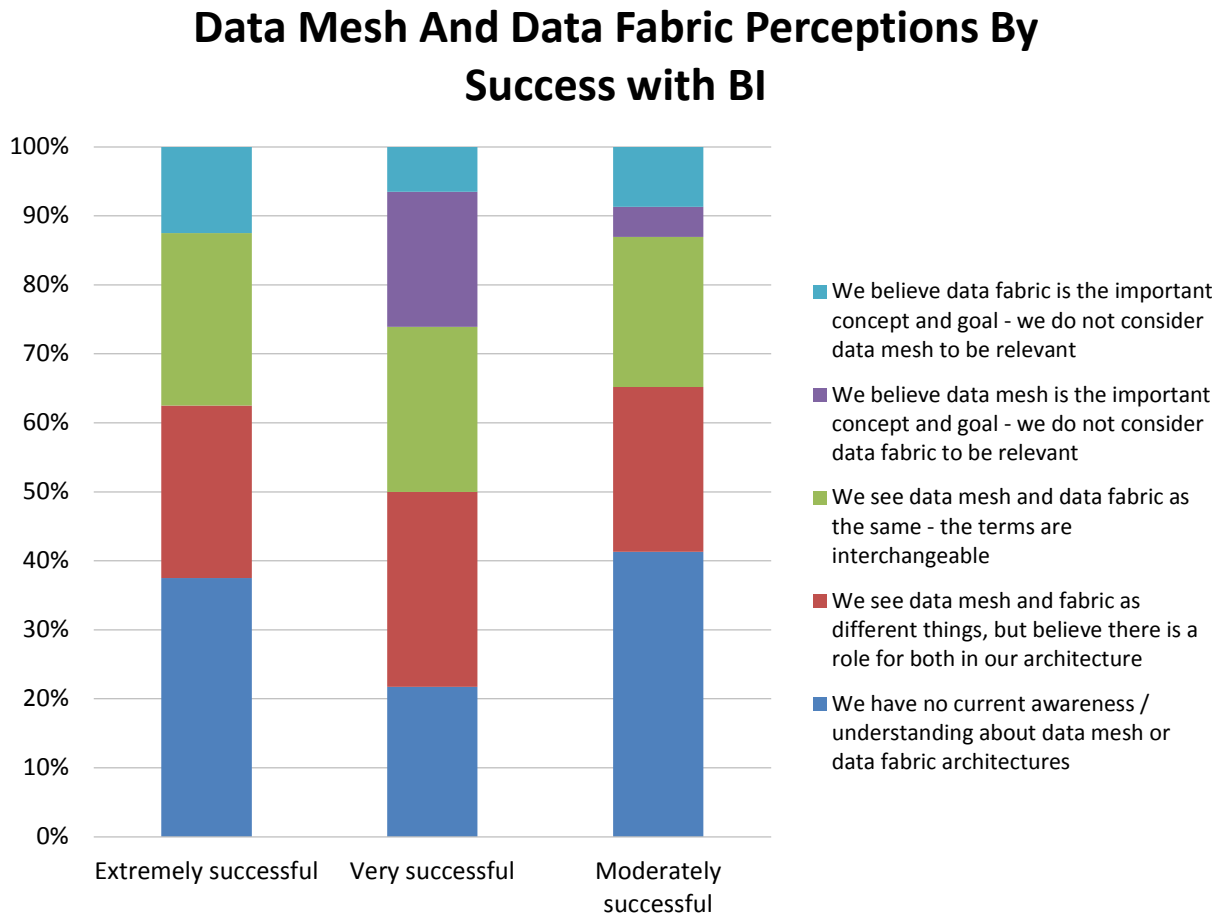


Figure 16 – Data mesh and data fabric perceptions by success with BI

Relevant Technologies for Building an Active Data Architecture

Because an active data architecture has wide-ranging capabilities and draws from technologies in various markets, it is interesting to assess what specific technologies organizations focus on in their early work in this area. Because of the distributed, interconnected, and dynamic nature of these architectures, it makes sense that semantic layer, data catalog, and data virtualization capabilities garner the most current and near-term investment. Nearly 60 percent of responding organizations indicate they have adopted or will adopt a semantic layer in the 12 months, and about 59 percent likewise for data catalog and 57 percent for data virtualization (fig. 17). Notably, data virtualization is already the most-adopted capability relevant for active data architecture, with over 38 percent of respondents indicating they have this capability in place today.

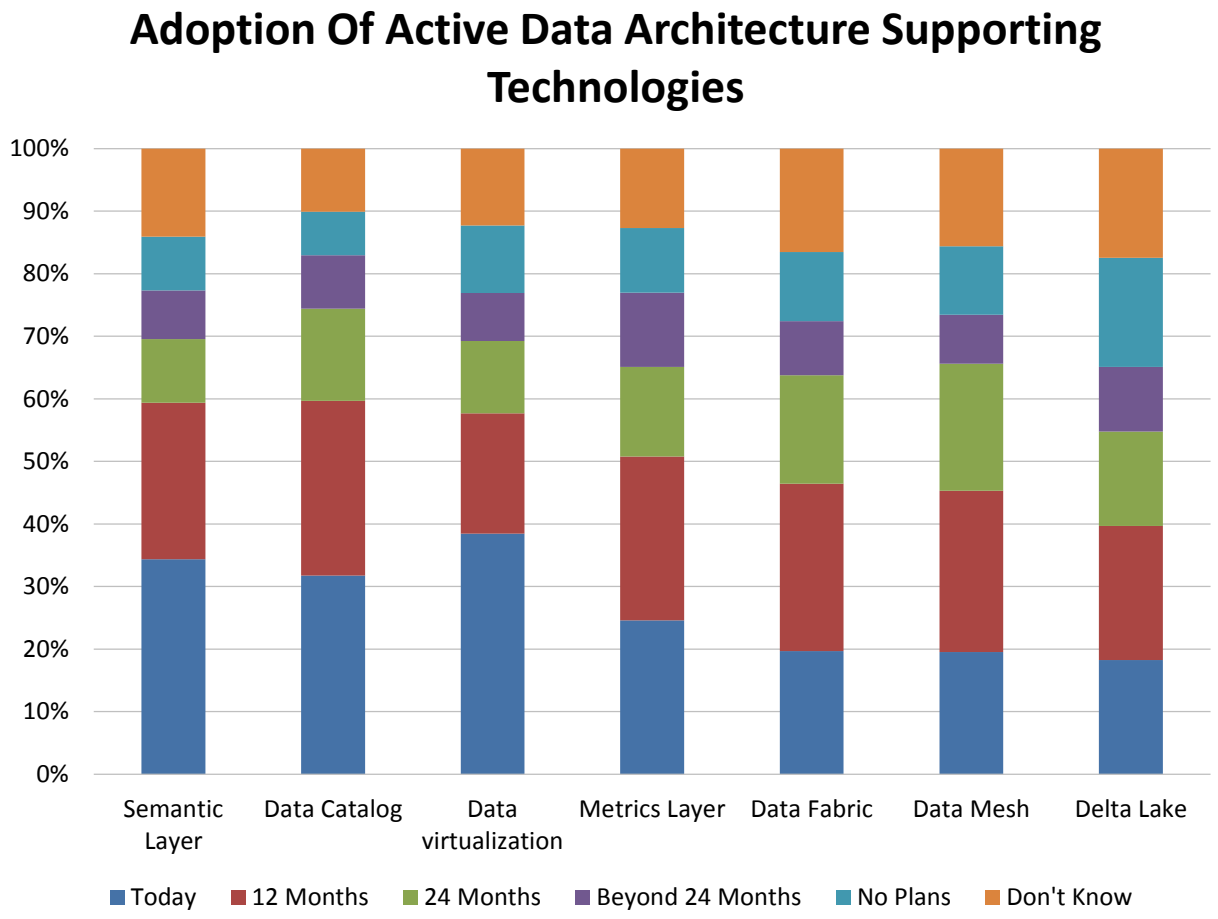


Figure 17 – Adoption of active data architecture supporting technologies

While data fabric and data mesh have been adopted by only about 20 percent of organizations today, these capabilities will exhibit the most acceleration in the next 12 months, based on survey data, with nearly 27 percent of organizations planning to adopt data fabric and 26 percent planning to adopt data mesh in that time frame.

Consistent with the overall market view, the adoption of technologies enabling active data architecture shows significant emphasis on semantic layer, data virtualization, and data cataloging. In particular, organizations in North America and EMEA prioritize these technologies, with over 30 percent of those in North America and over 25 percent of those in EMEA indicating they actively work with or imminently plan to deploy these capabilities (fig. 18).

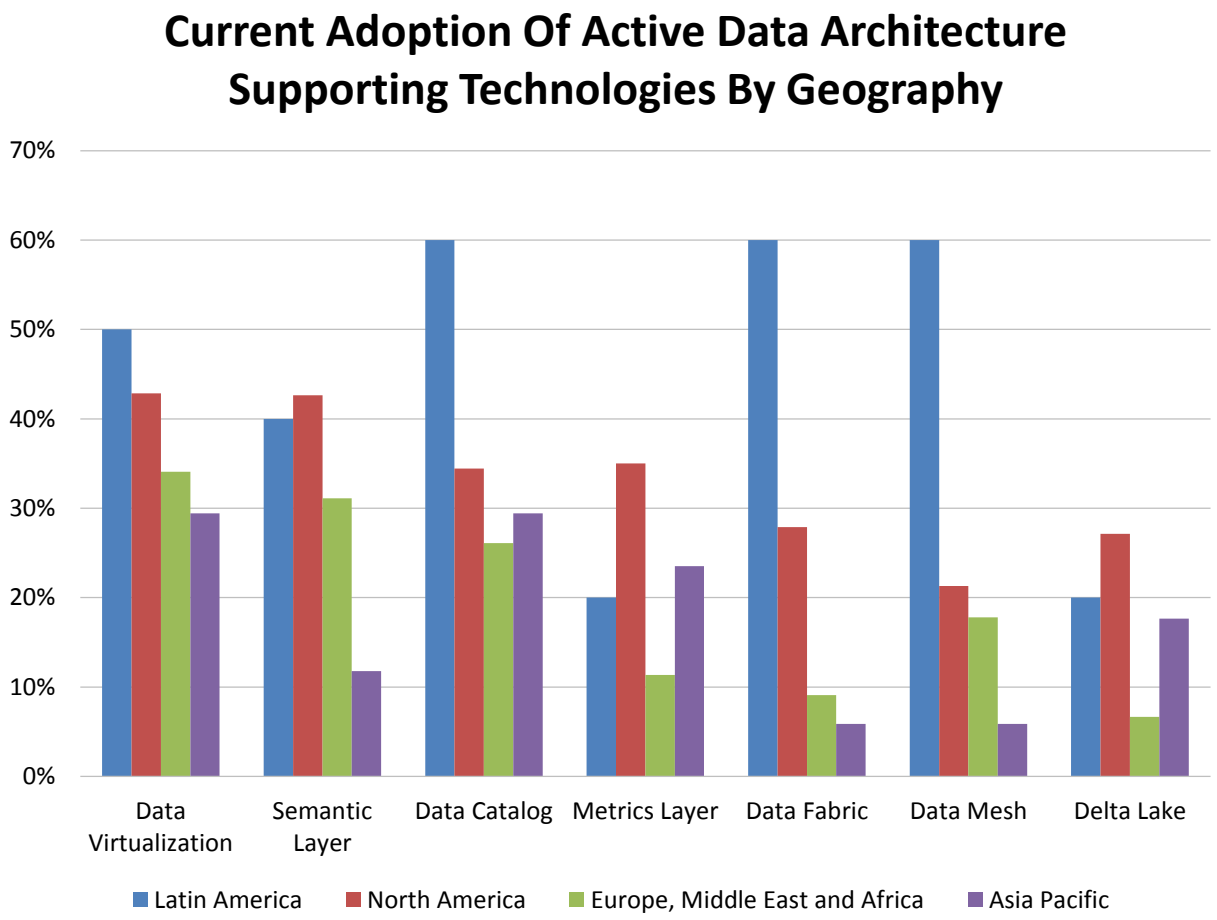


Figure 18 – Current adoption of active data architecture supporting technologies by geography

Various industries actively emphasize different sets of the enabling technologies relevant for active data architecture. While there is an overall emphasis on the same core technologies of data virtualization, semantic layer and data catalog, unique industry-specific patterns, emerge. For example, the accelerating focus on analytics, privacy, and connected data sources in healthcare drive to adopt data catalog (42 percent) and metrics layer (50 percent) at the highest rates among all industries (fig. 19). Organizations in the business services industry, naturally dealing with distributed (often external) data and fast business cycles exhibit among the highest adoption rates for data virtualization (53 percent), data mesh (32 percent), and data fabric (40 percent). In contrast, industries such as government lag in adoption of foundational capabilities such as data catalog (12 percent), metrics layer (12 percent), and semantic layer (12 percent), with no indication of adoption in the sample of data mesh and data fabric concepts.

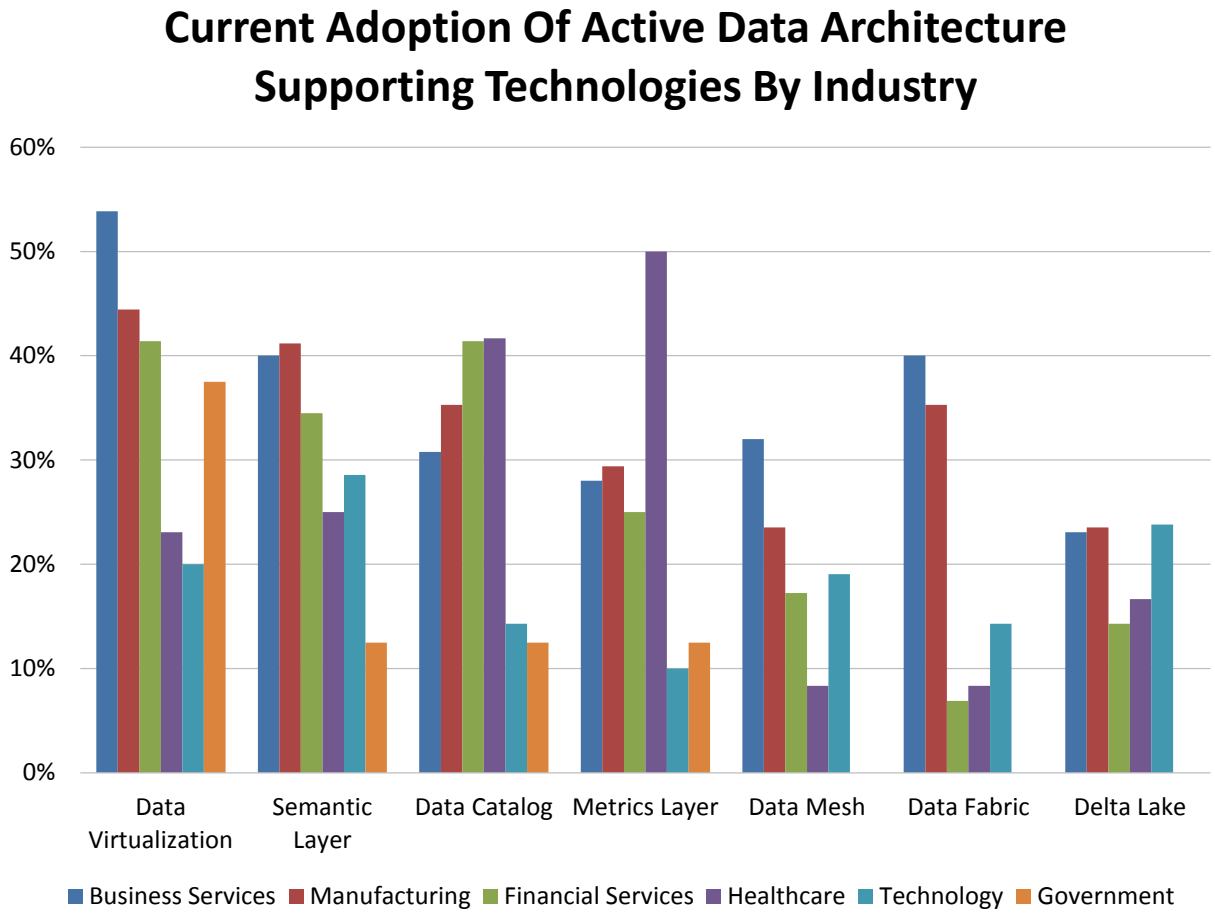


Figure 19 – Current adoption of active data architecture supporting technologies by industry

Adoption of technologies enabling active data architecture generally increases as organization size increases. Very large organizations (more than 10,000 employees) show the greatest rate of adoption across most of the enabling technologies, leading with the most substantial rates for data virtualization (48 percent), semantic layer (50 percent), and data catalog (42 percent) (fig. 20). In contrast, very small organizations (100 or fewer employees) show the greatest level of adoption of data fabric (34 percent) and data mesh (27 percent).

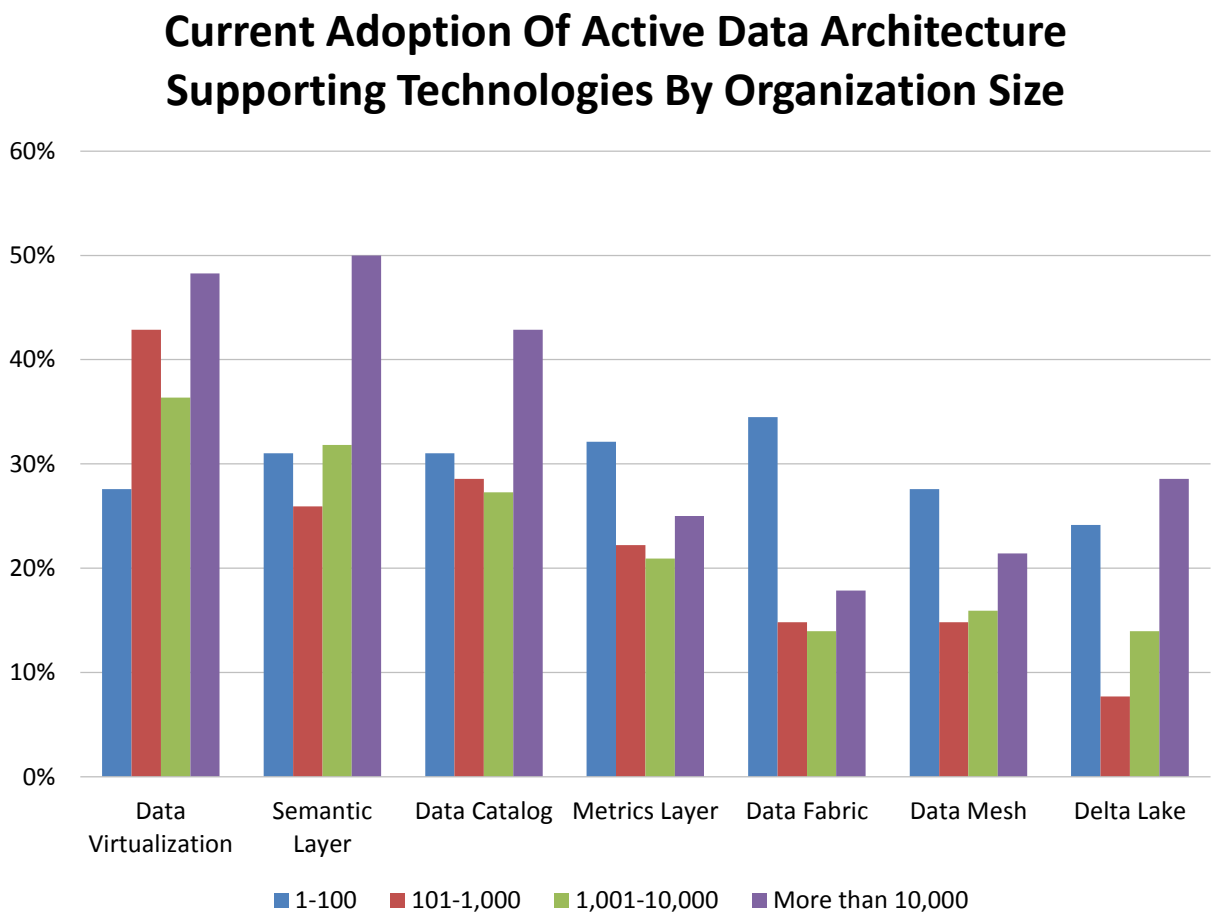


Figure 20 – Current adoption of active data architecture supporting technologies by organization size

Degree of success with BI investments appears well correlated with adoption of enabling technologies for active data architecture. Those organizations reporting they are extremely successful with BI also exhibit the highest adoption rates for the core enablers of data virtualization, semantic layer, data catalog, metrics layer, and data fabric (fig. 21). Also, there is a substantial gap in adoption between extremely successful and only moderately successful organizations in the range of 2-3X. For data virtualization, the rate of adoption is nearly 80 percent for extremely successful organizations versus only 30 percent for moderately successful organizations. For data catalog, the difference is 44 percent versus 19 percent, respectively.

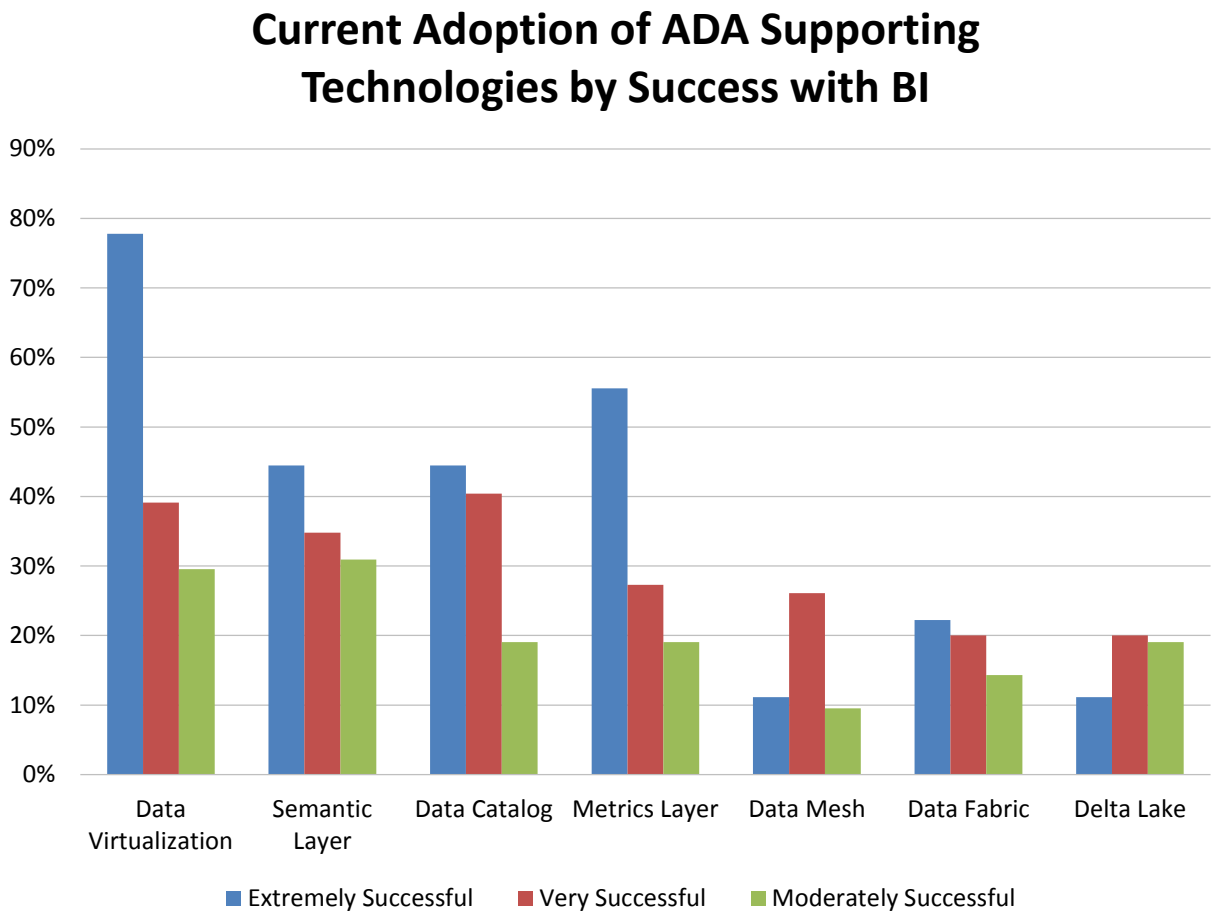


Figure 21 – Current adoption of ADA supporting technologies by success with BI

Active data architecture deployment options

We asked our survey respondents to indicate their preferences and expectations for deployment of active data architecture capabilities. As seen in many data and BI-related markets, there is a general propensity toward SaaS / cloud deployments, with nearly 60 percent of organizations indicating this is either critically important or very important to their plans, exceeding all other deployment options (fig. 22). Following closely is hybrid deployment (multi-cloud or cross data center integration) at about 55 percent; but it notably achieves an equal level of preference to SaaS / cloud when considering organizations indicating these deployment options are critically, very, or somewhat important. While on-premises deployments are substantially less preferred, a core of organizations indicate this is still critically important (over 25 percent).

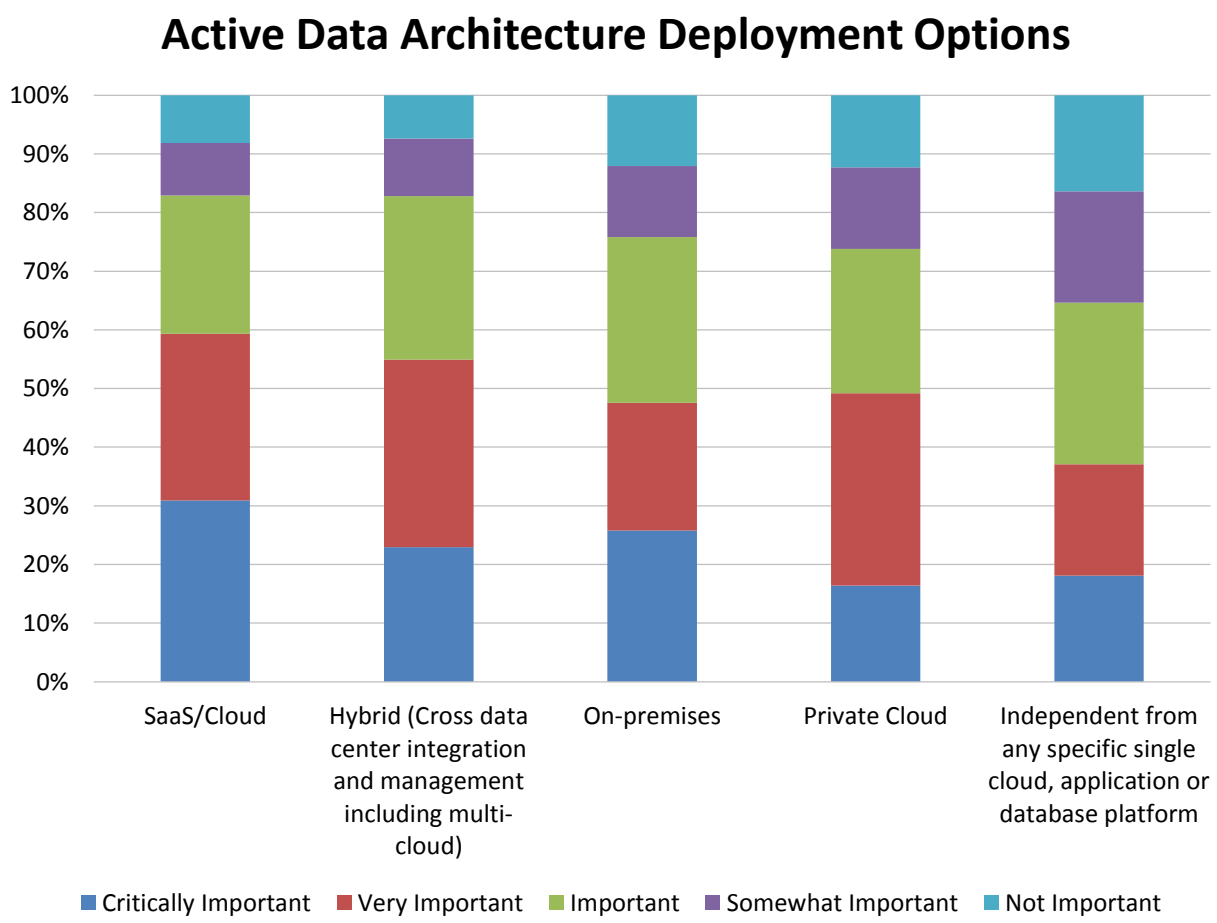


Figure 22 – Active data architecture deployment options

While most industries mirror the overall market trend toward SaaS / cloud deployments for active data architecture, the survey data show variances and industry-specific differences. For example, the government sector prioritizes on-premises and private cloud deployments over all other styles (fig. 23). Likewise, healthcare organizations show a top priority of on-premises deployment, even to the point of indicating SaaS / cloud is the lowest deployment option priority. More traditional industries like manufacturing and financial services appear to show great diversity, with nearly equal prioritization of all possible deployment options.

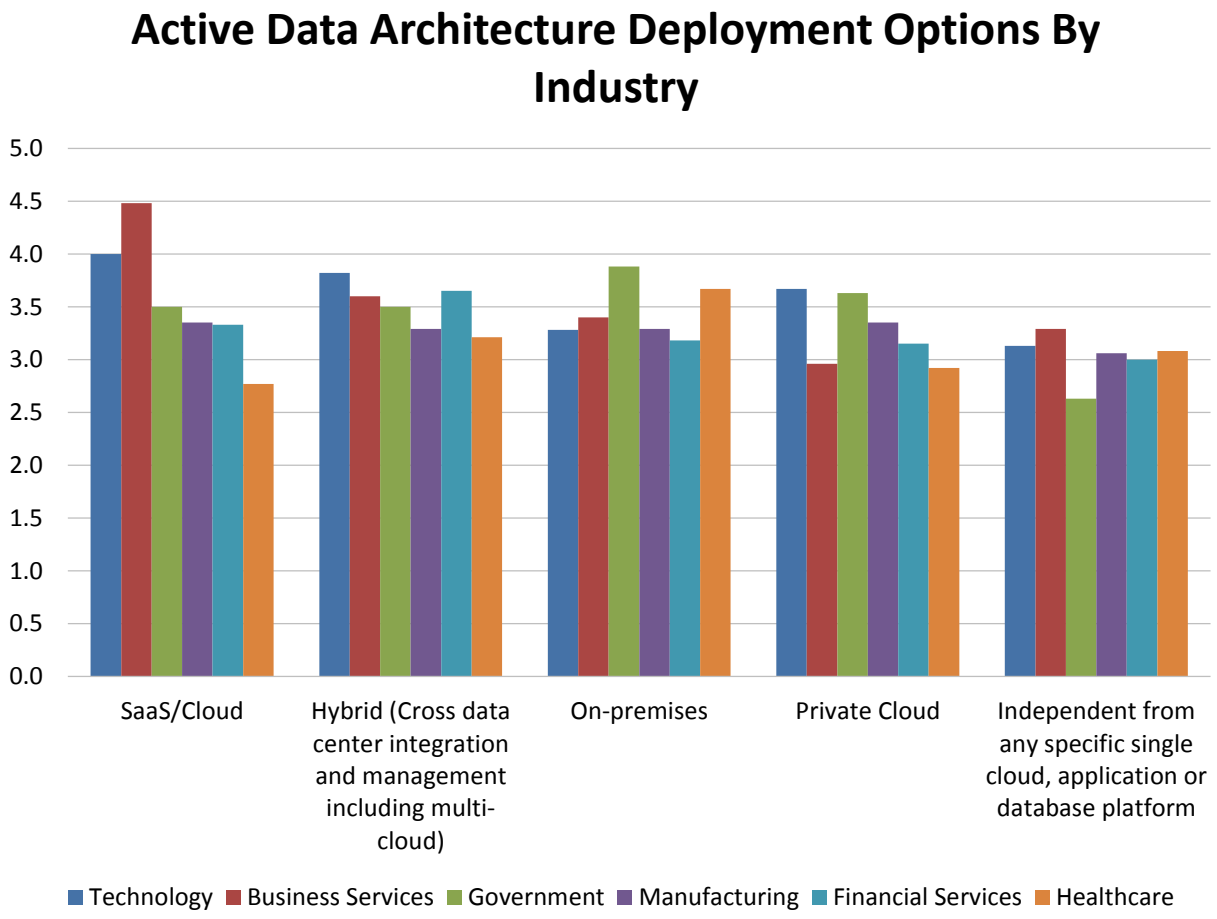


Figure 23 – Active data architecture deployment options by industry

Organizations prioritizing cloud-based deployments for active data architecture capabilities at the highest levels also report the greatest degrees of success with BI: those reporting they are extremely successful with BI most often indicate SaaS / cloud as their number-one deployment option priority, followed closely by hybrid (fig. 24). Organizations very successful or moderately successful with BI tend to also prioritize cloud deployment options, but much less substantially above on-premises deployments.

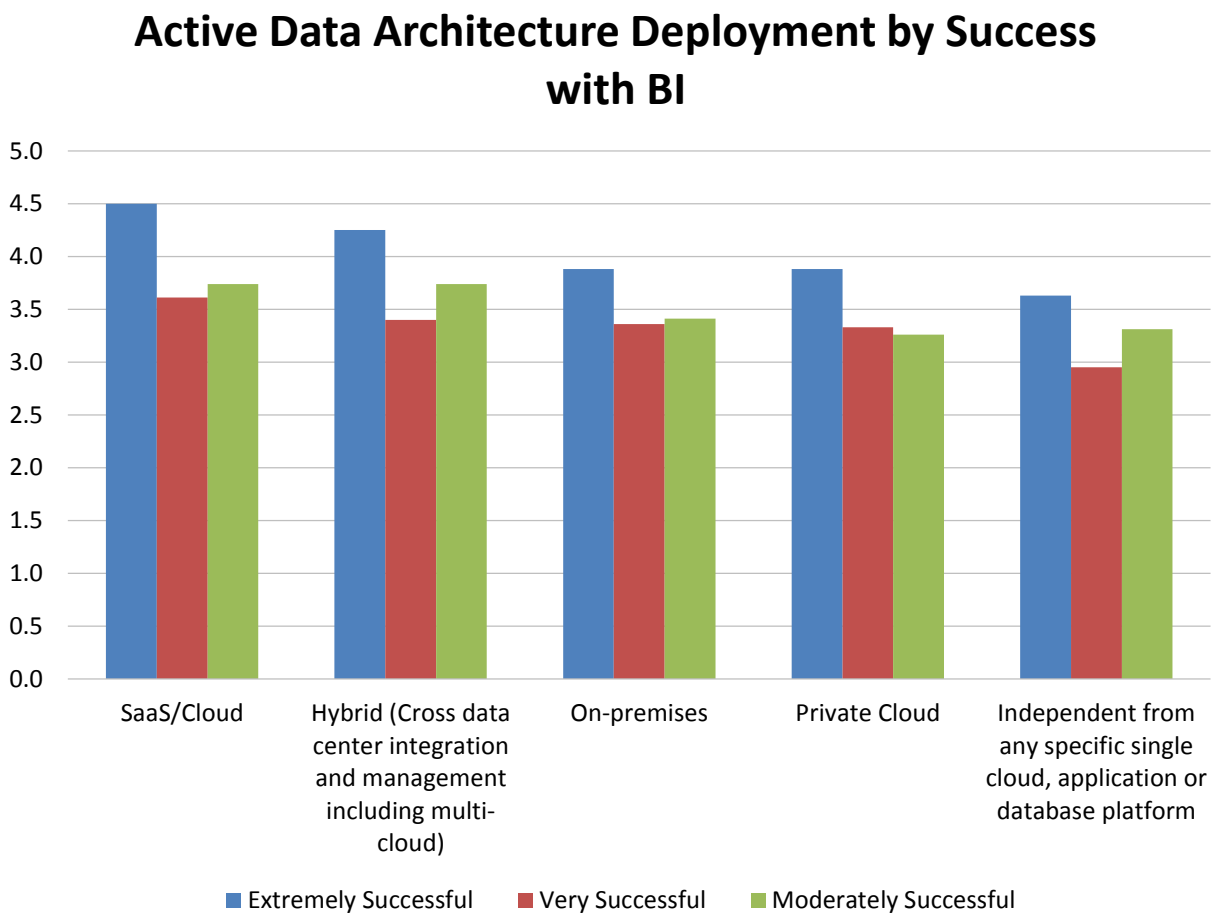


Figure 24 – Active data architecture deployment options by success with BI

Detailed Capabilities for Active Data Architecture Implementation

Active data architecture supports a platform-independent layer, which sits between physical data stores and points of data consumption. It is comprised of various data management capabilities including virtualized and distributed data access, data governance, and security. At its core, active data architecture is an abstraction layer translating business and physical structures. It is an architecture dynamically optimized for performance, scalability, and cost management.

Being an architecture, not a single product or piece of technology, an organization can build an active data architecture based on componentry coming from various existing markets that are well established. Specifically, the market focuses on core capabilities for data integration, data cataloging and metadata, governance, scalability and performance, dynamic optimization, and access and integration. This means that there is overlap and alignment with the evolving market for data engineering capabilities, including data integration, as well as the markets for governance and metadata technologies.

Both technology providers and buyers should recognize these overlaps and consider active data architecture to be, at the moment, a composite market that draws from others to serve a specific set of use cases, enabling the dynamic delivery of data products (data as a product).

Importance and Prioritization of Data Integration Capabilities

Active data architecture requires a diverse range of data integration styles. Most organizations are still evolving their data integration capabilities beyond bulk/batch (ETL or ELT) and towards more real-time and distributed styles (event streaming, message-oriented, and virtualized). The emphasis remains most heavily on bulk/batch, with over 70 percent of organizations indicating that style is critically or very important (fig. 25). This is followed closely by real-time event streaming, viewed as critically or very important by 67 percent of organizations. Data virtualization, a data integration style seeing rapid pace of adoption, approaches 50 percent of organizations considering it critically or very important. And message-oriented styles of data integration remain relevant but consume less of the focus, being viewed as critically or very important by only 45 percent of organizations.

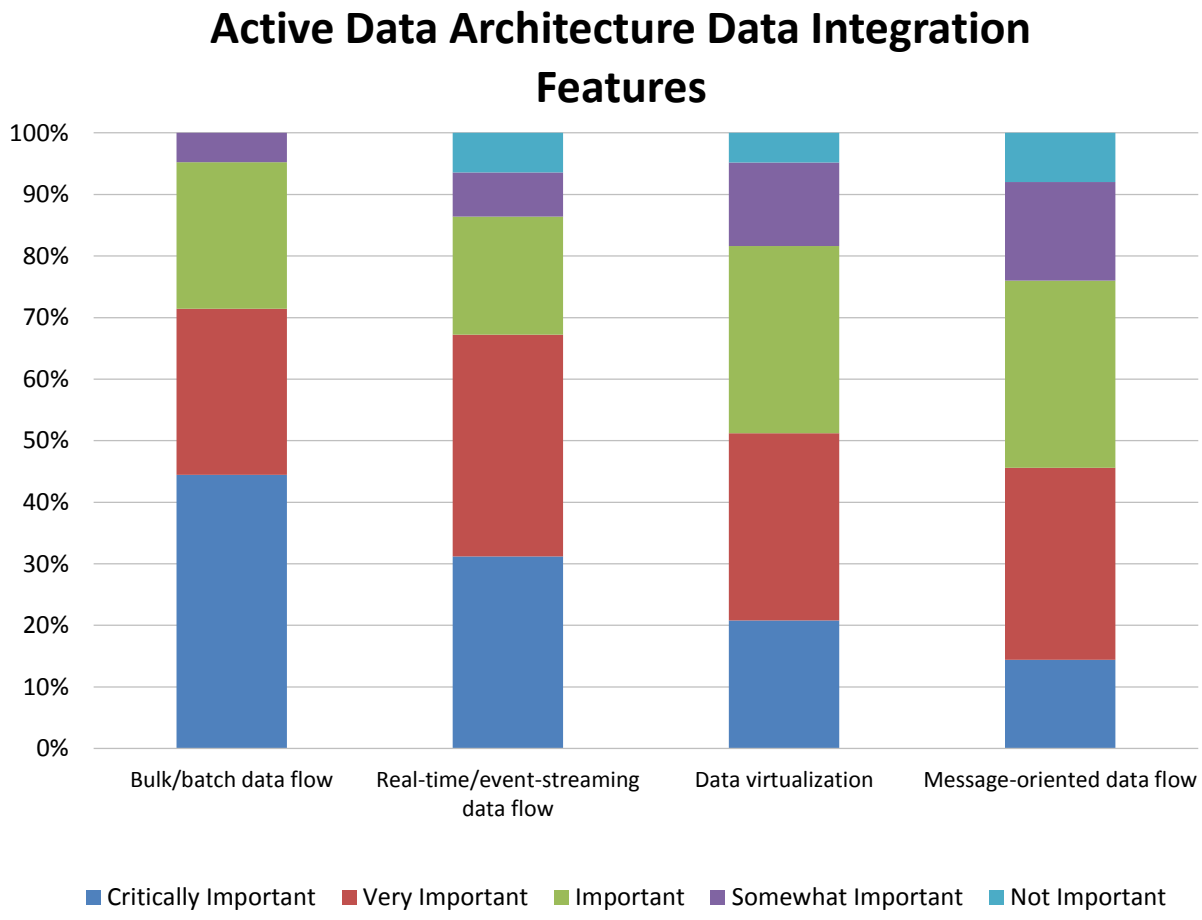


Figure 25 – Active data architecture data integration features

The reshaping of data integration capabilities in support of active data architecture is more advanced in certain global regions. Organizations in Asia Pacific express a higher level of importance for real-time event stream and data virtualization than those in North America and EMEA (fig. 26). At the same time, Asia Pacific organizations still identify bulk/batch as a high priority but at a lower level of frequency than their North American and EMEA peers. The result is that the data integration capabilities supporting active data architectures in Asia Pacific will be more diverse and batch-oriented than those in other regions.

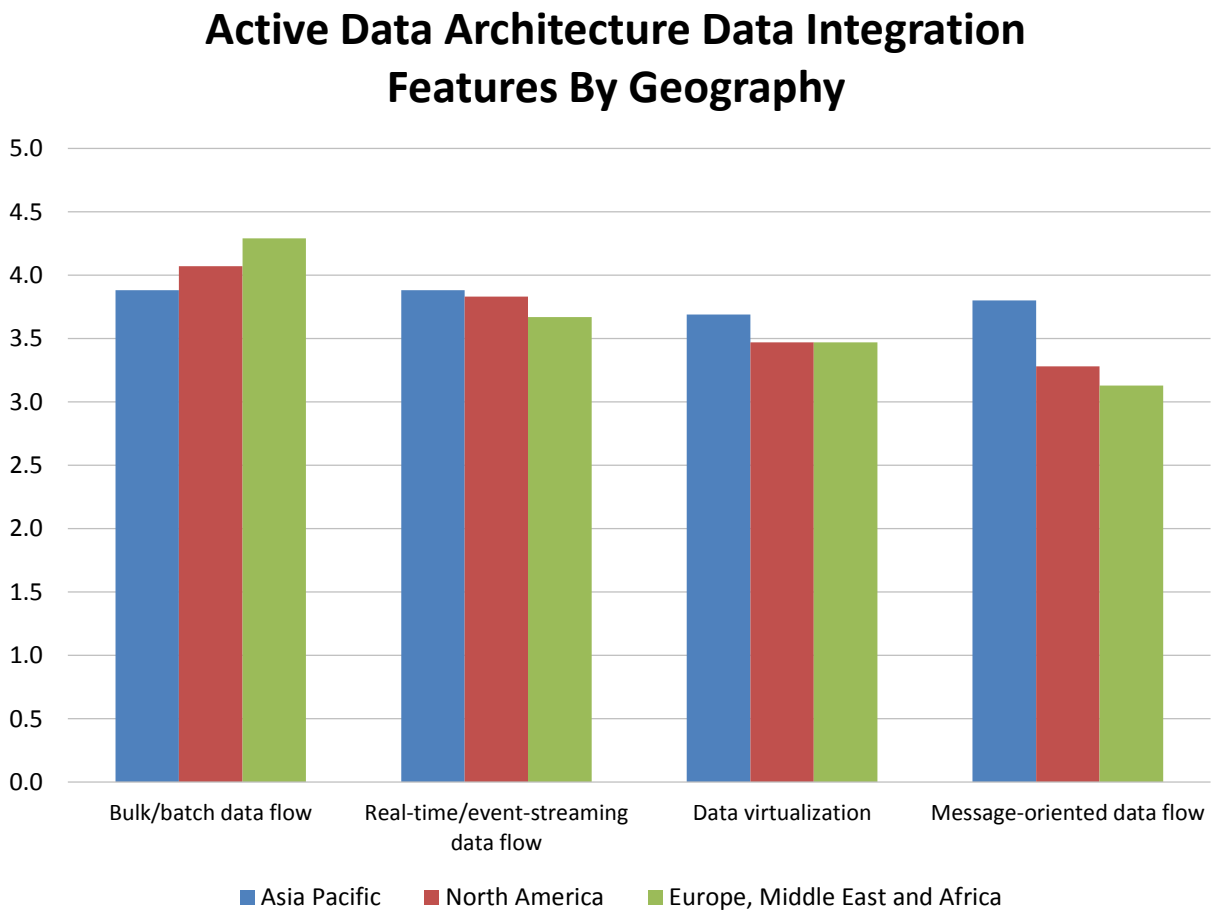


Figure 26 – Active data architecture data integration features by geography

All industry sectors generally follow the same pattern as the overall market, placing the highest priority on bulk/batch data flow, followed by real-time event streaming, then data virtualization, and message-oriented data flow as the lowest priority (fig. 27). However, some industries execute a more rapid rebalancing of styles than others. Business services organizations prioritize real-time event streaming at nearly the same level as bulk/batch, and technology organizations have more equal prioritization across all styles. Government organizations seem to be the slowest in executing a rebalancing away from purely bulk/batch, with that sector placing a much lower priority on the other styles than most other industries. Notably, while demand for data virtualization accelerates, the healthcare sector prioritizes it at a much lower level than others, which is likely due to the concerns over privacy that could arise from distributed and virtualized management of data.

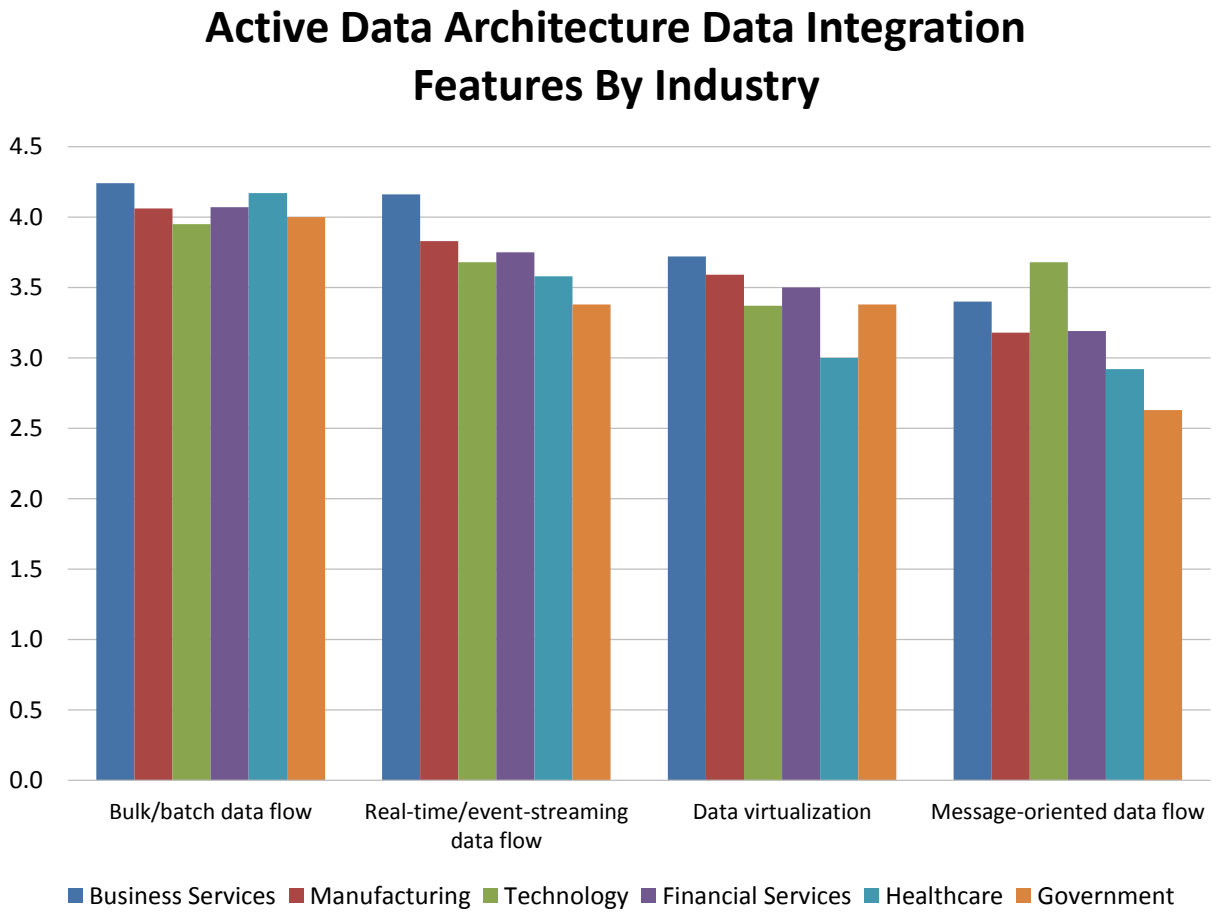


Figure 27 – Active data architecture data integration features by industry

The adoption of active data architecture capabilities will be critical for success of more complex and diversified BI initiatives. Survey data shows that organizations that are extremely successful with BI prioritize a diverse set of data integration styles, beyond traditional bulk/batch data flow (fig. 28). In fact, the most successful organizations prioritize real-time event streaming and data virtualization at nearly the same level, significantly higher than their peers that are somewhat less successful with BI.

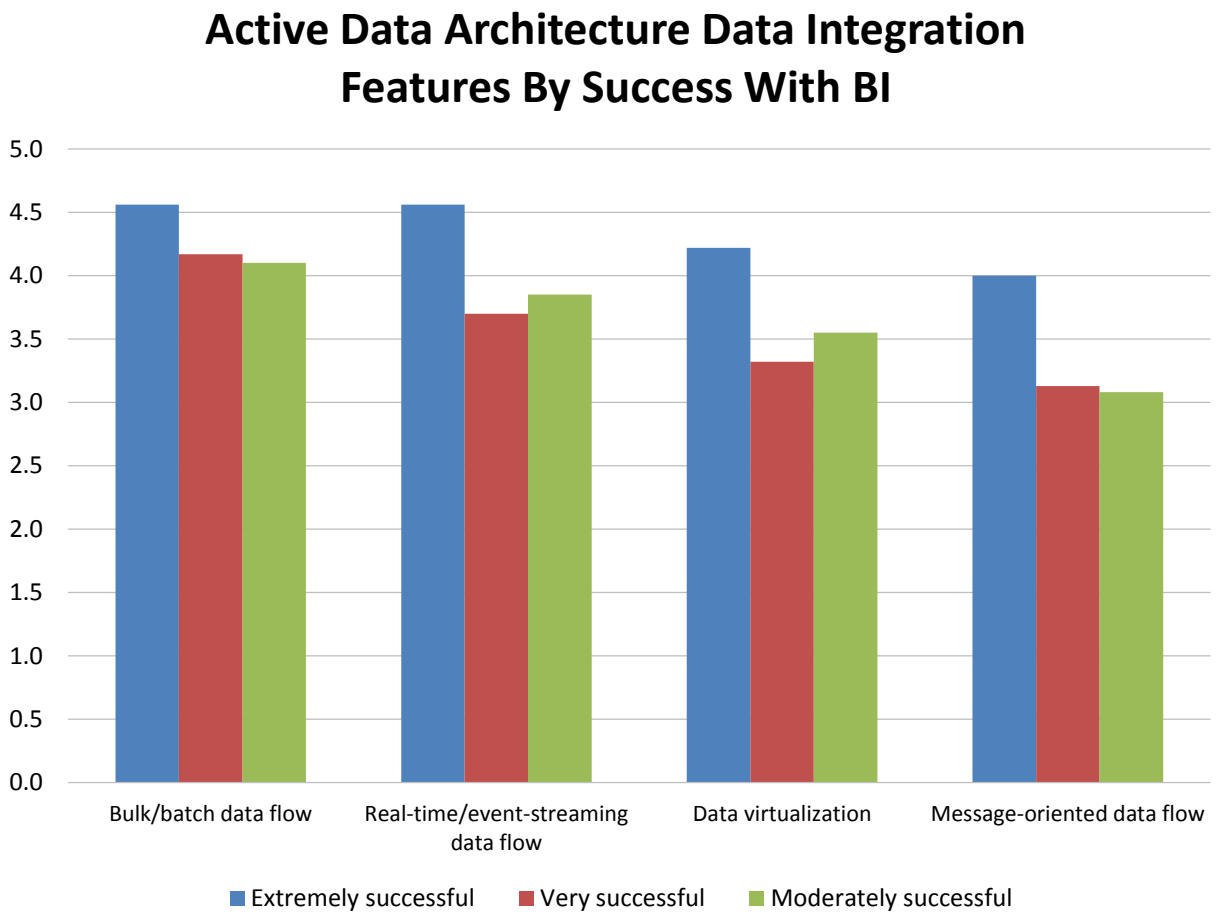


Figure 28 – Active data architecture data integration features by success with BI

Importance and Prioritization of Data Catalog and Metadata Capabilities

The diverse, distributed, connected, and dynamic nature of active data architecture requires capabilities to collect, understand, and leverage metadata describing relevant data sources, models, metrics, governance rules, and more. Organizations place significant focus on various types of metadata features in order to address these needs. They believe the ability to ingest metadata from a variety of sources is most important, with nearly 70 percent of organizations rating this feature as critically or very important (fig. 29). Impact analysis, which provides the ability to understand the impact of changes in a distributed and connected environment was deemed the next highest priority, with 57 percent of organizations rating it critically or very important. The remaining key metadata features, including lineage visualization, modeling of integrated views of data, modeling of the infrastructure (all the componentry of active data architecture) and optimization capabilities, are all viewed as critically or very important by approximately 50 percent of organizations.

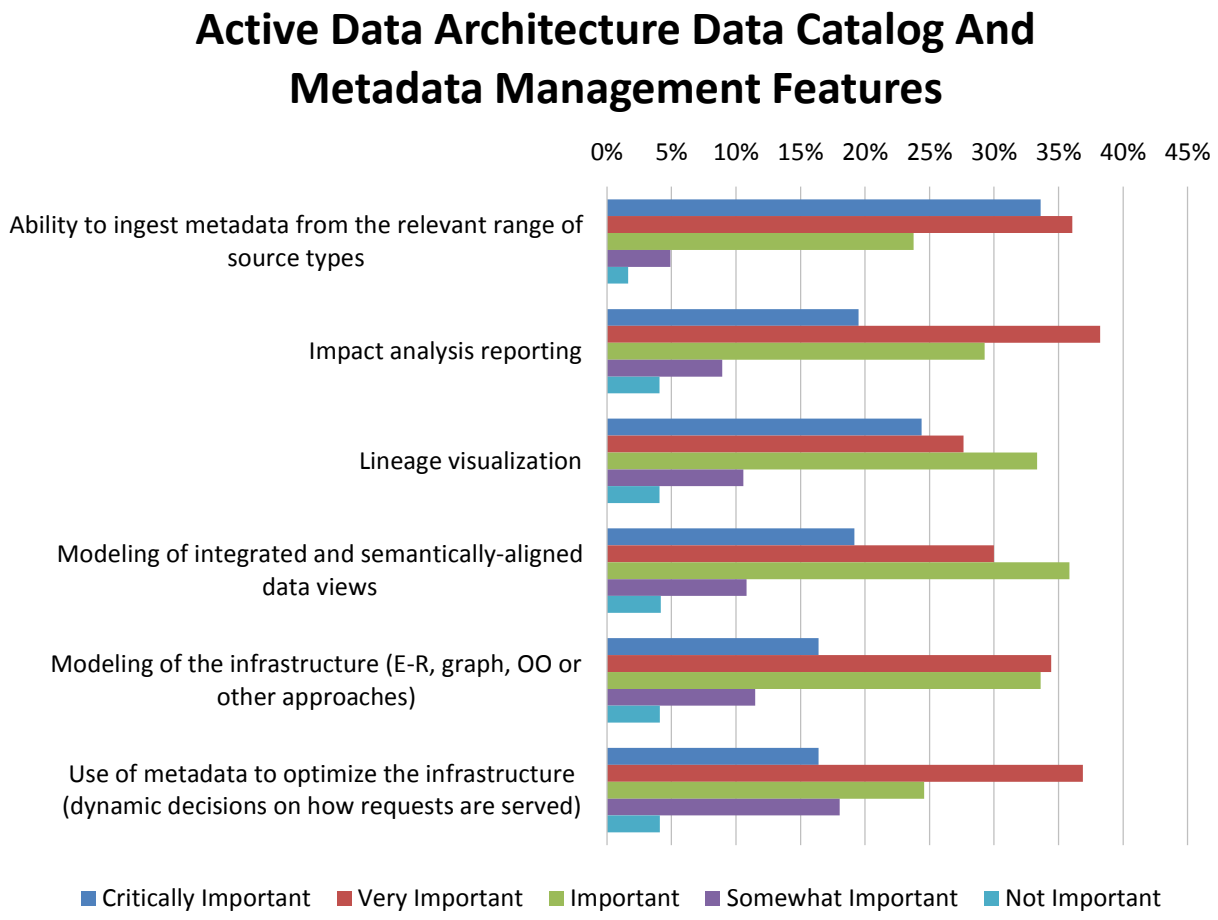


Figure 29 – Active data architecture data catalog and metadata management features

On a global basis, we observe significant differences in priorities about metadata management in support of active data architecture. While all regions identify metadata ingestion as their top priority, they view most other metadata features as much higher priorities for Asia Pacific and EMEA organizations than for those in North America (fig. 30). In particular, a feature crucial for truly making the data architecture "active" (modeling of the infrastructure and use of metadata for optimizing the environment) stands out as nearly twice as critical to Asia Pacific organizations as they are for EMEA and North America.

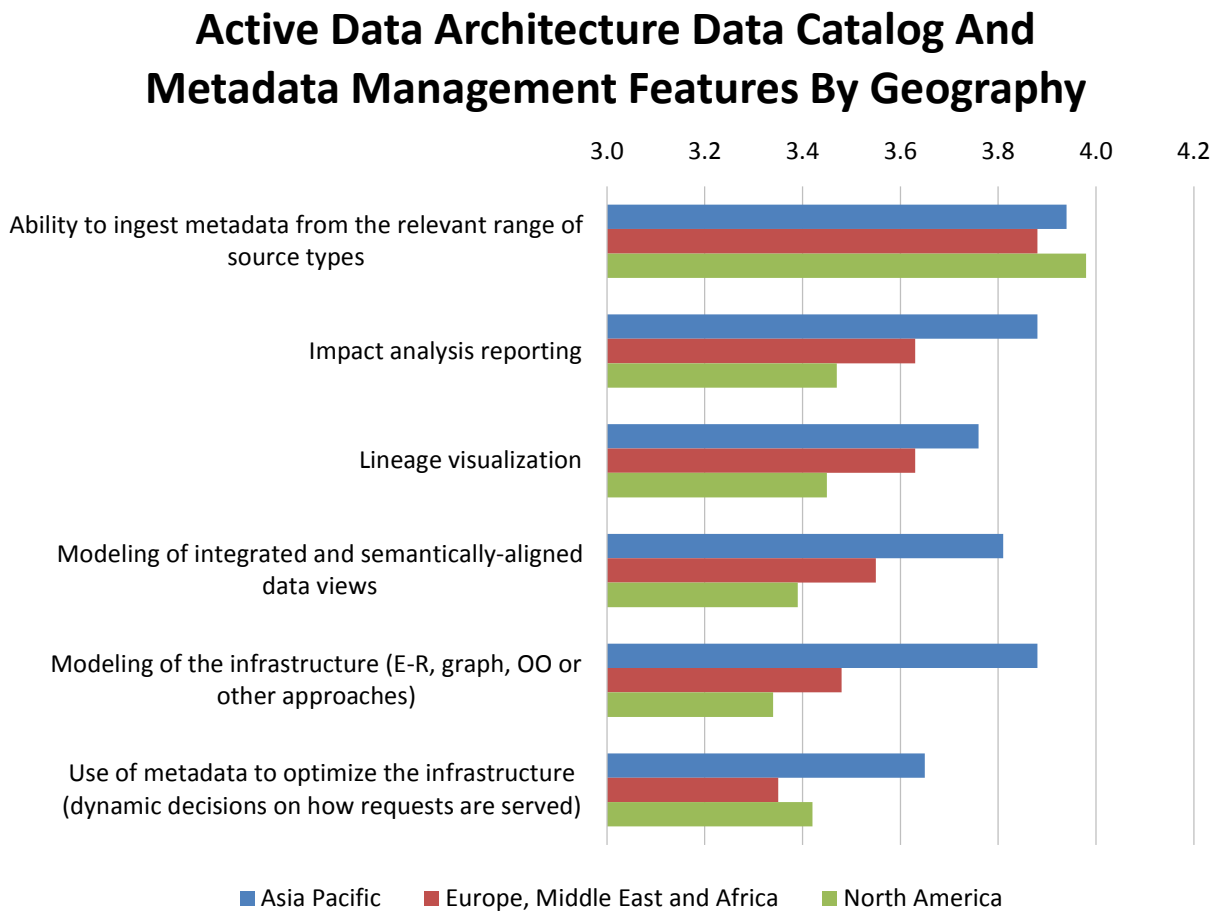


Figure 30 – Active data architecture data catalog and metadata management features by geography

On an industry basis, priorities across data catalog and metadata features in support of active data architecture follow a similar pattern to the overall market, with metadata ingestion being the top priority for most sectors (fig. 31). However, there are notable differences across industries. Business services, technology, and financial services prioritize the modeling and optimization features higher than other industries, positioning them for more complete and effective active data architecture implementation.

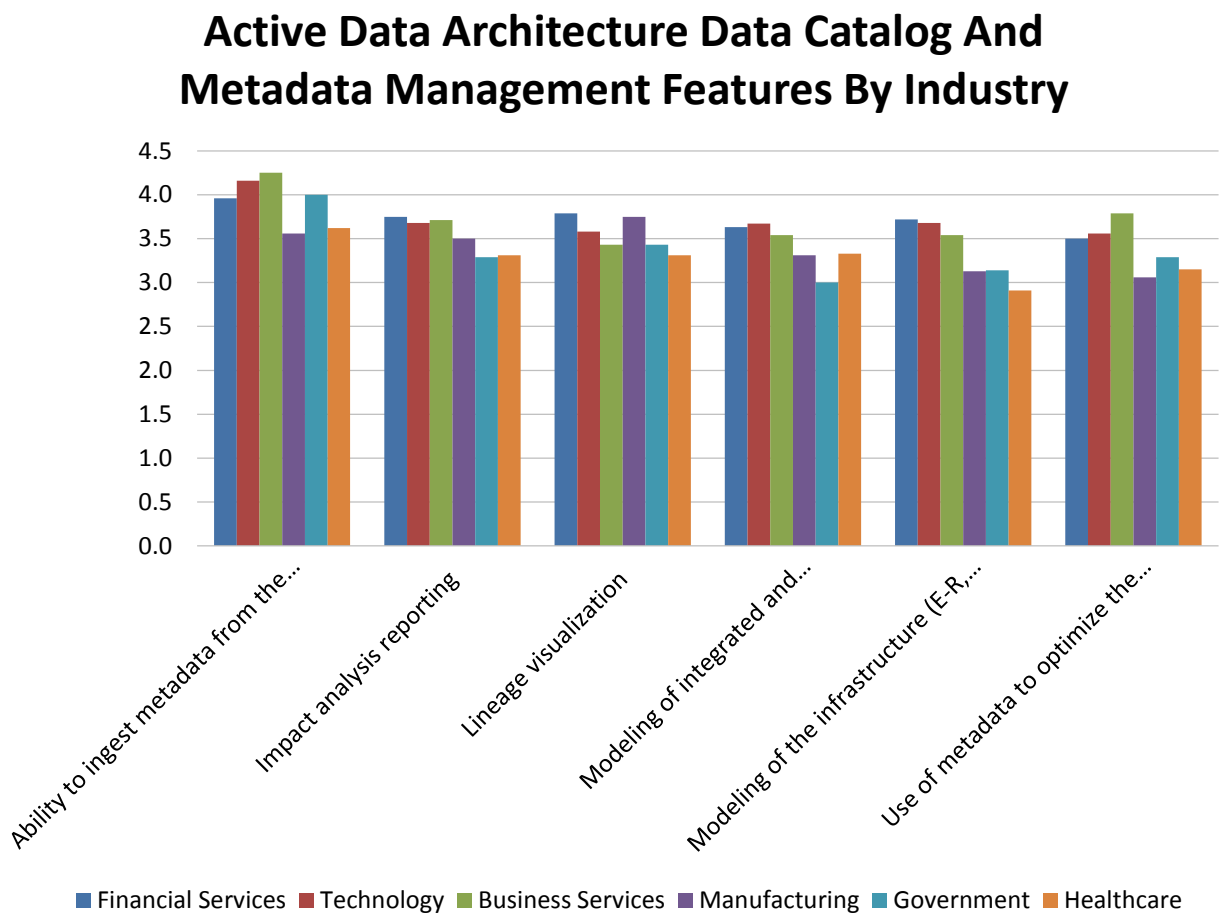


Figure 31 – Active data architecture data catalog and metadata management features by industry

As BI implementations become more complex and mission critical, the metadata features supporting active data architecture become more important. Survey data show a strong correlation between level of priorities of metadata features and degrees of BI success (fig. 32). Those organizations characterizing their BI efforts as extremely successful also express substantially higher priority levels for all key data catalog and metadata features that underpin active data architecture.

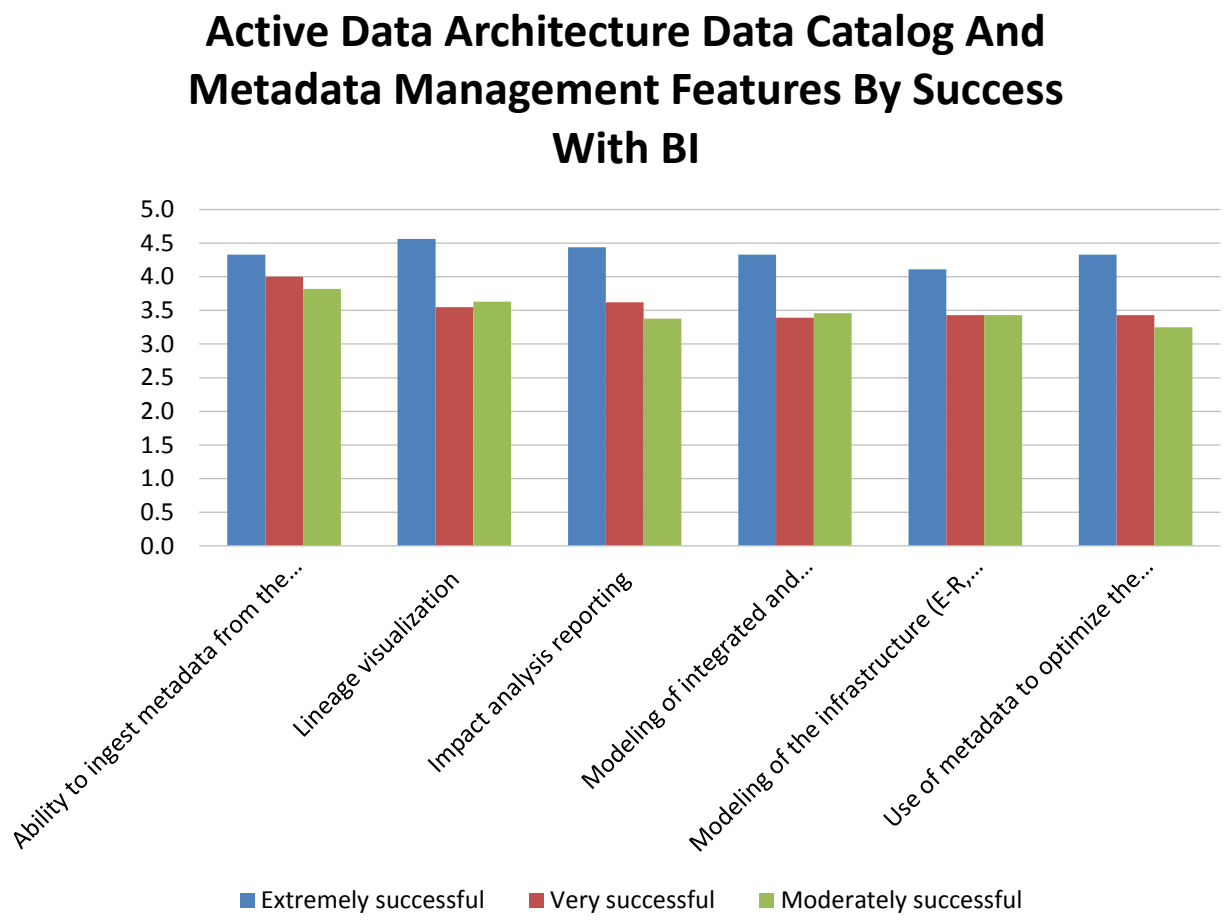


Figure 32 – Active data architecture data catalog and metadata management features by BI success

Importance and Prioritization of Governance Capabilities

Because active data architecture is about managing diverse and distributed data in a more automated manner, governance capabilities are crucial. There are six main dimensions of governance that organizations need to consider. Survey data show that organizations rank security as most important, with nearly 85 percent rating it as critically or very important, followed by privacy and quality, both about 82 percent; life cycle management at 67 percent; models and definitions at 60 percent; and, finally, cost management at 59 percent (fig. 33). These data reflect a strong focus on the fundamentals of governance, where most organizations struggle today. More advanced thinking on governance, represented by a focus on models, definitions, and cost management, is not yet mainstream.

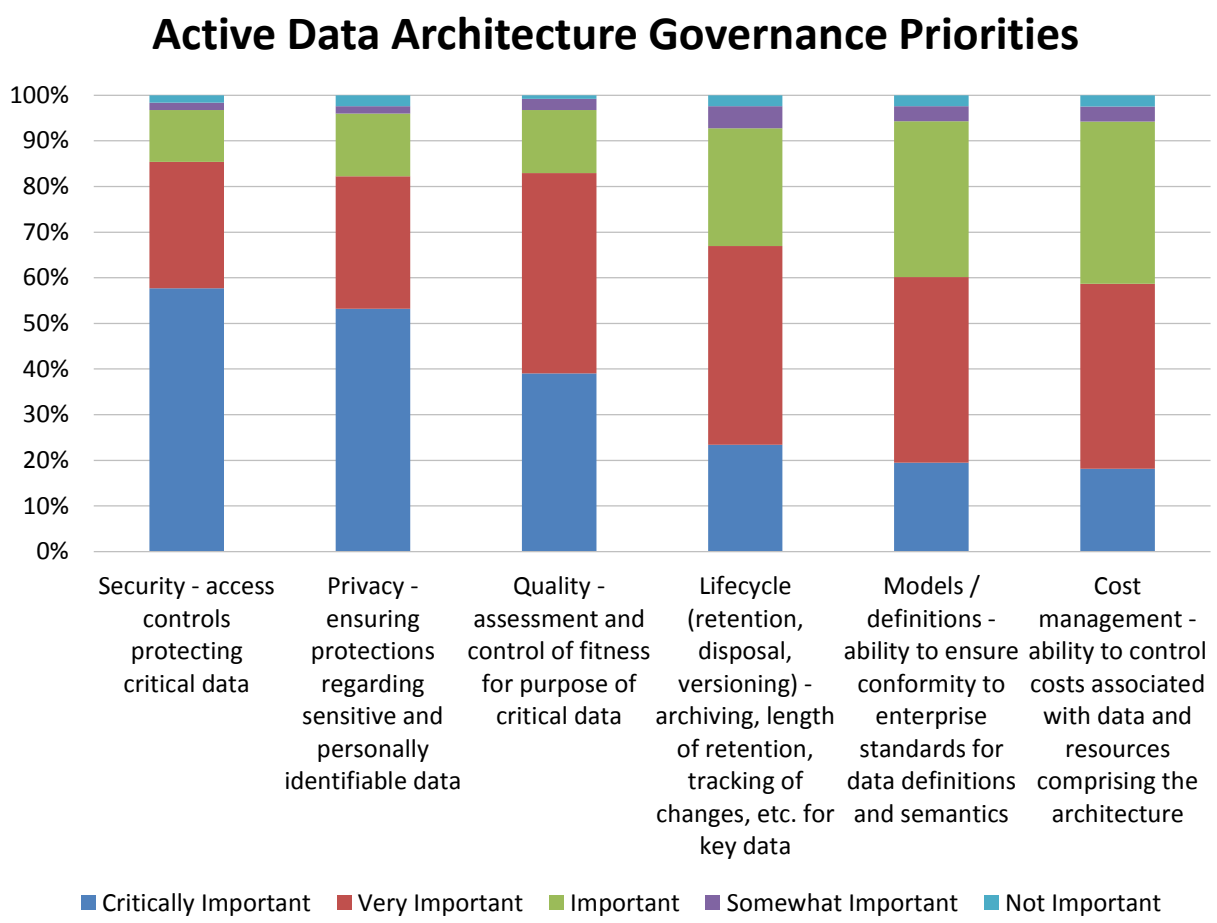


Figure 33 – Active data architecture governance priorities

As with other core active data architecture capabilities, priorities for governance features across industries in general follow the overall market trends. however, differences appear in the industry segmentation of the data (fig. 34). Government organizations prioritize security and privacy much higher than all other features and other industries, due to the sensitivity of data they often manage. At the same time, government prioritizes models, definition, and cost management at a much lower level than all other industries, which may place organizations in that sector at a disadvantage when trying to achieve the full benefits of active data architecture. Business services prioritize these same features, as well as quality, at a higher level, likely due to the emphasis on customer and B2B data.

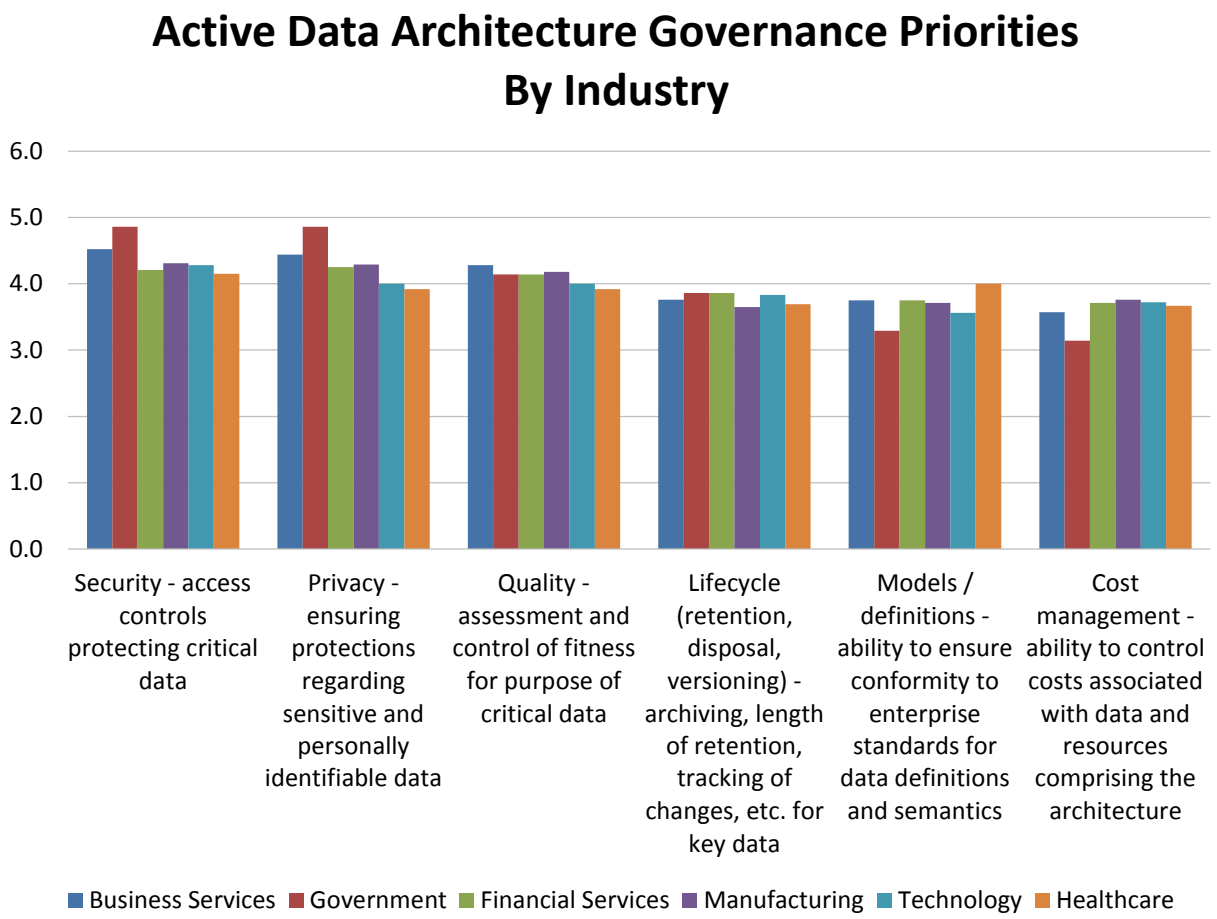


Figure 34 – Active data architecture governance priorities by industry

Survey data show that organizations achieving the highest levels of BI success also prioritize governance features at the highest level among their peers (fig. 35). The converse also emerges from the data, where those organizations reporting only moderate success with BI tend to prioritize governance features at a lower level than their (more successful) peers. Notably, security, privacy, and quality, the mainstays of governance, are the highest priorities, regardless of degree of BI success.

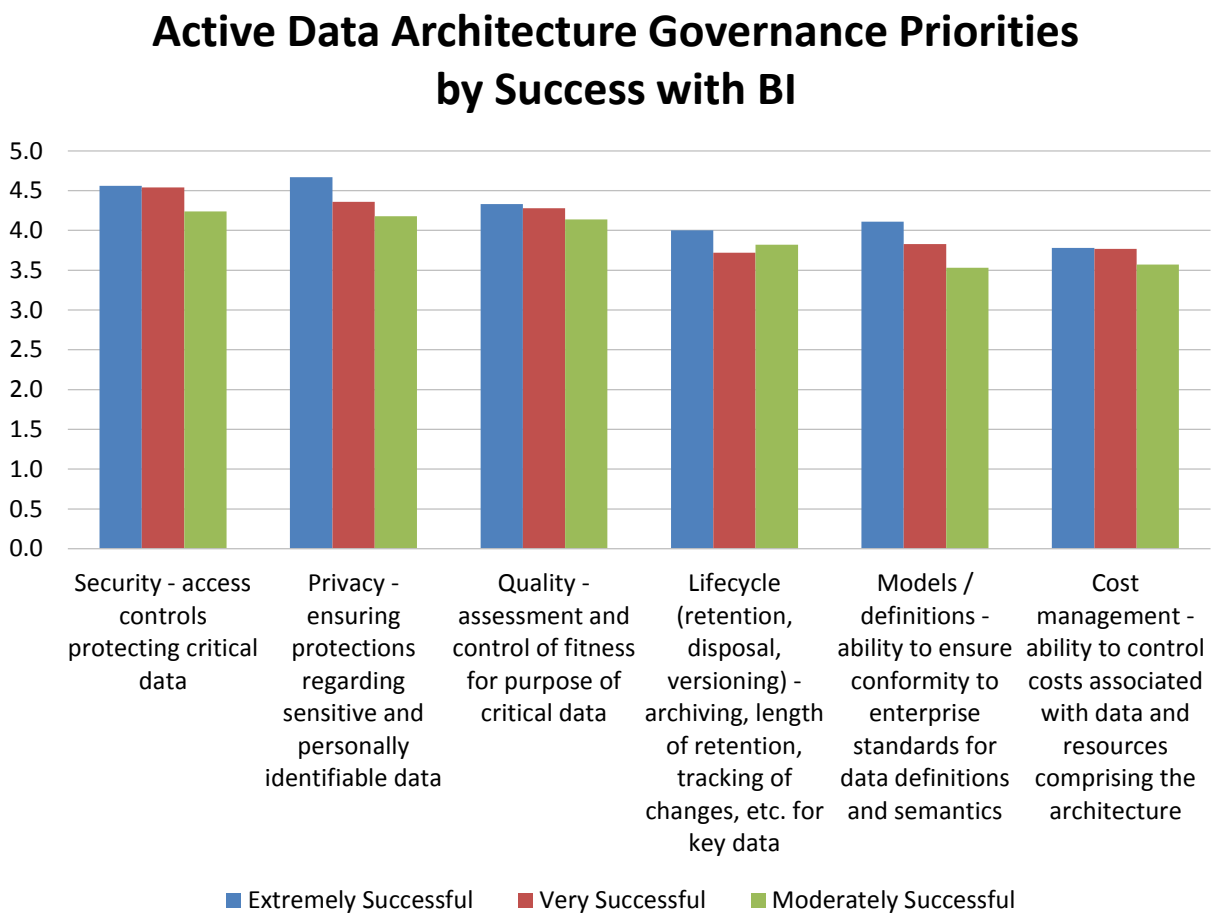


Figure 35 – Active data architecture governance priorities by success with BI

Importance and Prioritization of Scalability and Performance Capabilities

With the trend toward distributed, connected, and virtualized data management techniques, the ability to ensure good performance and scale over larger and more complex data landscapes will be key to success with active data architecture. Our survey data show that organizations believe certain scalability and performance features are more important than others. Nearly 95 percent of organizations rate data persistence and caching as important, very important, or critically important (fig. 36). This falls to 85 percent for pre-fetch and pre-transform capabilities, 87 percent for automated data placement, and about 83 percent for distributed query optimization. The lowest level of importance for distributed query optimization appears to conflict with the accelerating demand for data virtualization, which requires distributed query optimization to achieve suitable performance. This represents an area where technology providers can differentiate and further educate the market. However, a substantial number of organizations (18 percent) view distributed query as critically important.

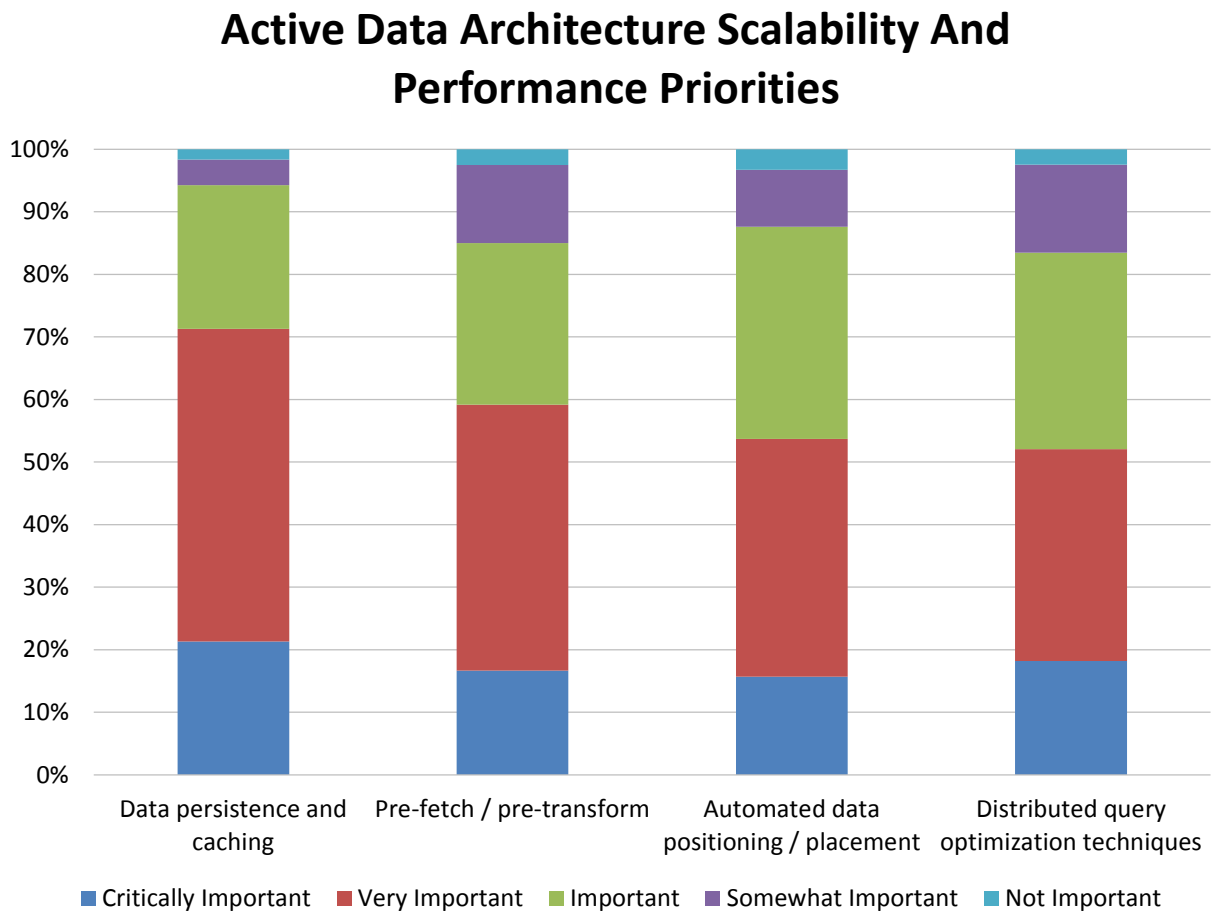


Figure 36 – Active data architecture scalability and performance priorities

On a geographic basis, in general, Asia Pacific organizations place a higher priority on all scalability and performance features for active data architecture than organizations in North America and EMEA (fig. 37). Notably, Asia Pacific organizations tend to equally prioritize each of the features. In contrast, North America and EMEA organizations place a higher priority on data persistence and caching and a much lower priority on distributed query optimization. This implies that more persistence-centric and physical architectures, rather the more distributed, dynamic, and virtualized vision of active data architecture, remain at the forefront of many organizations' strategy.

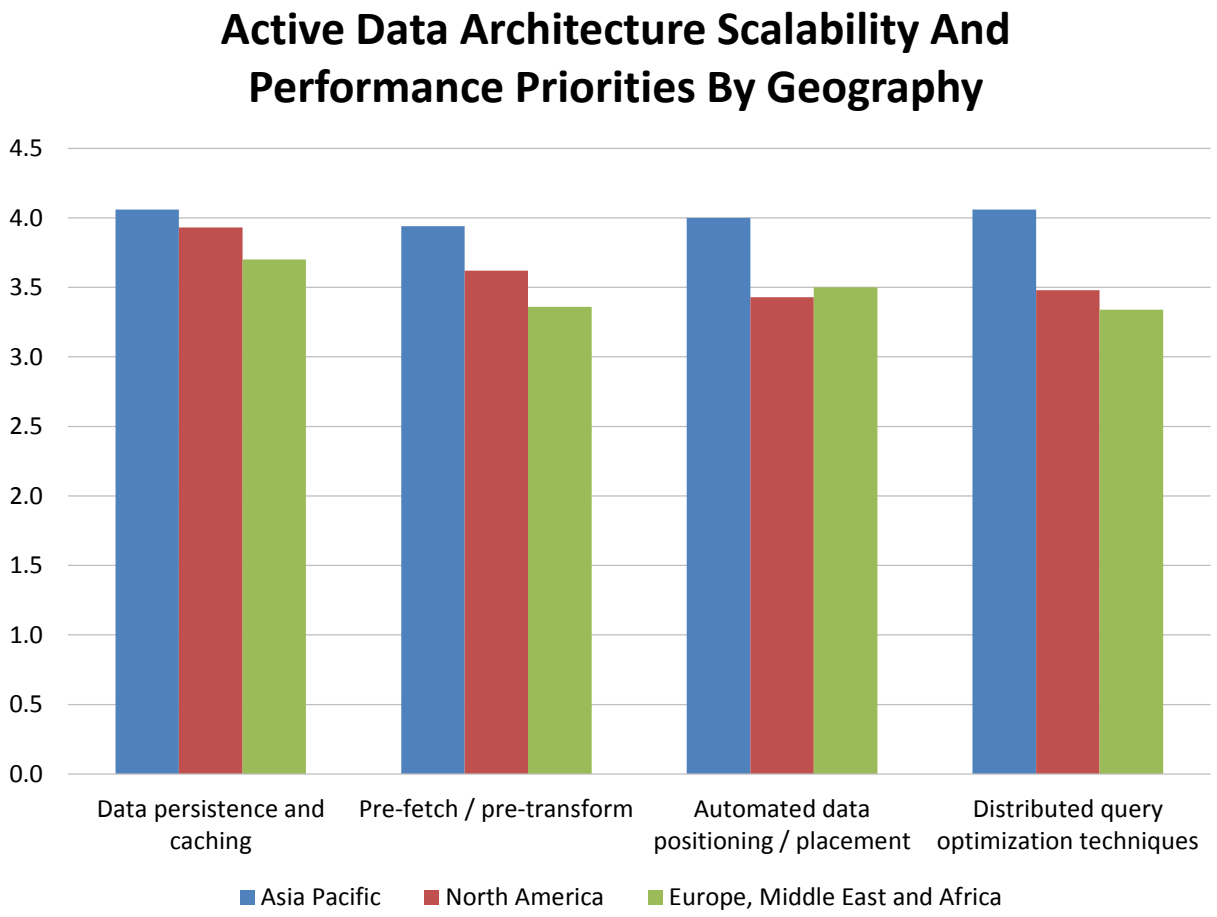


Figure 37 – Active data architecture scalability and performance priorities by geography

Slight variations between industry prioritize are visible from the survey data. For example, business services organizations prioritize data persistence and caching at a significantly higher level than other industries (fig. 38). Technology organizations, likely aware of active data architecture trends, reflect relatively equal priorities across all features. And government organizations express lower priorities than all other industries for all features except data persistence and caching; again, this is likely due to momentum for existing architectures that are largely about physical data movement and single-source data access.

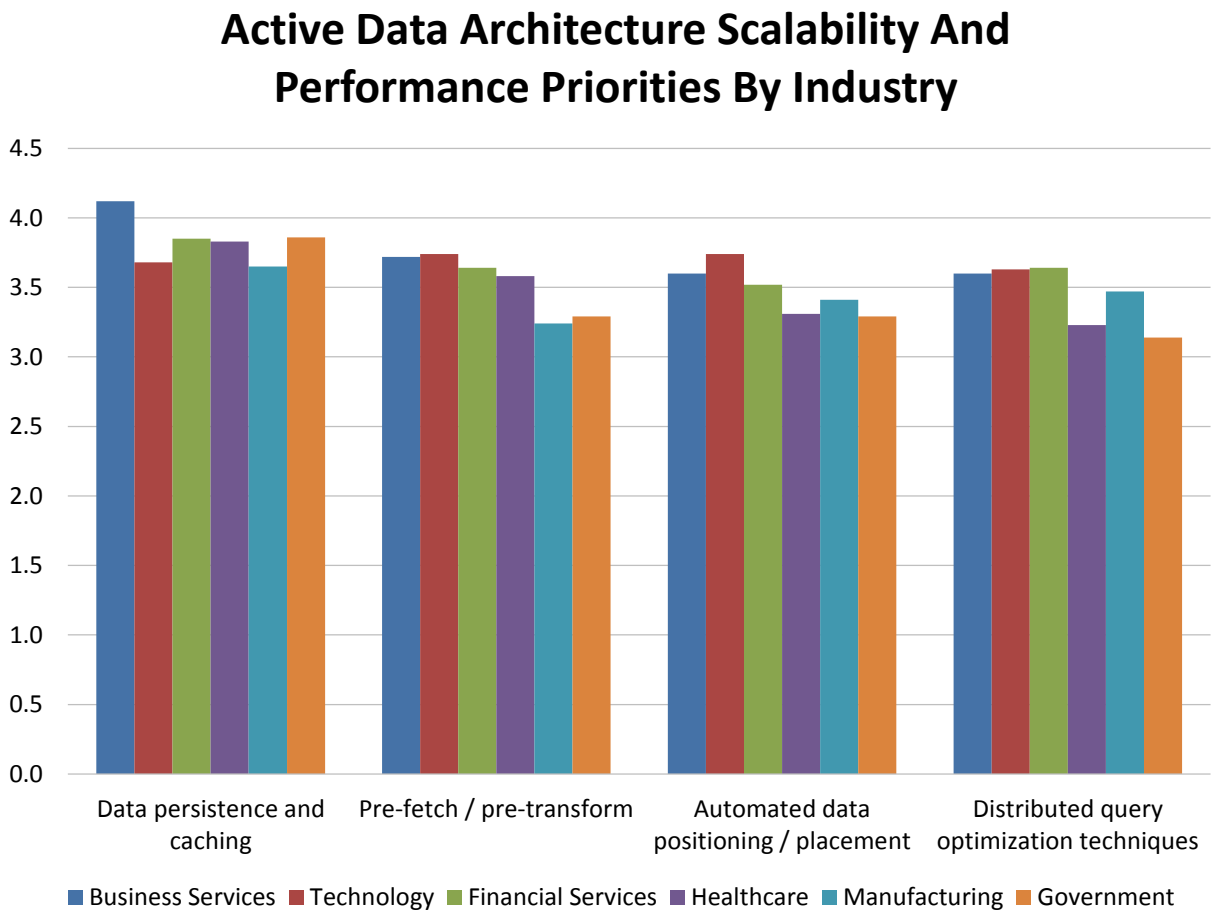


Figure 38 – Active data architecture scalability and performance priorities by industry

Prioritization of scalability and performance features is generally related to degree of BI success, with organizations reporting their BI efforts are extremely successful also prioritizing these features higher than their peers (fig. 39). Organizations with greater maturity in their BI efforts likely see the need for a different approach to data architecture that will support more complex, diverse, and distributed data requirements.

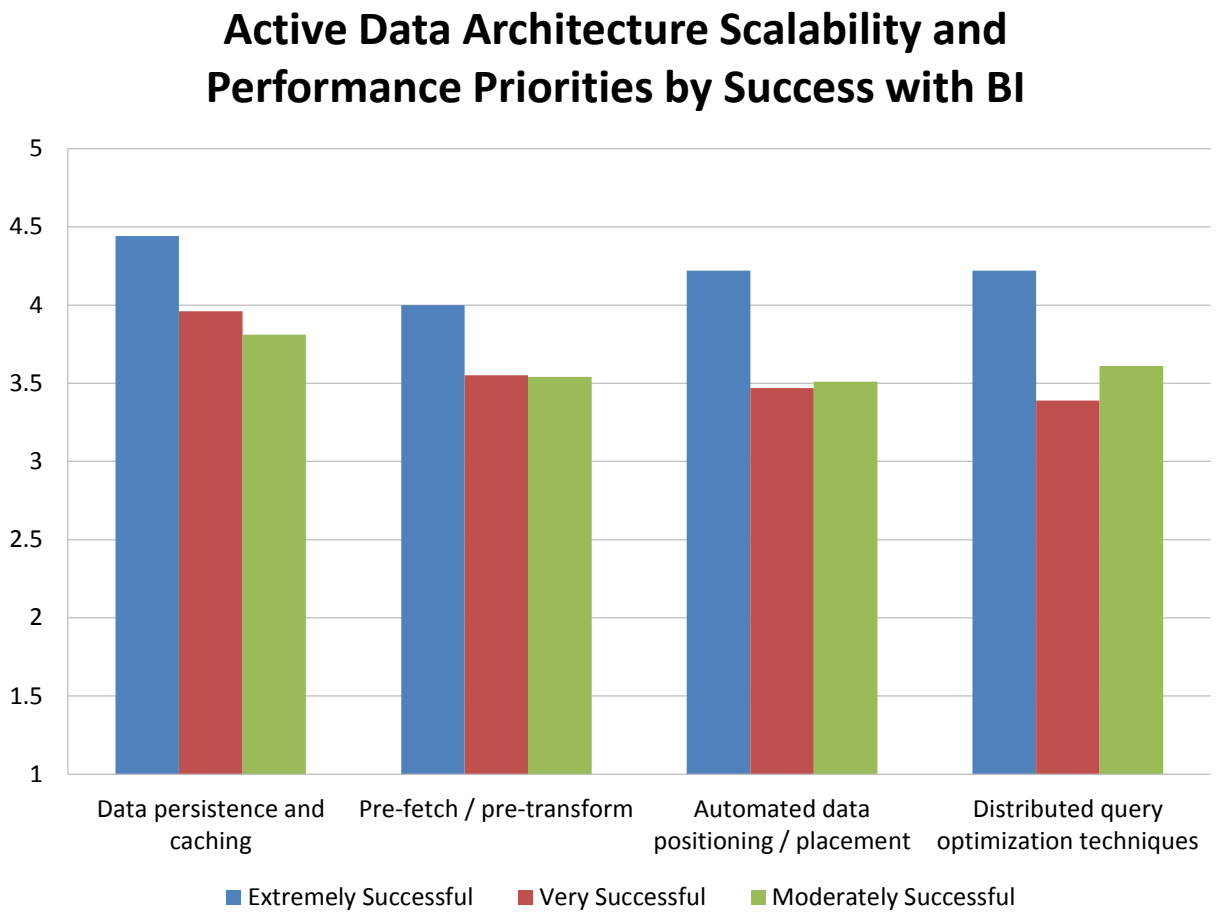


Figure 39 – Active data architecture scalability and performance priorities by success with BI

Importance and Prioritization of Dynamic Optimization Capabilities

Inherent in the concept of active data architecture is the idea of dynamism, meaning that the architecture can dynamically change its behavior to optimize levels of service in light of current conditions and business requirements. This may mean adjusting data placement or choosing an alternate style of data integration or reprioritizing workloads in light of available resources. In order to optimize, monitoring behavior of the architecture is the starting point. Hence, organizations indicate that monitoring of KPIs for performance, cost, and capacity are the most important features in their view. Nearly 75 percent of organizations indicate that monitoring of performance KPIs is critically important or very important (fig. 40). Capabilities for optimizing the architecture, either manually (via recommendations) or in an automated fashion, are subsequent steps in organizations' plans, with 46 percent and 37 percent rating these two features as either critically or very important.

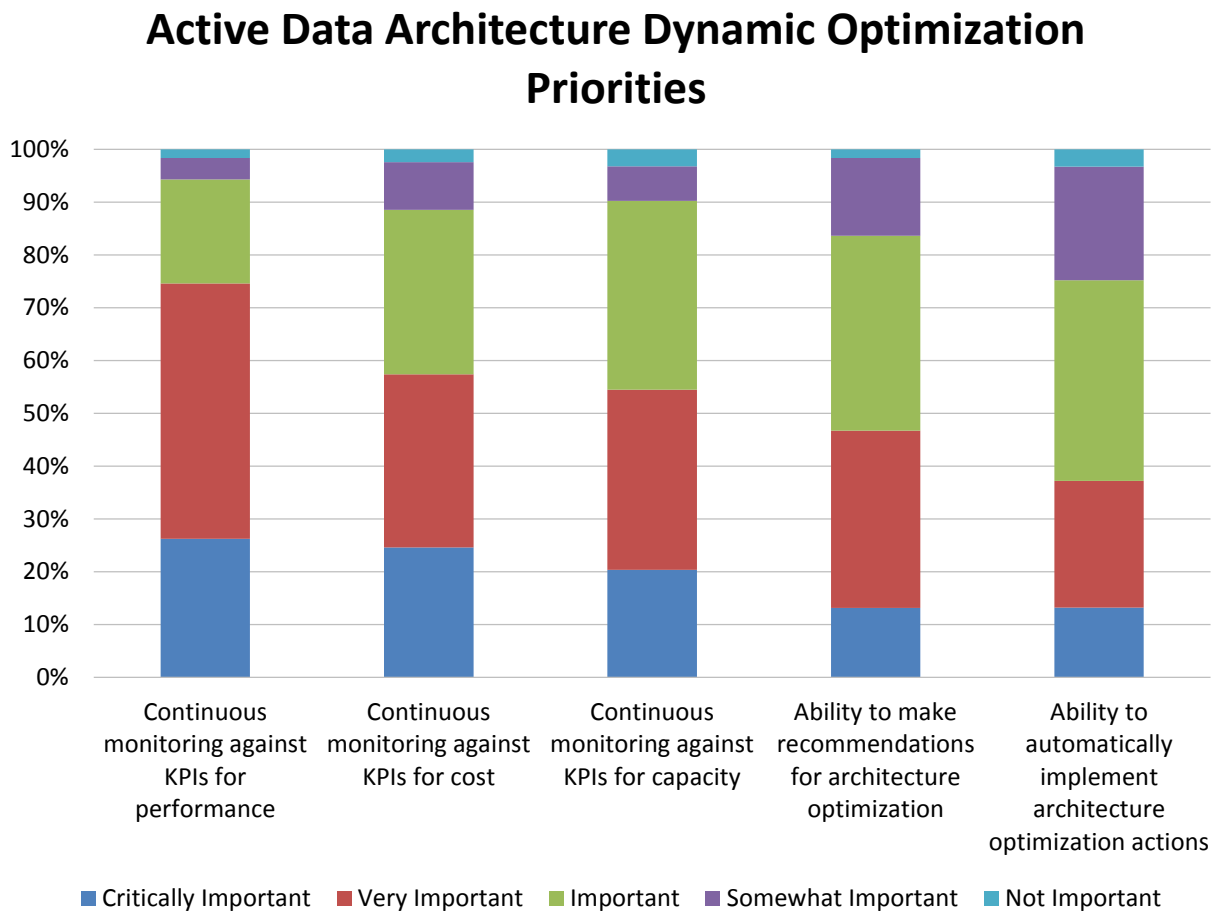


Figure 40 – Active data architecture dynamic optimization priorities

In this feature category, Asia Pacific organizations again exhibit more vision of the importance, rating most features as a higher priority than other regions (fig. 41). The gap relative to North America and EMEA organizations is most substantial for the two features that are about actively optimizing the architecture: optimization recommendations and automation of optimization actions. This implies Asia Pacific may arrive at the full benefits of active data architecture, via dynamic behavior, sooner than counterparts in other regions.

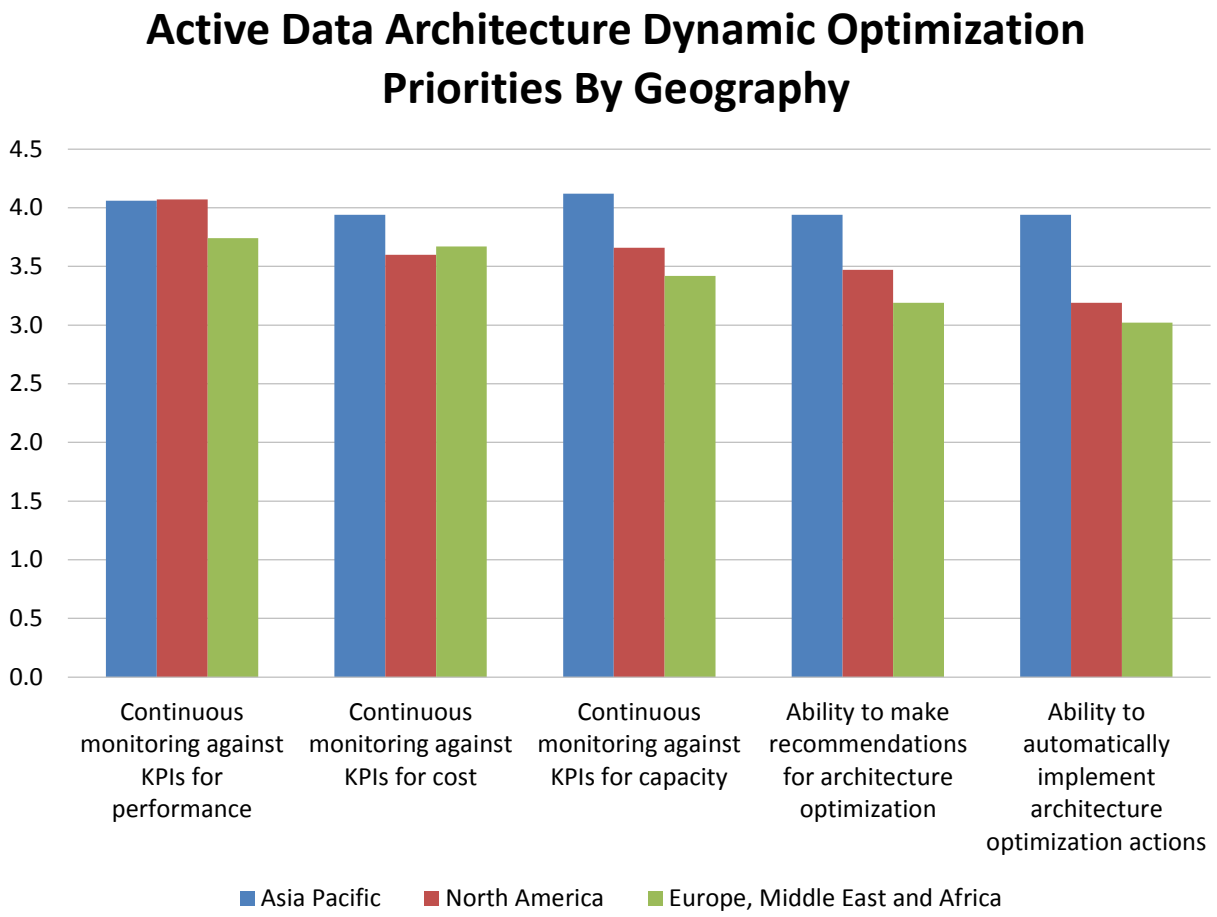


Figure 41 – Active data architecture dynamic optimization priorities by geography

In a pattern similar to that seen for the scalability and performance features, industries including business services, technology, and financial services place a higher priority on almost all the dynamic optimization features compared to other industries (fig. 42). The business services sector in particular stands out for the broader vision of active data architecture, rating the priority of optimization recommendations and automated optimization actions substantially higher. While the government sector places a high priority on monitoring of performance KPIs, that industry rates most other feature substantially lower, implying a longer journey toward the goals of active data architecture.

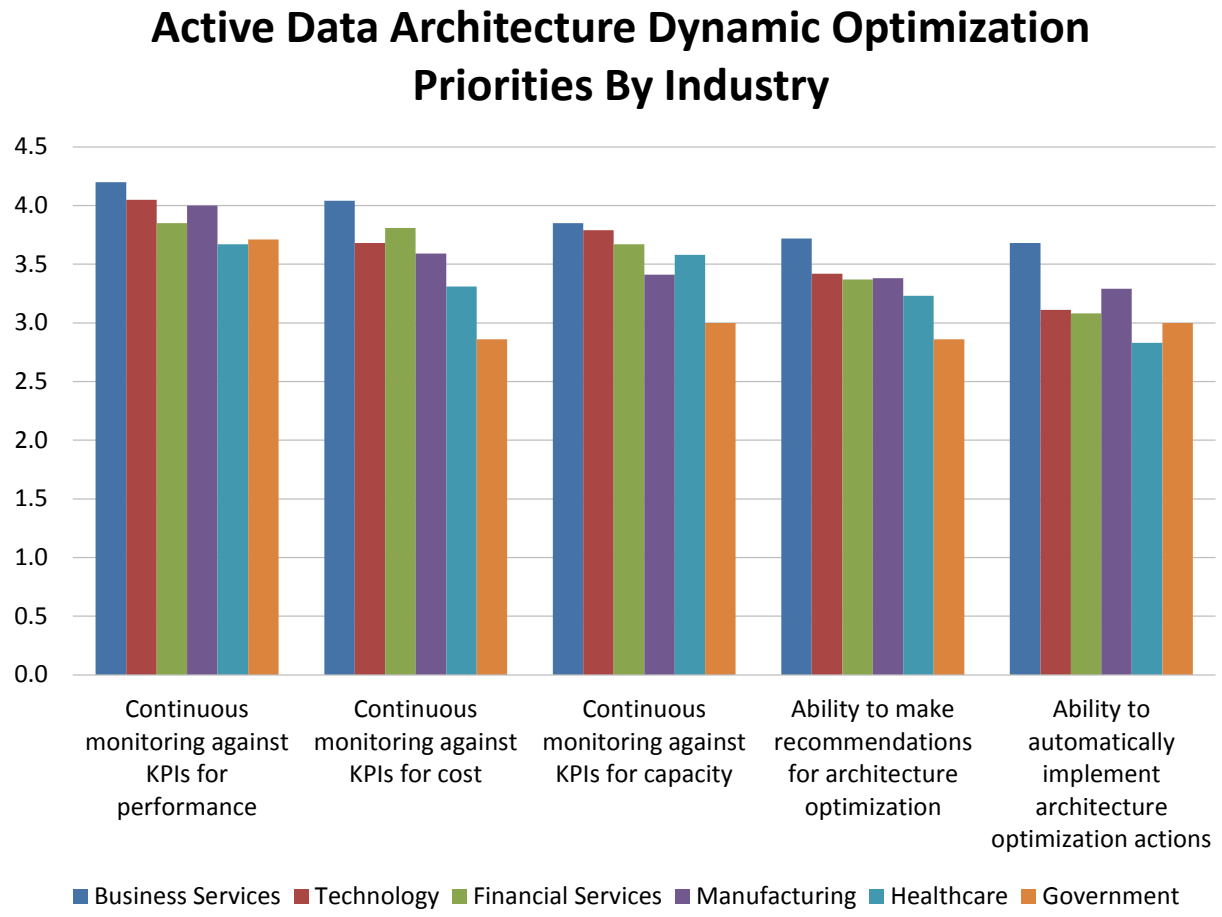


Figure 42 – Active data architecture dynamic optimization priorities by industry

Consistent with other feature categories, organizations with BI investments deemed extremely successful tend to prioritize dynamic optimization features at a much higher level than those experiencing lesser degrees of BI success (fig. 43). The gap is most prominent for the optimization recommendations and automatic implementation of optimization actions features, implying those organizations with the most BI success are also most likely to capture the full potential of active data architecture.

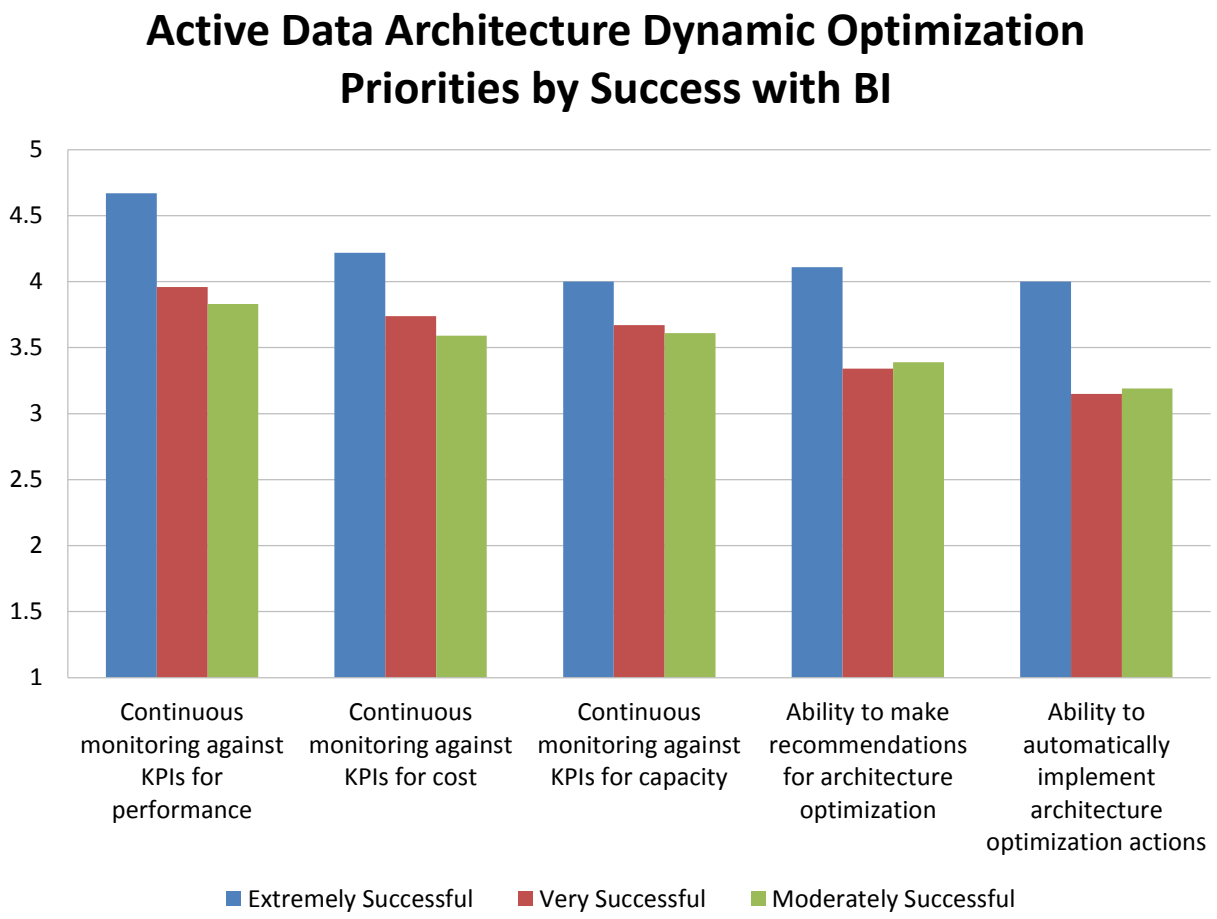


Figure 43 - Active data architecture dynamic optimization priorities by success with BI

Importance and Priorities for Access Styles of Active Data Architecture

Because the vision for active data architecture is that it is a pervasive and dynamic layer supporting a range of data product needs, organizations must access the core capabilities of the architecture in a variety of ways: programmatically from various systems and tools using APIs, orchestrated in any sequence or combination, and designed into business processes by using packaged workflow or business process design and execution tools. Most organizations rate the importance of APIs highest, with nearly 88 percent stating this feature is critically important, very important, or important (fig. 44). This is followed by arbitrary execution of sequences and combinations of capabilities (77 percent) and execution via workflow tools (75 percent).

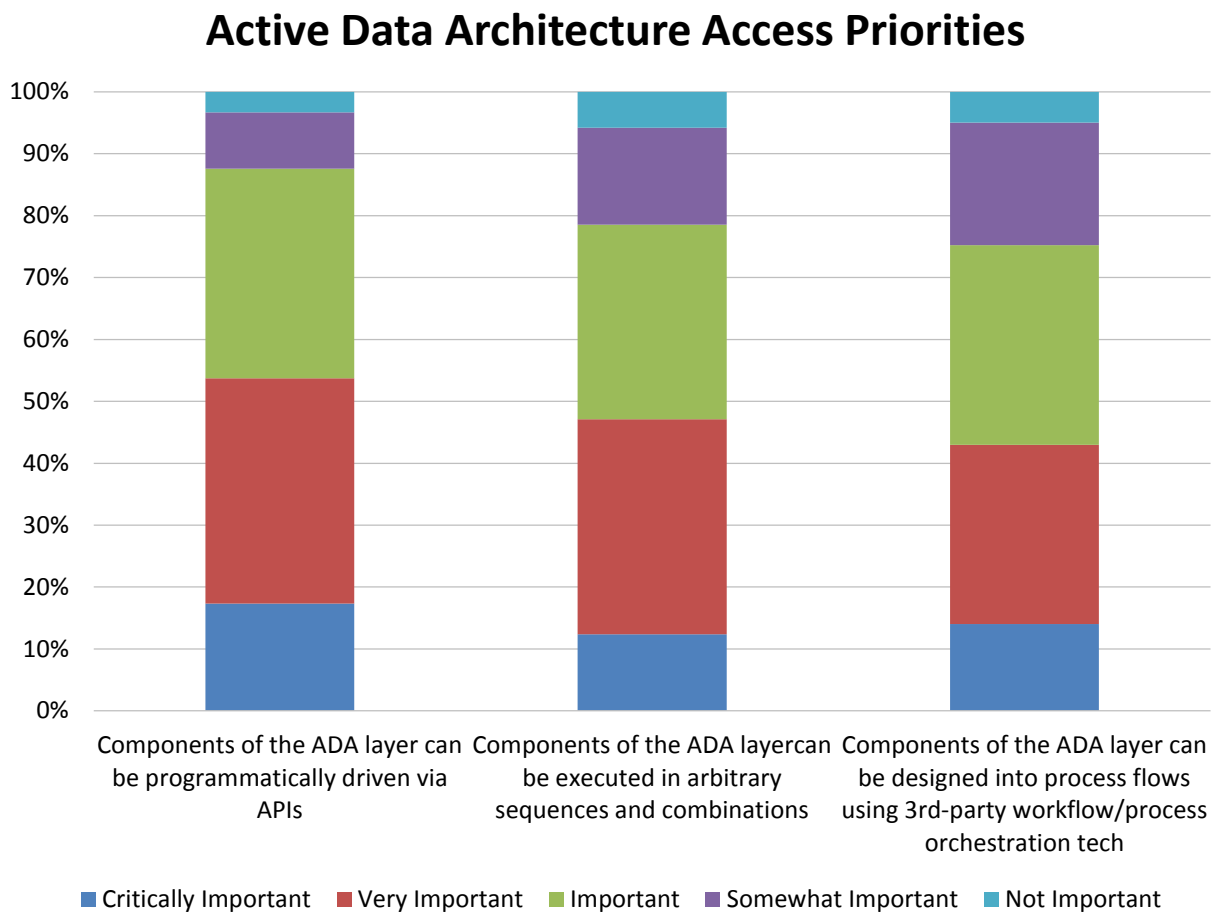


Figure 44 – Active data architecture access priorities

On a regional basis, substantial differences in priorities for active data architecture access features are clearly visible. Asia Pacific organizations rate each of the features at an almost equally high priority and substantially higher than peers in other regions (fig. 45). The gap is most substantial for the ability to orchestrate active data architecture capabilities in the context of third-party workflow or process orchestration tools.

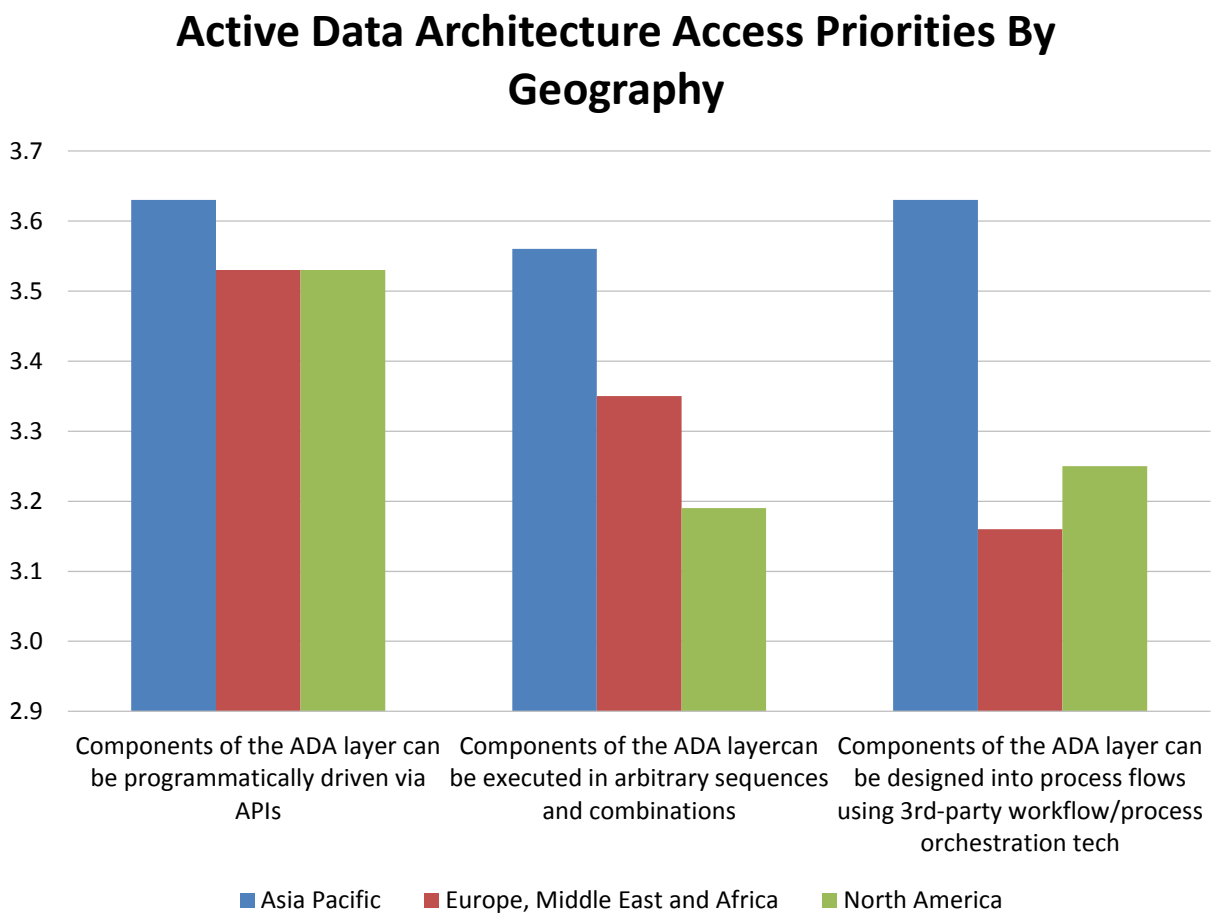


Figure 45 – Active data architecture access priorities by geography

Similar to other feature categories, business services, technology, and financial services exhibit the strongest preferences for each of the active data architecture access features (fig. 46). The healthcare sector also places high priority on each of the features, actually exceeding all industries in priority ranking for the ability to execute active data architecture capabilities in arbitrary sequences and combinations. Here, too, government organizations assess a much lower priority to all features, implying the full vision for active data architecture lags in that sector.

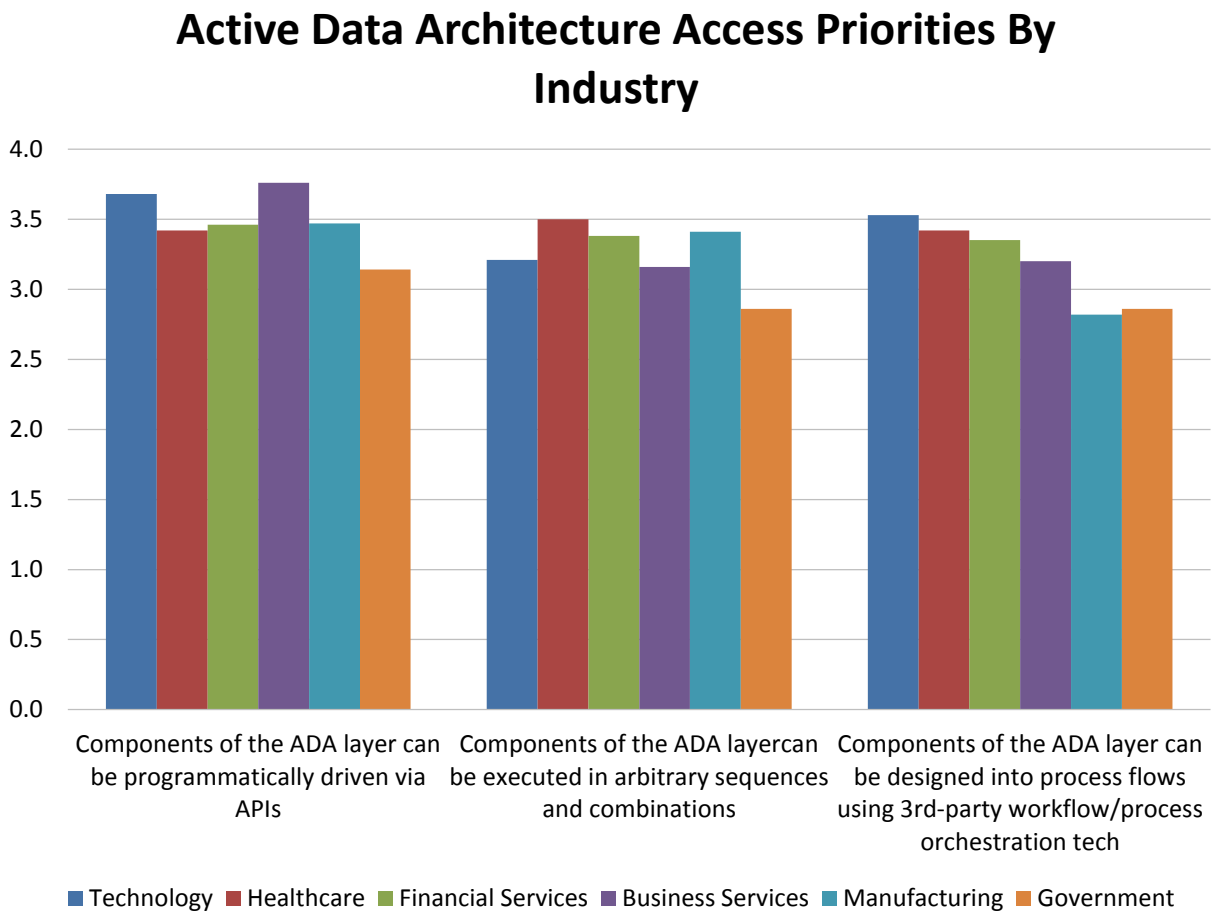


Figure 46 – Active data architecture access priorities by industry

Surveyed organizations with greater success in their BI initiatives tend to more heavily prioritize more heavily these features that enable more automation in an active data architecture. Those reporting their BI investments are *extremely successful* place a 10-15 percent higher priority, on average, on these capabilities with the most significant gaps over their less-successful peers relative to APIs and the ability to execute componentry in arbitrary sequences and combinations (fig. 47). There is only a very slight difference in average priorities among those organizations where BI is very successful or moderately successful.

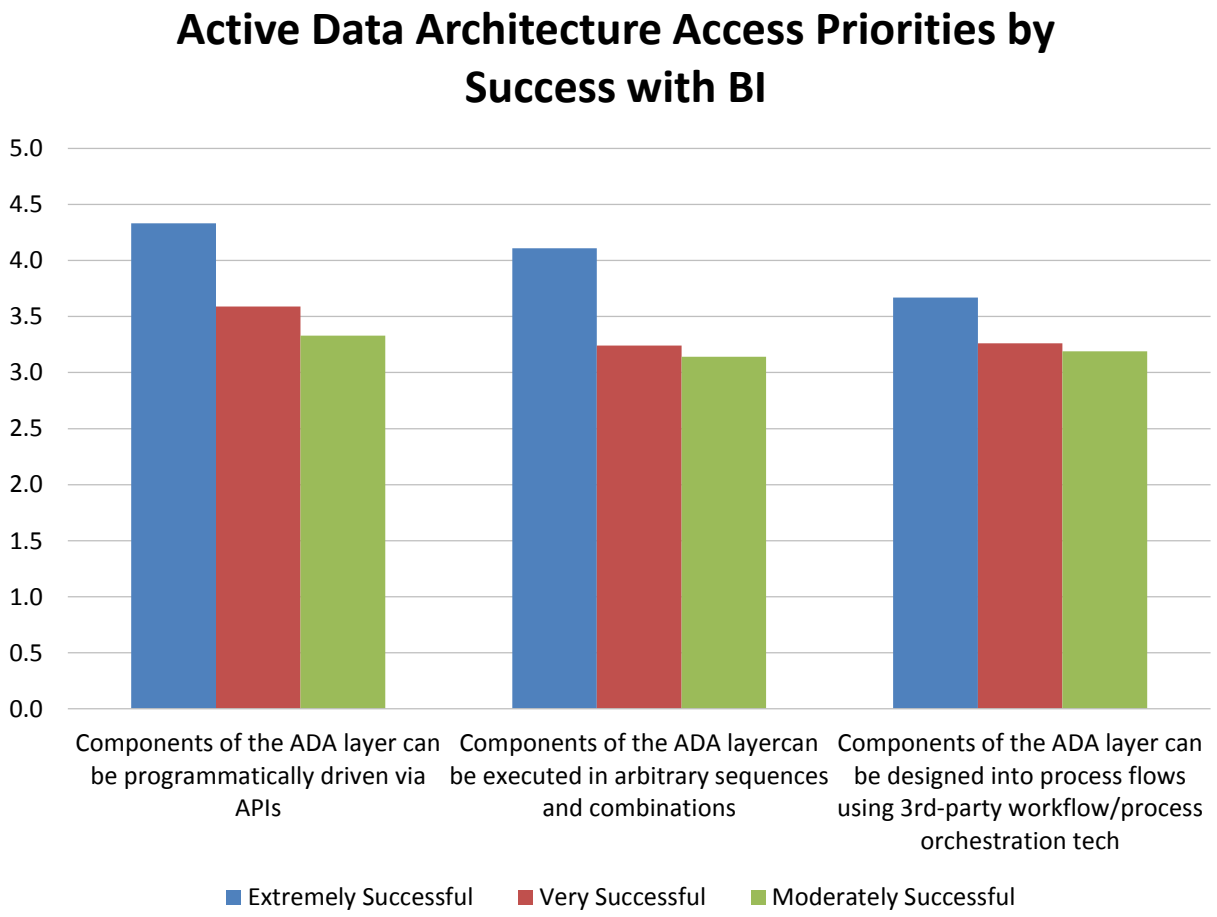


Figure 47 – Active data architecture access priorities by success with BI

Sources of Technology to Support Active Data Architecture

Since active data architecture draws on technology residing in various markets, customers have a lot of choices for sourcing the components most critical for their specific vision and road map. Possible providers range from those in the data engineering / data integration tools market, to emerging vendors of data governance technology, metadata management solutions, and database / data persistence platforms. Even the prevalent providers of BI and analytics tools provide capabilities relevant in the context of active data architecture, and their footprint in the market also makes them a relevant source.

We asked survey respondents to specify their preferred source for technology to enable their active data architecture plans. Because of the diversity of capabilities needed for active data architecture, multiple sources can be relevant for a single organization. Slightly over 58 percent indicate they would leverage technology from vendors of data integration tools. Forty-eight percent indicate that prominent BI and analytics providers would be a source (fig. 48). Thirty-eight percent indicate they would look to vendors specifically positioning themselves as providers of data mesh or data fabric technology. Somewhat smaller segments of the sample plan to look toward vendors of data catalog / metadata (36 percent), cloud infrastructure (34 percent), database technology (34 percent), or data governance tools (34 percent).

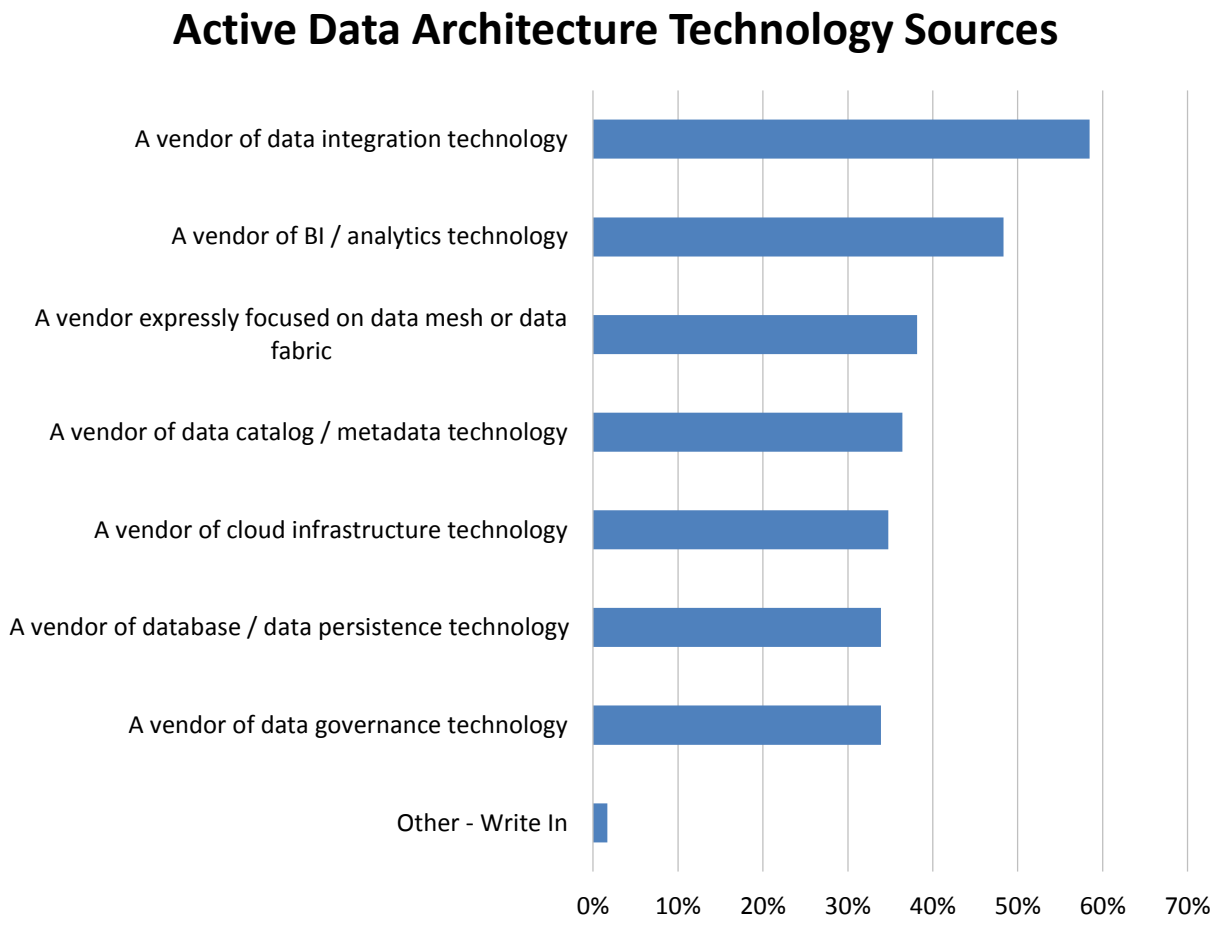


Figure 48 – Active data architecture technology sources

While most regions see data integration technology vendors as their primary choice for supporting an active data architecture, variances exist. Organizations in North America and EMEA are much more likely to leverage technology from their preferred vendors of BI tools than those in Asia Pacific (fig. 49). In contrast, Asia Pacific organizations plan to prioritize working with vendors that are data mesh / fabric focused or metadata focused more heavily than their peers in other regions. This, again, speaks to the broader and more dynamic vision for active data architecture: organizations emphasizing application / visualization-independent capabilities are likely to gain more benefits from active data architecture concepts.

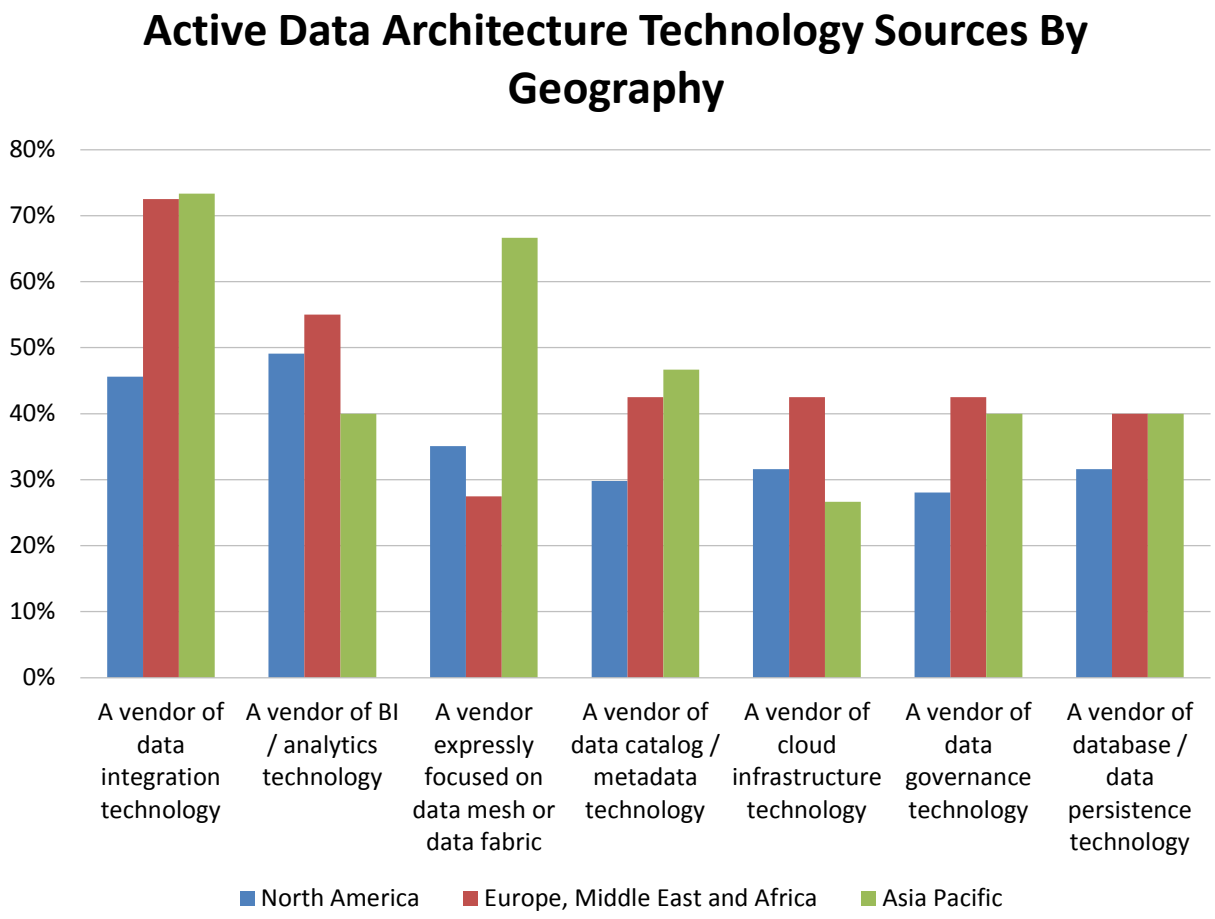


Figure 49 – Active data architecture technology sources by geography

When looking at the survey data from an industry view, the focus remains most heavily on vendors of data integration technology, particularly so for organizations in financial services, technology, and government (fig. 50). However, technology organizations de-emphasize BI tools vendors as a choice and place a stronger emphasis on providers of data governance technology. Organizations in the business services sector are more likely to choose vendors of cloud infrastructure services over most other vendor types. Manufacturing organizations reflect a very different set of preferences, with a substantially reduced focus on providers of data mesh and fabric technology but a strong focus on metadata management and governance vendors.

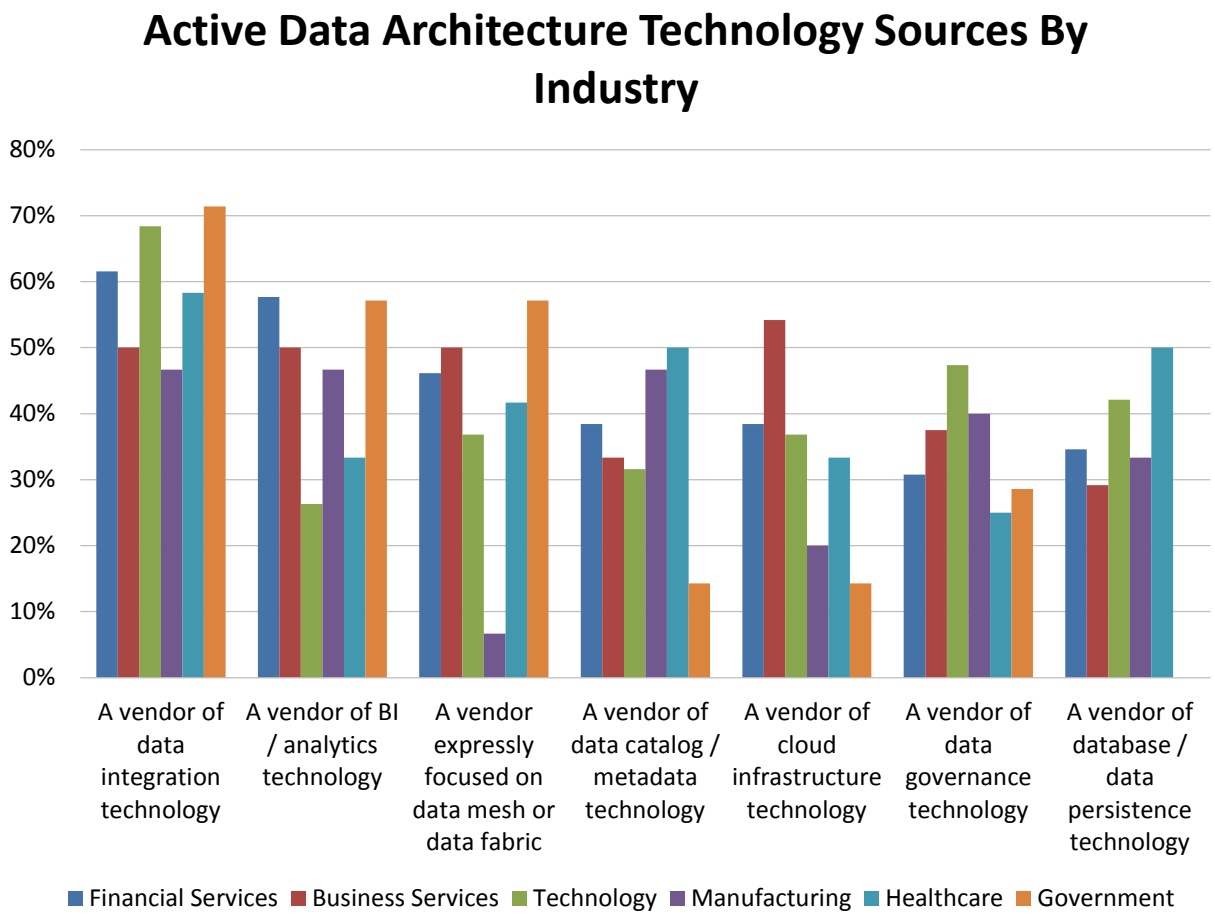


Figure 50 – Active data architecture technology sources by industry

Clear and different preferences are visible based on organizations' degree of success in their BI initiatives. Interestingly, those organizations with the greatest degree of success (*extremely successful*) express the lowest preference for providers of BI and analytics technology to support their active data architecture needs (fig. 51). In contrast, they focus much more heavily on vendors of data integration technology, fabric/mesh-specific providers, and vendors of data governance technology. In contrast, organizations that are mostly or *moderately successful* with BI indicate BI and analytics tools vendors as their strongest preferences, followed by vendors of data integration technology. These organizations prioritize vendors of metadata and governance technology at a lower level, which may inhibit their ability to capture the full benefits of active data architecture.

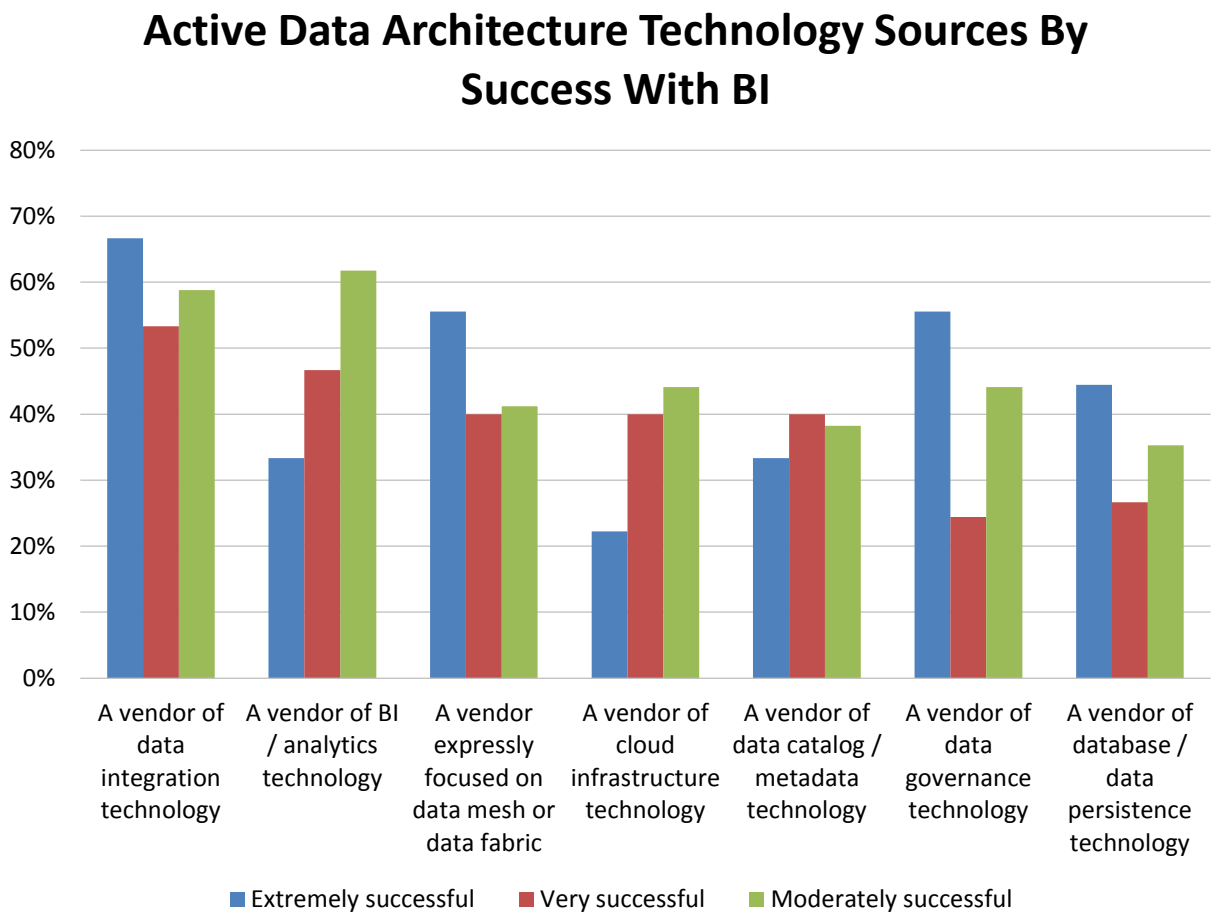


Figure 51 – Active data architecture technology sources by success with BI

Industry and Vendor Analysis

Industry and Vendor Analysis

Active Data Architecture Industry Importance

We asked industry respondents to indicate the importance of active data architecture. Respondents include vendors that do not have solutions specifically targeted at active data architecture. The greatest majority (82 percent) report that active data architecture is critically important, followed by very important at 31 percent, with 6 percent indicating it is somewhat important (fig. 52).

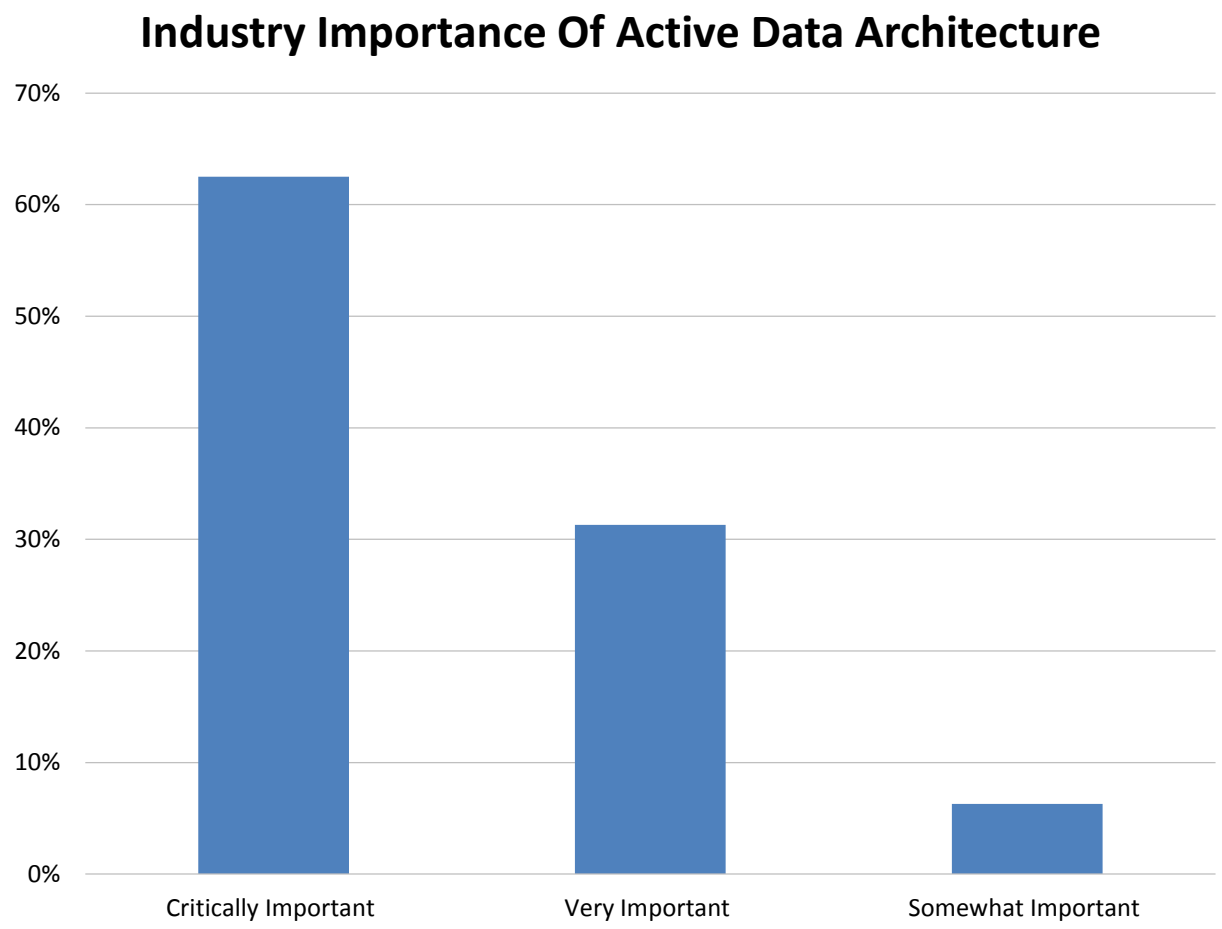


Figure 52 – Industry importance of active data architecture

Industry Support for Active Data Architecture Deployment Options

As with users, vendors made the shift to cloud-based deployment, with SaaS / cloud support currently at 80 percent (fig. 53). Eighty-five percent of vendors indicate support for active data architecture capabilities in a completely cloud-, application- and database platform-independent manner. Seventy-eight percent support on-premises deployment, which is the lowest level of any of the deployment options.

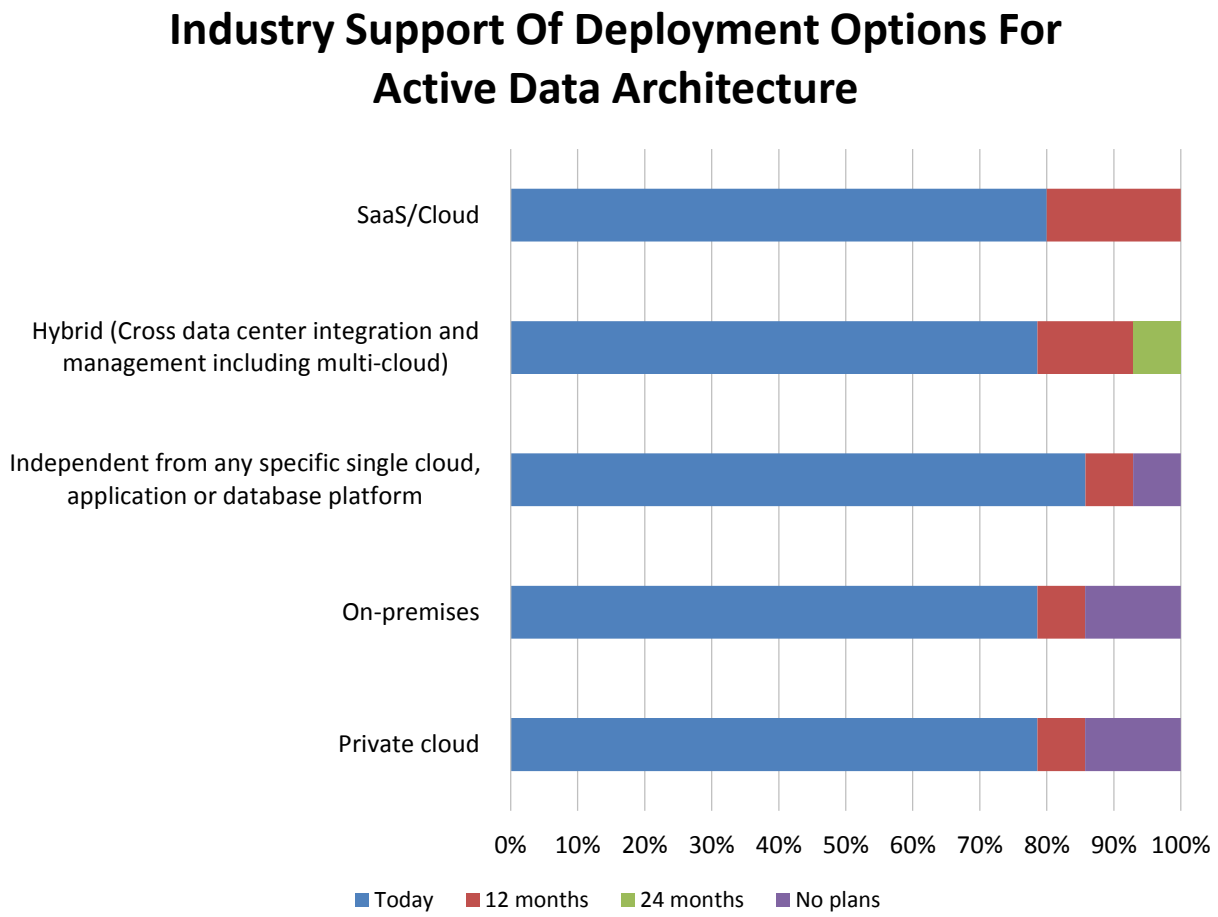


Figure 53 – Industry support of deployment options for active data architecture

Active Data Architecture Pricing Models

In 2024, the data suggest that vendor organizations emphasize named user licensing, which appears to be somewhat at odds with the idea of active data architecture as a "layer" rather than a specific piece of user-facing technology (fig. 54). However, this is significant activity toward pricing models more aligned with this idea, including platform utilization, query workload, and capacity metrics.

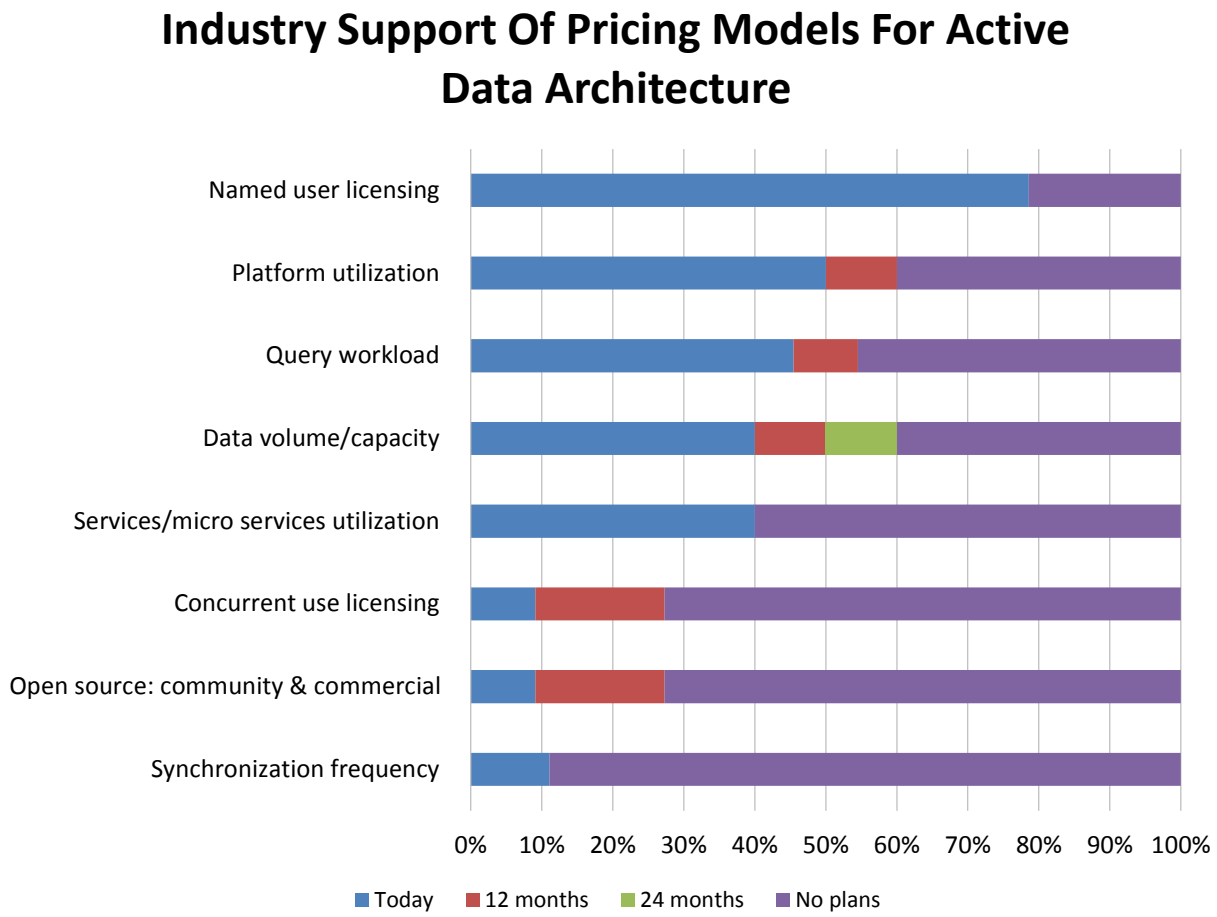


Figure 54 – Industry support of pricing models for active data architecture

Industry Support for Data Integration Features

In large part, users can expect a majority of vendors to meet their data integration requirements for active data architecture because all data integration styles are well supported (fig. 55), although potential gaps exist in particular for data virtualization, which will be a key capability for many users.

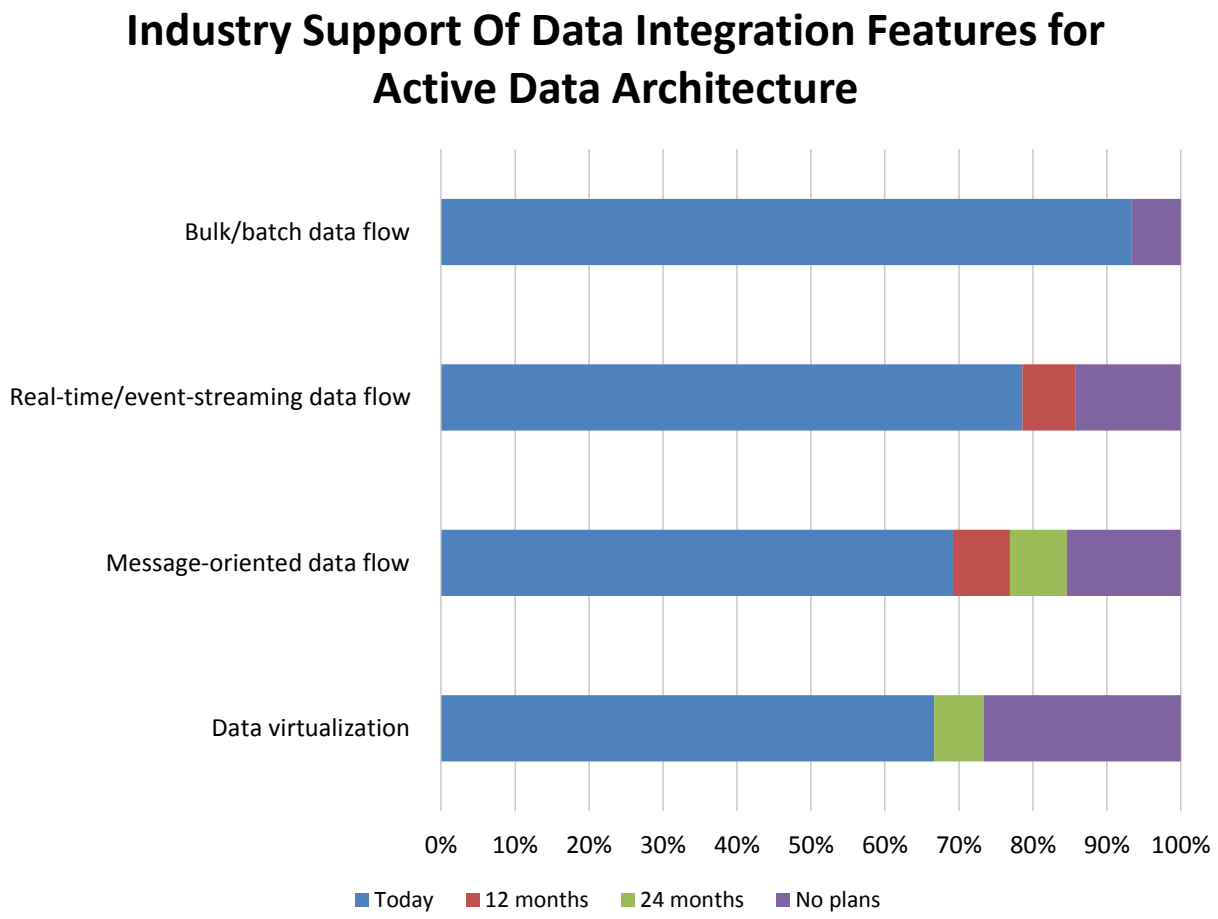


Figure 55 – Industry support of data integration features for active data architecture

Industry Support for Data Cataloging and Metadata Management Features

Vendor support for metadata management capabilities is much more variable than for data integration features. While metadata ingestion is well supported, users may find many other features lacking for many vendors, where current support is provided by less than 50 percent of vendors (fig. 56). In particular, many vendors lack advanced capabilities to enable dynamic optimization.

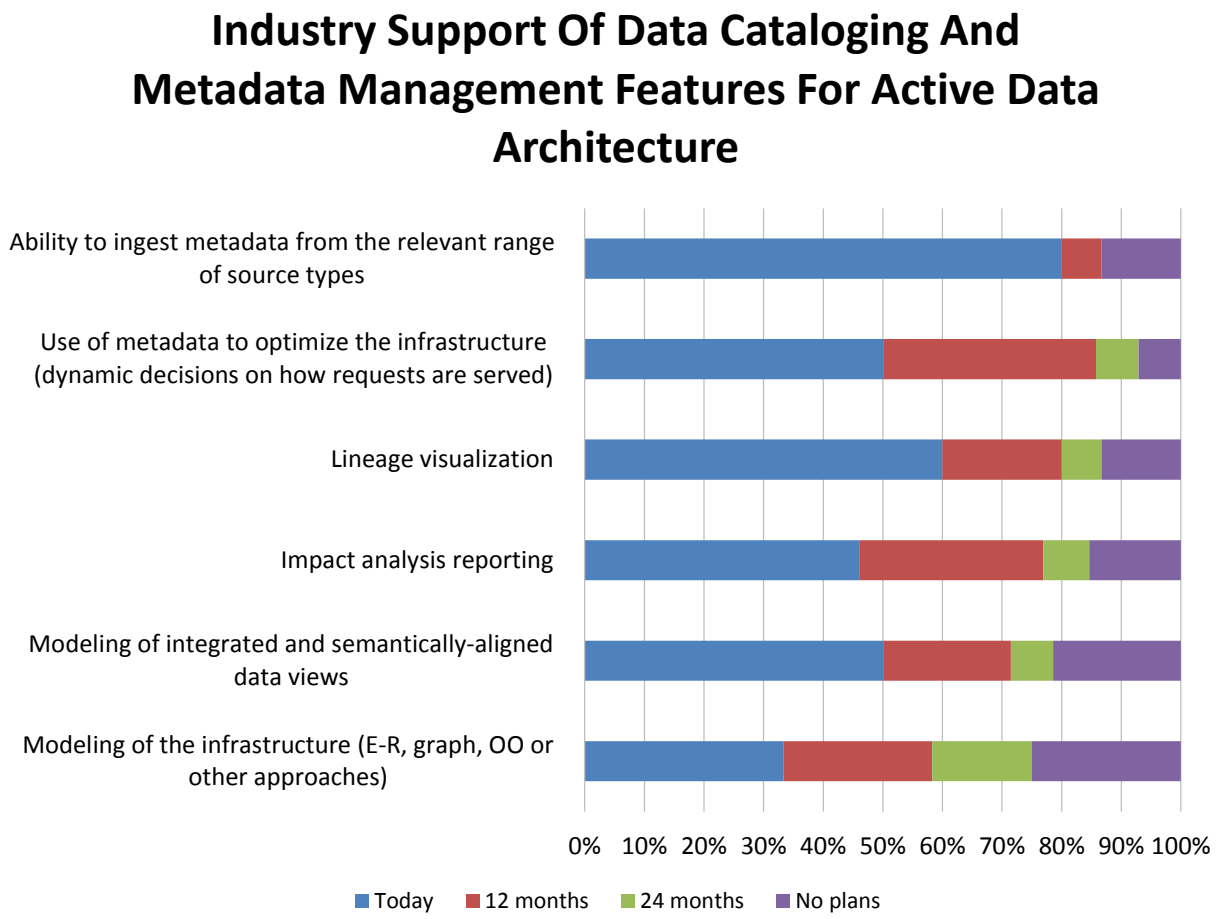


Figure 56 – Industry support of data catalog and metadata management features for active data architecture

Industry Support for Governance Features

The security and privacy dimensions of governance have the highest levels of current vendor support, although vendor plans for the next 12 months will bring life cycle and quality capabilities up to similar levels (fig. 57). Newer governance areas of modeling and metadata / definitions will be somewhat lacking for users over the next 24 months, with many vendors indicating mid-term plans in these areas.

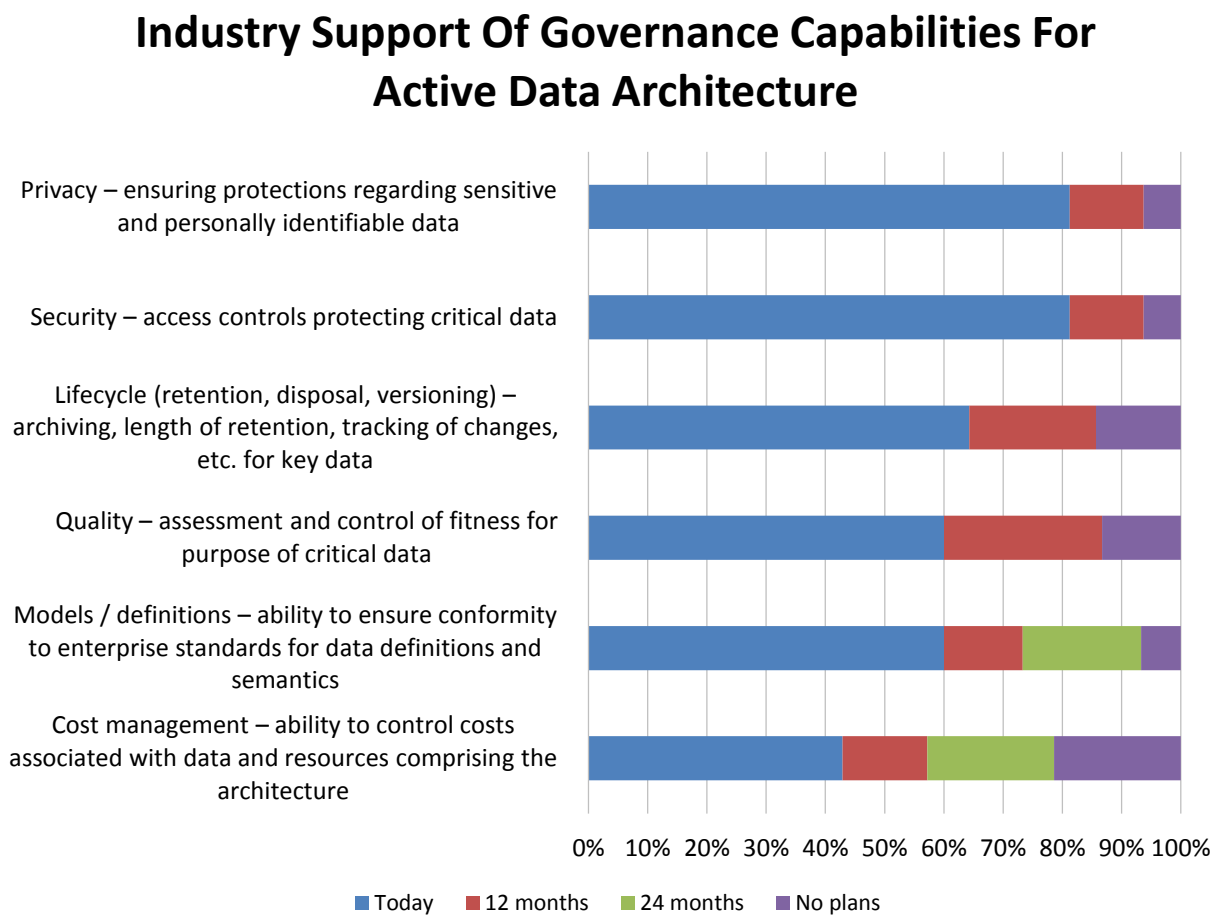


Figure 57 – Industry support of governance features for active data architecture

Industry Support for Scalability and Performance Capabilities

User demands for enabling strong scalability and performance in their active data architecture work are well supported currently, with over 90 percent of vendors indicating current support for distributed query optimization and data persistence / caching (fig. 58). Scalability and performance capabilities that will enable more dynamic and automated behavior of an active data architecture, specifically pre-fetch/transform and automated data positioning, are not as fully prevalent in the market, although some vendors have plans to close this gap in the next 12-24 months.

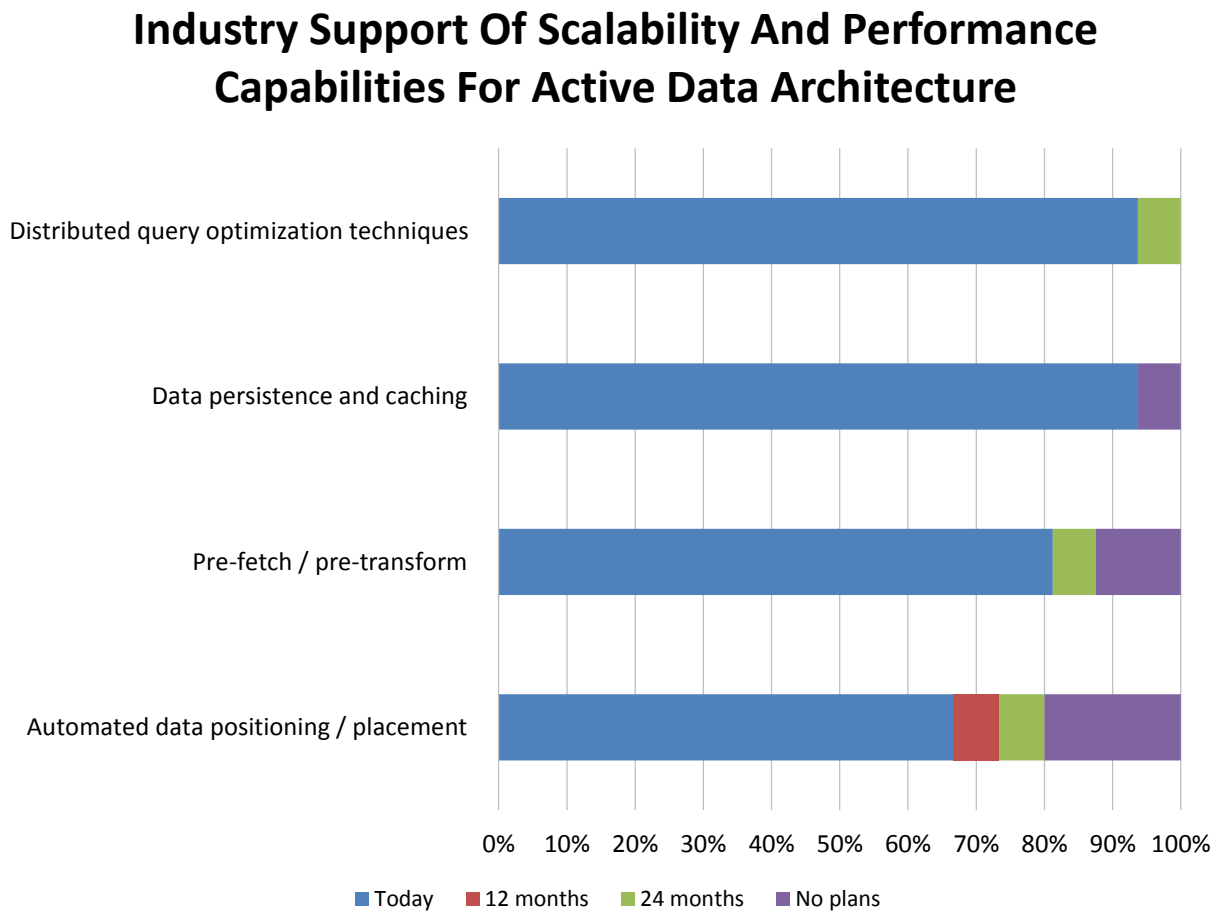


Figure 58 – Industry support of scalability and performance capabilities for active data architecture

Industry Support for Dynamic Optimization Capabilities

Dynamic optimization is what truly makes the active data architecture "active;" yet, this is an area where both user understanding and available capabilities from vendors are lacking. While some vendors provide support for tracking metrics (57 percent support metrics for performance, 50 percent support metrics for capacity, and 30 percent support metrics for cost), less than 30 percent have features for making optimization recommendations or automatic optimization actions (fig. 59). However, these areas appear to be a focus for many vendors over the next 12-24 months.

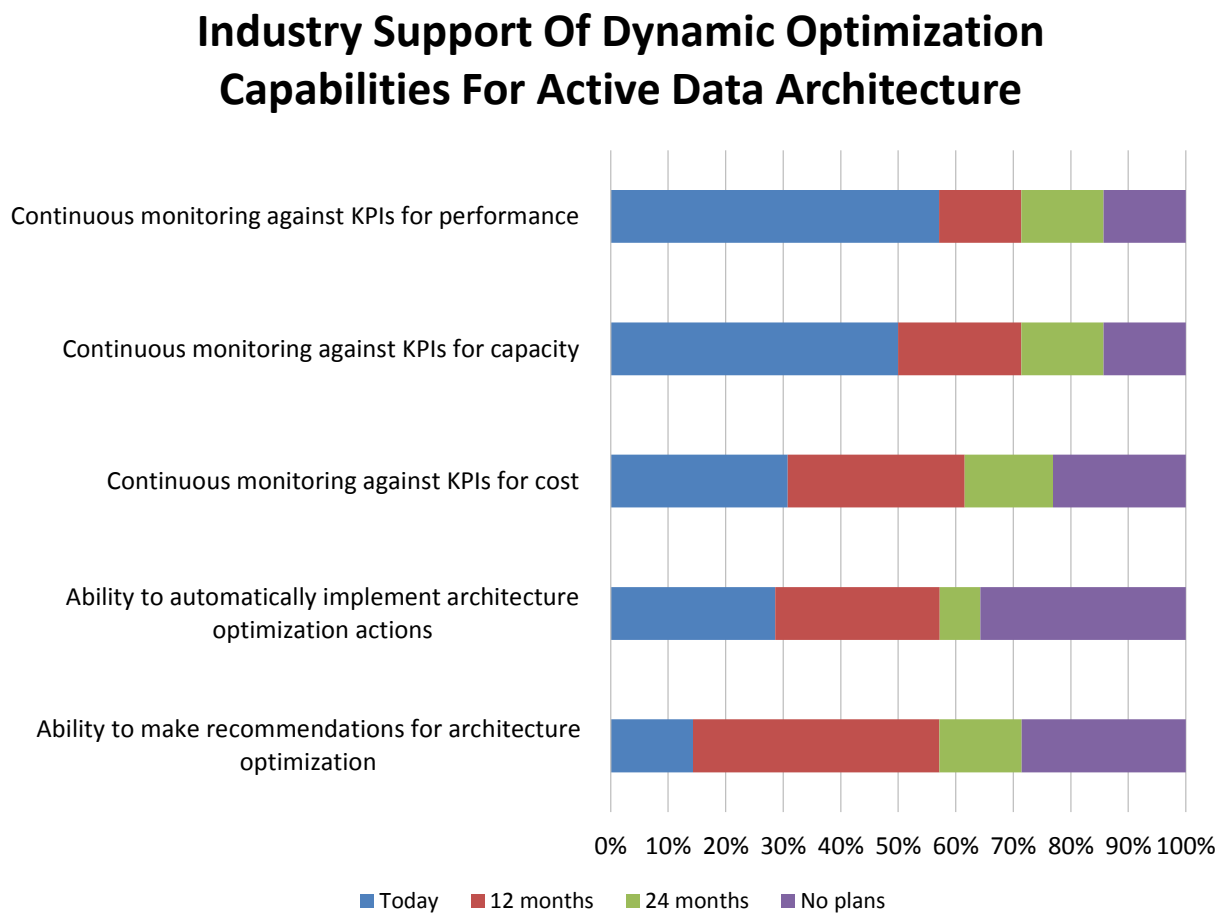


Figure 59 – Industry support of dynamic optimization capabilities for active data architecture

Industry Support for Access Capabilities

Vendors generally support API access to their active data architecture capabilities (86 percent currently support this) (fig. 60). Ability to execute capabilities in arbitrary combinations and sequences, as well as interoperability with third-party workflow and process optimization tools remains currently limited (73 percent and 60 percent, respectively). Vendors will need to evolve these areas to fulfill the full vision for active data architecture.

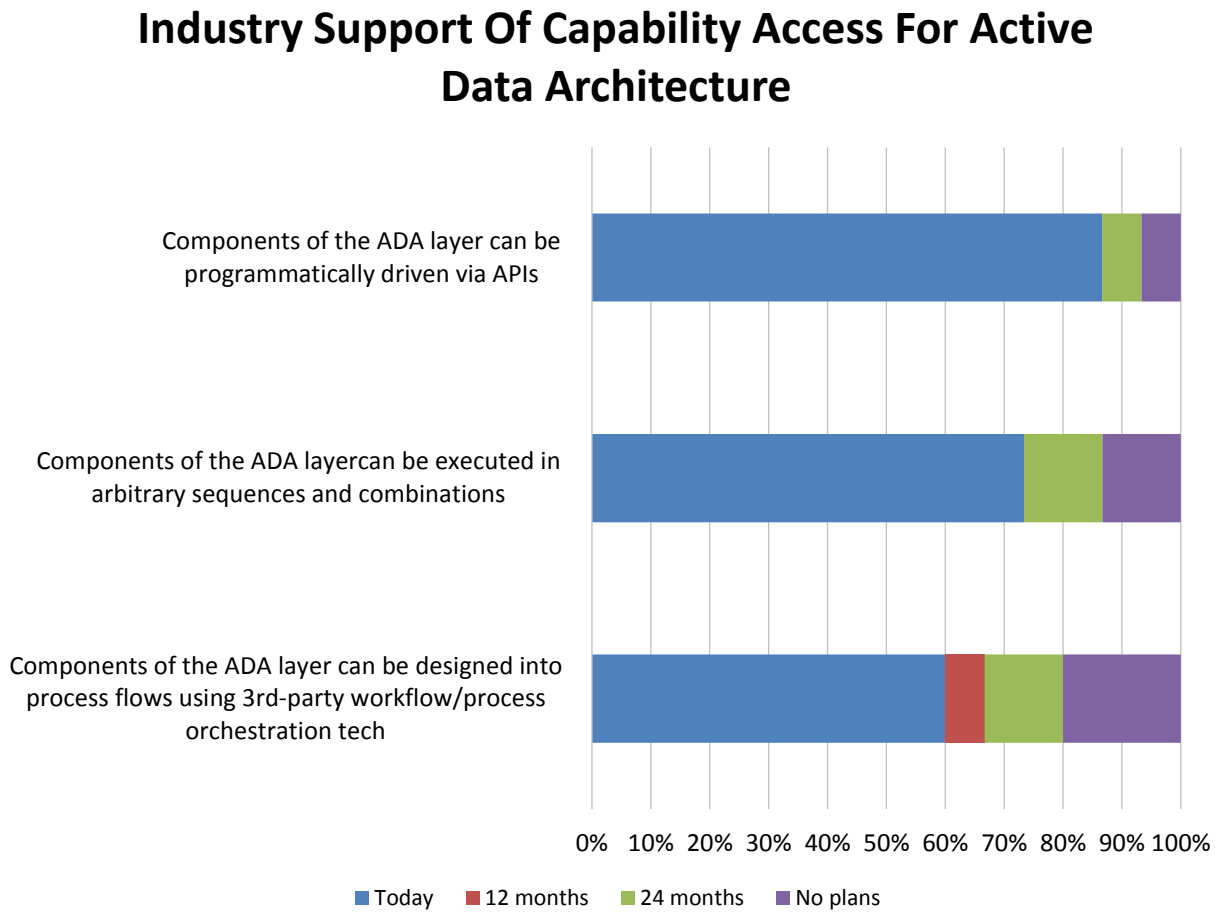


Figure 60 – Industry support of capability access for active data architecture

Vendor Ratings

Vendor Ratings

We include 19 vendors in our 2024 active data architecture ratings (fig. 61). For each vendor, we consider the following features: integration, catalog/metadata management, governance, scale/performance, and ADA access. In this report, we include only vendors that score 50 percent or greater.

The top ten vendors had similar scores, with minor differences among them and with multiple ties. The variations within the next five were somewhat greater, but still relatively modest.

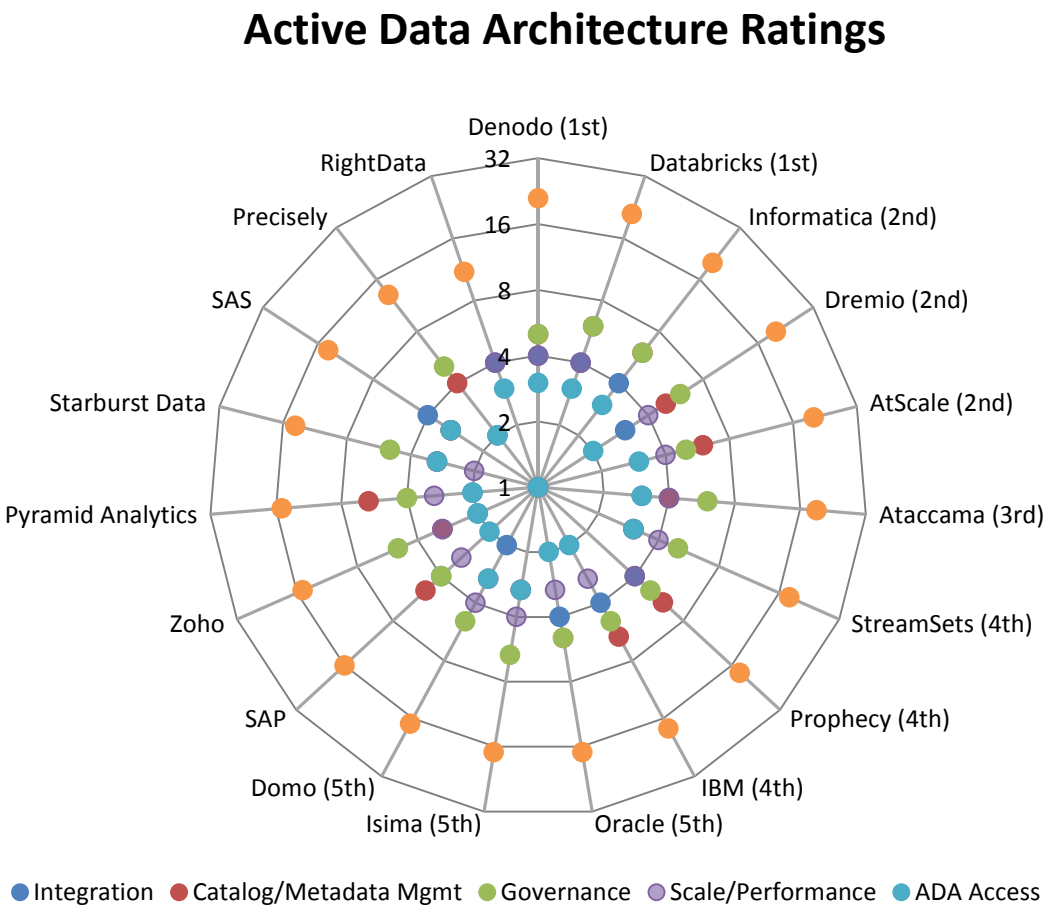


Figure 61 – Active data architecture vendor ratings