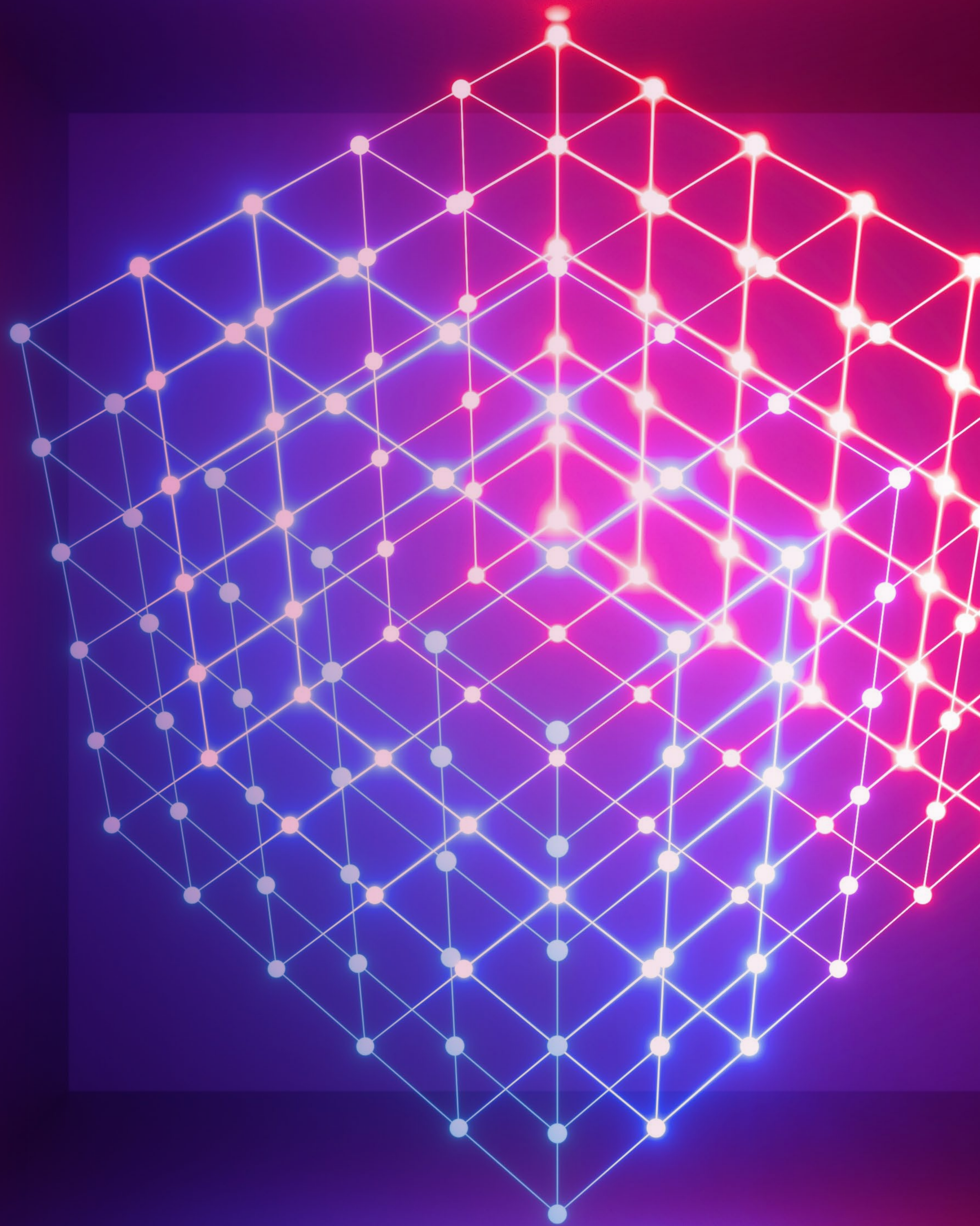




Data Mesh

laying the foundations
for future data architecture



The challenges of monolithic platform architecture

Data architecture is about choosing trade-offs. No single insights platform is optimised for all use cases. While there are design patterns that can co-exist on a platform, experience tells us that a monolithic platform-centric architecture is only ever optimised for one or two types of use cases.

For instance, consider a data lake which is generally designed for ad-hoc data discovery and data science use cases.

Data lakes are optimised for fast ingestion of data and making it available as quickly as possible for insight generation by data practitioners.

It is, however, unfit for real-time integration or application of business rules to derive repeatable trusted management reporting. On the other hand, a data warehouse is useful for structured analysis of data, but cannot handle real-time streaming events and complex event processing. Monolithic platform architecture, regardless of whether it is a data lake or data warehouse (or even a hybrid data 'lakehouse'), isn't setting organisations up for success.

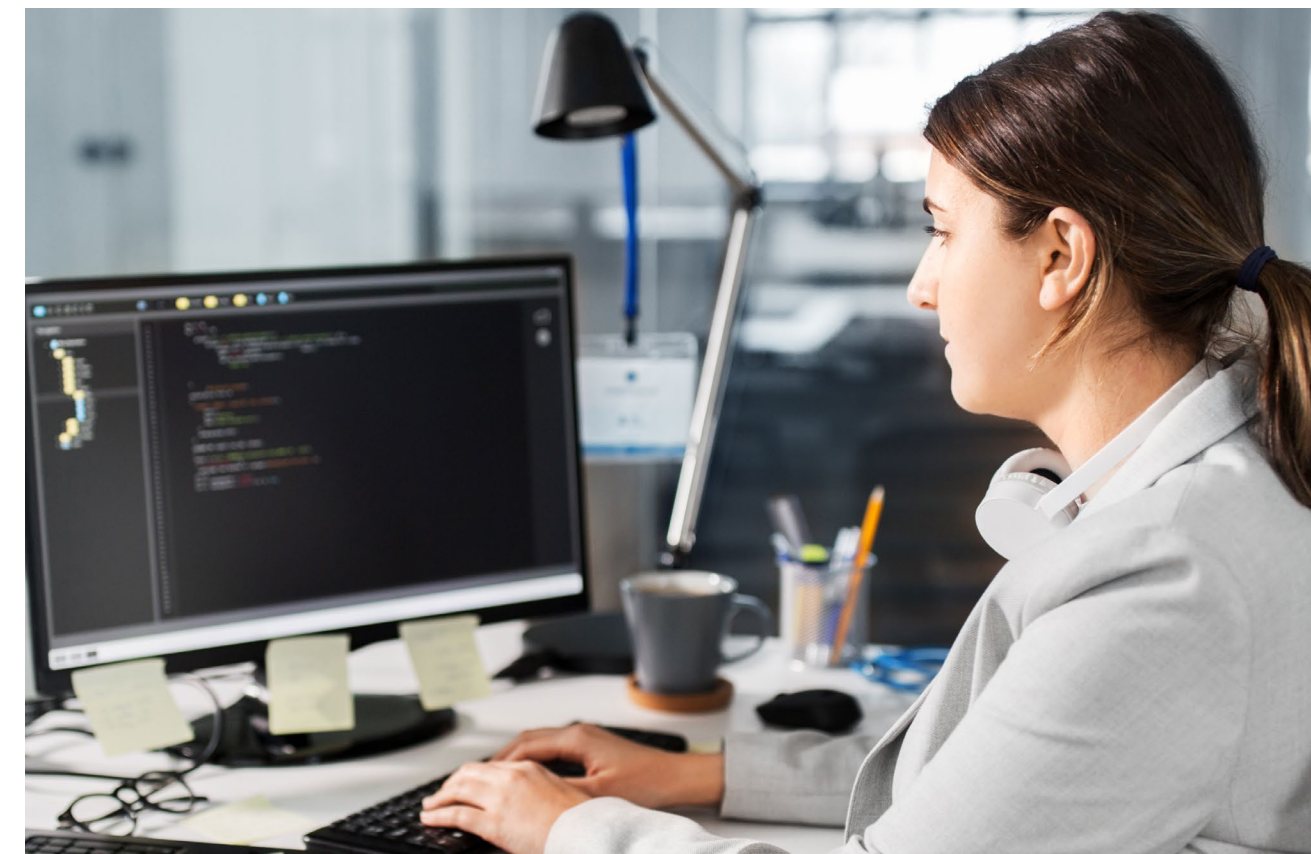


The single source of truth

Building a single source of truth to enable business insights is conventional wisdom and a well-accepted goal for most data strategies. After all, if there isn't a single source of truth then we risk duplicating effort, increasing costs and fuelling a general lack of trust in data and insights.

The traditional federated model where data and analytics teams aim to build a single source of truth, and business teams analyse this data, is being increasingly challenged as the available technology matures. This is caused by a few factors.

- 1. Firstly, the number of technologies that make up any enterprise architecture these days is increasing exponentially.** There causes a never-ending barrage of data ingestion needing to be done. Central teams tasked with building this are never really entirely up-to-date with the ingestion of the data, let alone delivering any insights or leveraging the power of machine learning.
- 2. Secondly, available technology to analyse, visualise and crunch data is getting easier for average citizens to use and replicate.** This results in business teams resorting to their own data manufacturing and refinement processes using a myriad of self-service tools because they can generate insights faster than it would take if they waited for the single source of truth to be made available.



What is data mesh?



Data mesh is an architecture that allows for different data domains – think of these as business units – to directly connect to the various data sources that an organisation has.



New Zealand businesses are increasingly looking to upgrade their on-premises systems to cloud-based platforms to allow for more flexible data architecture.

One potential modern state for data architecture that has become a popular goal for many businesses is the concept of 'data mesh', a decentralised architecture first described by Zhamak Dehghani, a principal consultant at ThoughtWorks. Here, we explore the challenges and assess the potential of a data mesh. These data domains are responsible for the creation and maintenance of data products (outputs that can be consumed by an external on internal audience).

The data domains across a business are designed to have the capability to utilise data to provision data products that are shaped by specialisations of the domain. For instance, the customer experience team may focus their development efforts on the CX data they are connected to, creating data products for use within their team or across the wider organisation. The data products may be prepared with the intention to provide insights or interactivity, or to transform and curate the data in a way that that will make it consumable for downstream data products.



What are the benefits?

This bypasses the requirement for data to be heavily processed by a centralised data team prior to it being analysed. Additionally, with training throughout the organisation in provisioning data products, there is less need for business intelligence requests to be sent to the centralised data team. This reduces the time to insight and increases the specialised team knowledge used in the creation of data products.

Data mesh is set up to be a flexible, innovative architecture, and as technologies are purpose-built to help facilitate a data mesh framework, the architecture will be easier to implement within organisations. For businesses that are not yet far enough along the maturity curve to fully embrace data mesh, steps can still be taken towards preparing for a future state data mesh. These steps could include implementing organisational data governance or performing a cloud-based migration in preparation for future technology.

How might data mesh work in practice?

Case study: data mesh governance in a Microsoft Azure environment

Azure Purview is an overarching data governance product that allows for data discovery and the classification of data in a data mesh environment. A central data integration team is given the responsibility of performing data quality checks and completing the registration of metadata for data that enters an organisation. The data integration team is also given a bird's eye view of where the data is stored, who has access to what data, and where the data is moved. This makes sure that the data used for data products is ubiquitous across the organisation, allowing for smoother interactivity of data products down the line. Limited or no transformations are applied to the data at this stage as data domains are given the opportunity to create products from untouched data sources.

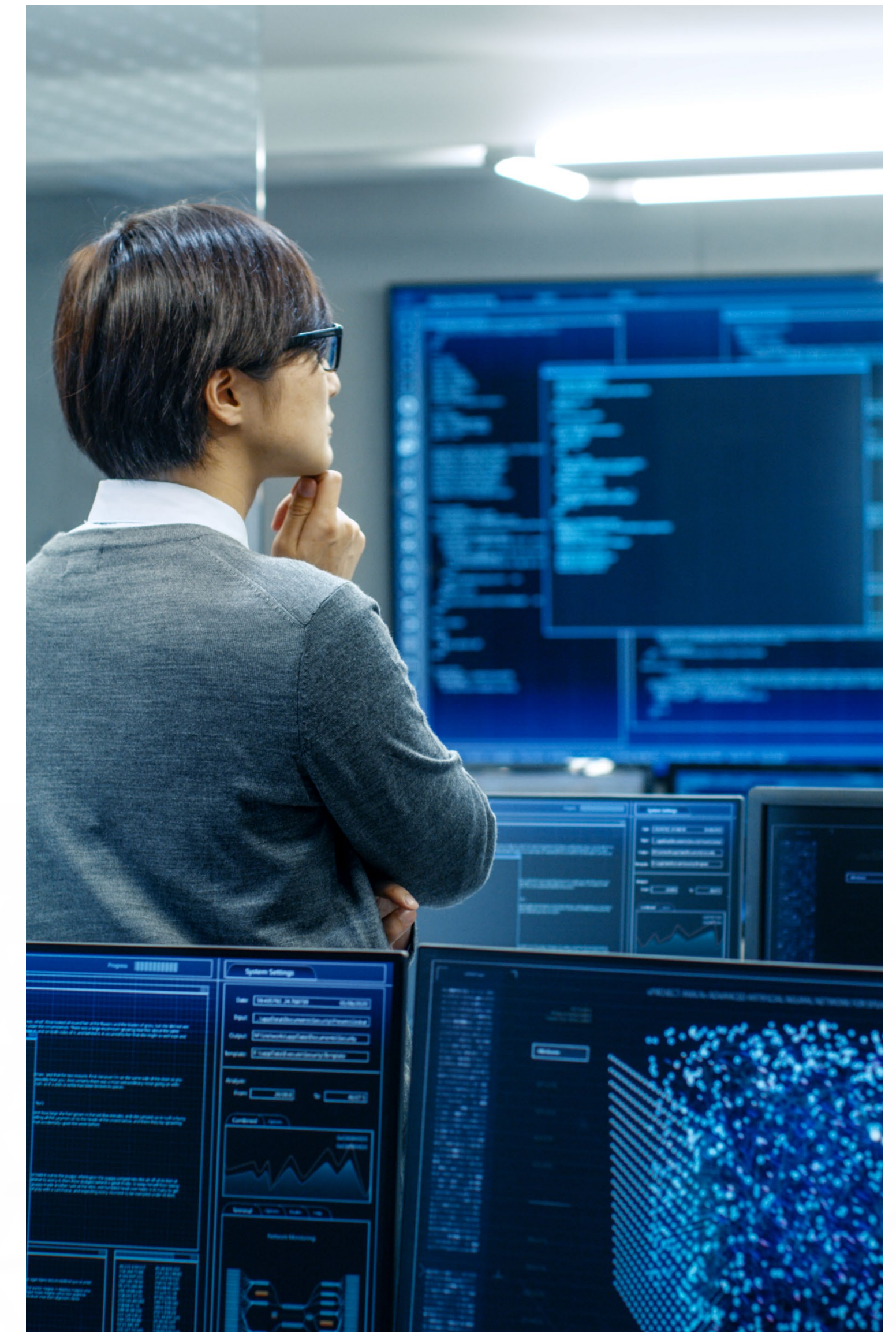
While Azure Purview allows for overall data governance, Azure AD and defined resource groups allow for the segmentation of privileges within domains. Resource groups are logical partitions when considering each data domain. The individuals who make up the data domain can be given individual or security group based permissions on the resource group based on their required level of access. The Microsoft [Zero Trust Security](#) model is recommended, allocating privileges on the principle of least privilege. As organisations scale up in size or requirements, data domains may be partitioned further through the use of differing Azure Subscriptions, with the individual data products within the domain requiring individual resource groups.

What are the governance risks to be aware of?

Strong data governance is required in organisations implementing data mesh as there are key risks that result from poorly established data governance or a lack of visibility across the organisation.

1. Data duplication – Without clear visibility from domain to domain of data products across a business, there runs the risk of multiple similar products being created. This creates a drain on resources, increased cost of operations, and wasted effort. Therefore, a centralised planning system, where product purposes, outputs and timelines are recorded should be implemented.
2. Dependencies – Data mesh shifts the responsibility of product upkeep from a centralised data team on to teams across the various data domains. The responsibilities for upkeep and monitoring that reside within the data domains and products must be clearly defined as the outputs of the data product, if not standalone BI products, may be used in downstream data products. Without the correct upkeep, this could result in a waterfall effect of poor data product quality spilling across an organisation.

Data mesh provides the framework for a heavily decentralised data architecture system with emphasis on speed, specialisation, and volume of insight. In organisations where the creation of products is held to a centralised data team, delays can result from a lack of specialised knowledge, resource constraints, and data processing time. The partitioning of an organisation into data domains allows for the simultaneous production of multiple valuable data products within an organisation. The effective implementation of data mesh requires excellent planning and data governance policy development. Thorough preparation for the exciting architecture of data mesh will lead organisations to a fast-paced, data-centric future.





Stephen Hastings

Partner
KPMG Lighthouse
Auckland

E: stephenhastings@kpmg.co.nz



Alistair Evans

Director
KPMG Lighthouse
Auckland

E: aevans@kpmg.co.nz



Daniel Glidden

Associate Director
KPMG Lighthouse
Auckland

E: dglidden@kpmg.co.nz



Dean Franklet

Senior Advisor
KPMG Lighthouse
Auckland

E: dfranklet@kpmg.co.nz



The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavour to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act upon such information without appropriate professional advice after a thorough examination of the particular situation.

© 2022 KPMG, a New Zealand Partnership and a member firm of the KPMG global organisation of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved