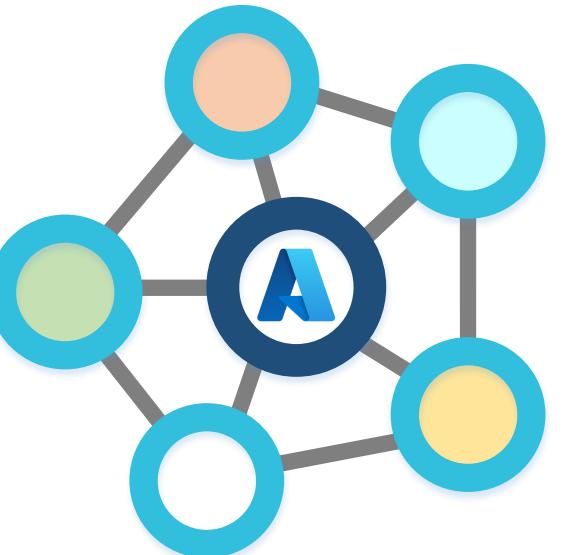
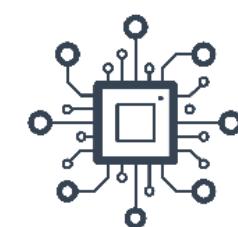


Building a Data Mesh Architecture in Azure

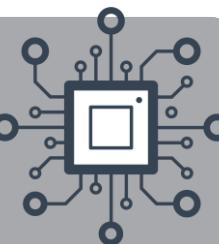
Theory vs Practice



Paul Andrew | Technical Architect in Azure CoE



Mr Paul Andrew
Consulting Ltd



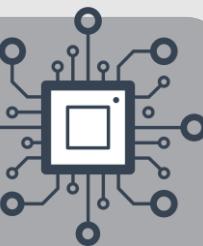
<https://github.com/mrpaulandrew>

CommunityEvents

Demo code, content and slides from various community events.

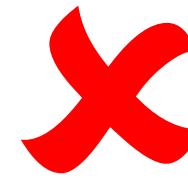


[{Event/Location}-{Month}-{Year}](#)



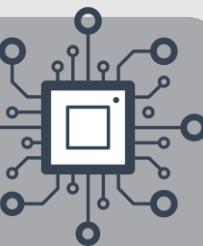
What is the answer to life, the universe and everything?

Answer:
42



Answer:
It depends!





What is big data?

Answer:

It depends!



Answer:

“Any data that you cannot process
in the time that you have/want
using the technology you have.”



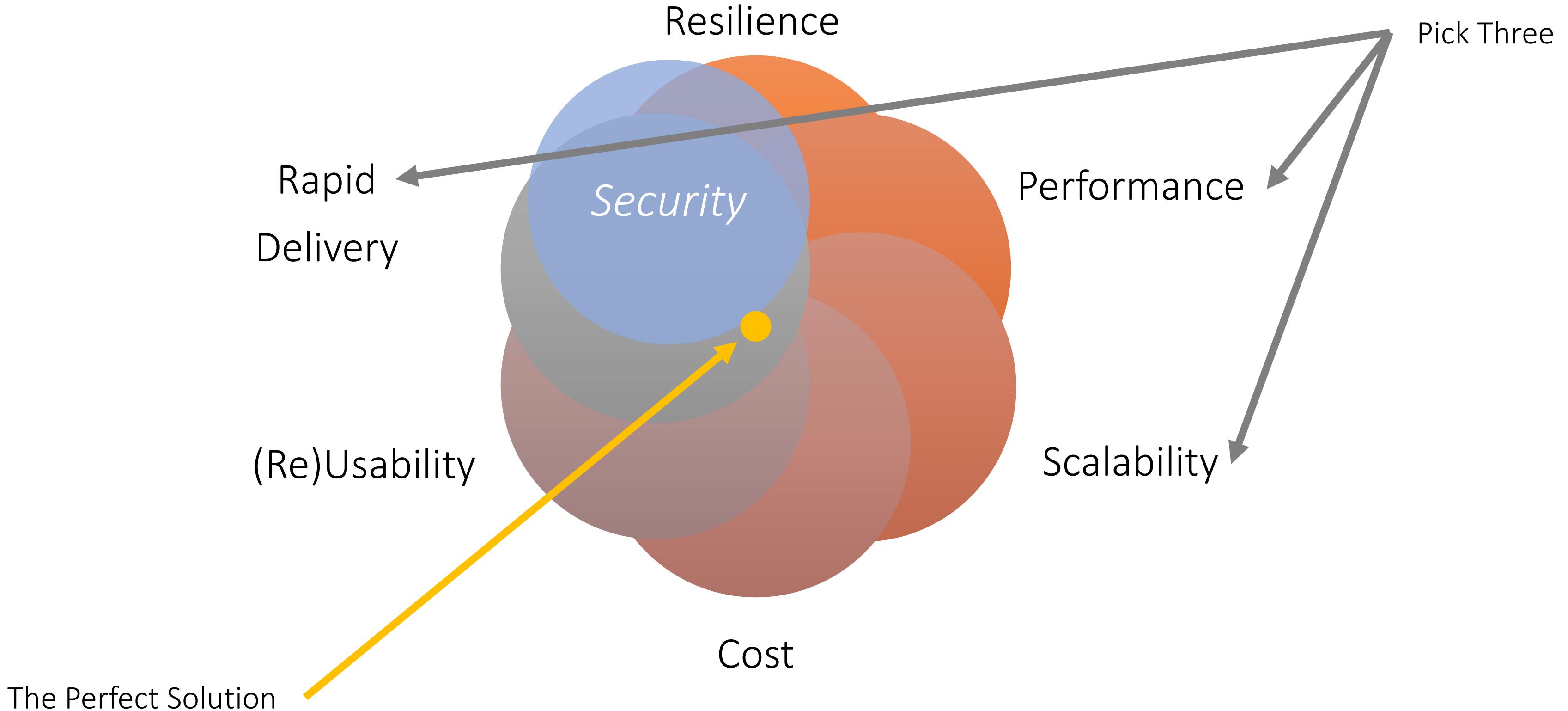
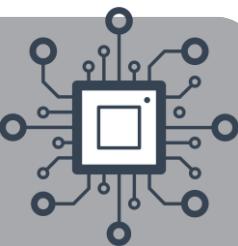
- Buck Woody

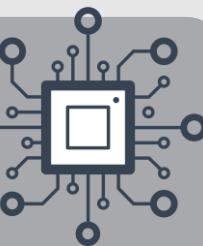
@BuckWoodyMSFT

Volume
Velocity
Variety
Veracity
Value



What is our primary design focus?





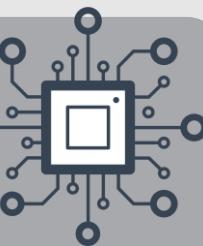
What is the goal of our data solutions?

Data
Collection

Data
Sources

*Paul's Magic Box -
From the Hogwarts School of
Witches & Wizardry*

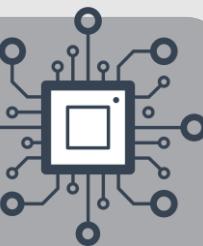
Data
Insight



What is the goal of our data solutions?



Data = Information = Knowledge = Power/Insights

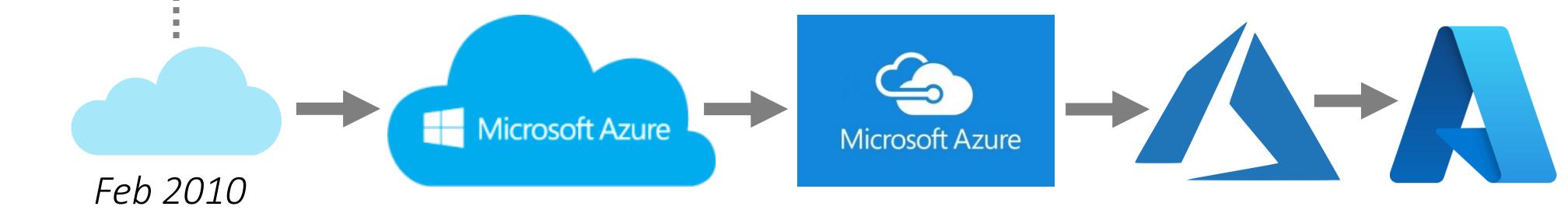
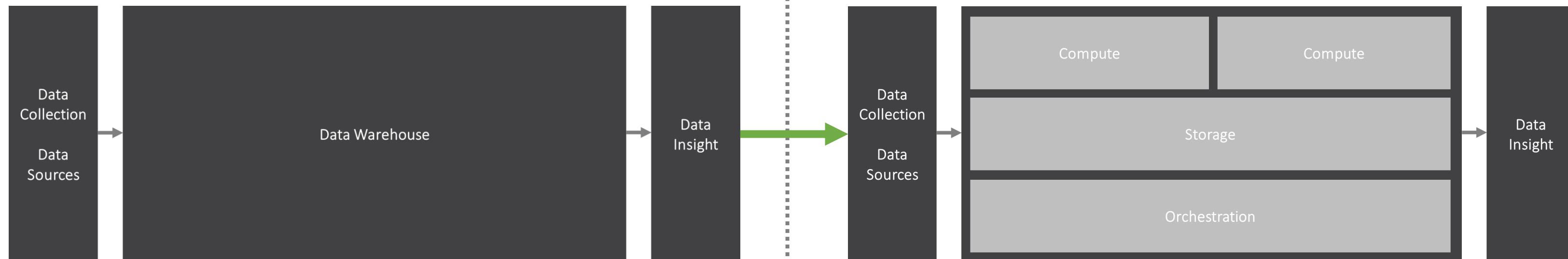
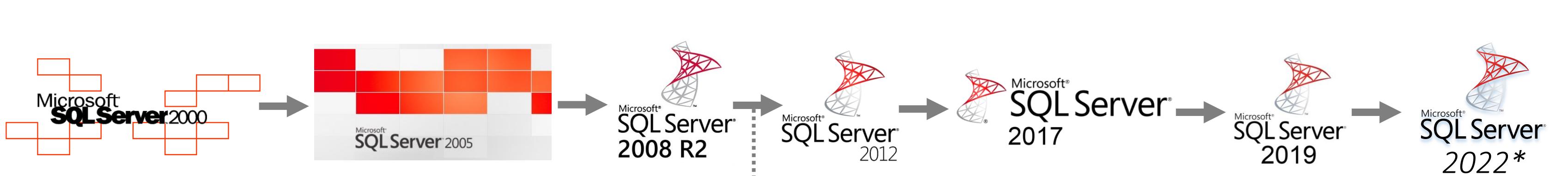
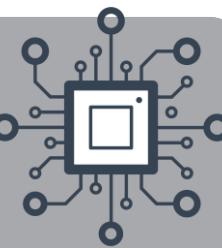


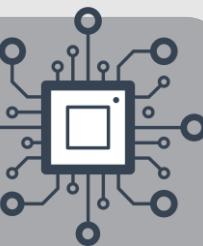
What is the goal of our data solutions?



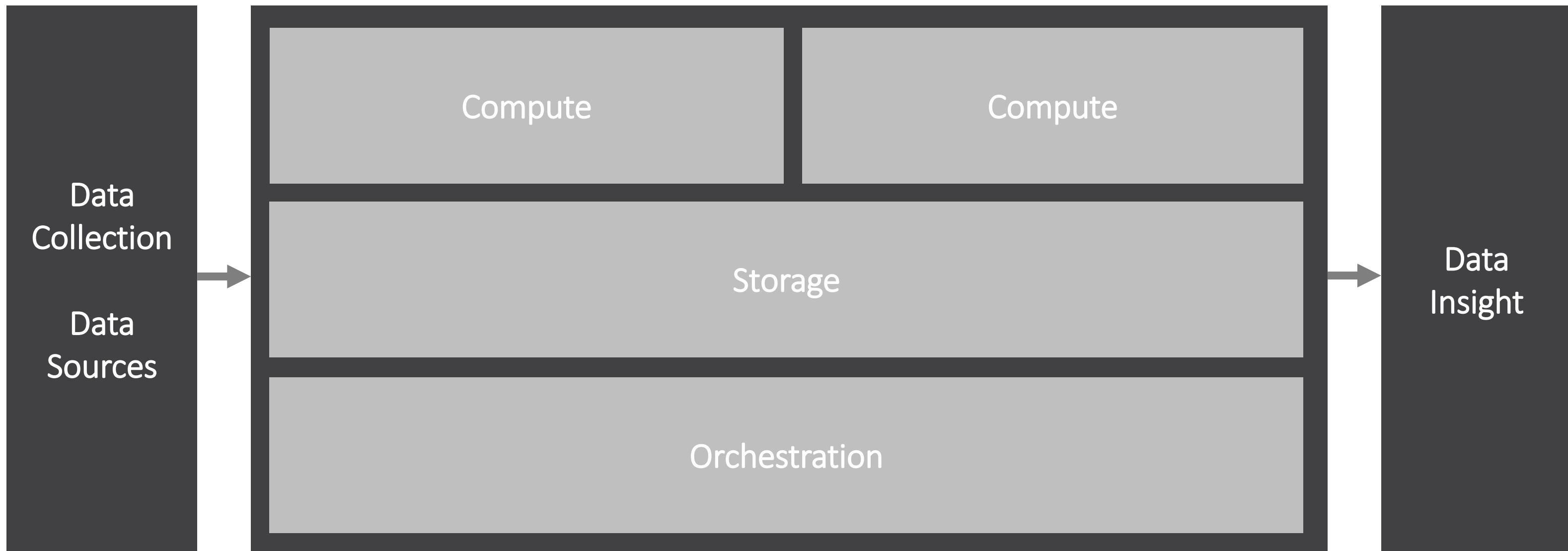
Data = Information = Knowledge = Power/Insights

Data Platform Evolution





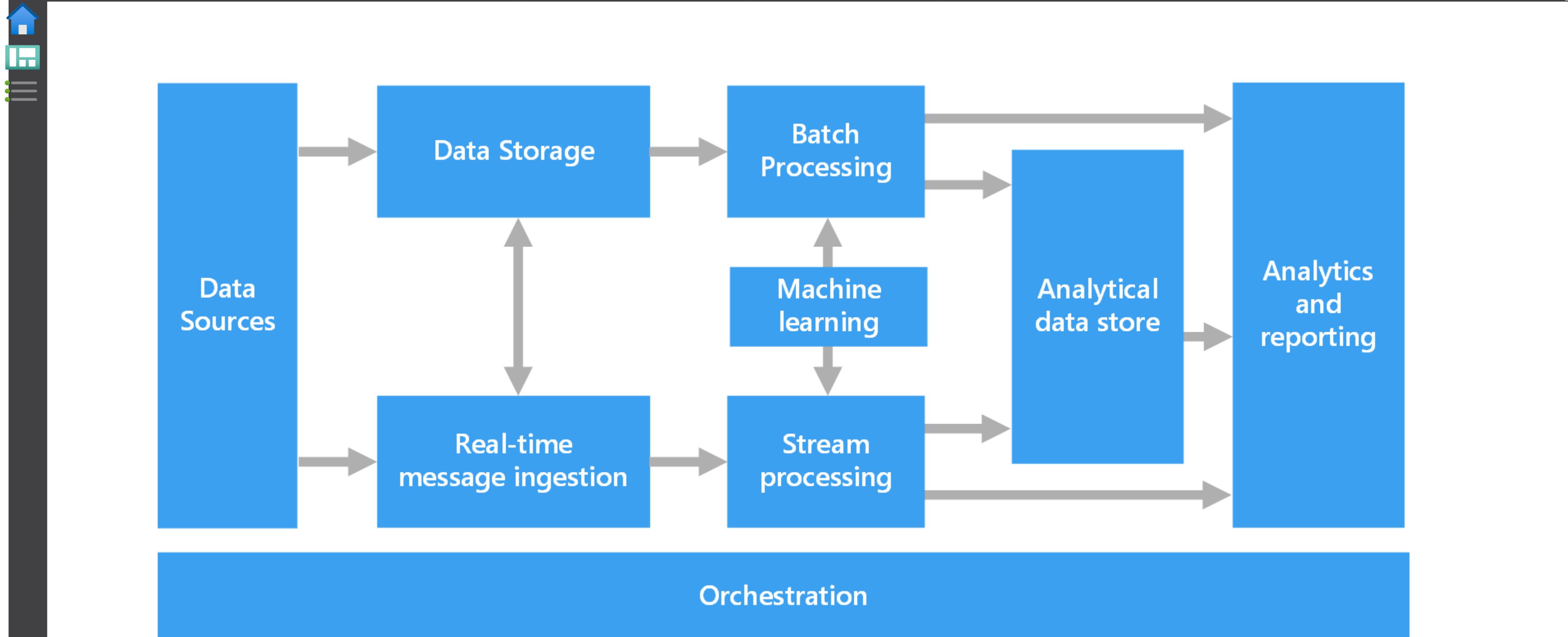
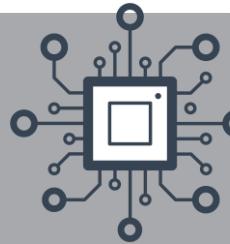
What is the goal of our data solutions?



Data = Information = Knowledge = Power/Insights

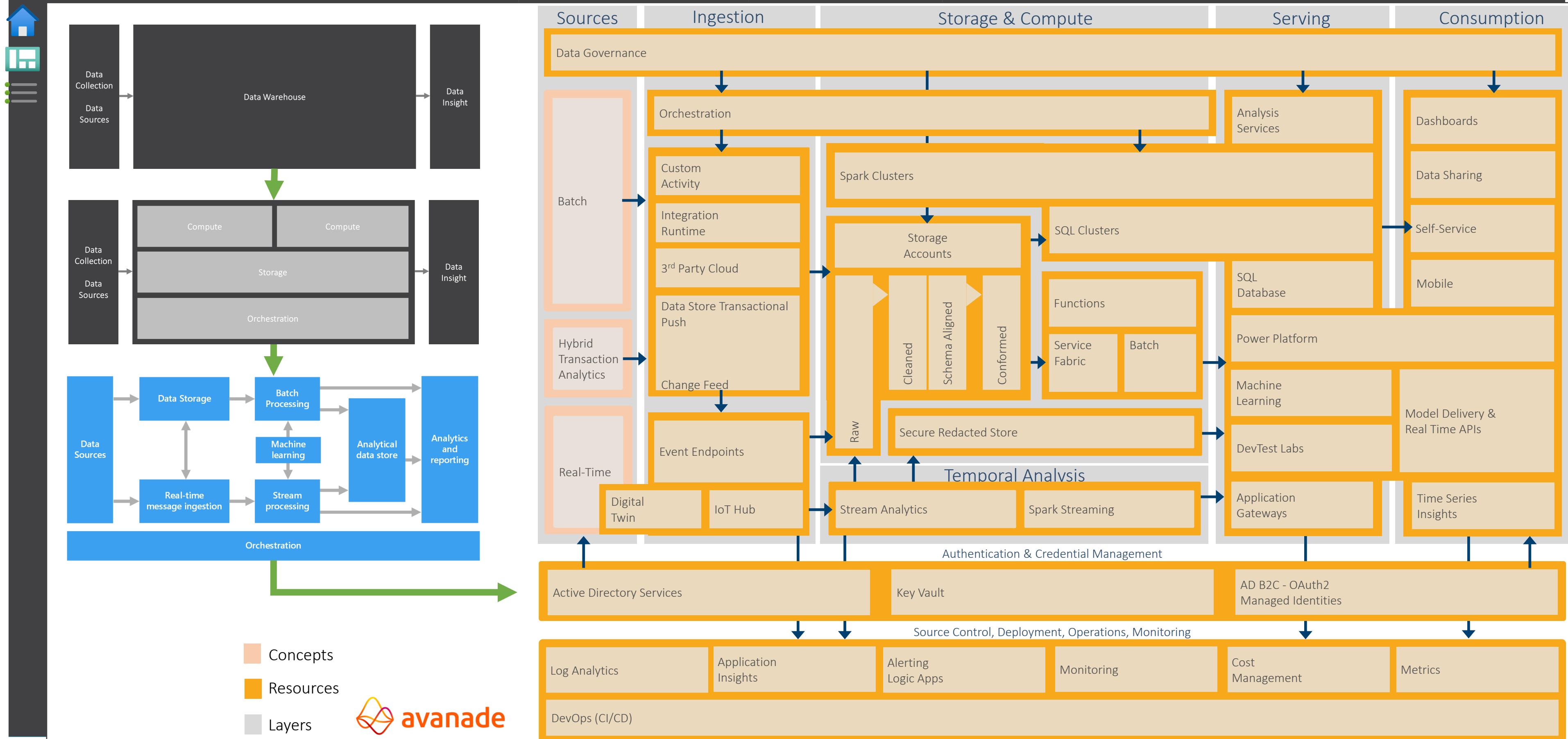
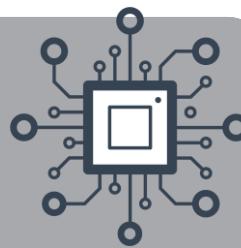


Microsoft's Components of a Big Data Architecture



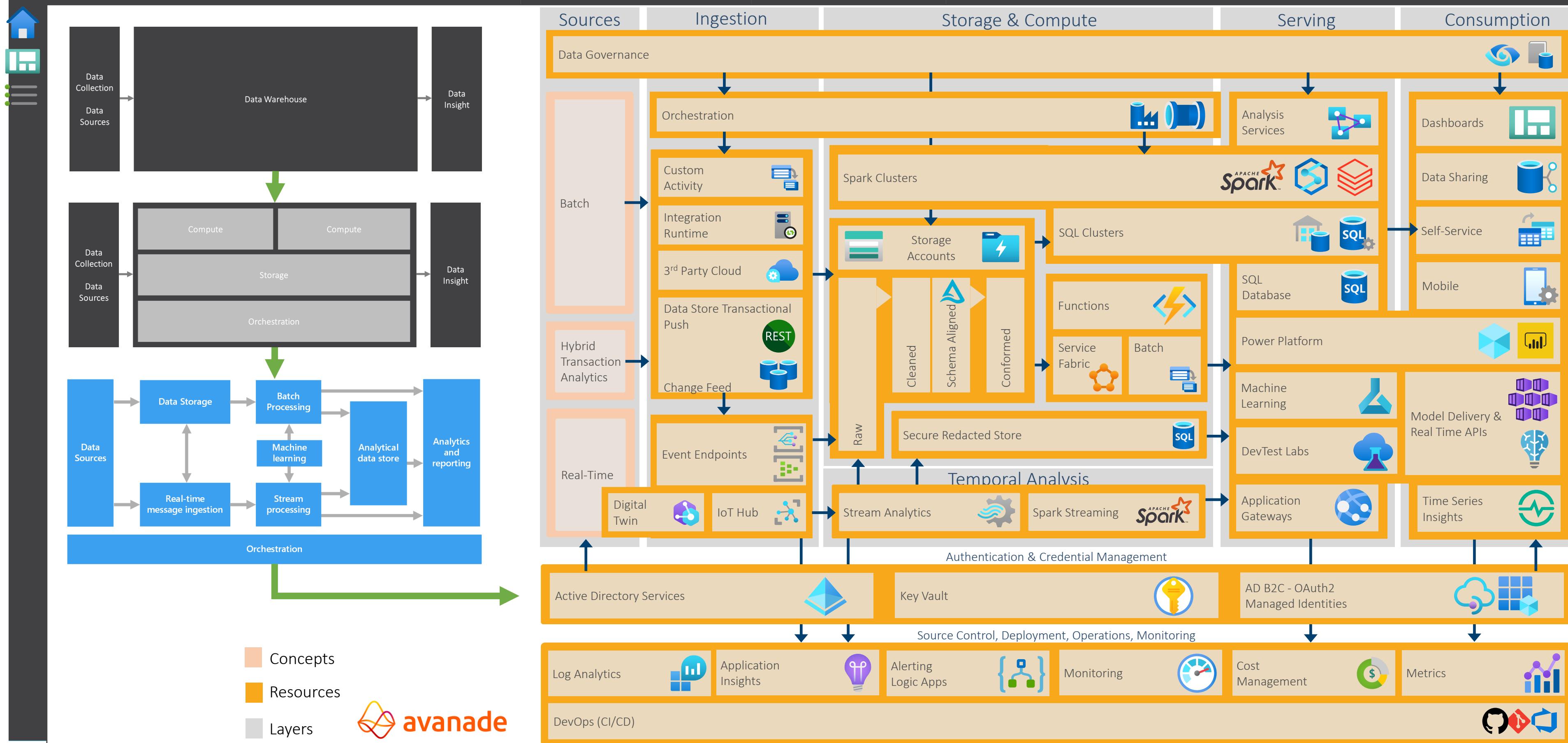
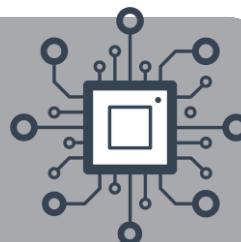


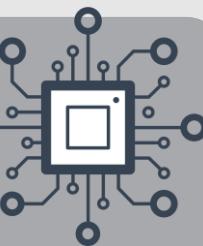
A Logical Data Architecture



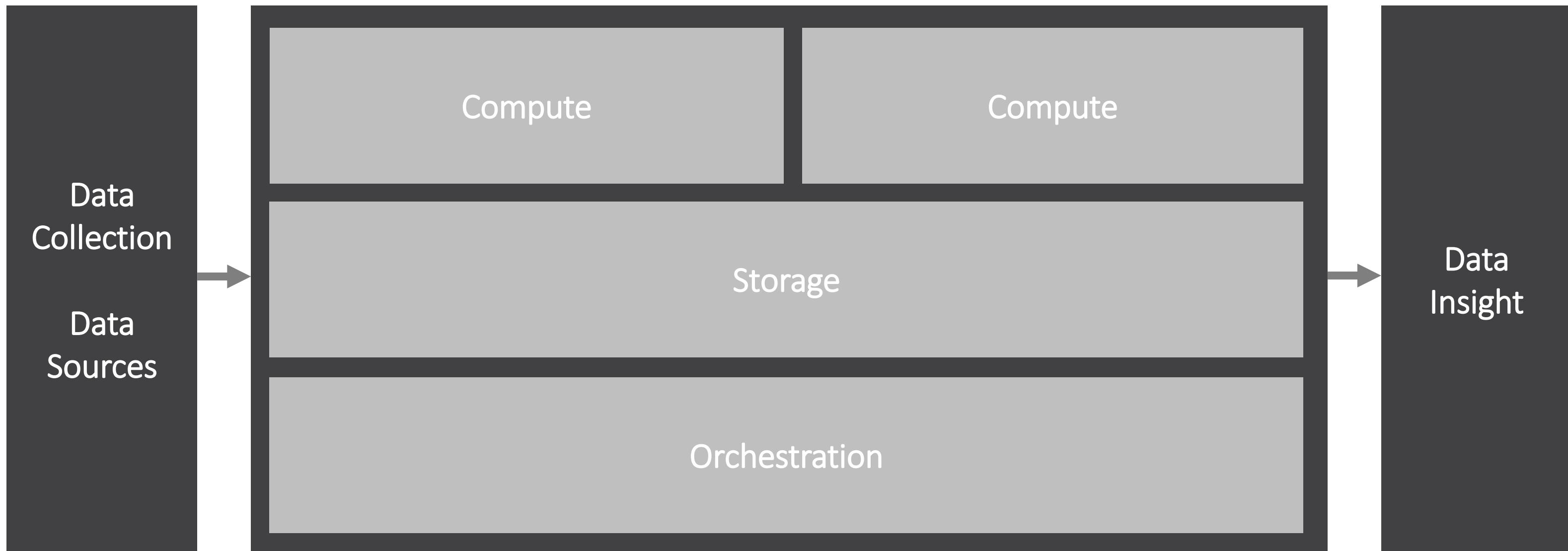


A Logical Data Architecture

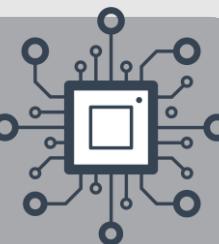




What is the goal of our data solutions?



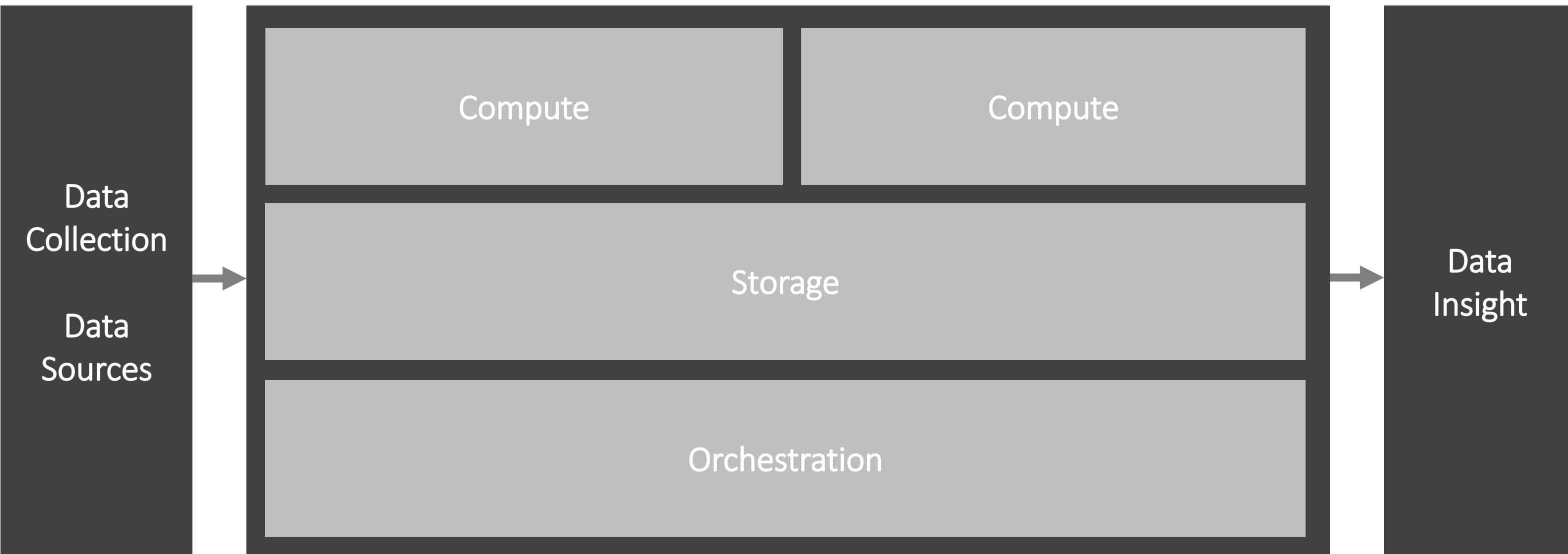
Data = Information = Knowledge = Power/Insights



Extract

Transform

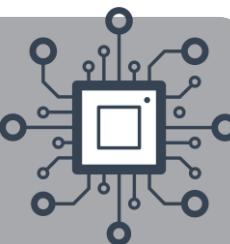
Load



Data = Information = Knowledge = Power/Insights



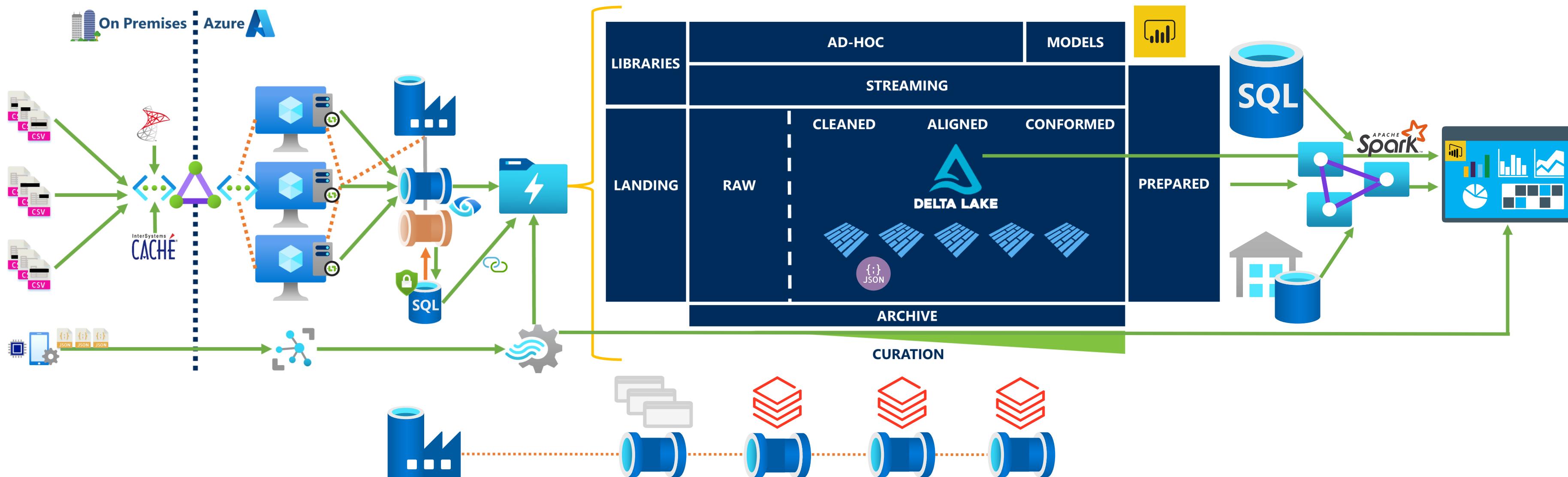
One Data Product



Extract

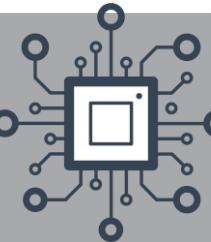
Transform

Load





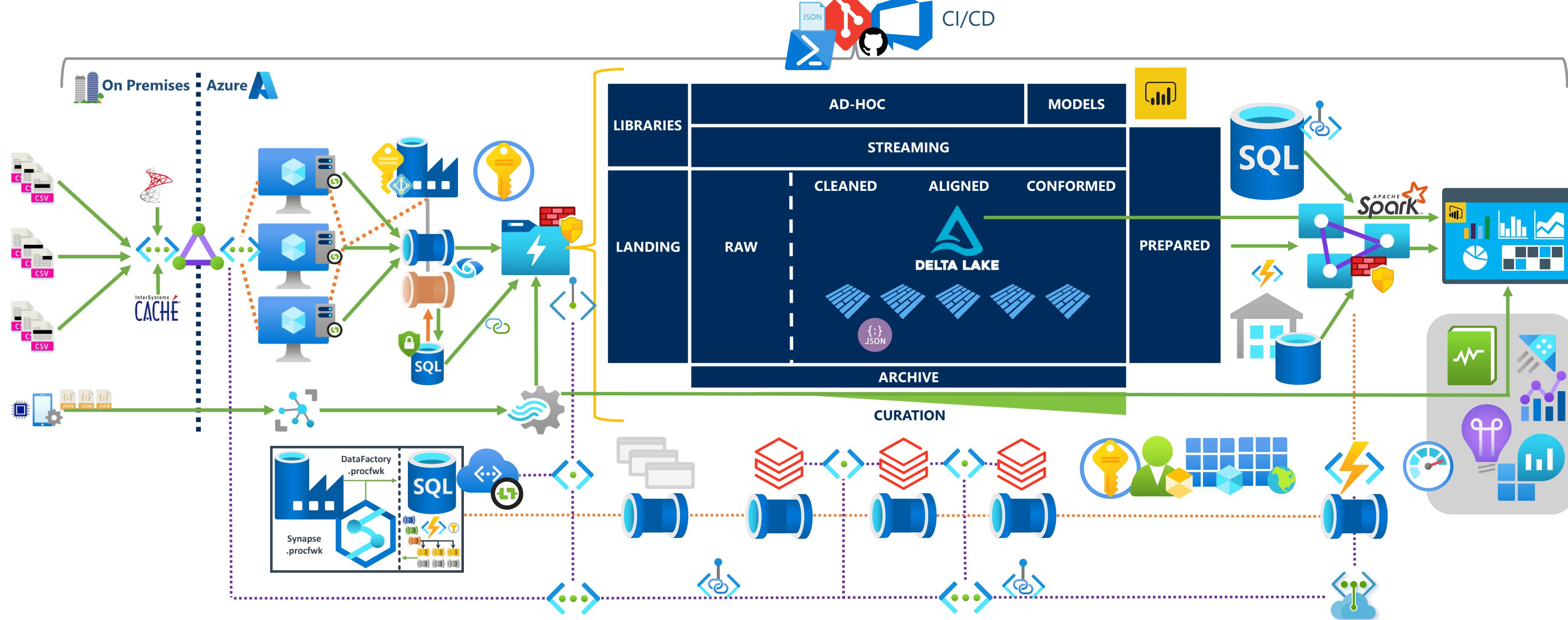
One Data Product



Extract

Transform

Load



Data Mesh

- Zhamak Dehghani

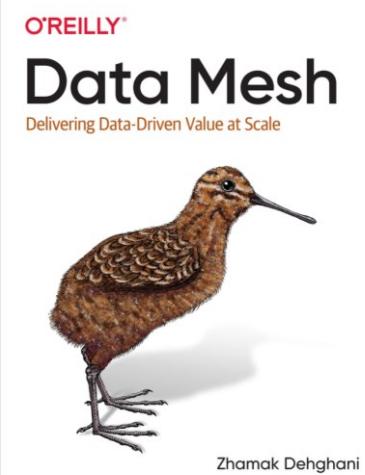
@zhamakd

<https://martinfowler.com/articles/data-mesh-principles.html>



ISBN-10
1492092398

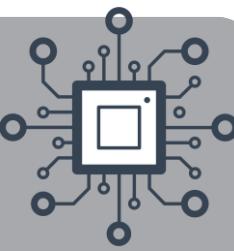
ISBN-13
978-1492092391



1. Domain-oriented decentralised data ownership and architecture.
2. Data as a product.
3. Self-serve data infrastructure as a platform.
4. Federated computational governance.



Theory vs Practice

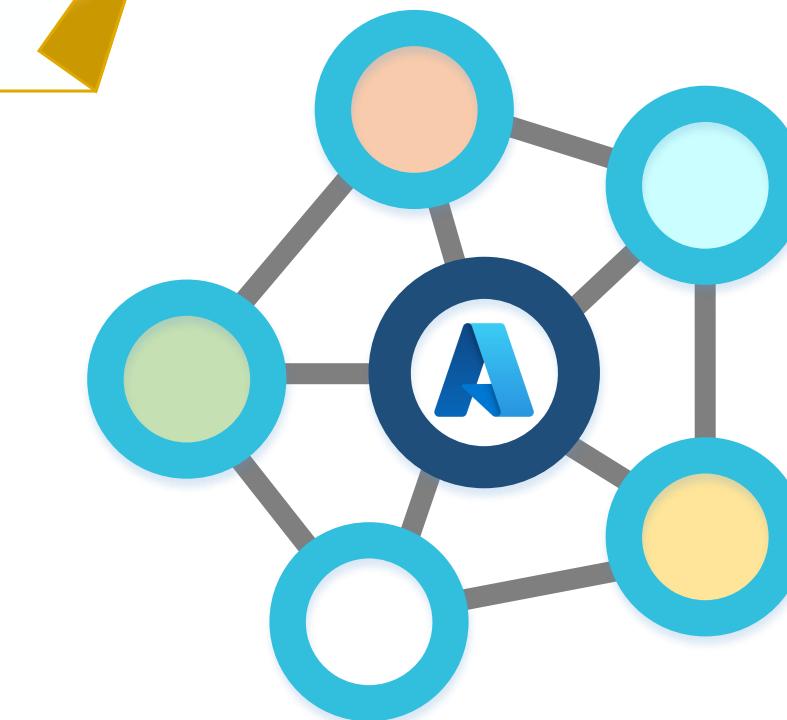


Data as a product.

Federated computational governance.

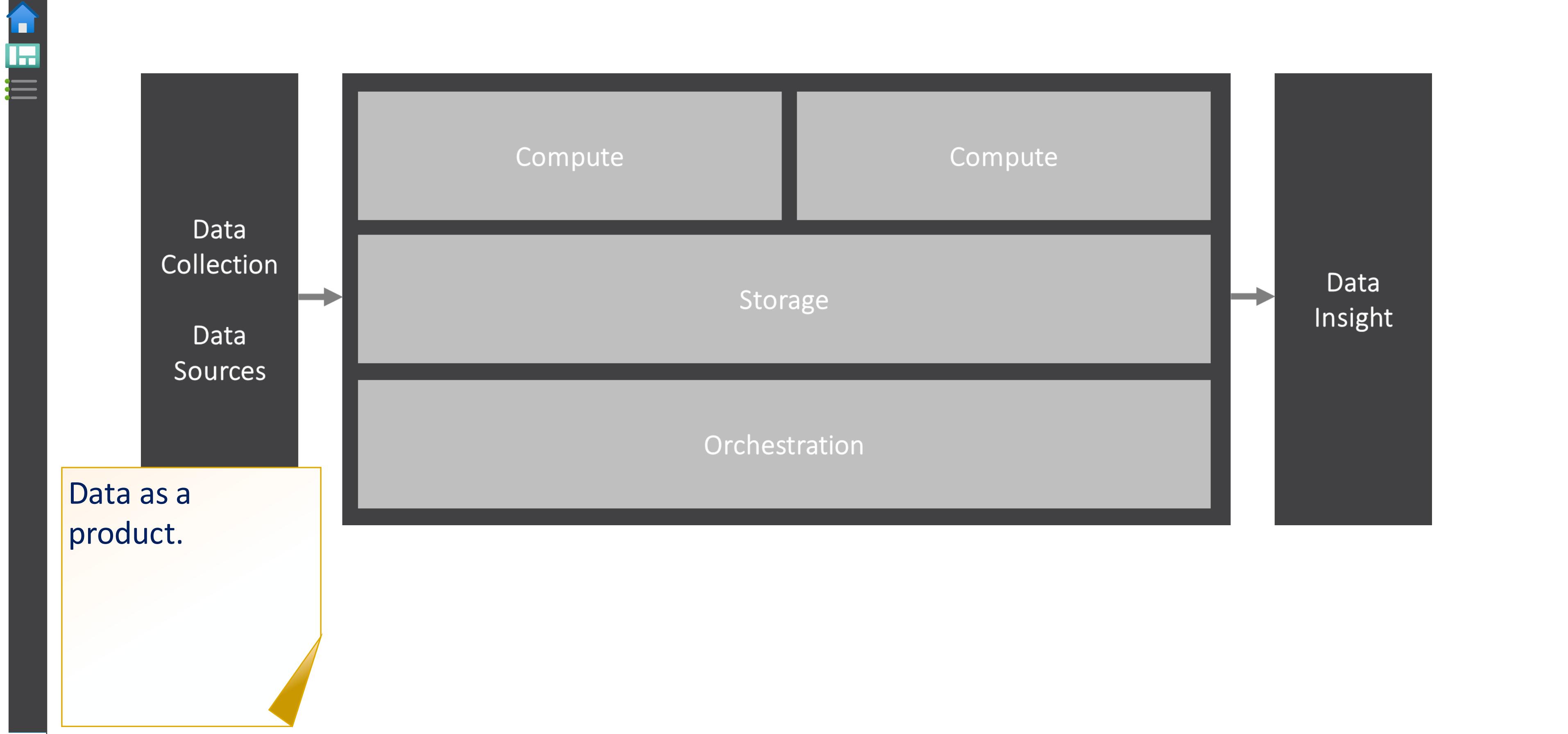
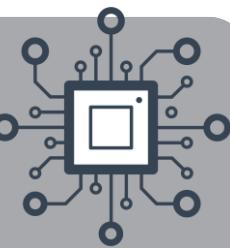
Domain-oriented decentralised data ownership and architecture.

Self-serve data infrastructure as a platform.

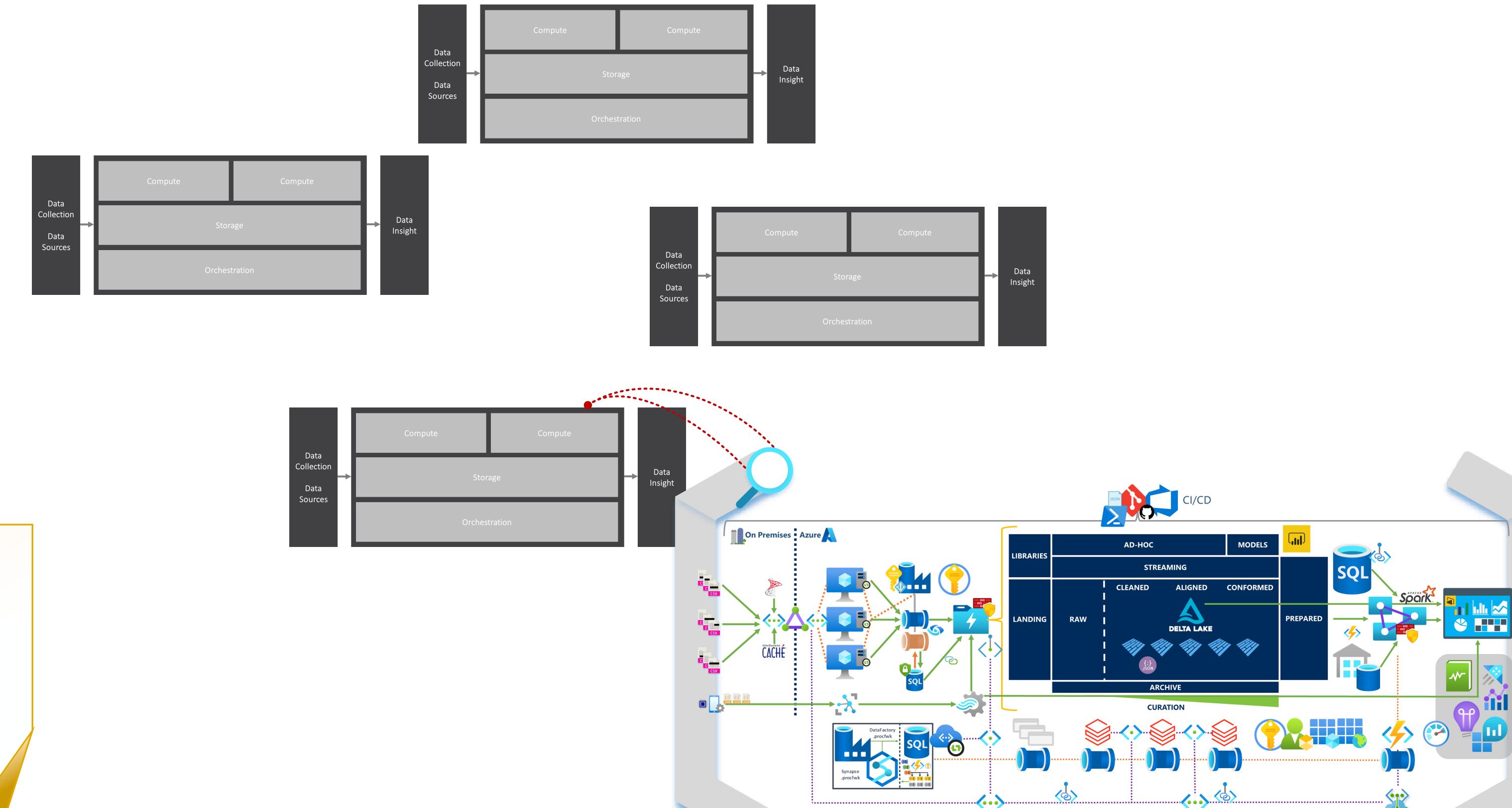
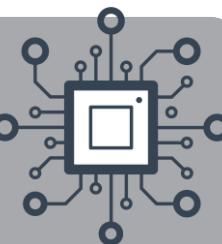




Data Products

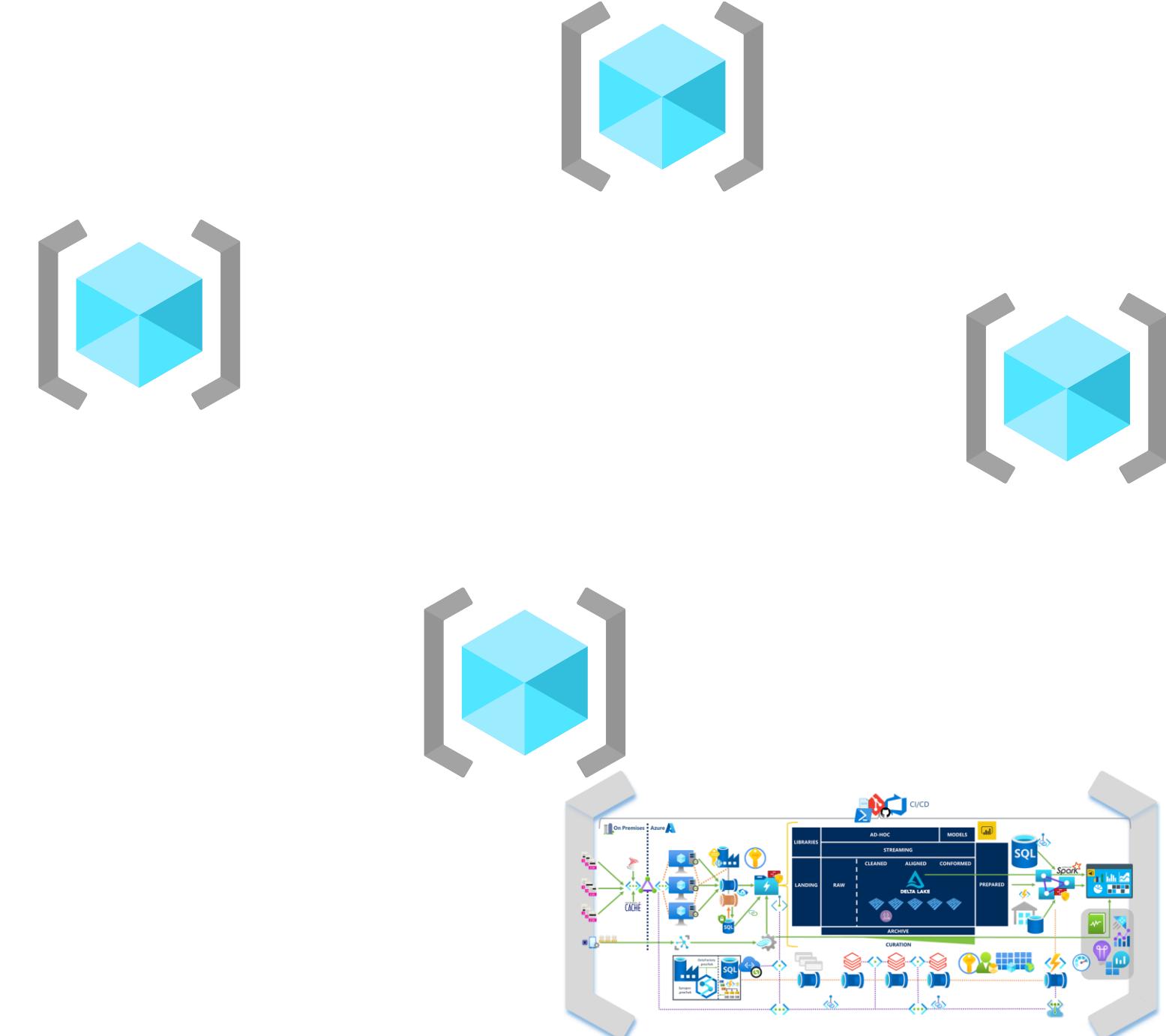
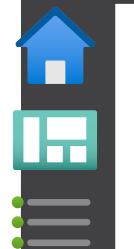
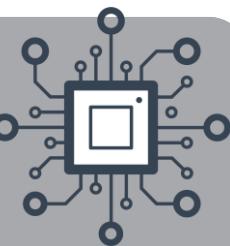


Data Products

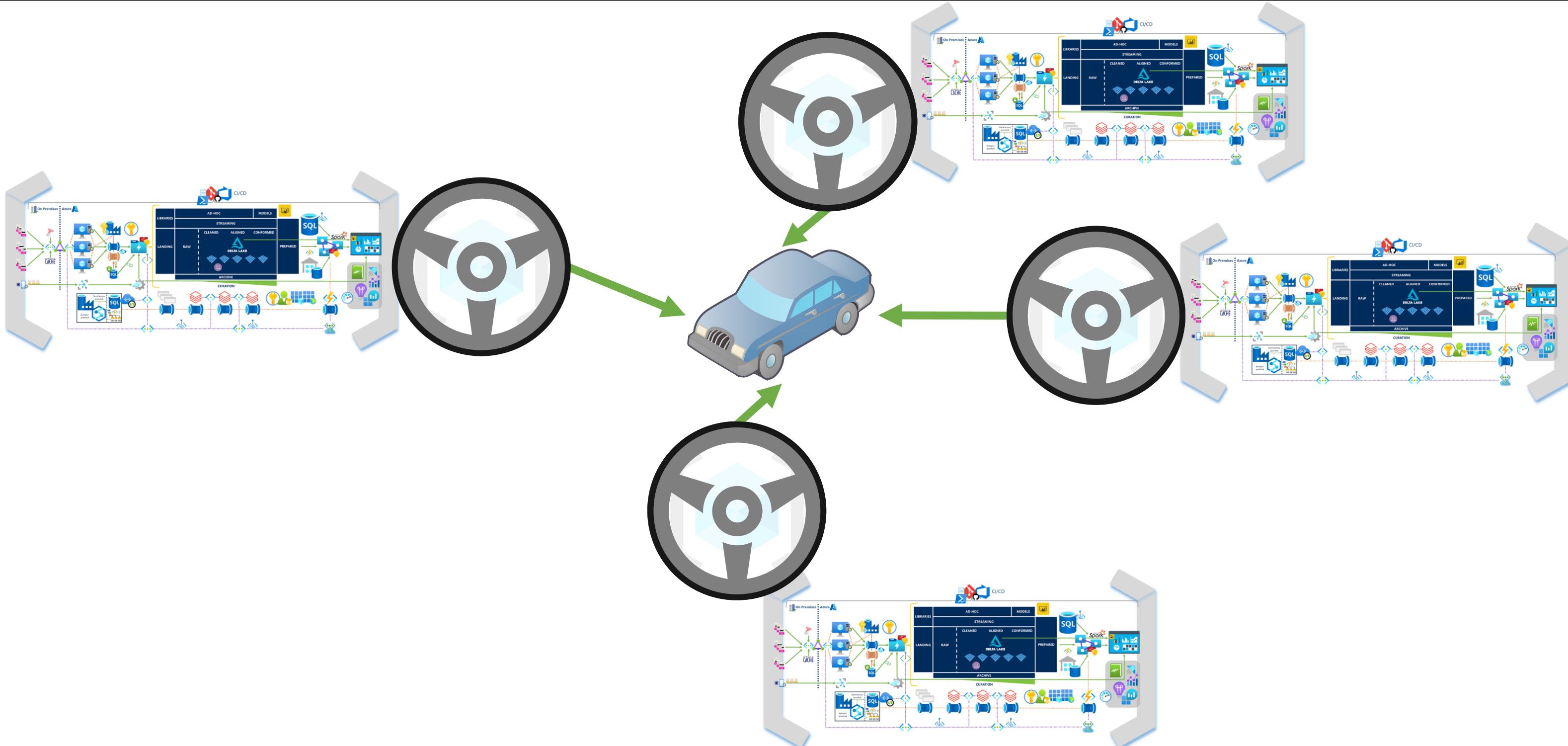
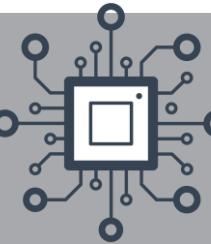




Data Products in Azure



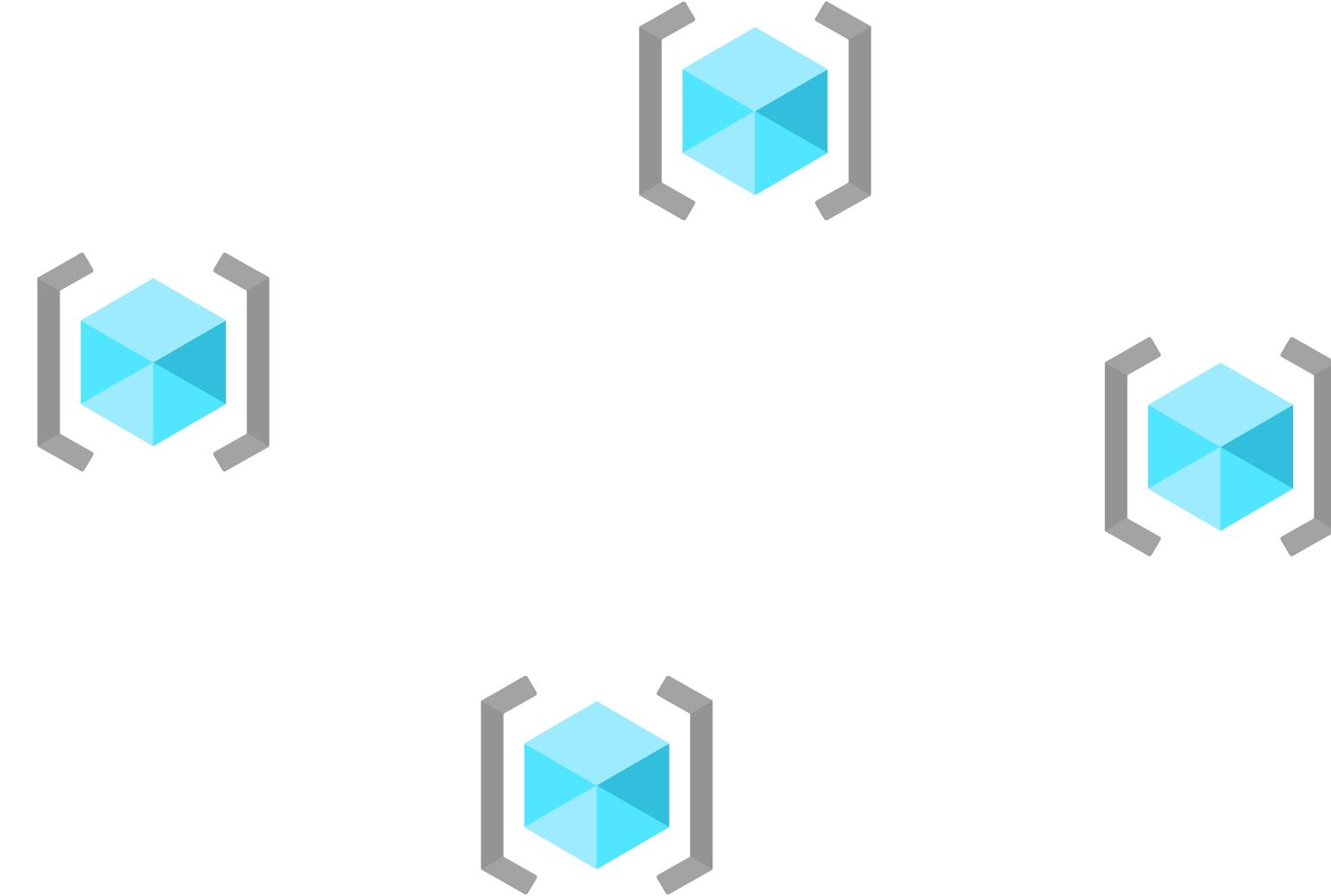
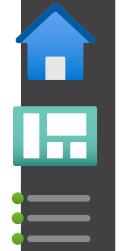
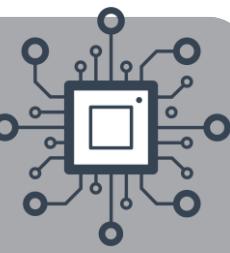
Data Products in Azure



What does our minimum viable Data Mesh contain?



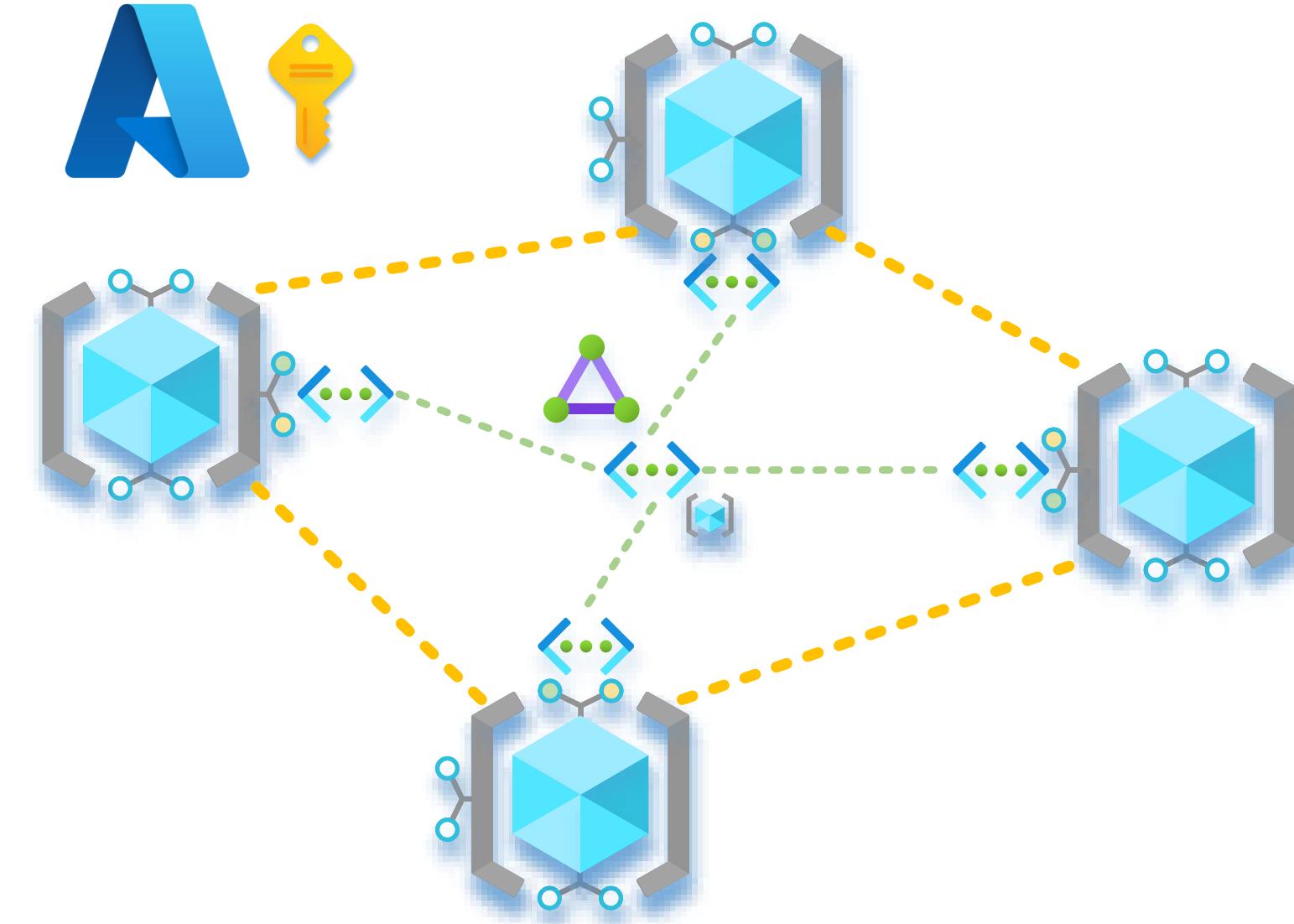
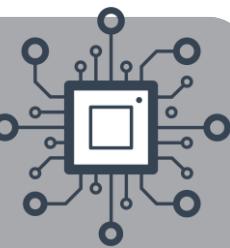
Data Products in Azure



What does our minimum viable Data Mesh contain?

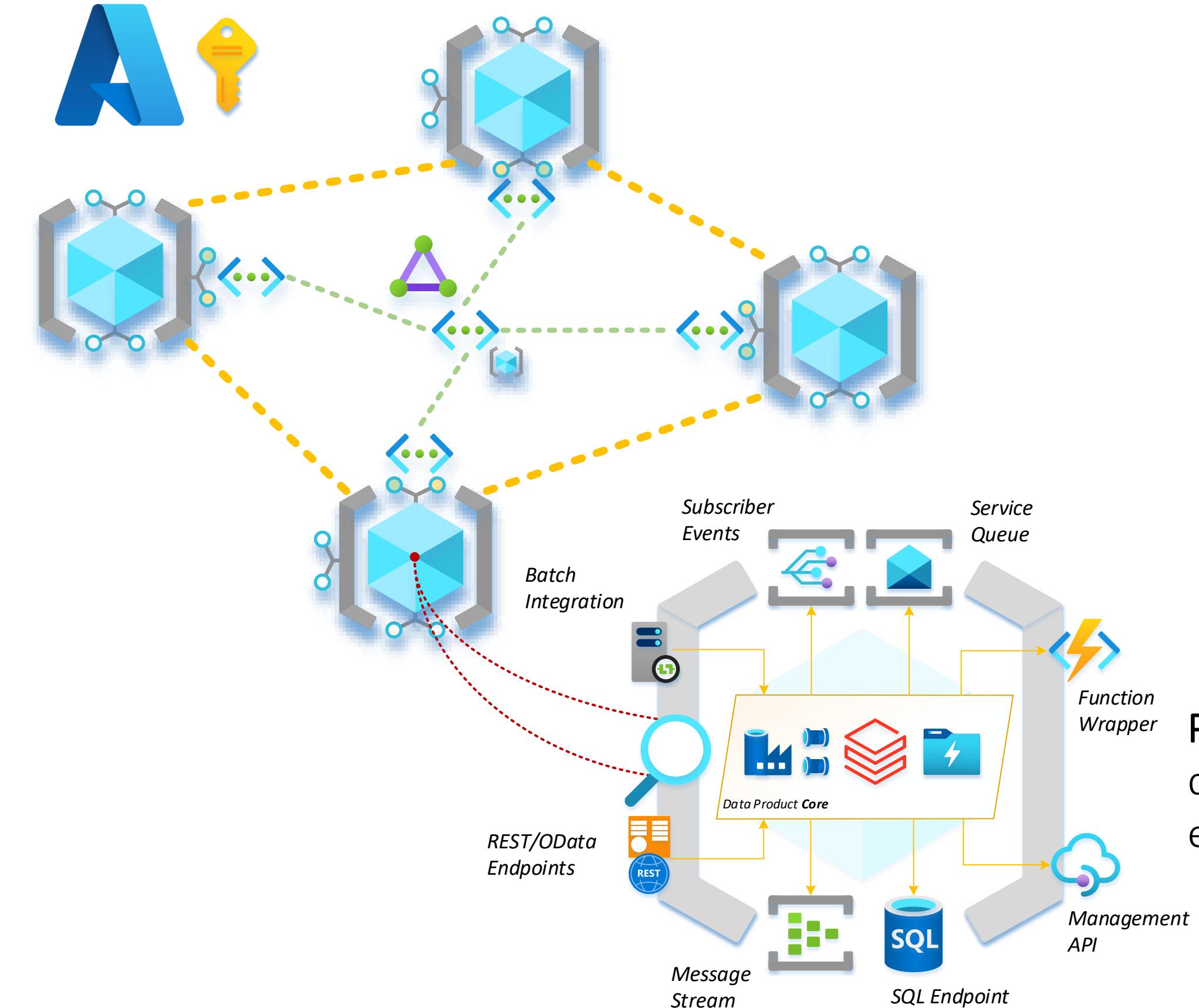
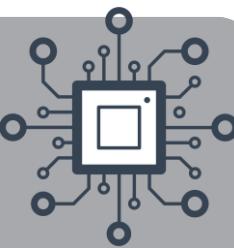


Data Products in Azure with Interfaces



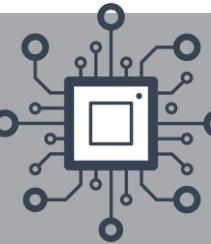
What does our minimum viable Data Mesh contain?

Data Products in Azure with Interfaces



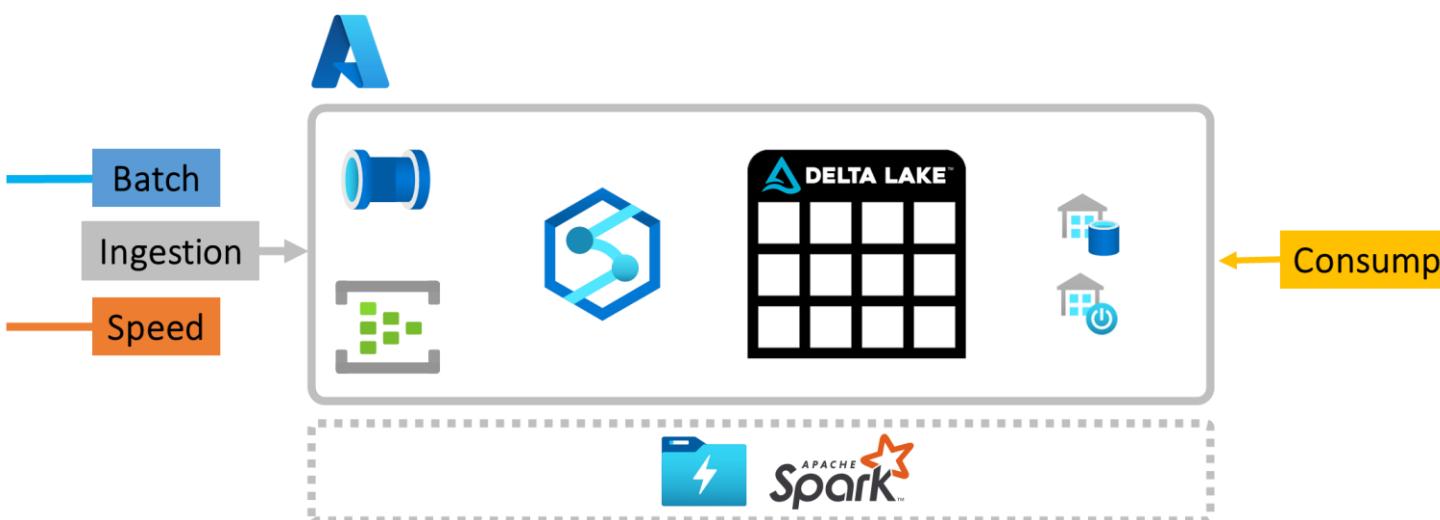
Primary Interfaces –
data integration and
exchange.

Data Products in Azure with Interfaces

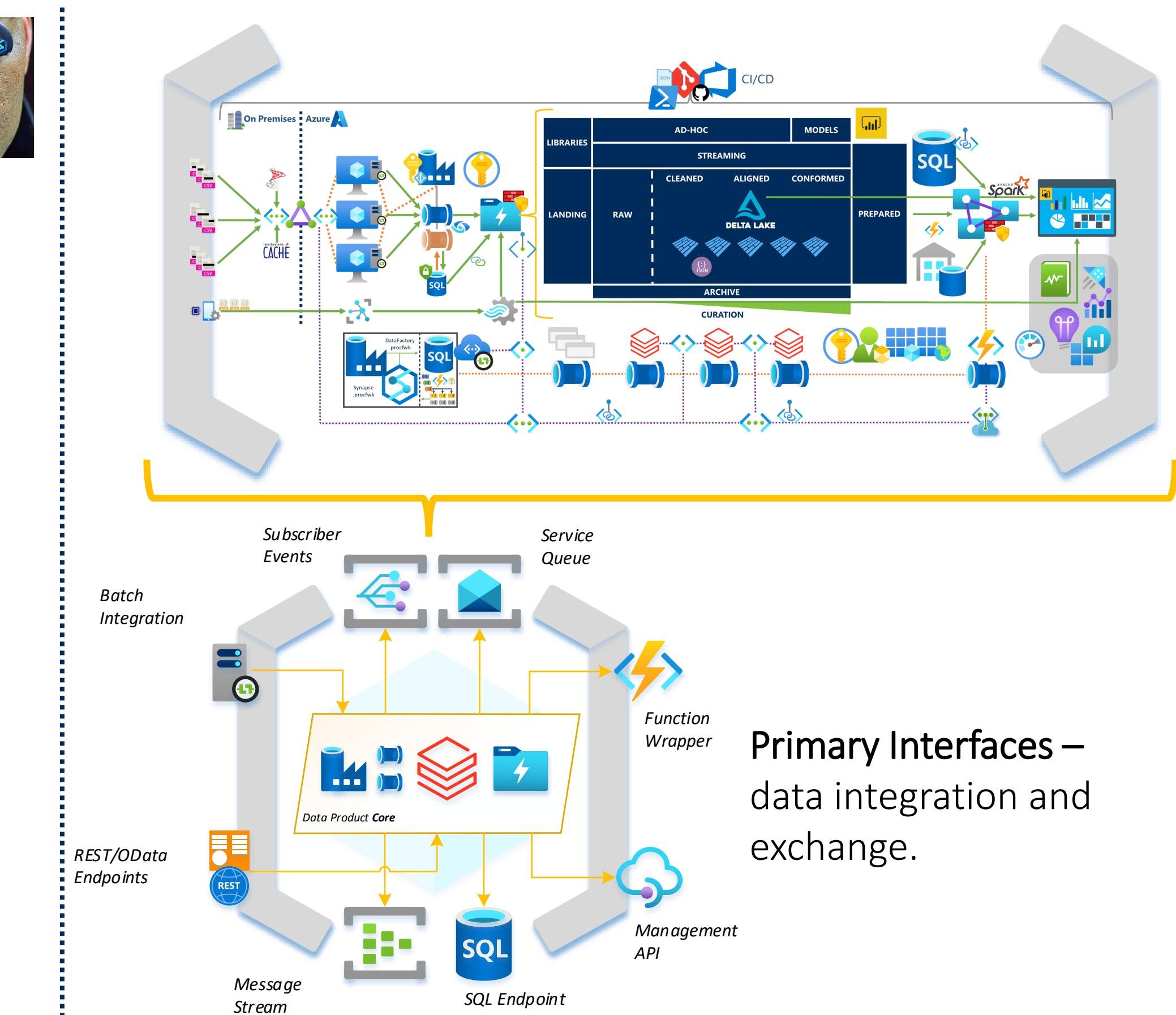
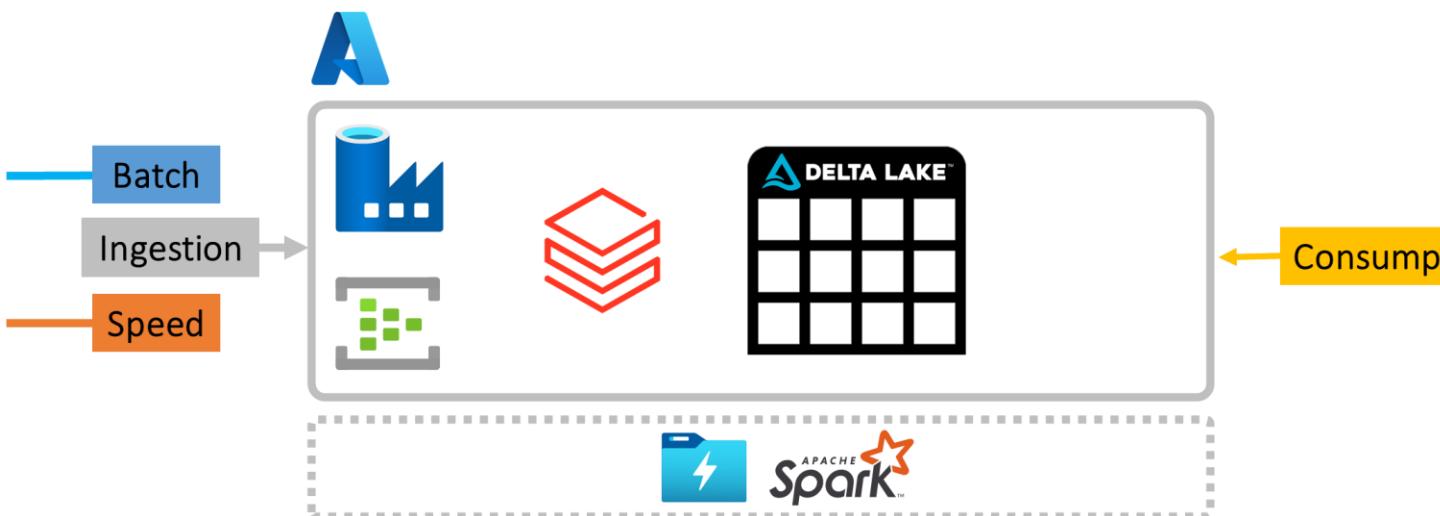


Kappa Data Handling Pattern

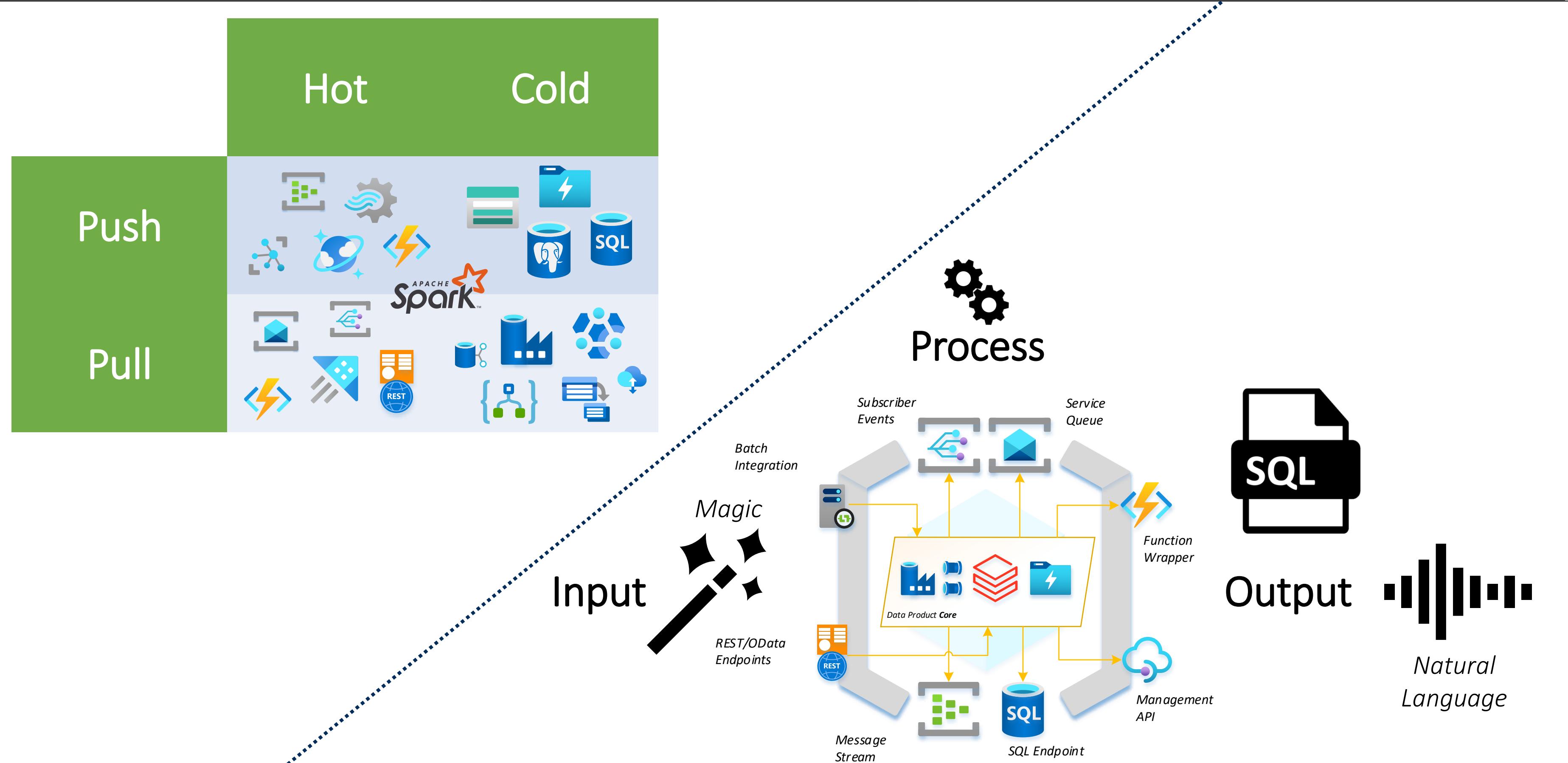
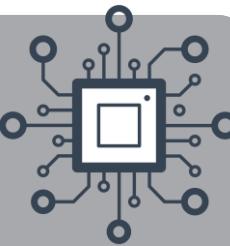
Using Synapse:



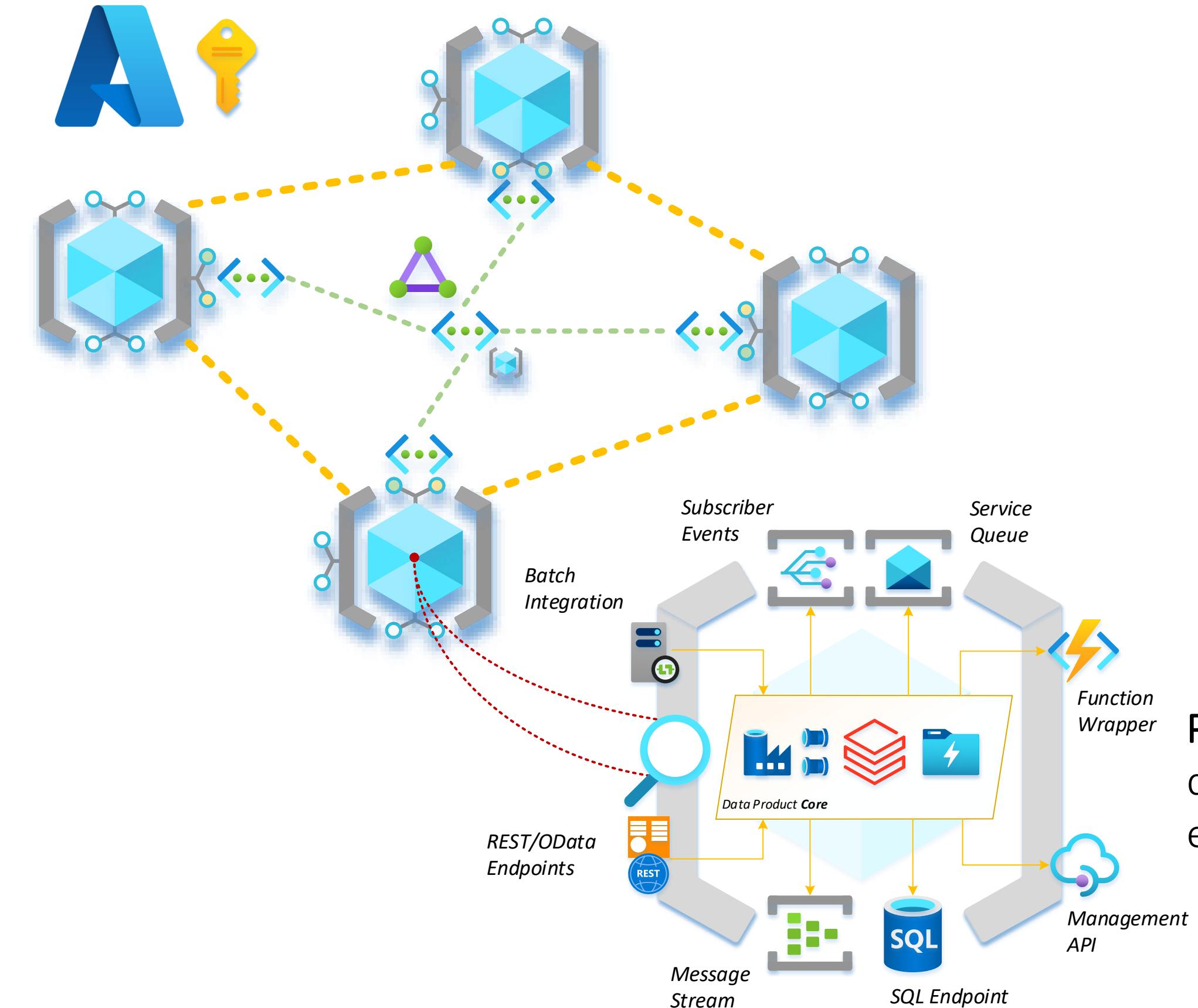
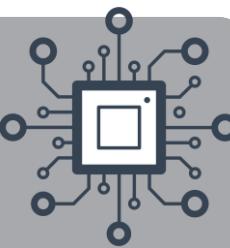
Using Databricks:



Data Products in Azure with Interfaces

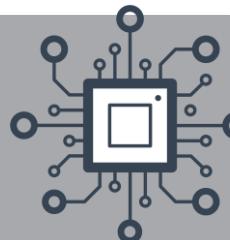


Data Products in Azure with Interfaces

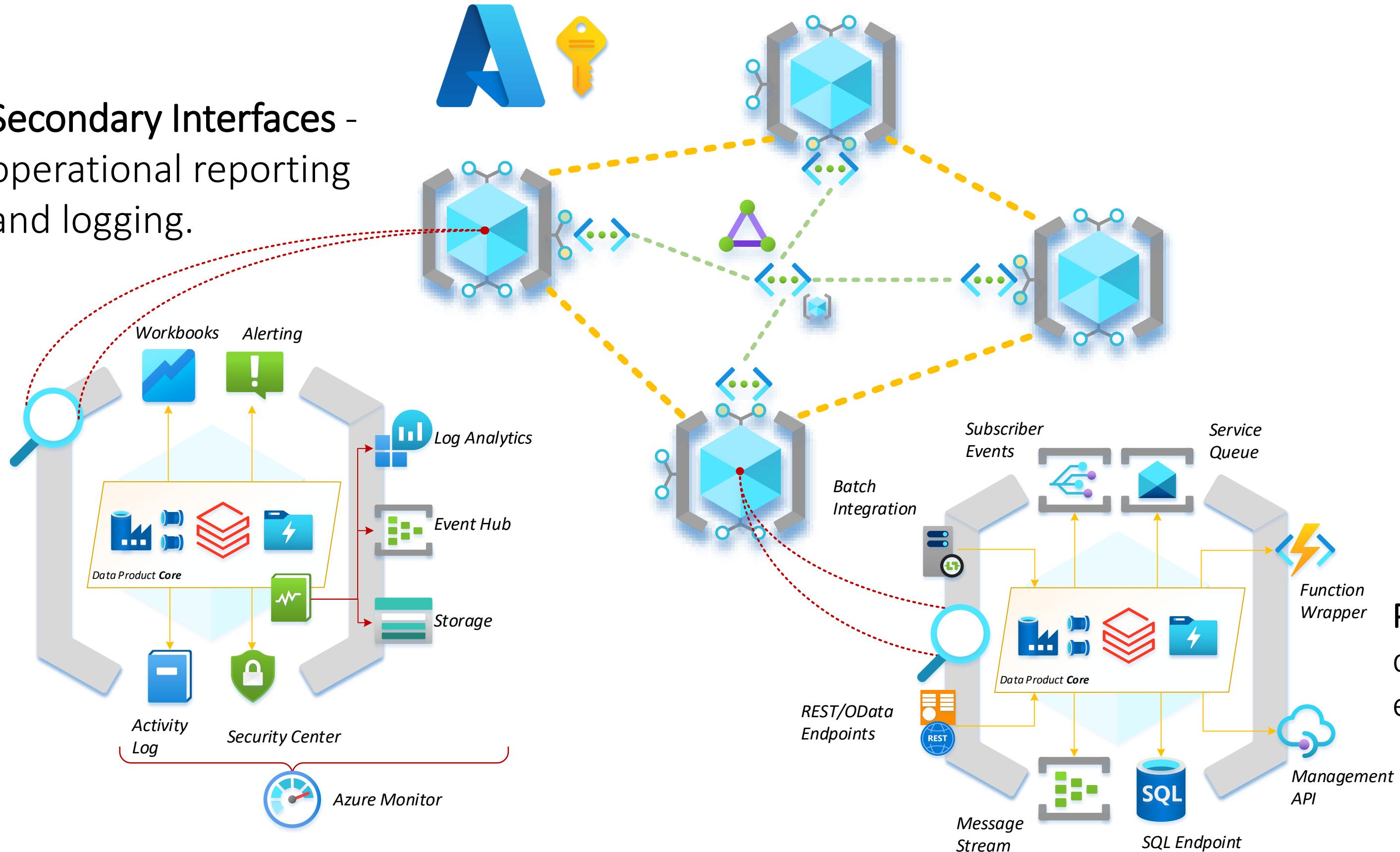




Data Products in Azure with Interfaces

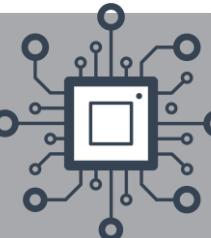


Secondary Interfaces –
operational reporting
and logging.

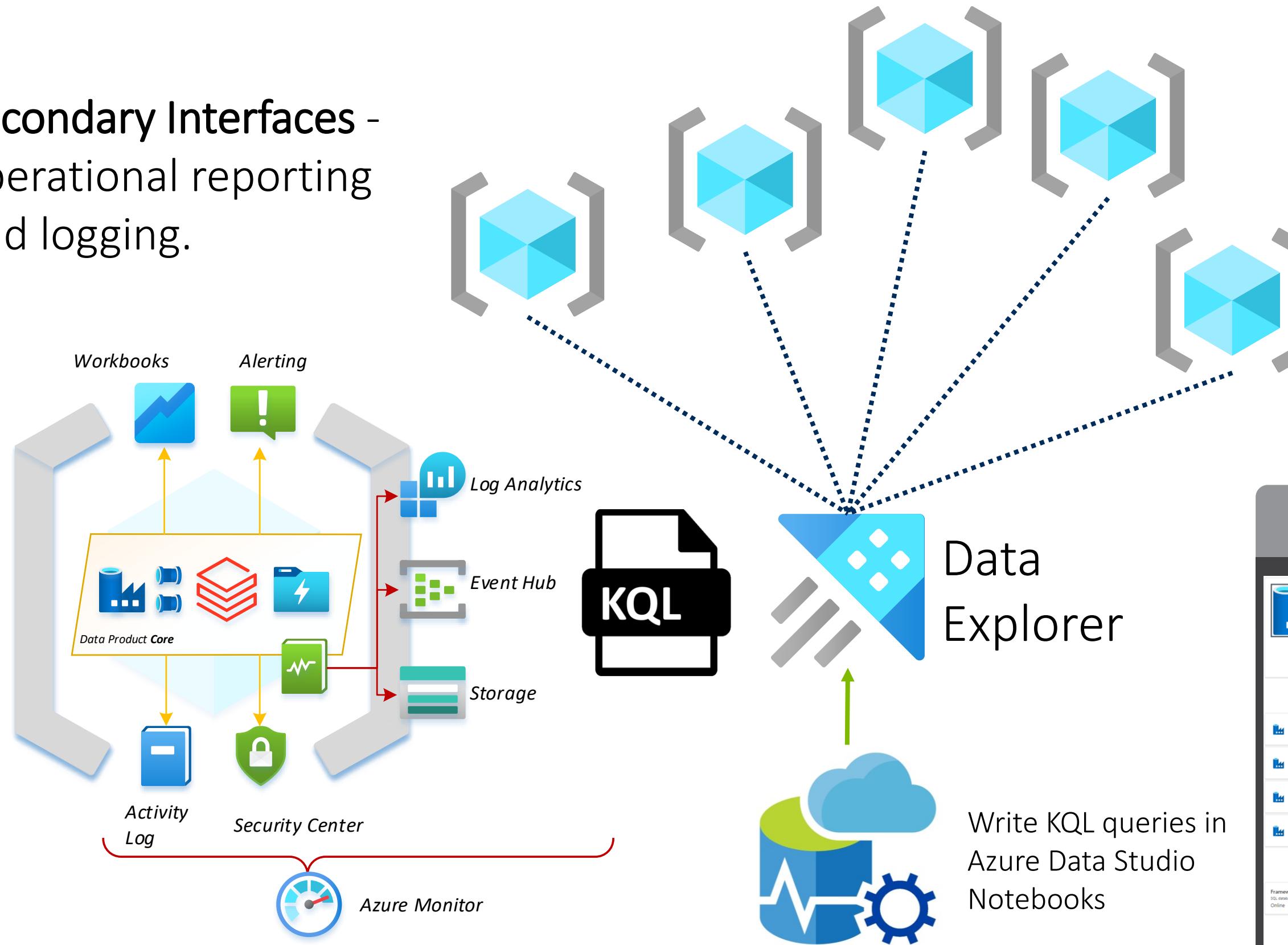


Primary Interfaces –
data integration and
exchange.

Data Products in Azure with Interfaces

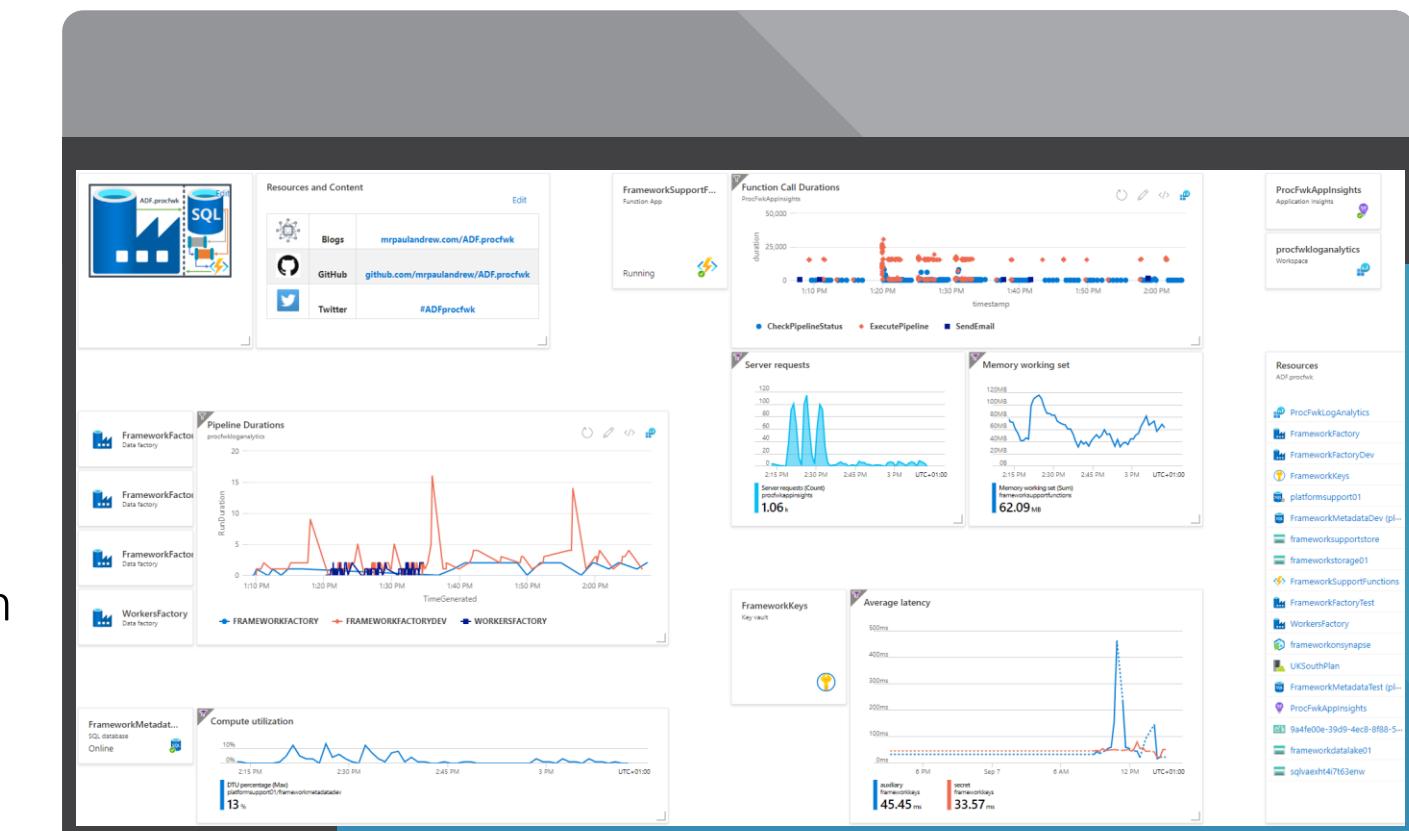


Secondary Interfaces - operational reporting and logging.



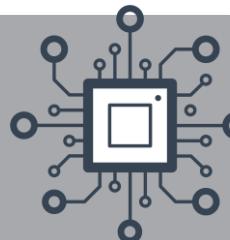
Write KQL queries in
Azure Data Studio
Notebooks

Should all Data Mesh capabilities be de-centralised?

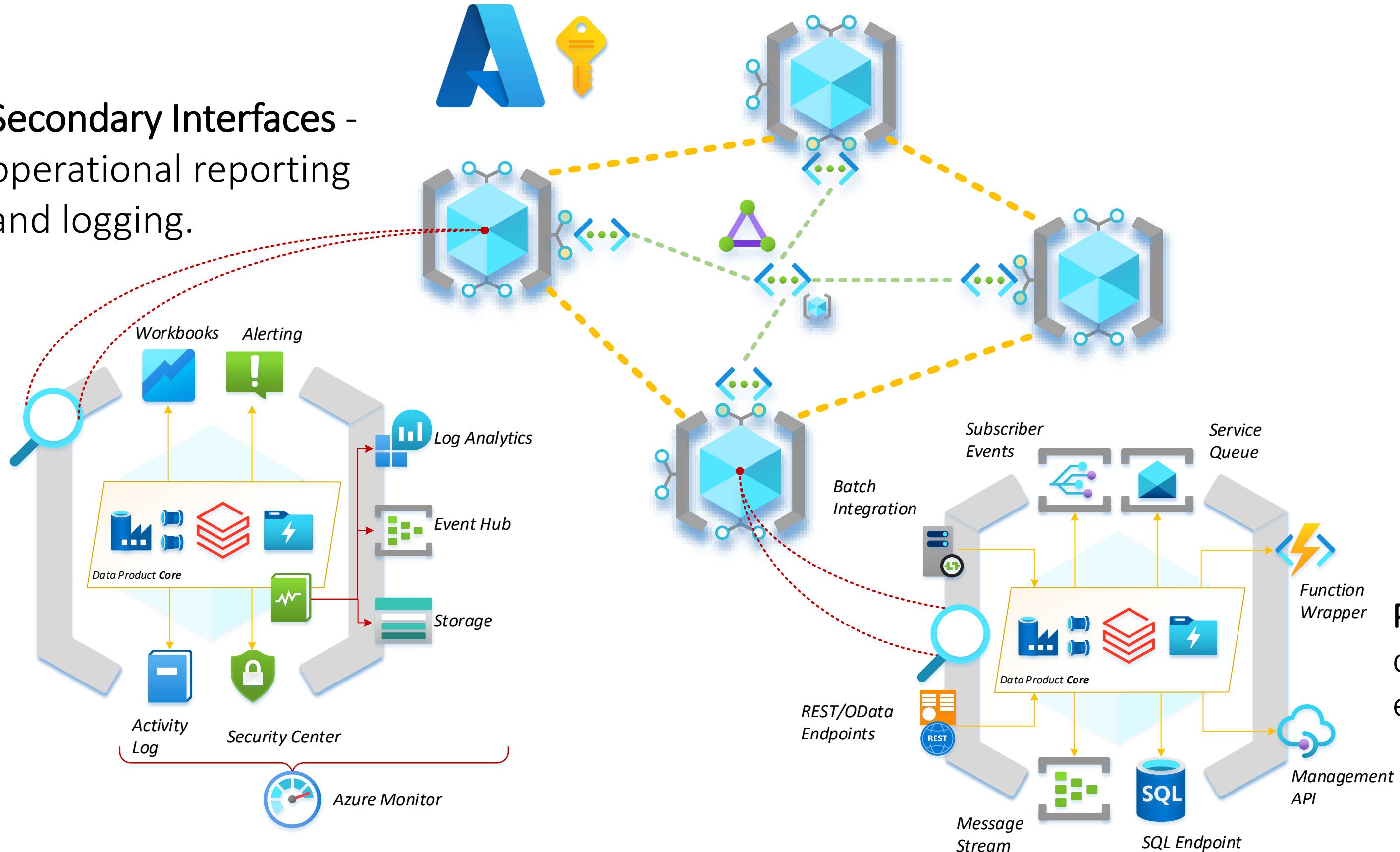




Data Products in Azure with Interfaces

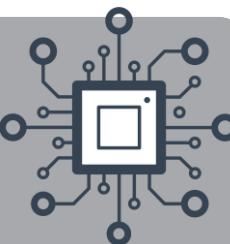


Secondary Interfaces –
operational reporting
and logging.

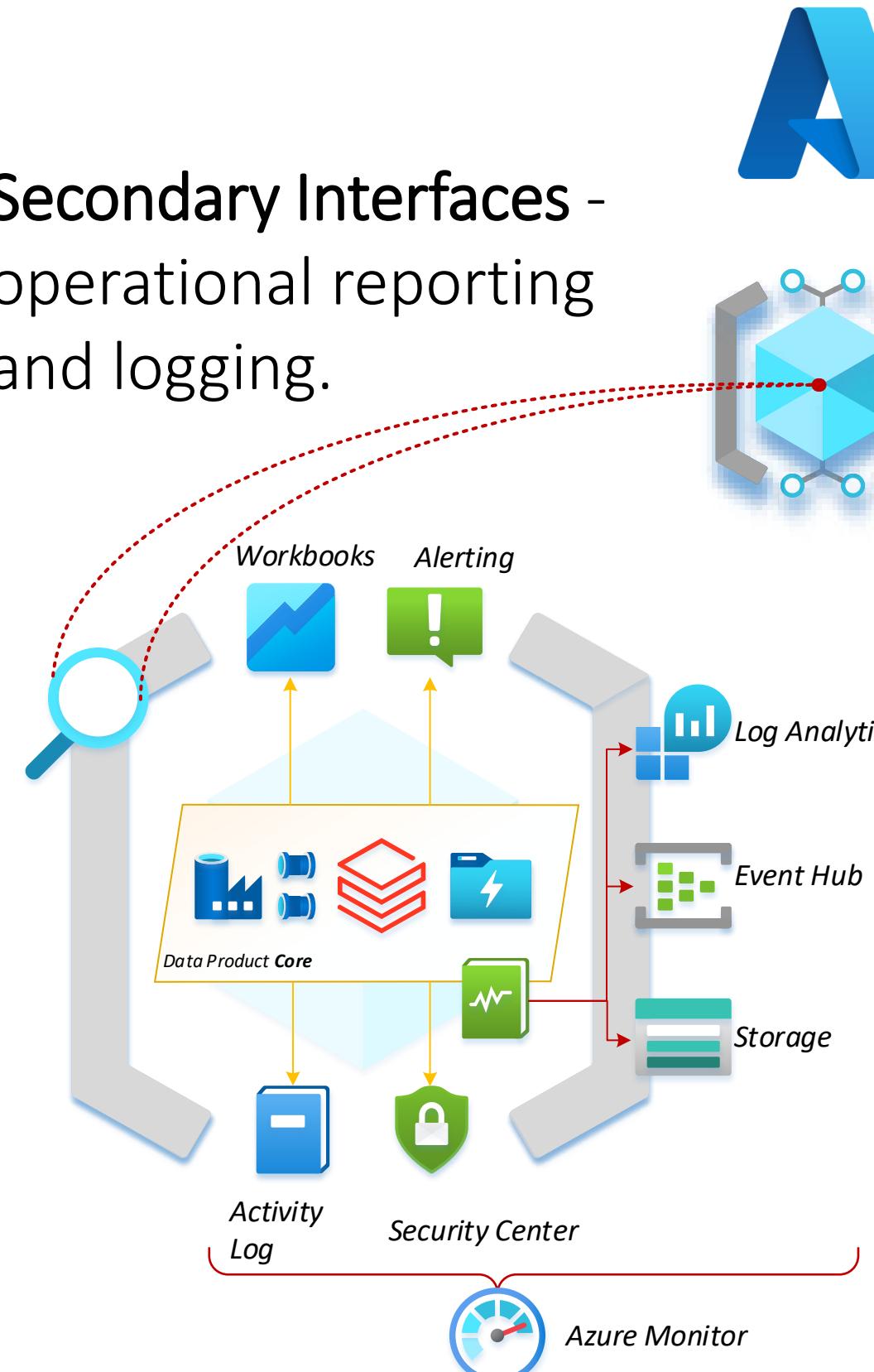


Primary Interfaces –
data integration and
exchange.

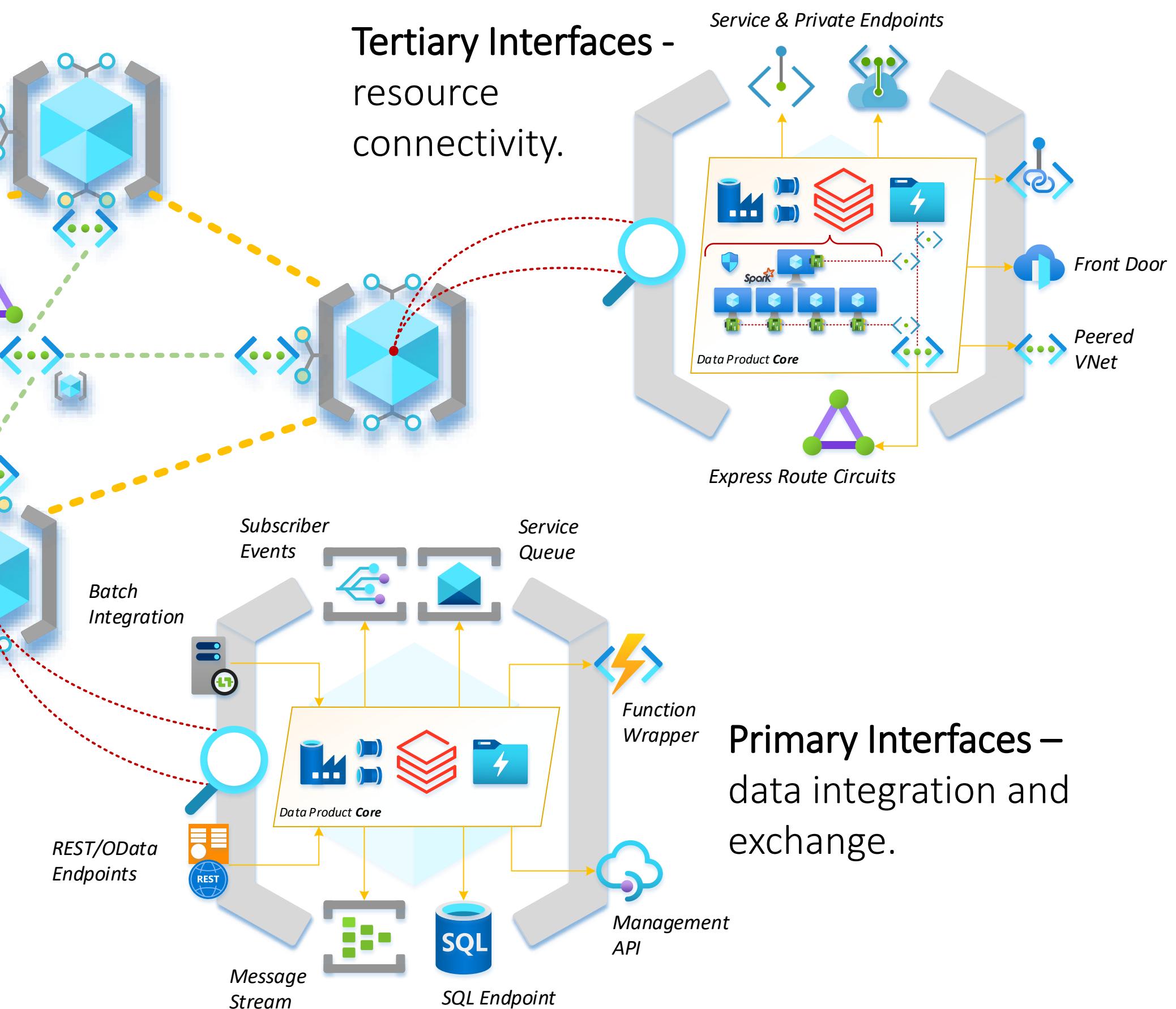
Data Products in Azure with Interfaces



Secondary Interfaces –
operational reporting
and logging.

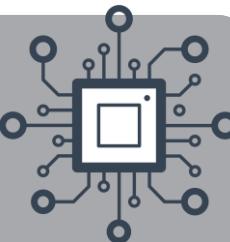


Tertiary Interfaces –
resource connectivity.

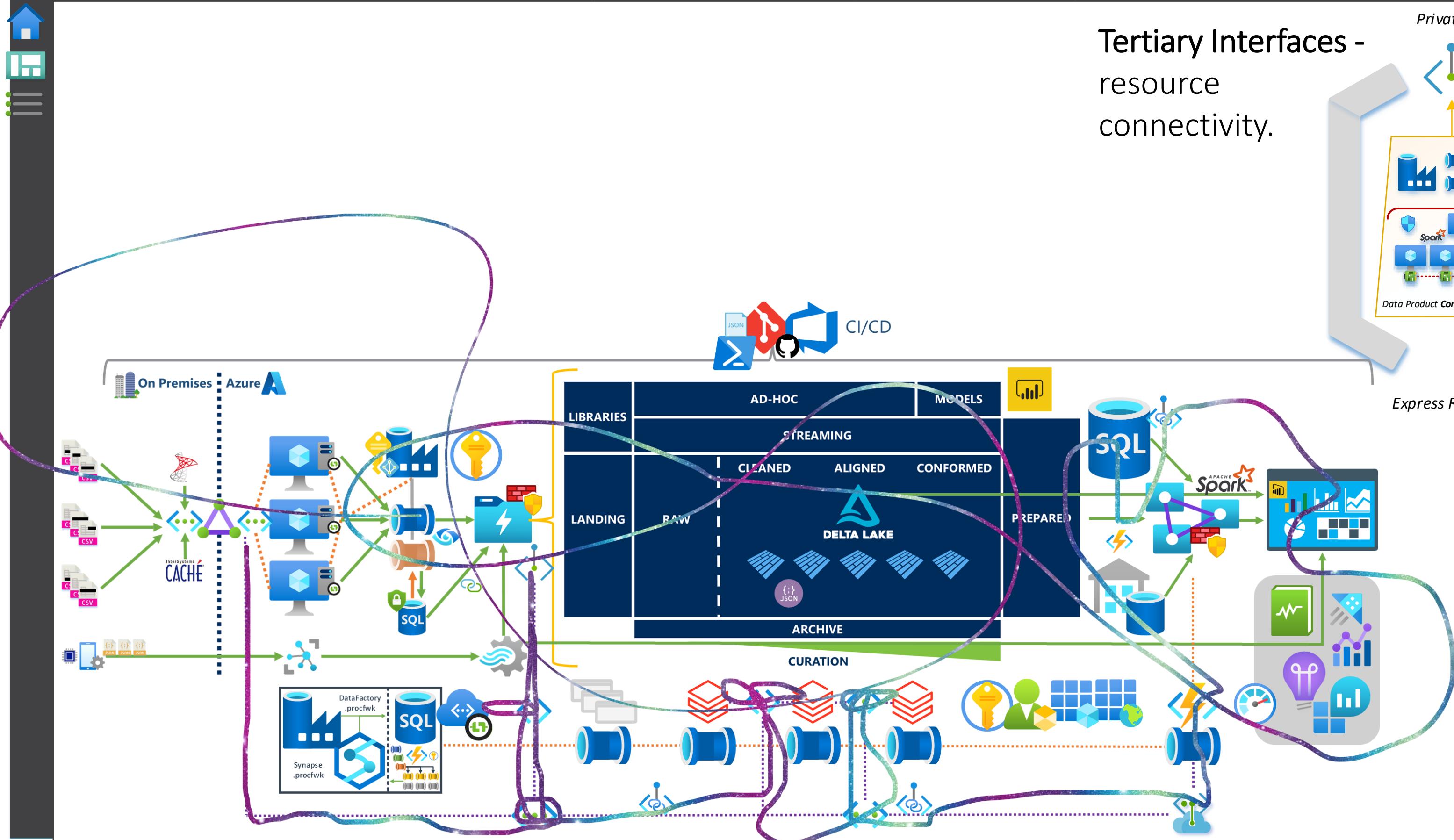
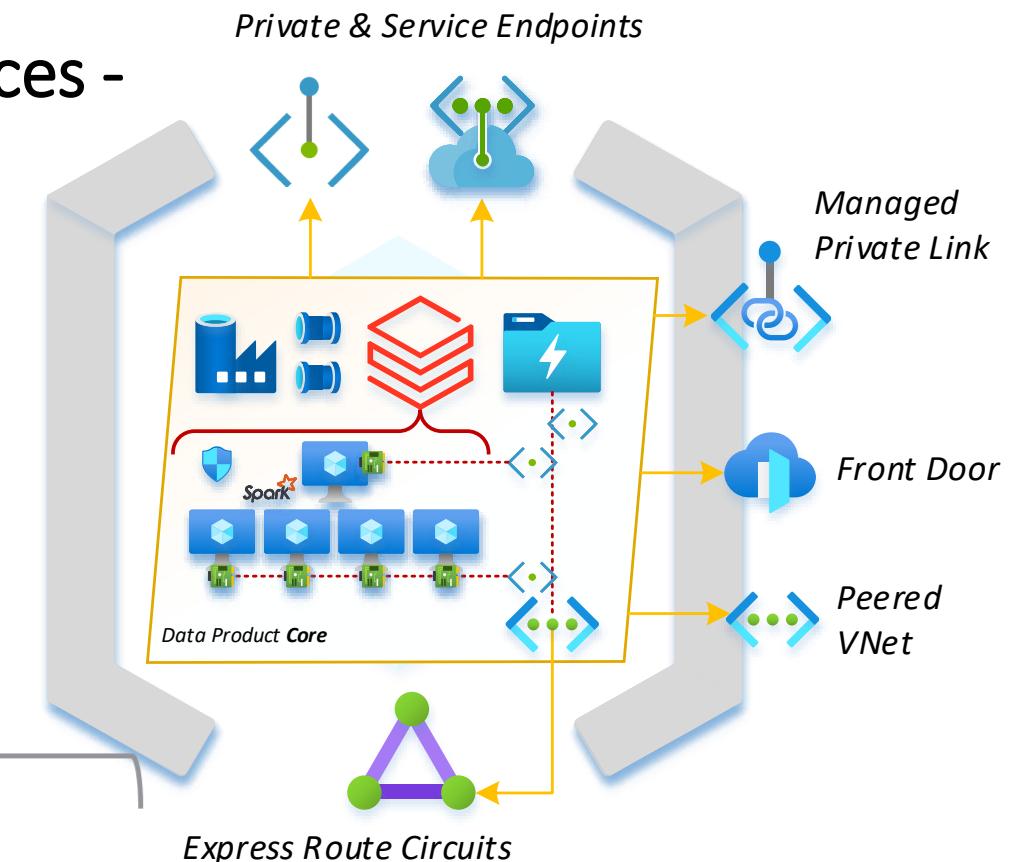


Primary Interfaces –
data integration and
exchange.

Data Products in Azure with Interfaces

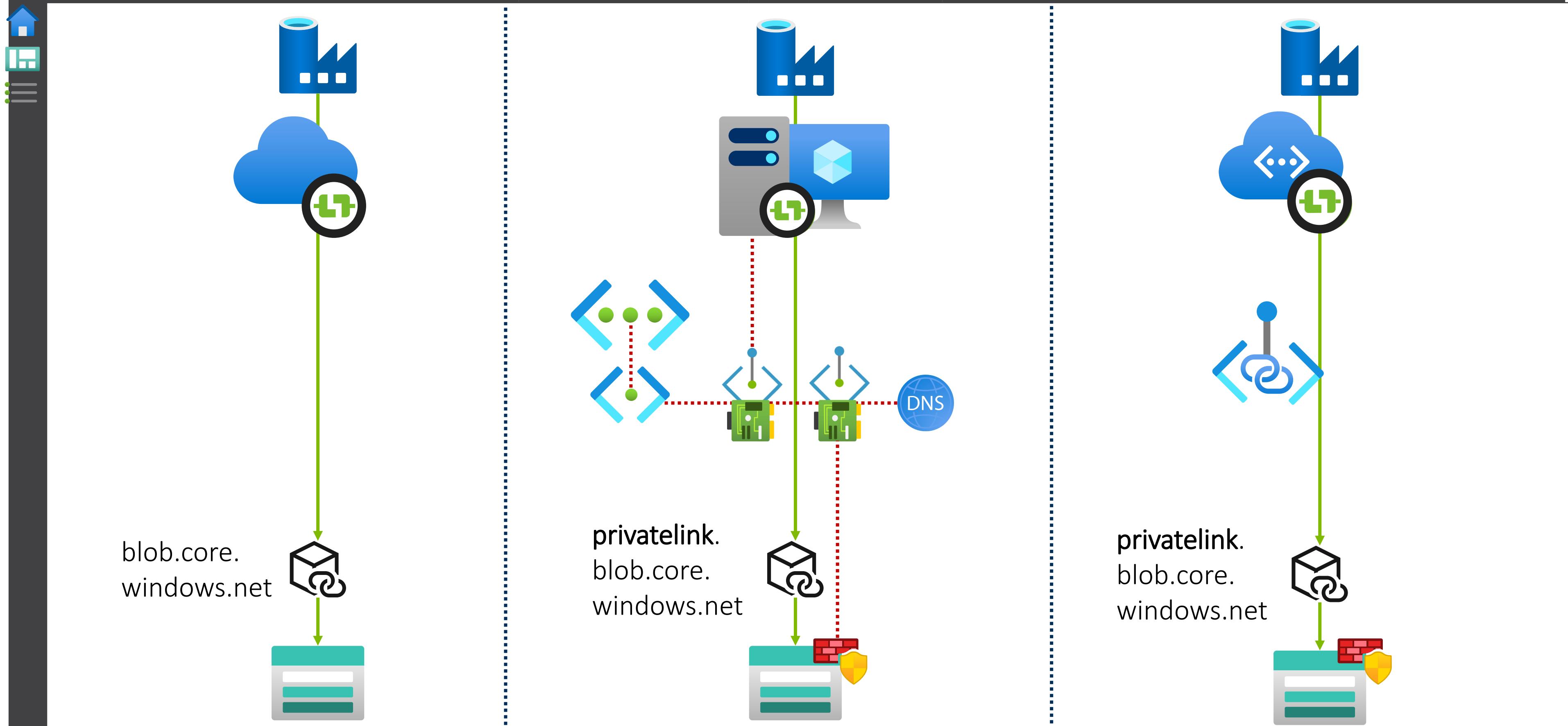
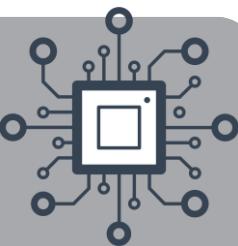


Tertiary Interfaces -
resource
connectivity.

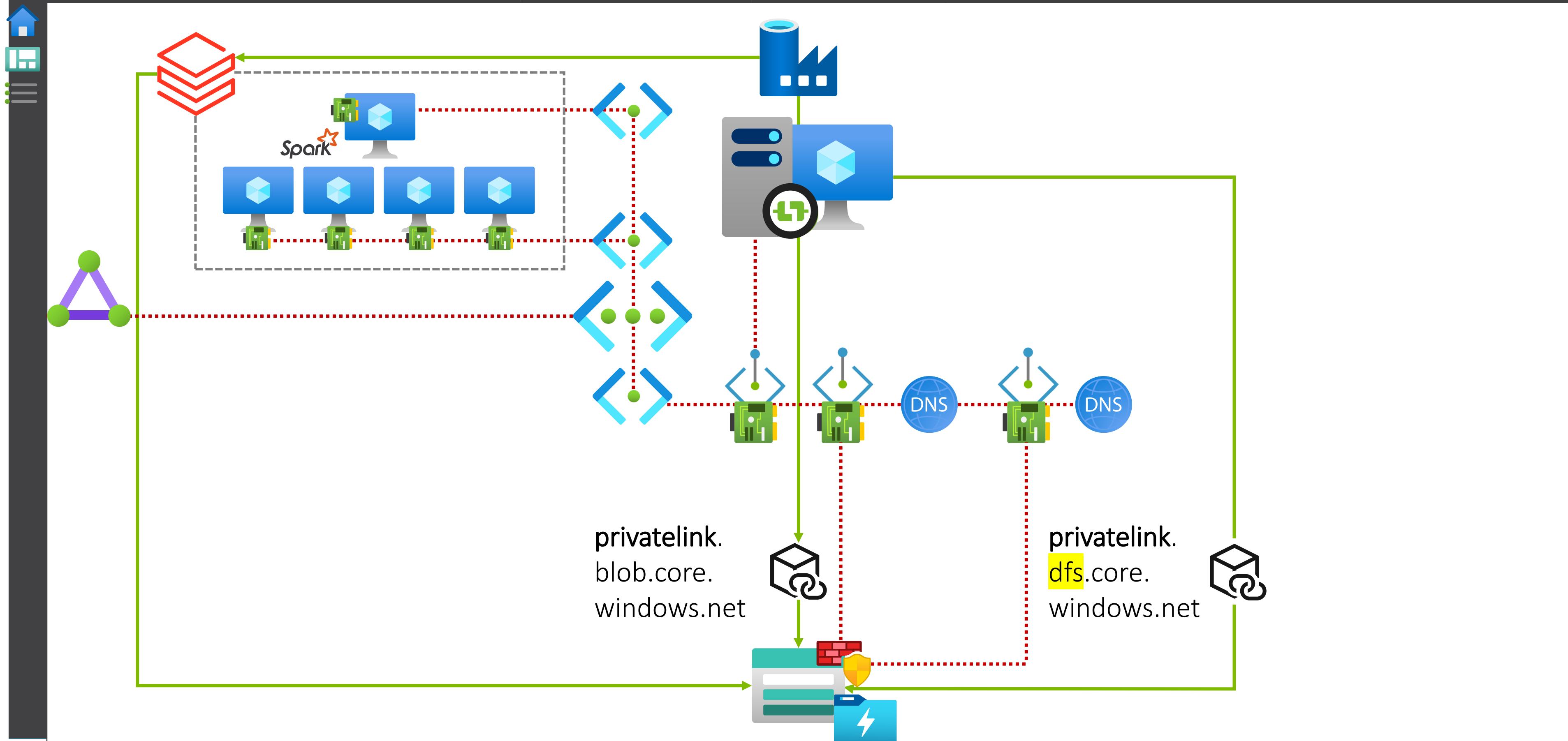
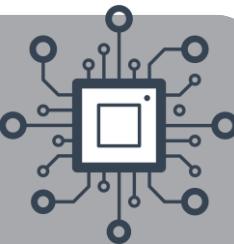




Managed vs Unmanaged Connections



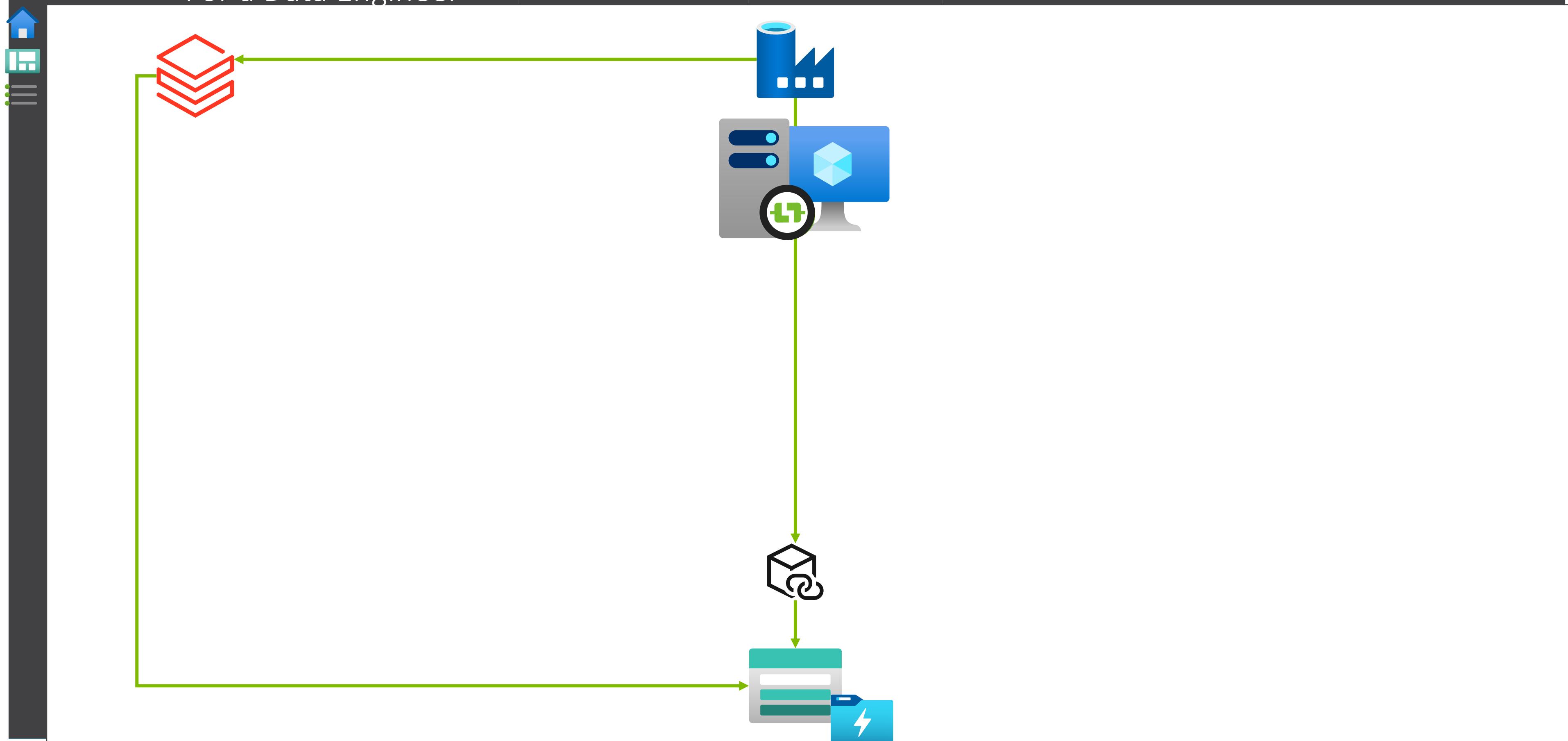
Further VNet Connections





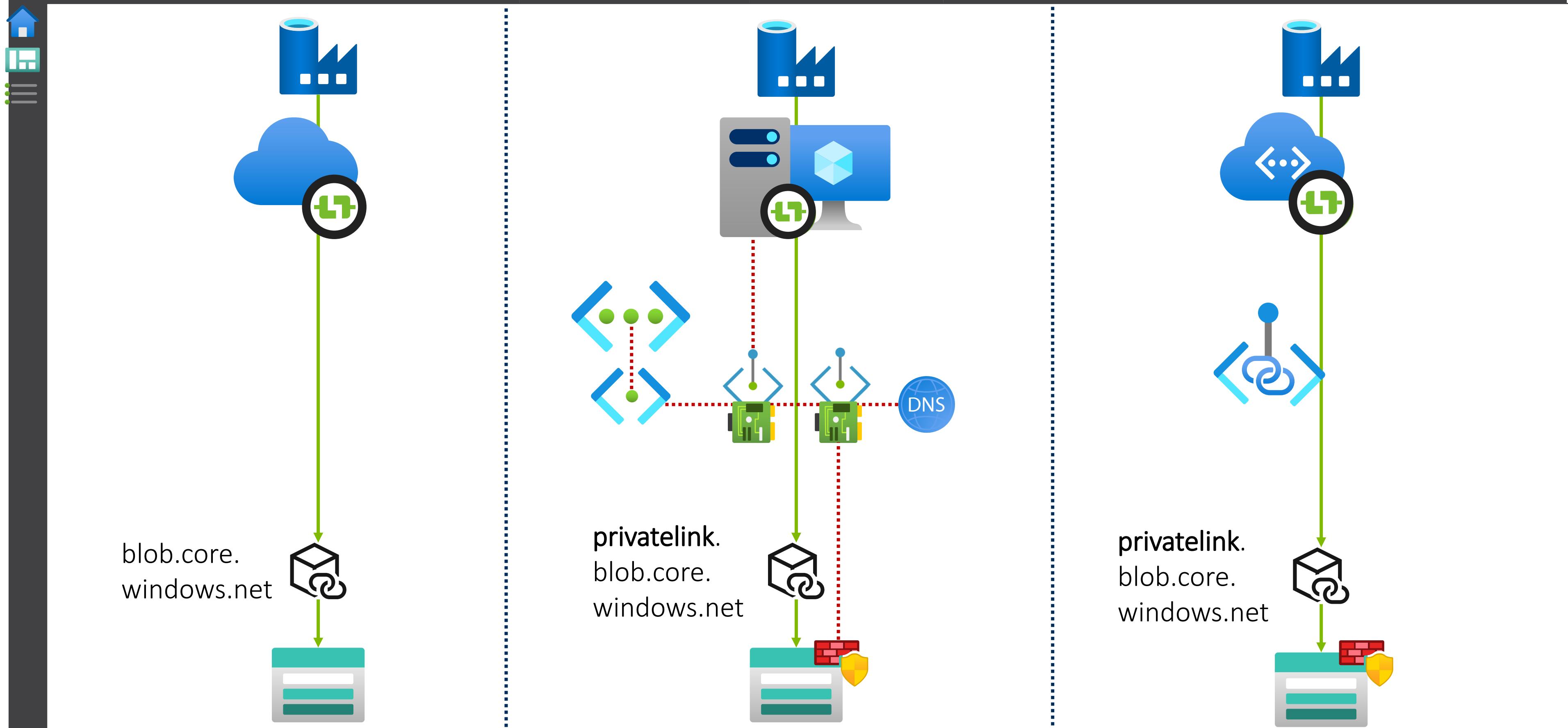
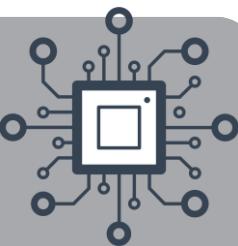
Further VNet Connections

For a Data Engineer

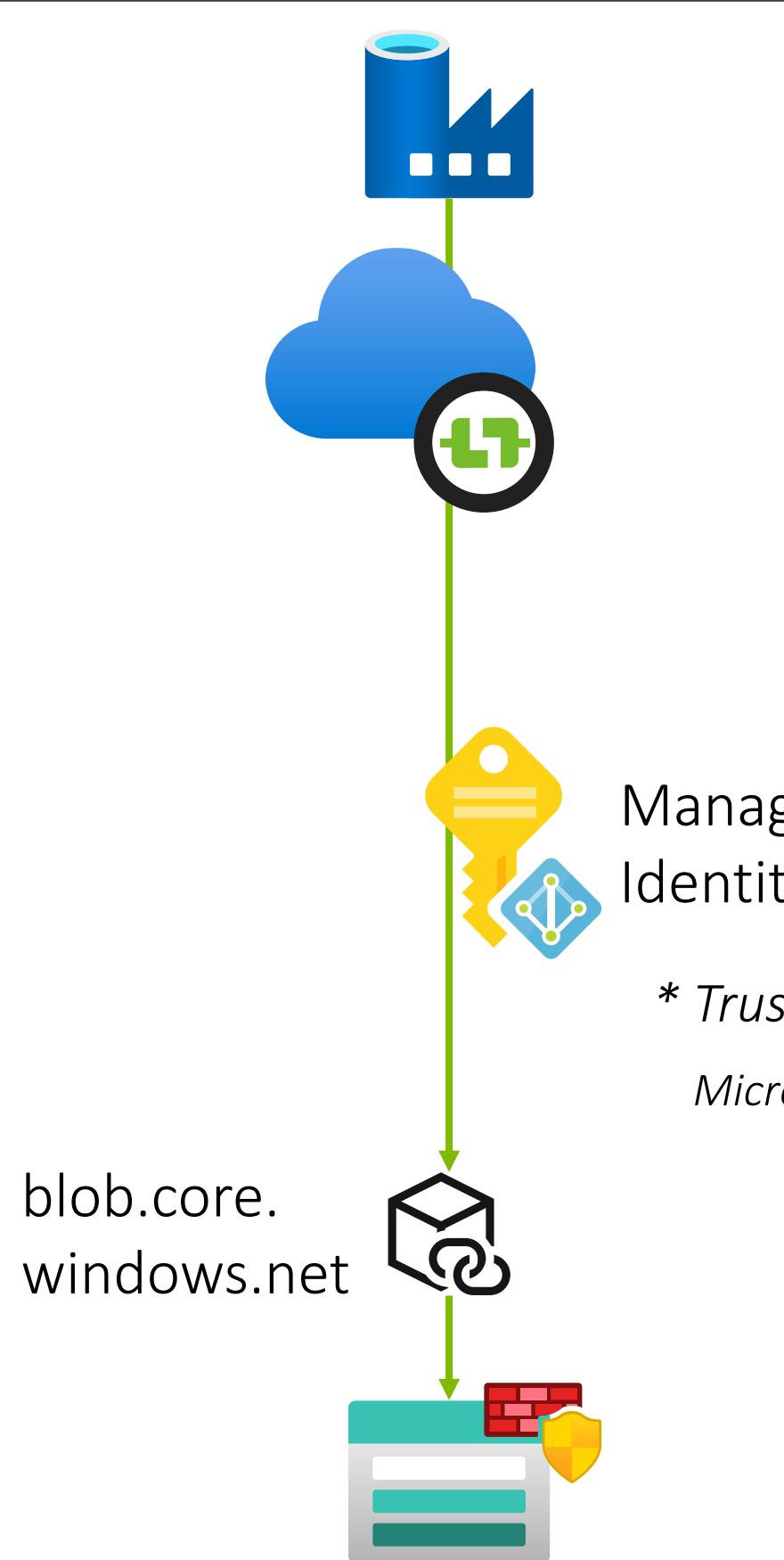
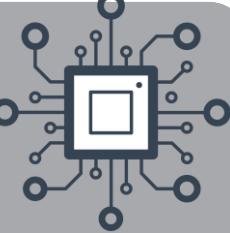
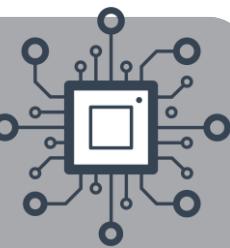




Managed vs Unmanaged Connections



Public Connections



Managed
Identity

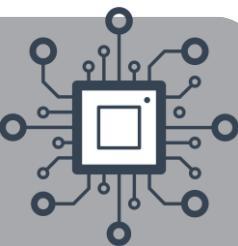
** Trusted access based on system-assigned managed identity.*

Microsoft.DataFactory/factories

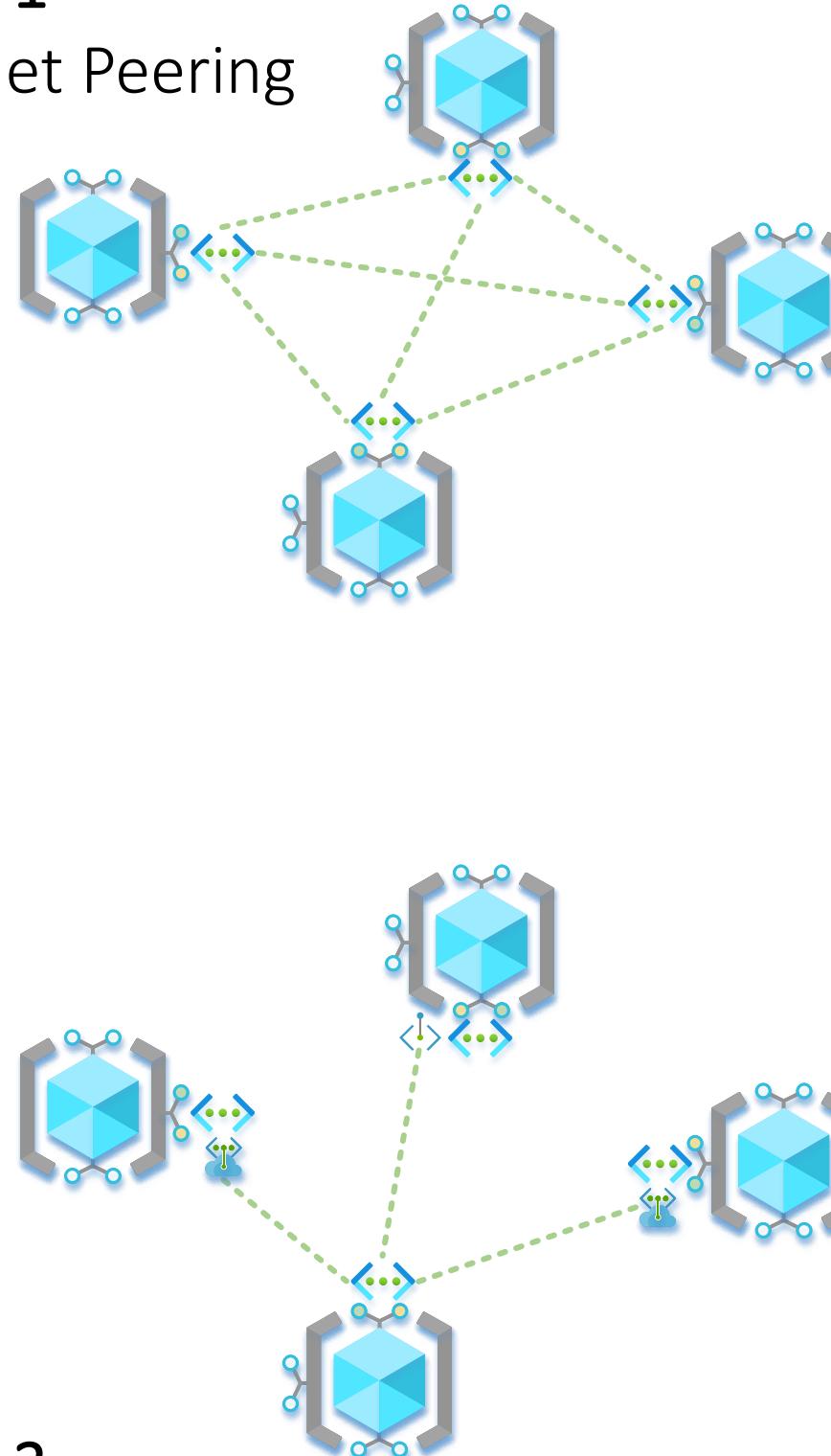
`blob.core.
windows.net`



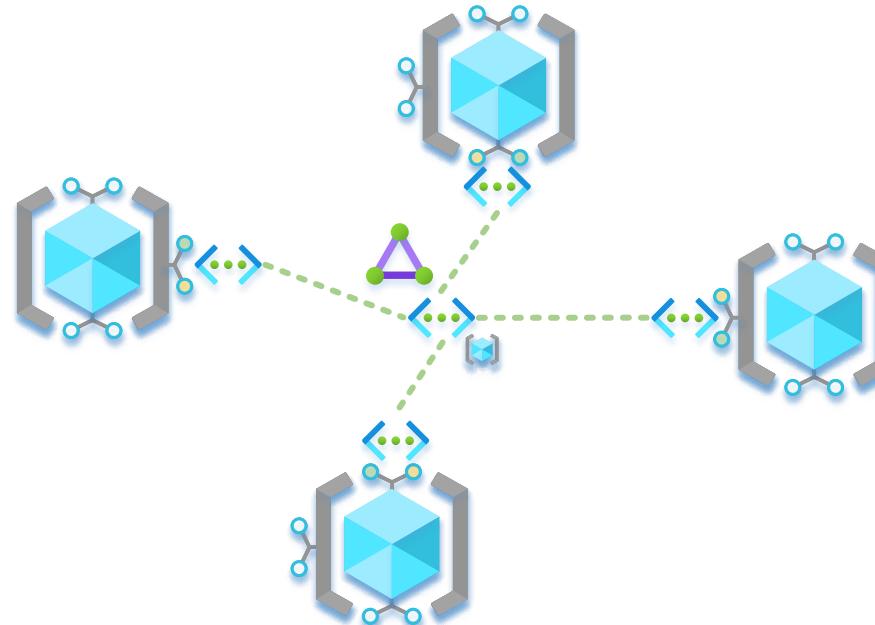
Data Products in Azure with Interfaces



Option 1
Full VNet Peering

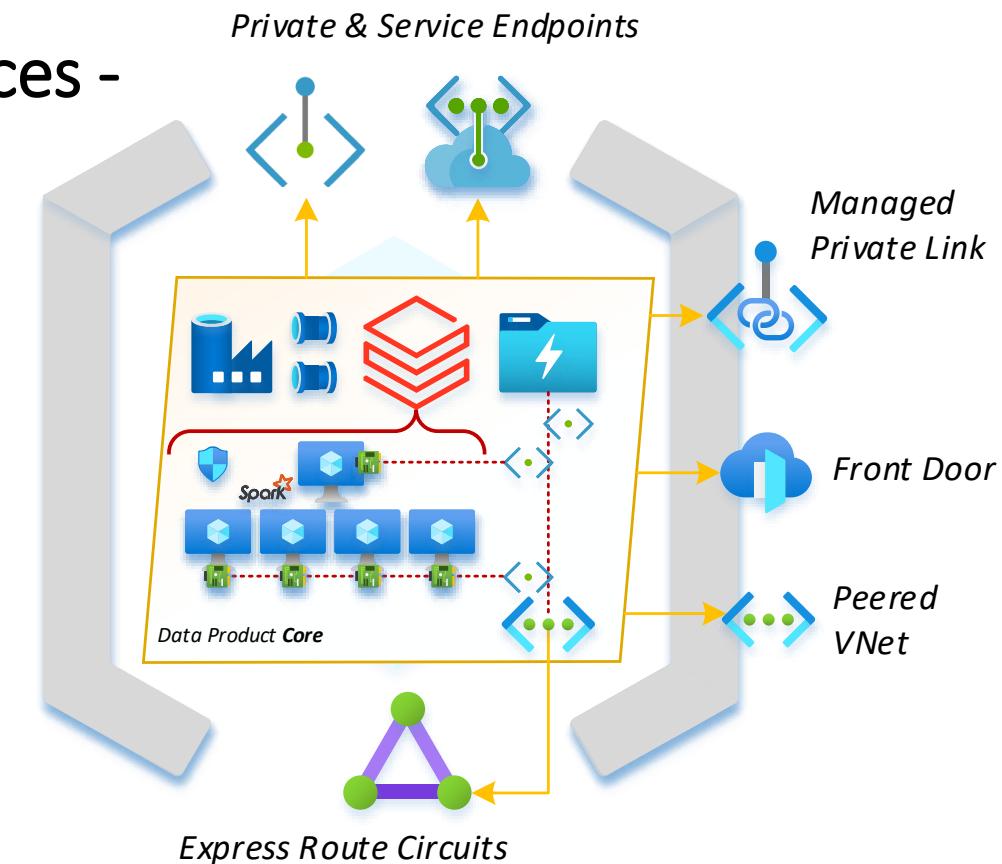


Option 3
Targeted Service/Private Endpoints



Option 2
Hub and Spoke VNet Peering

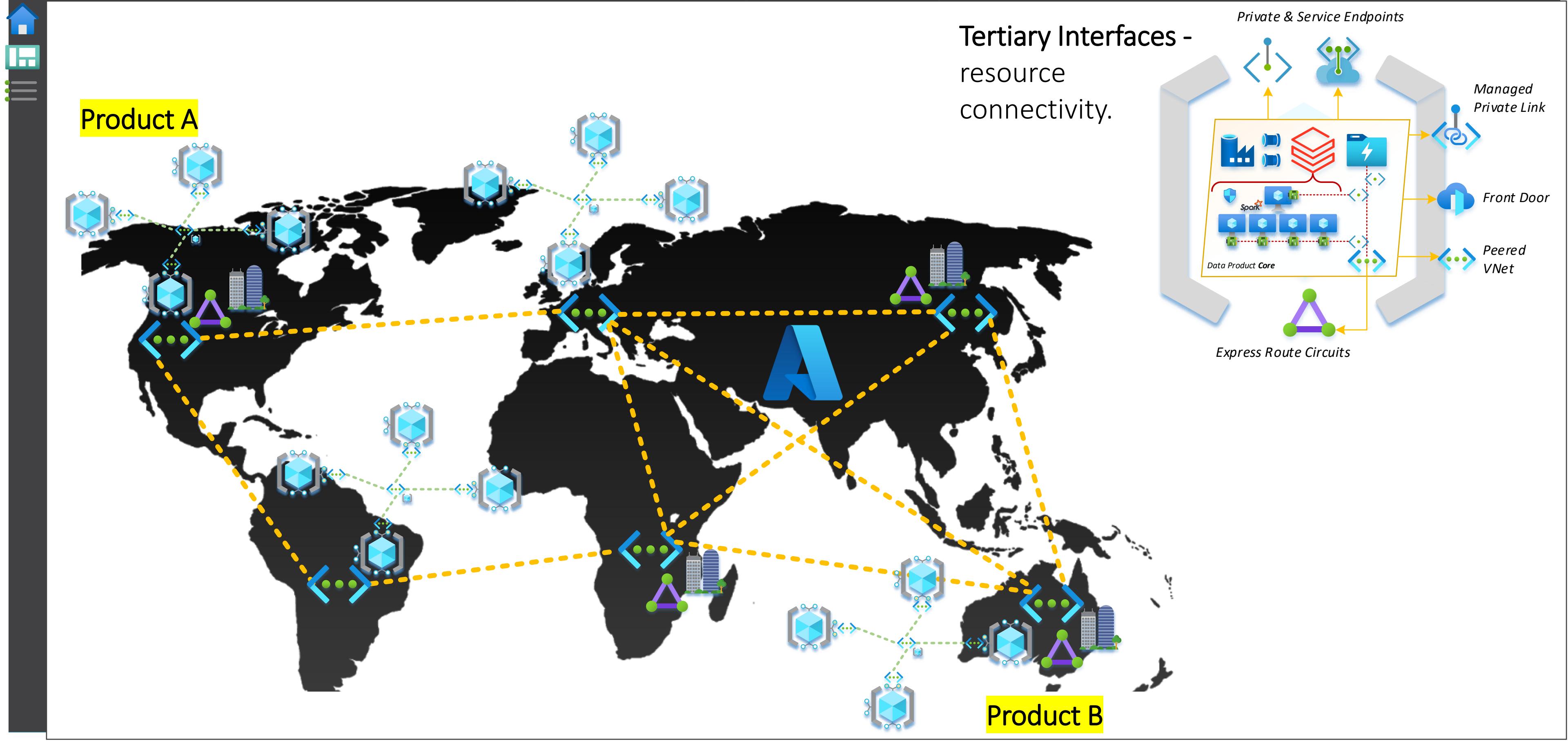
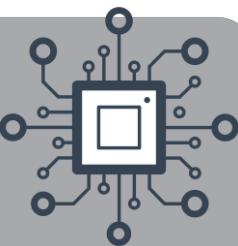
Tertiary Interfaces -
resource
connectivity.



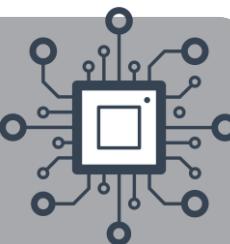
Other Options
Use public endpoints.
Just use local resource firewalls.
Use Managed Private Links



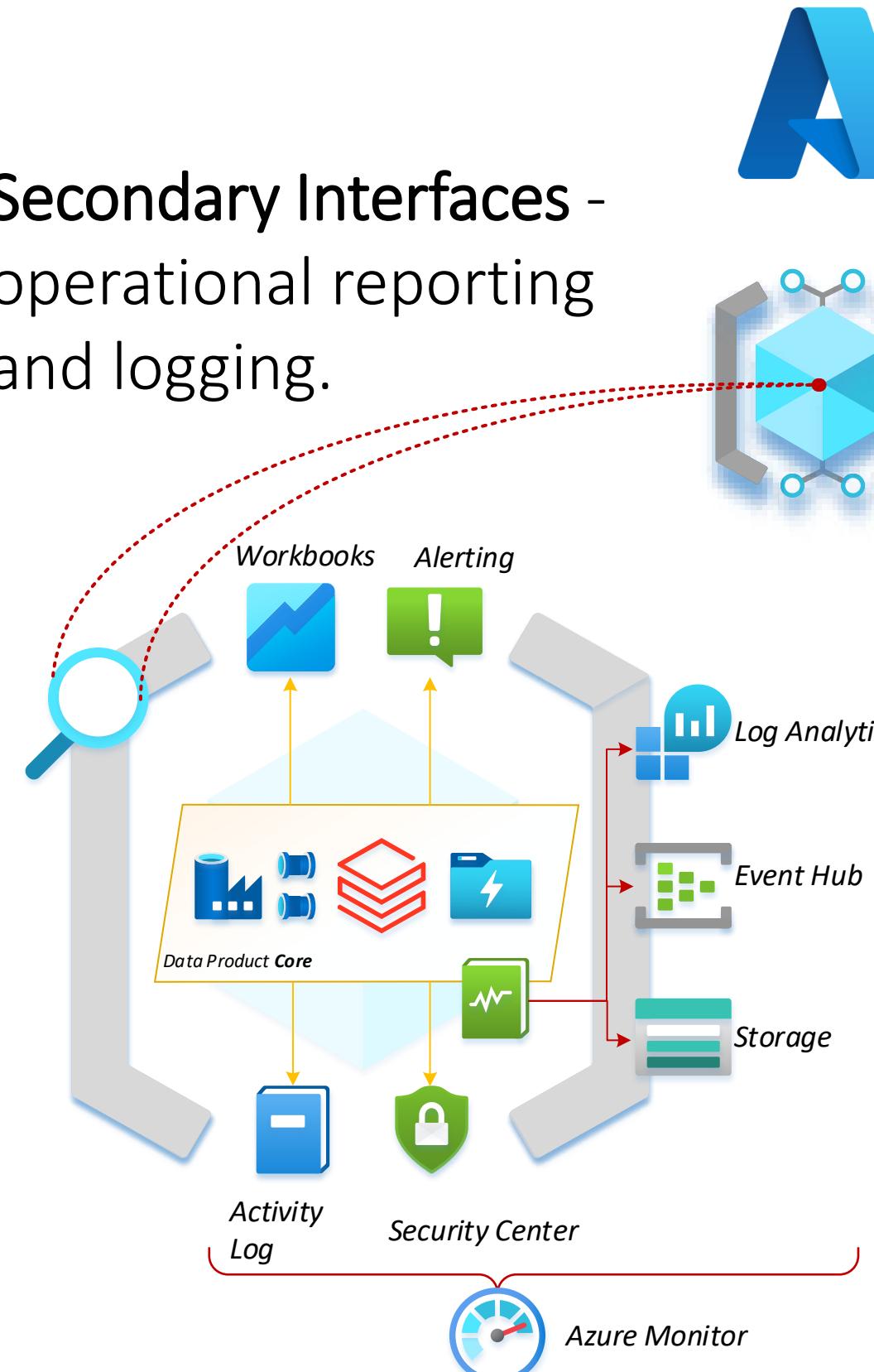
Data Products in Azure with Interfaces



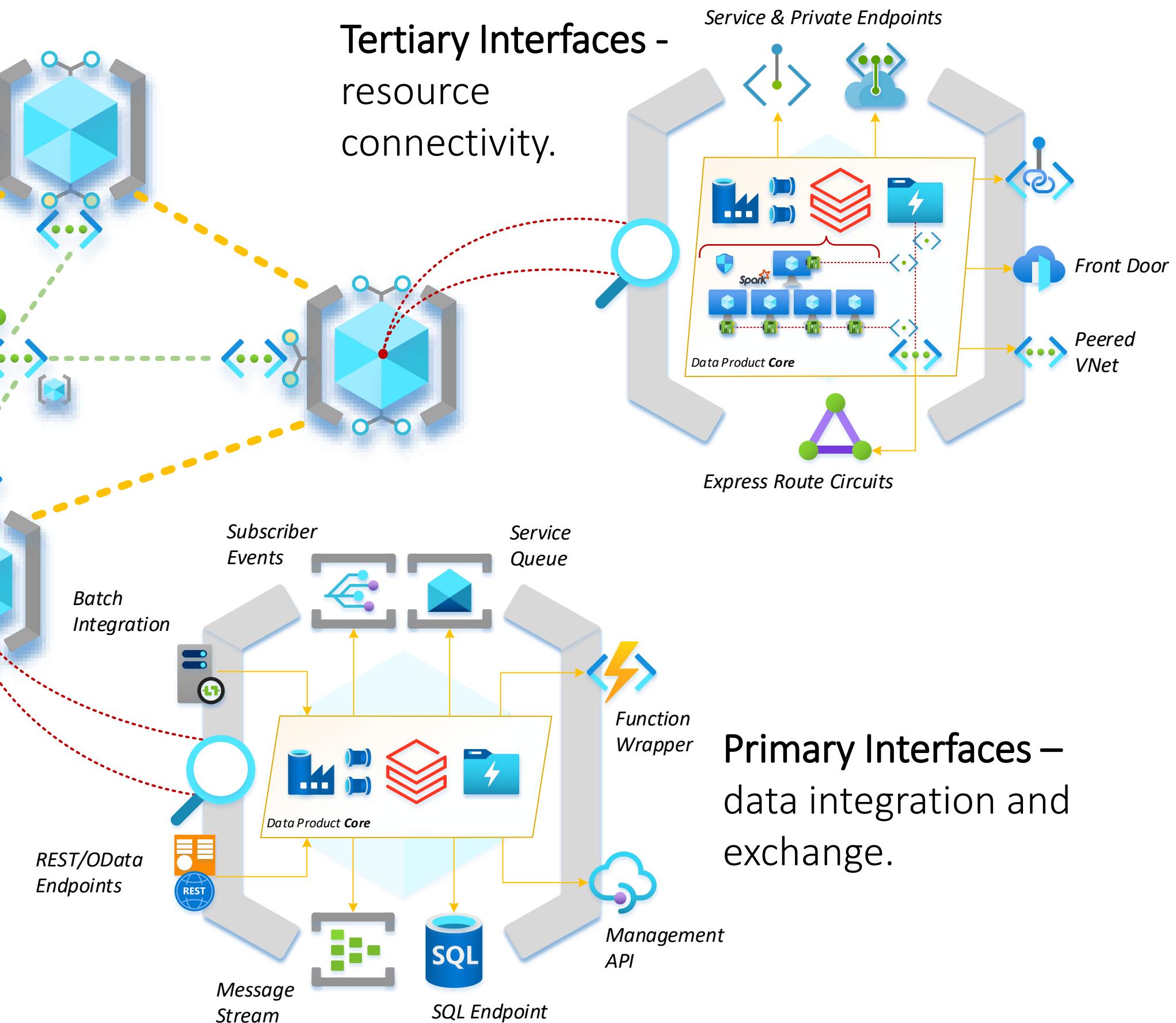
Data Products in Azure with Interfaces



Secondary Interfaces –
operational reporting
and logging.



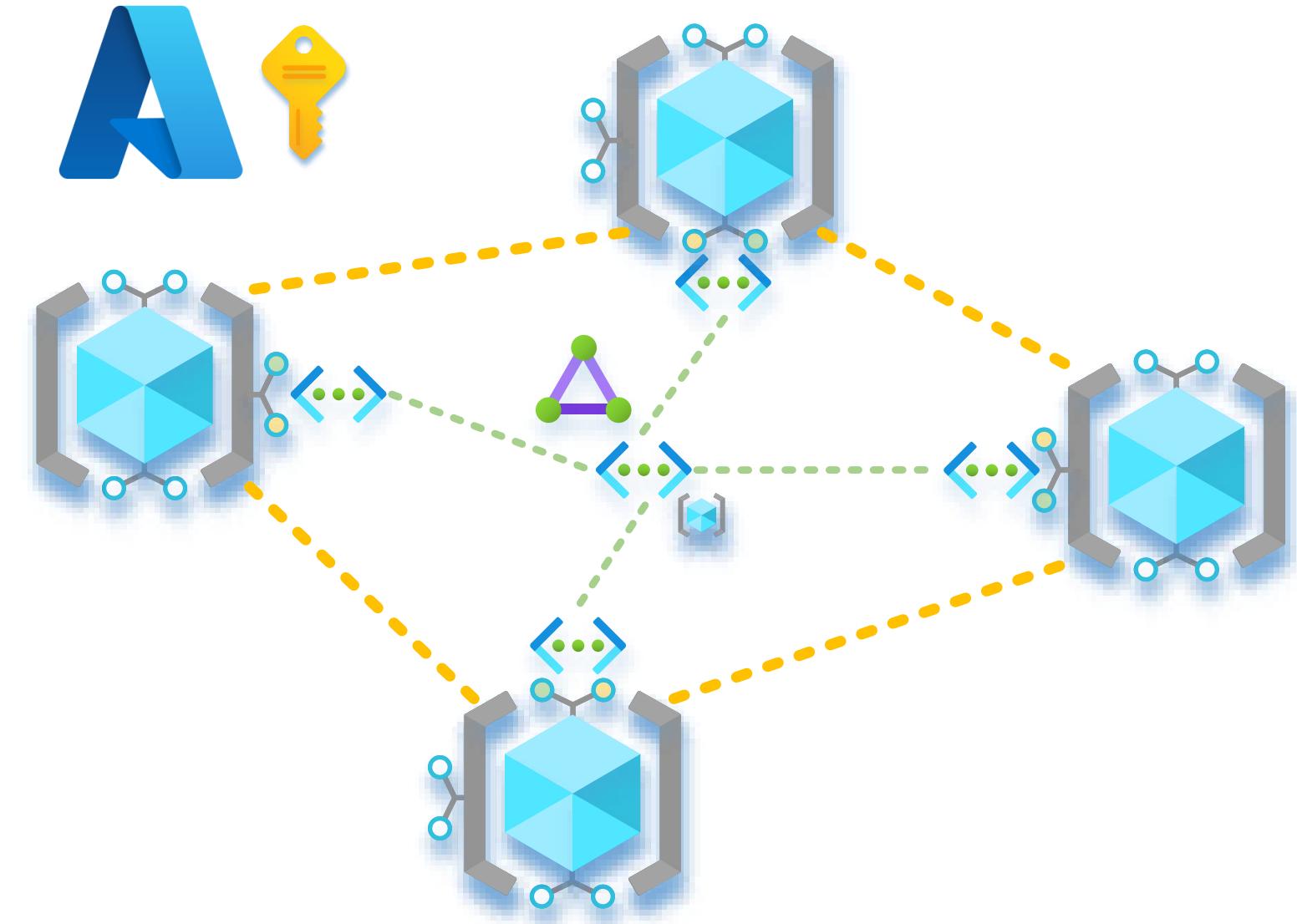
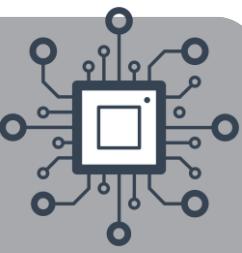
Tertiary Interfaces –
resource
connectivity.



Primary Interfaces –
data integration and
exchange.

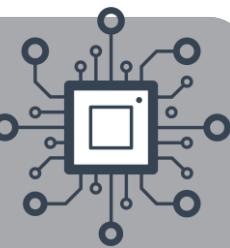


Data Domains in Azure

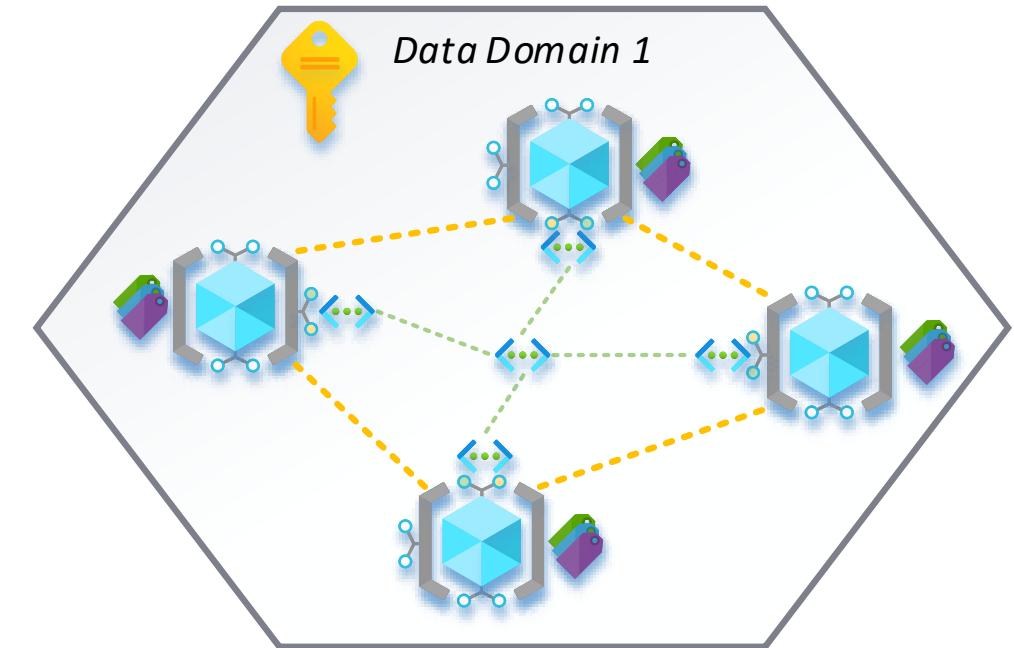




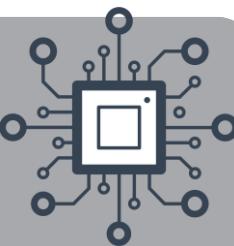
Data Domains in Azure



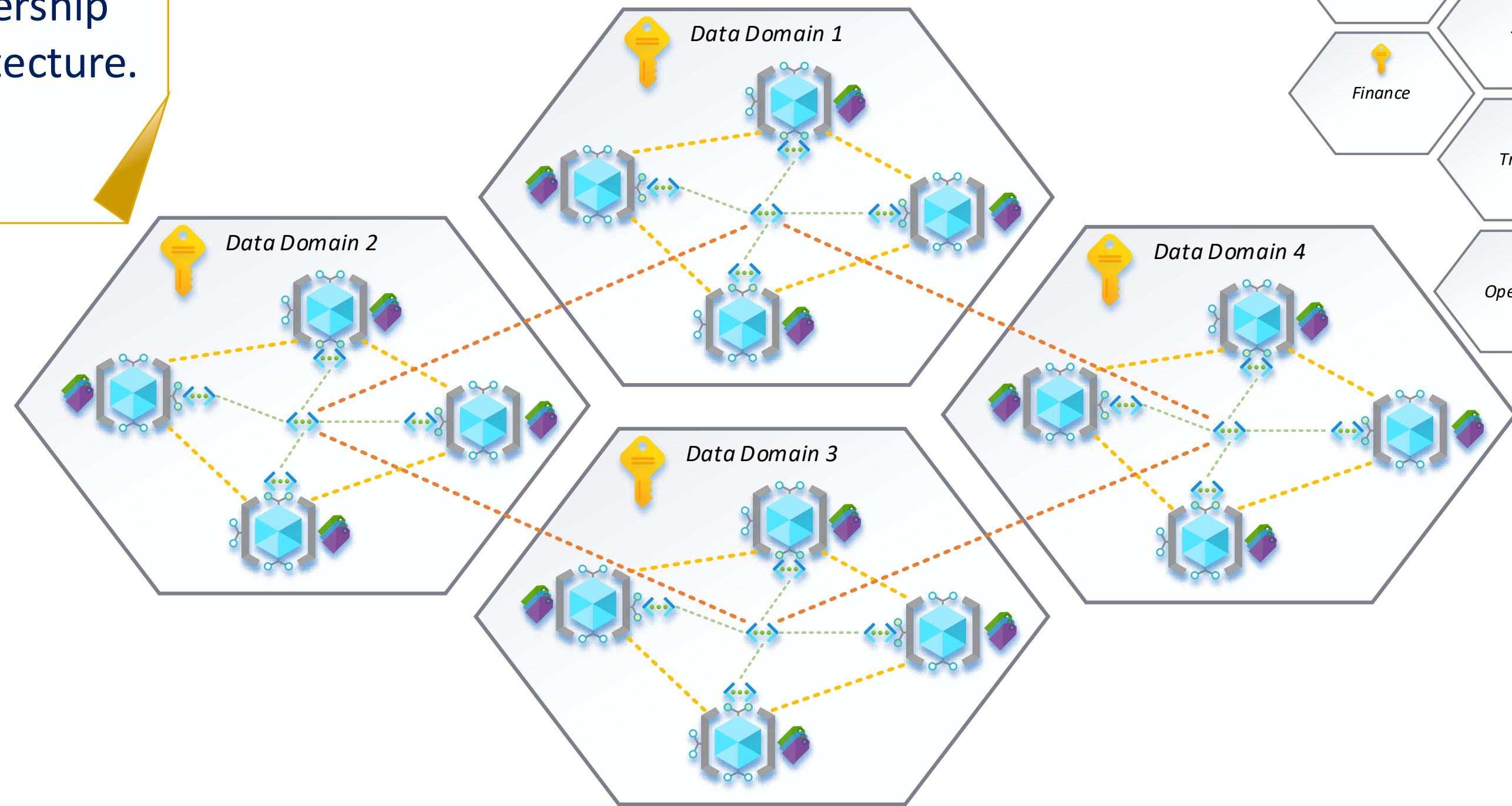
Domain-oriented
decentralised
data ownership
and architecture.



Data Domains in Azure

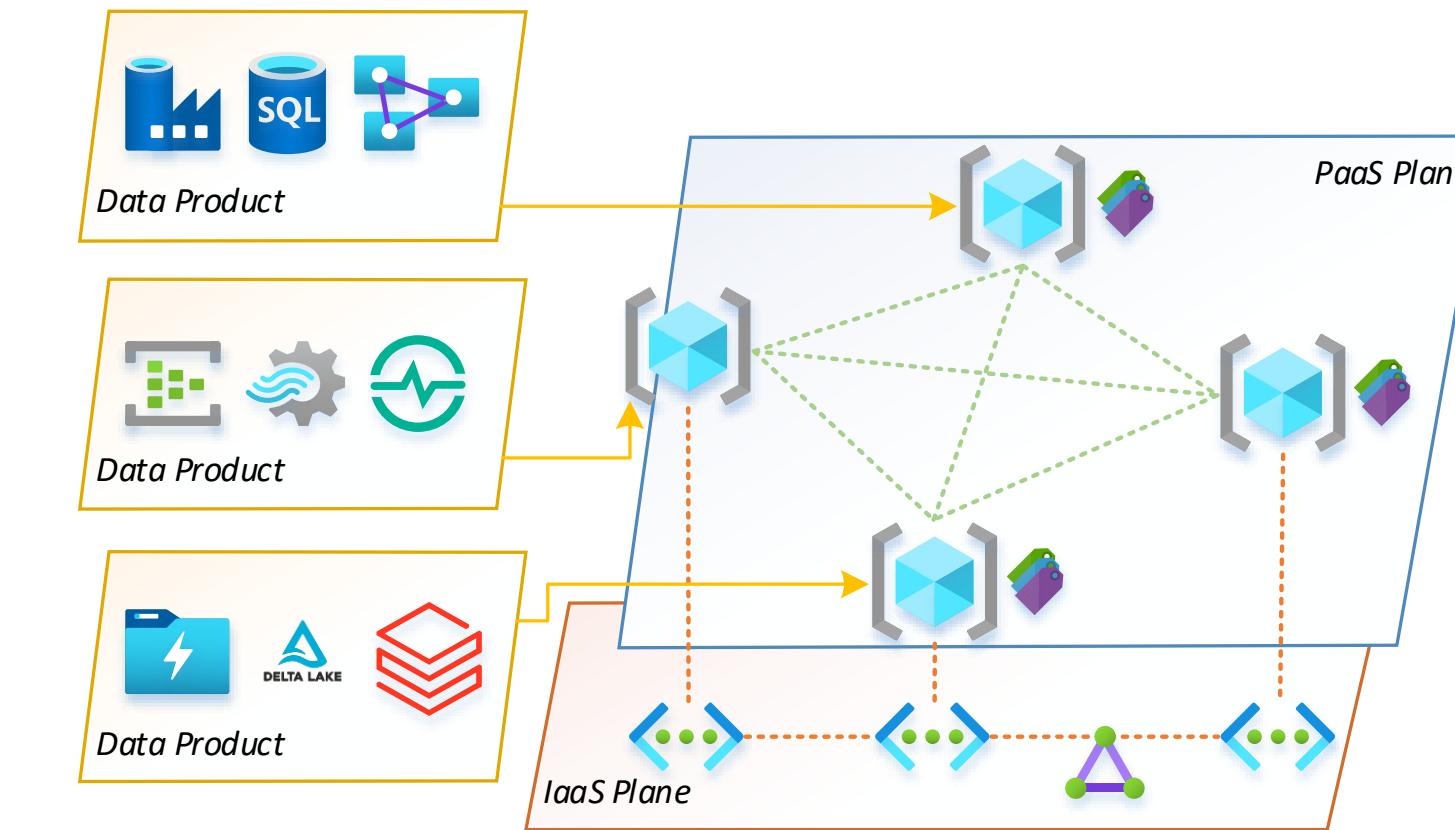
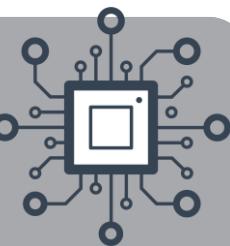


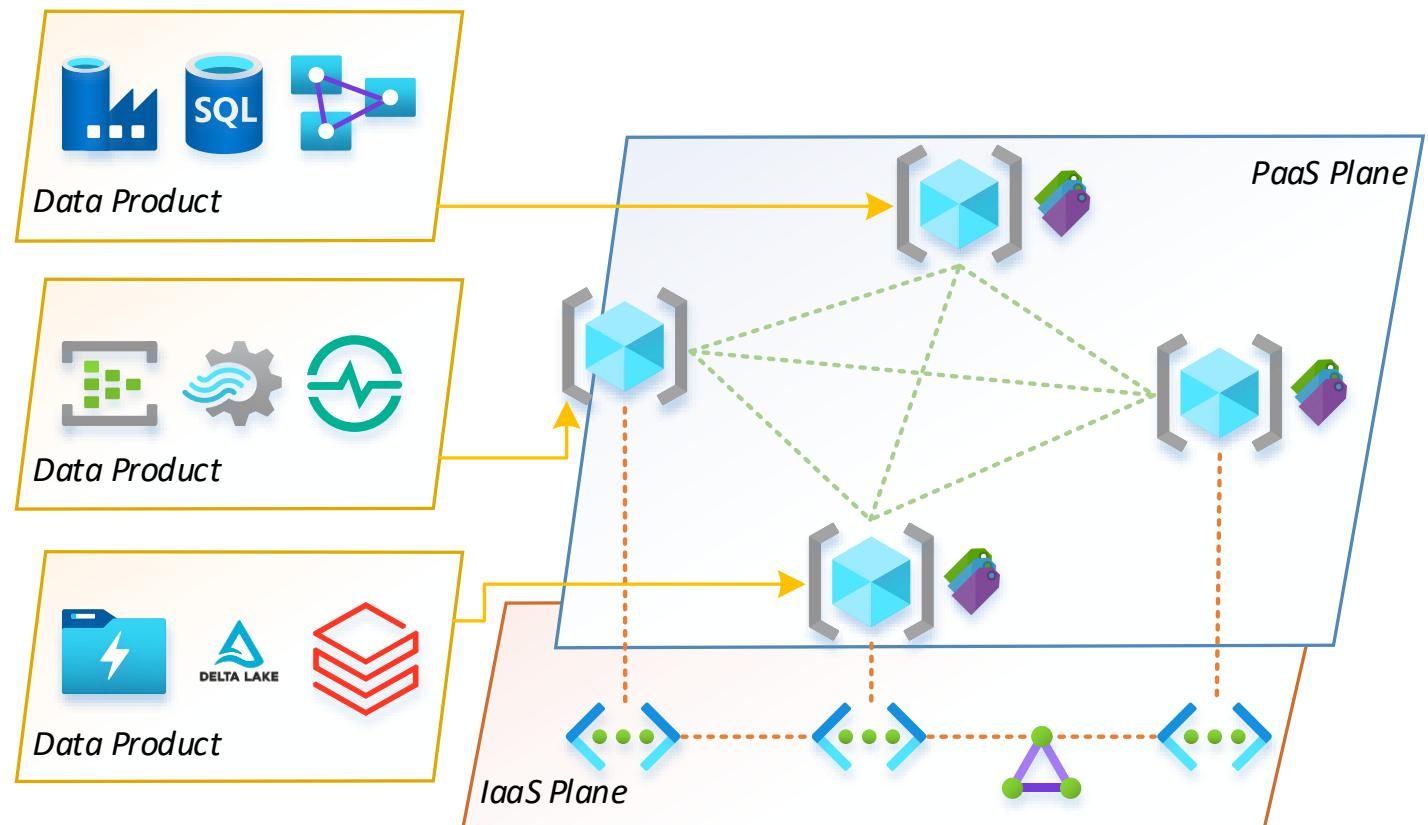
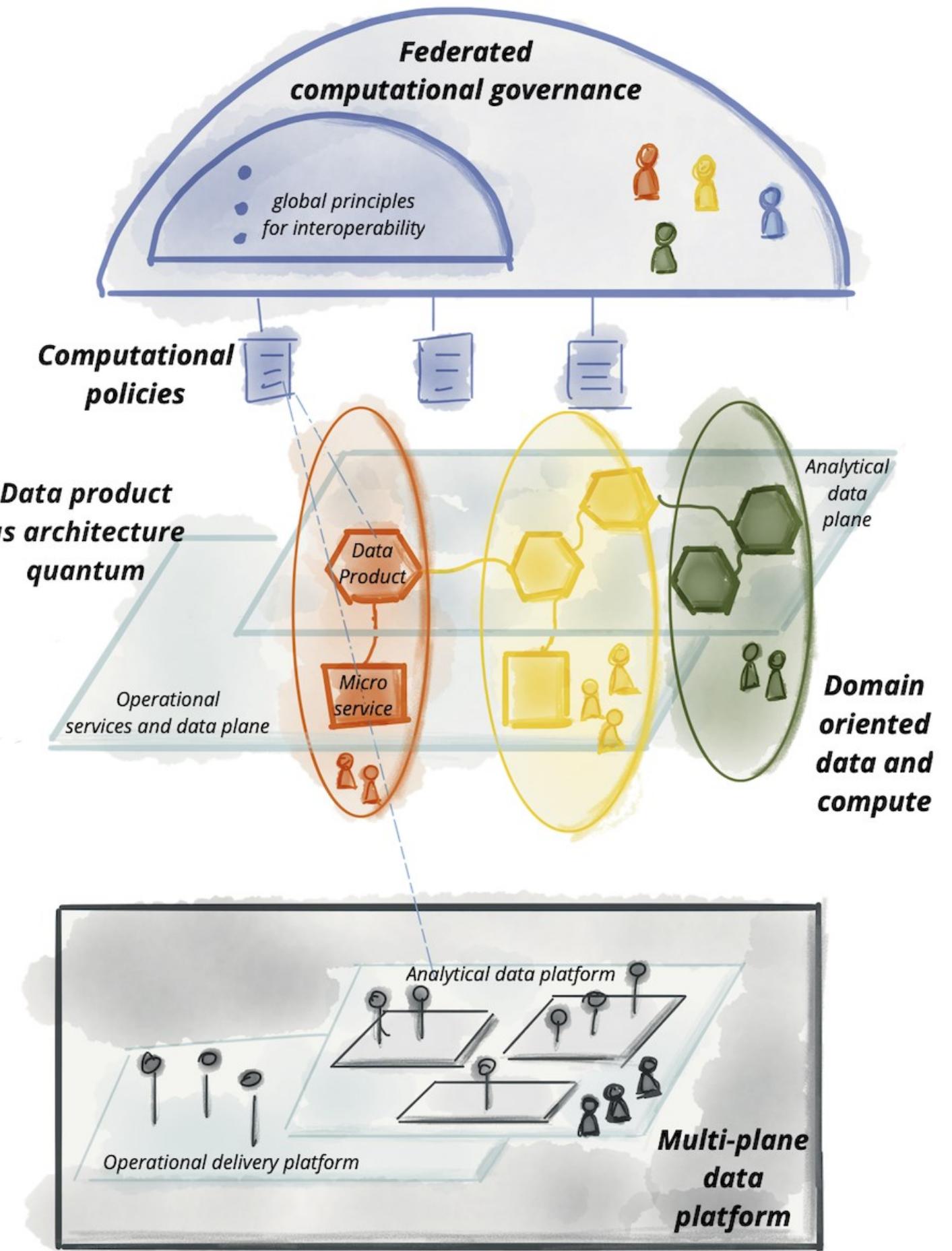
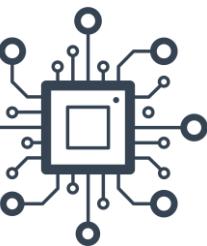
Domain-oriented decentralised data ownership and architecture.

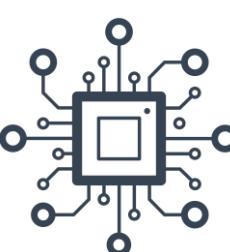
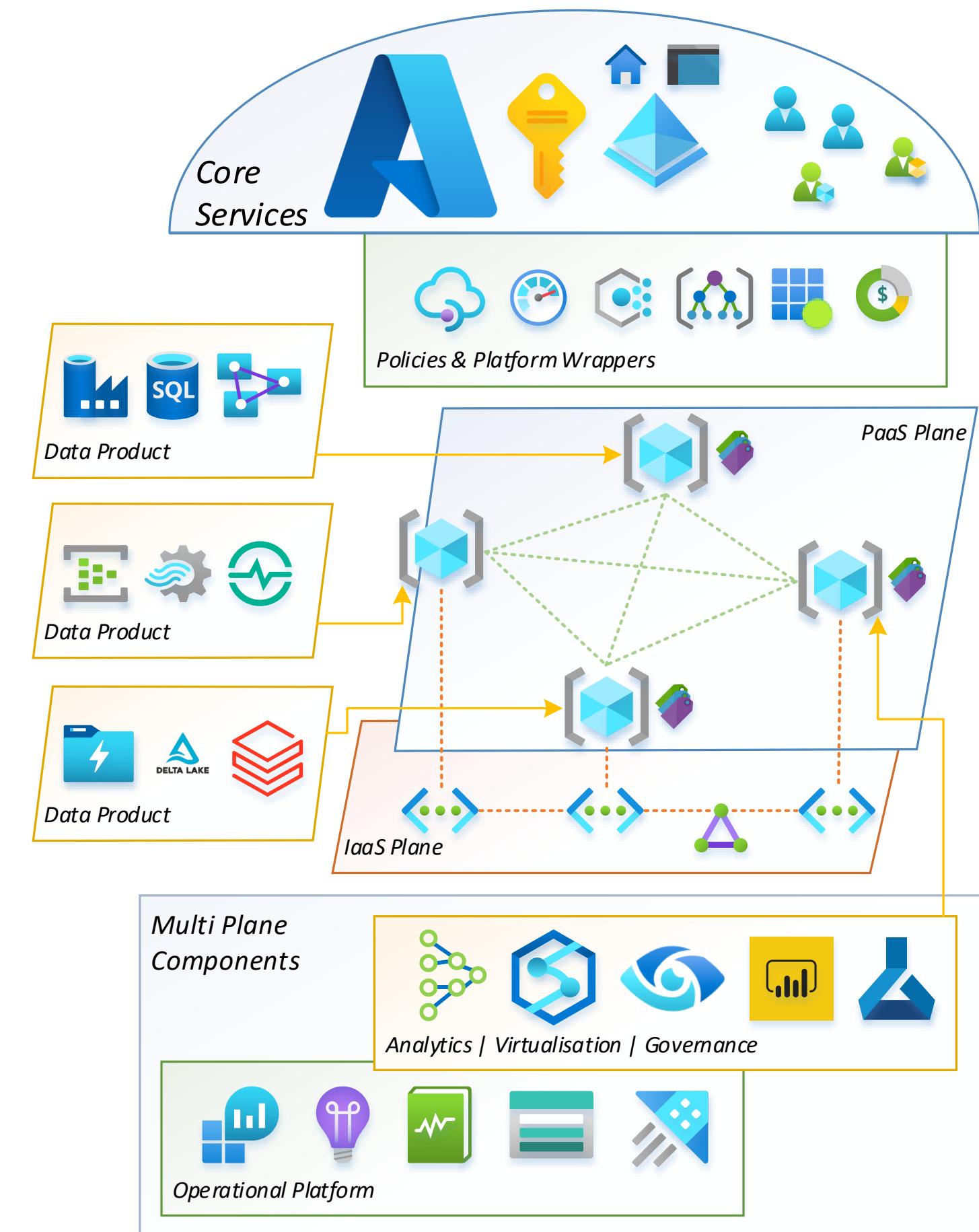
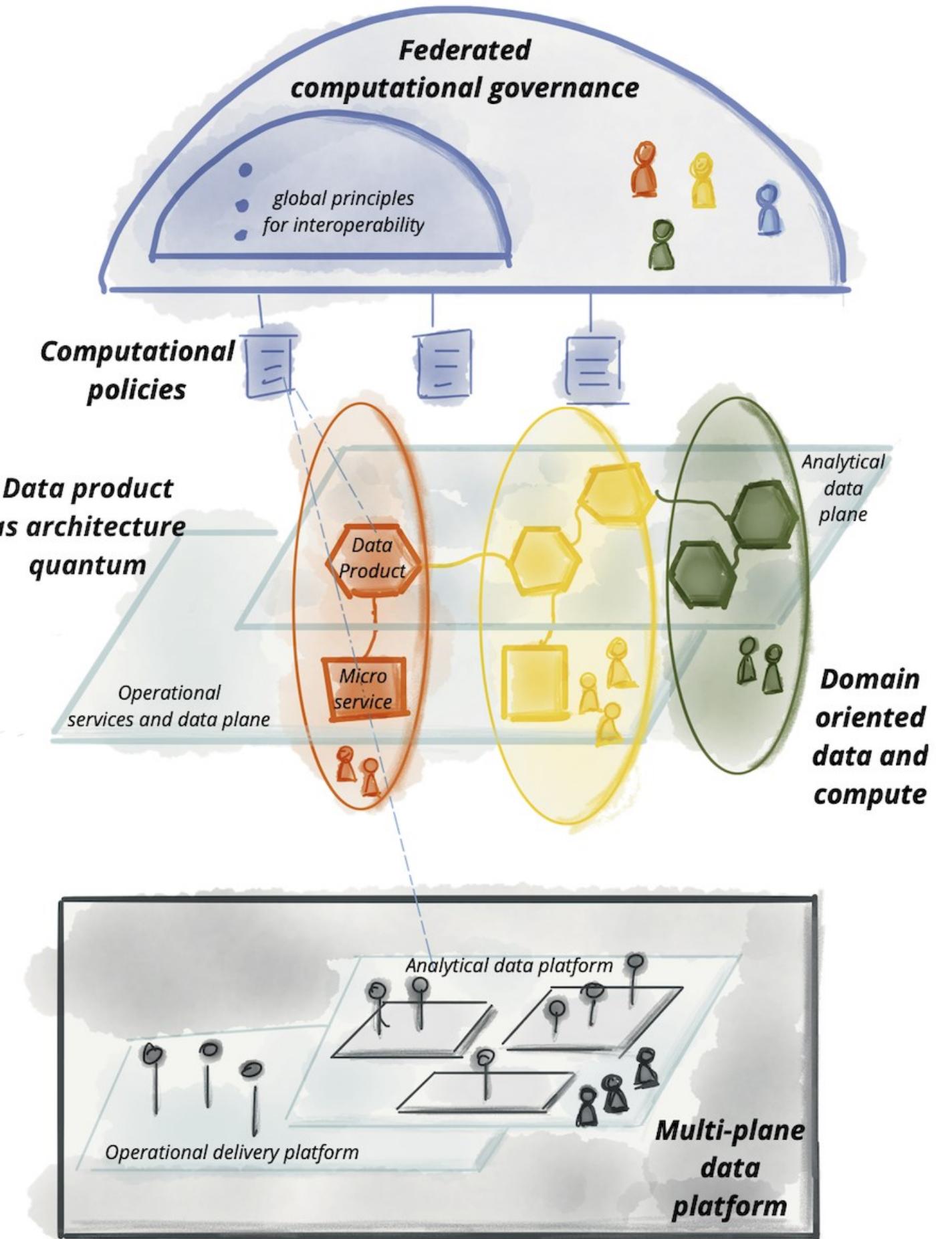


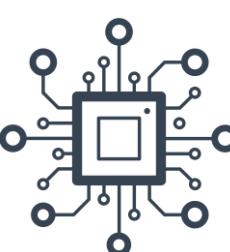
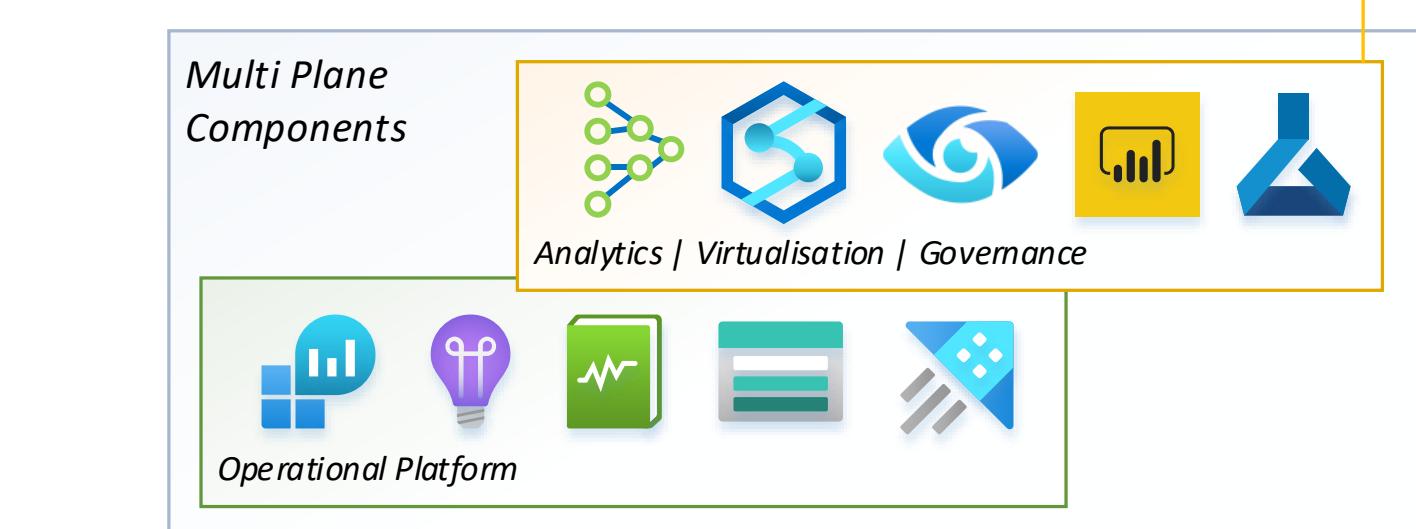
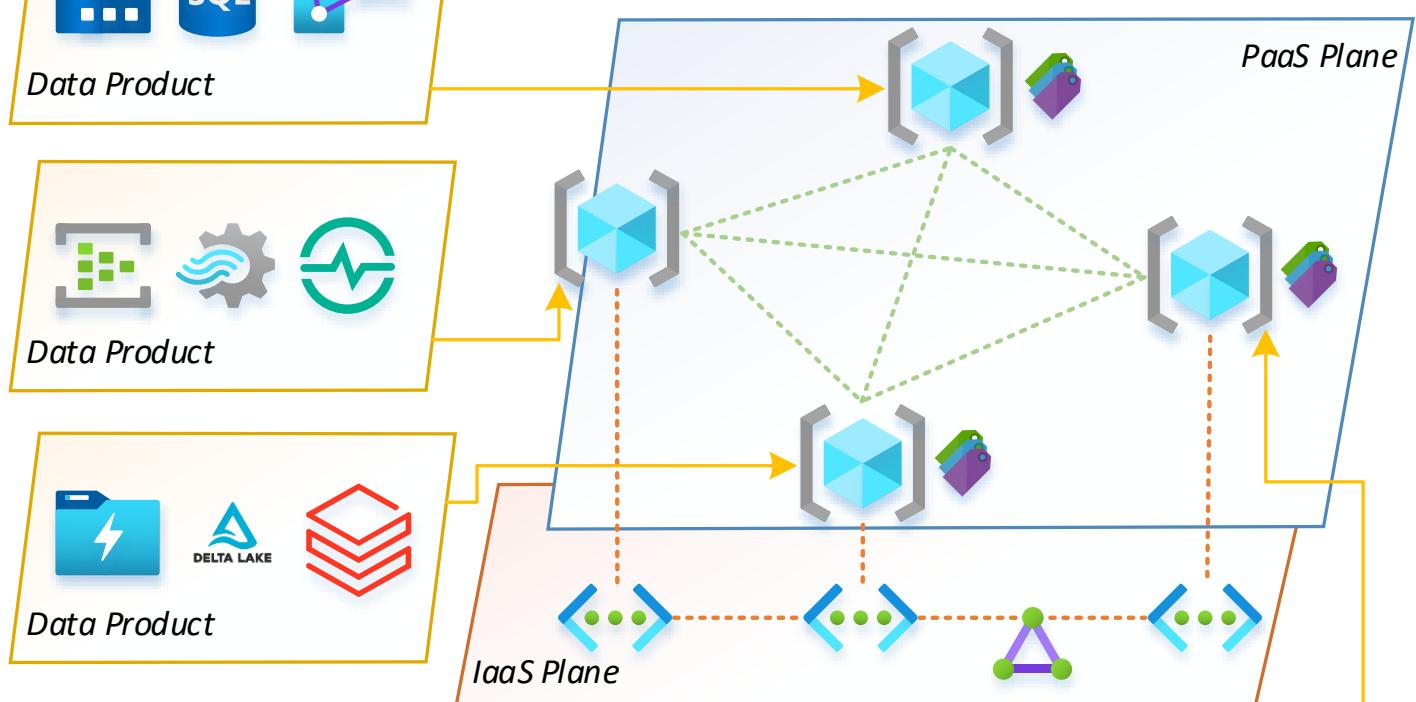


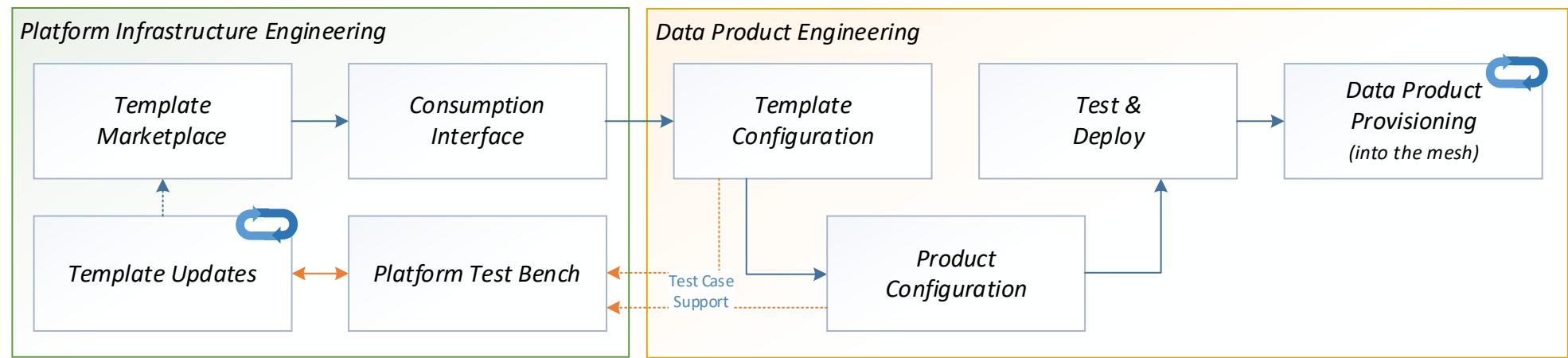
Data Domains in Azure



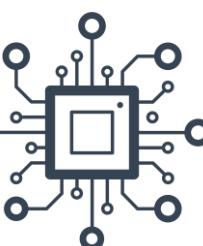
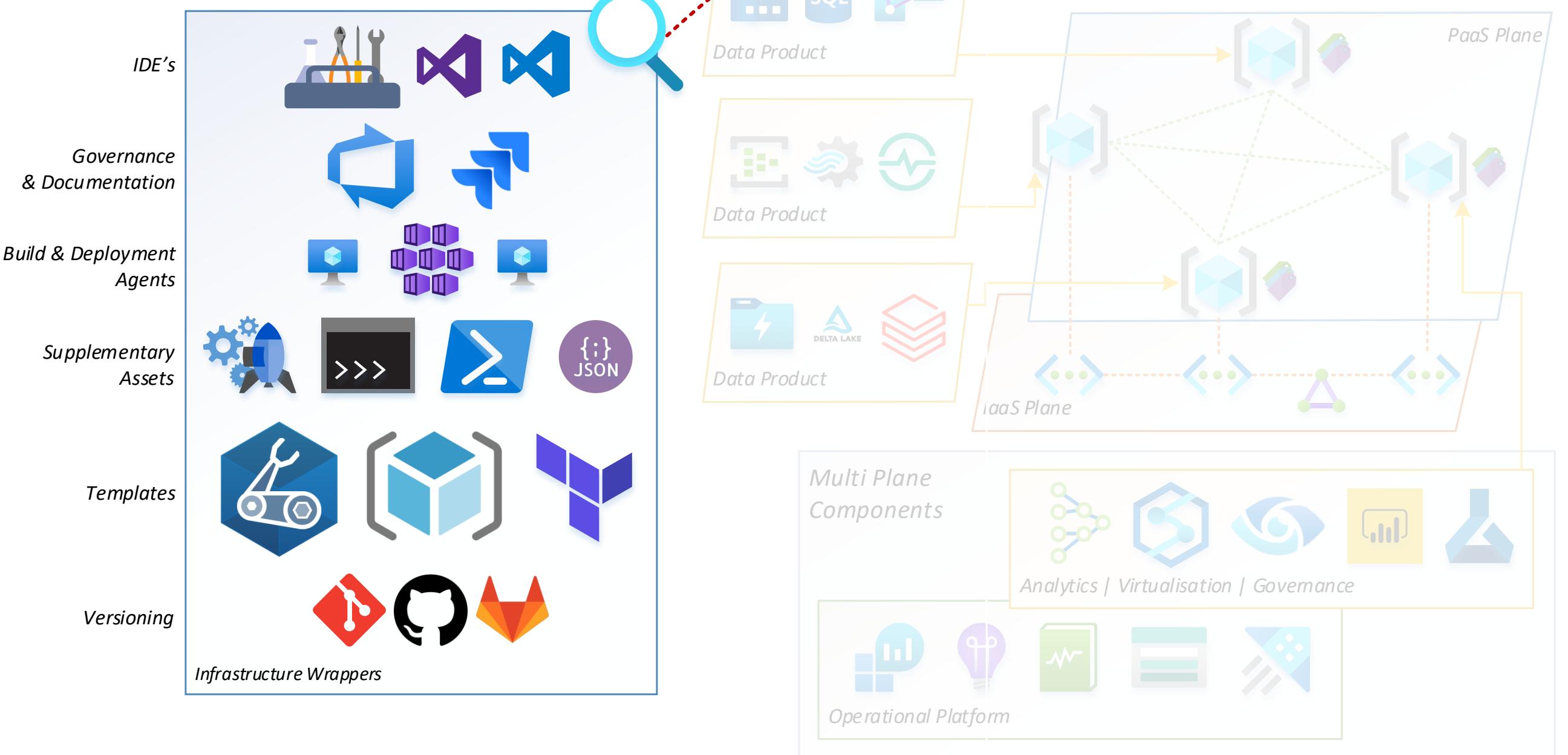


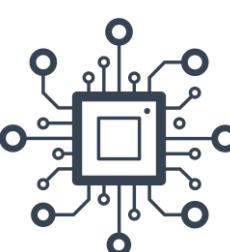
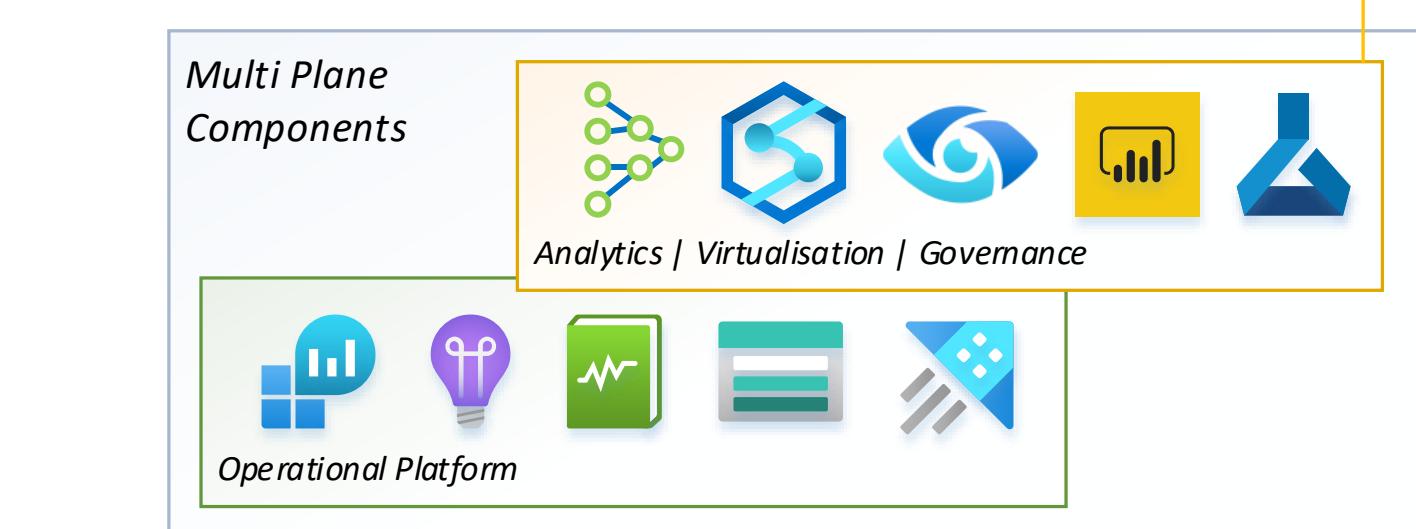
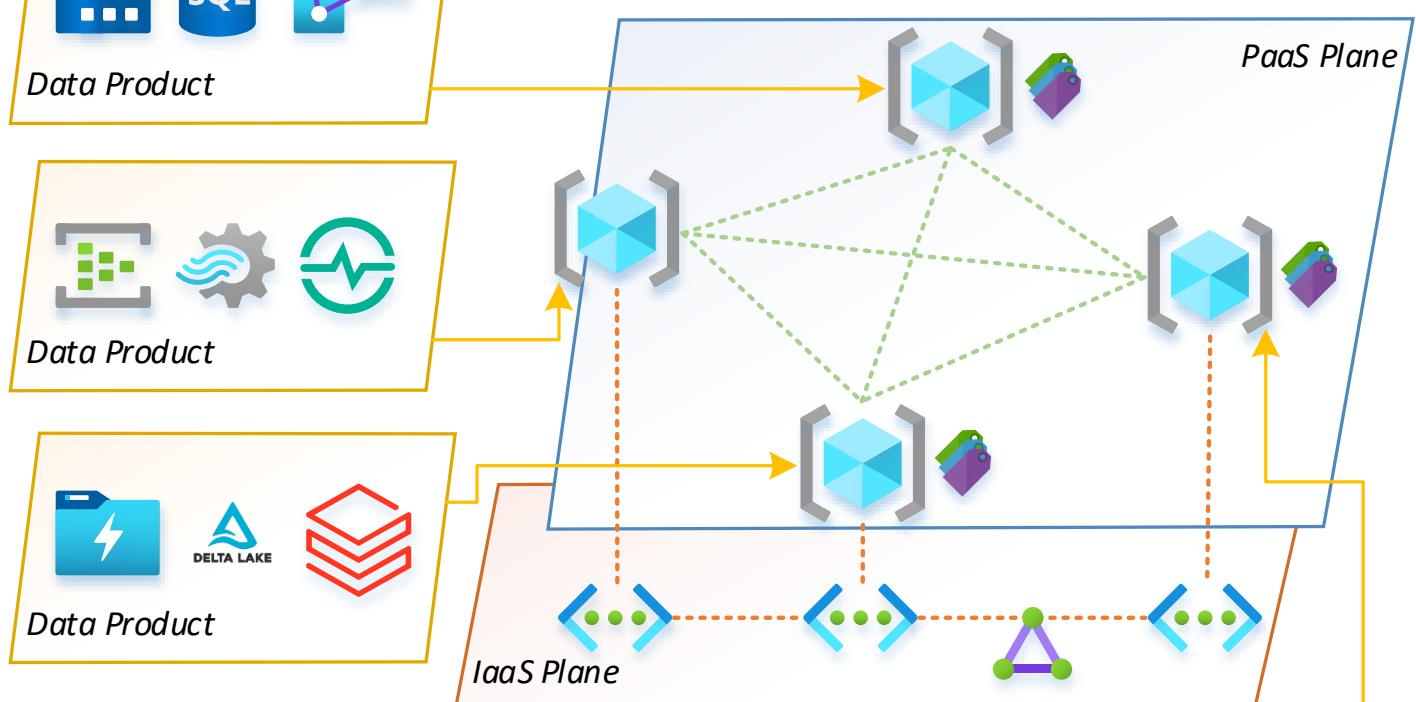




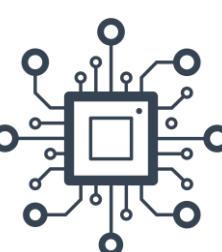
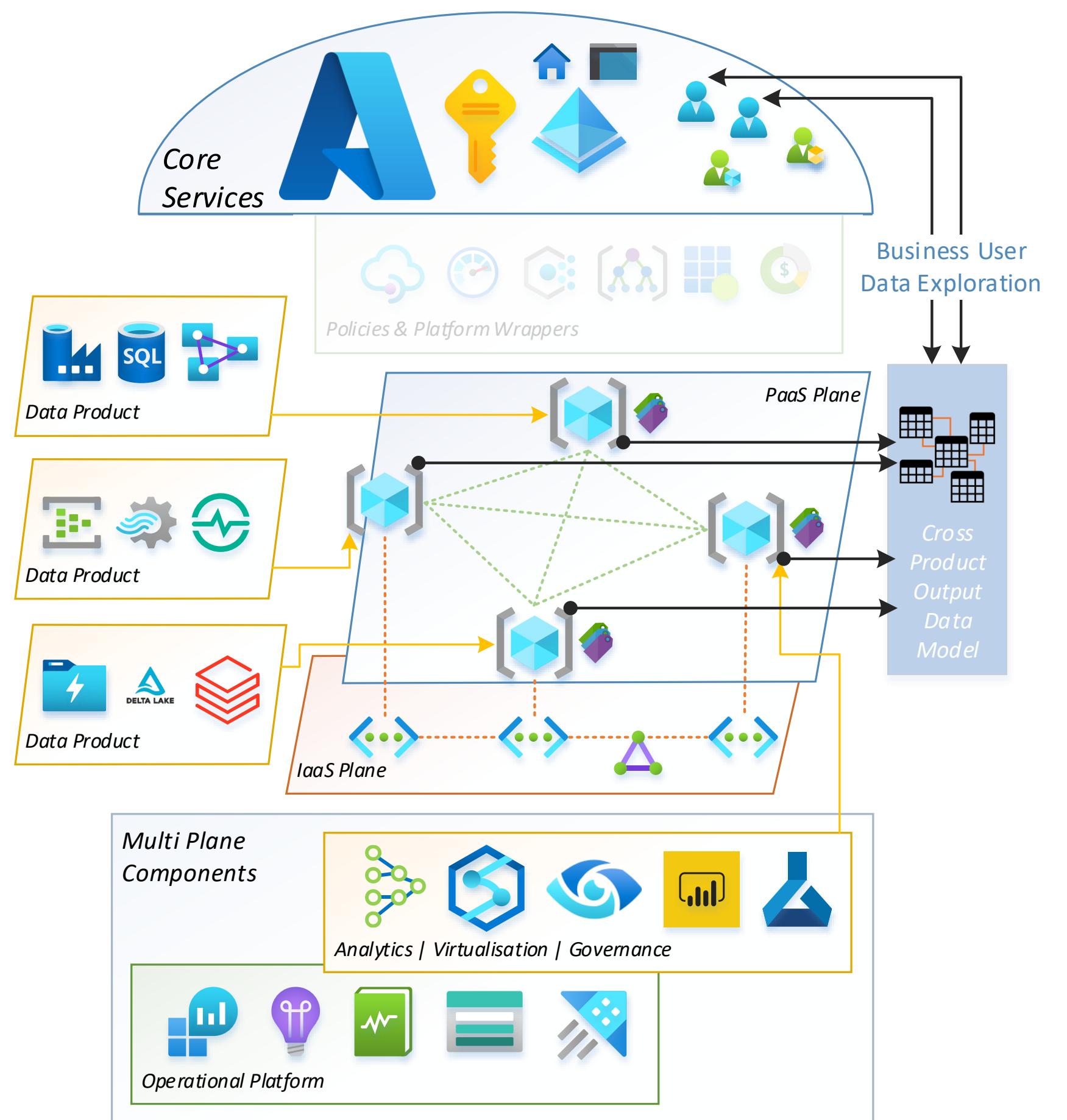


Self-serve data infrastructure as a platform.

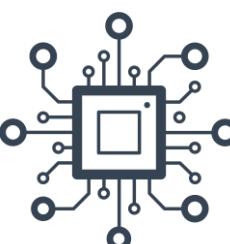
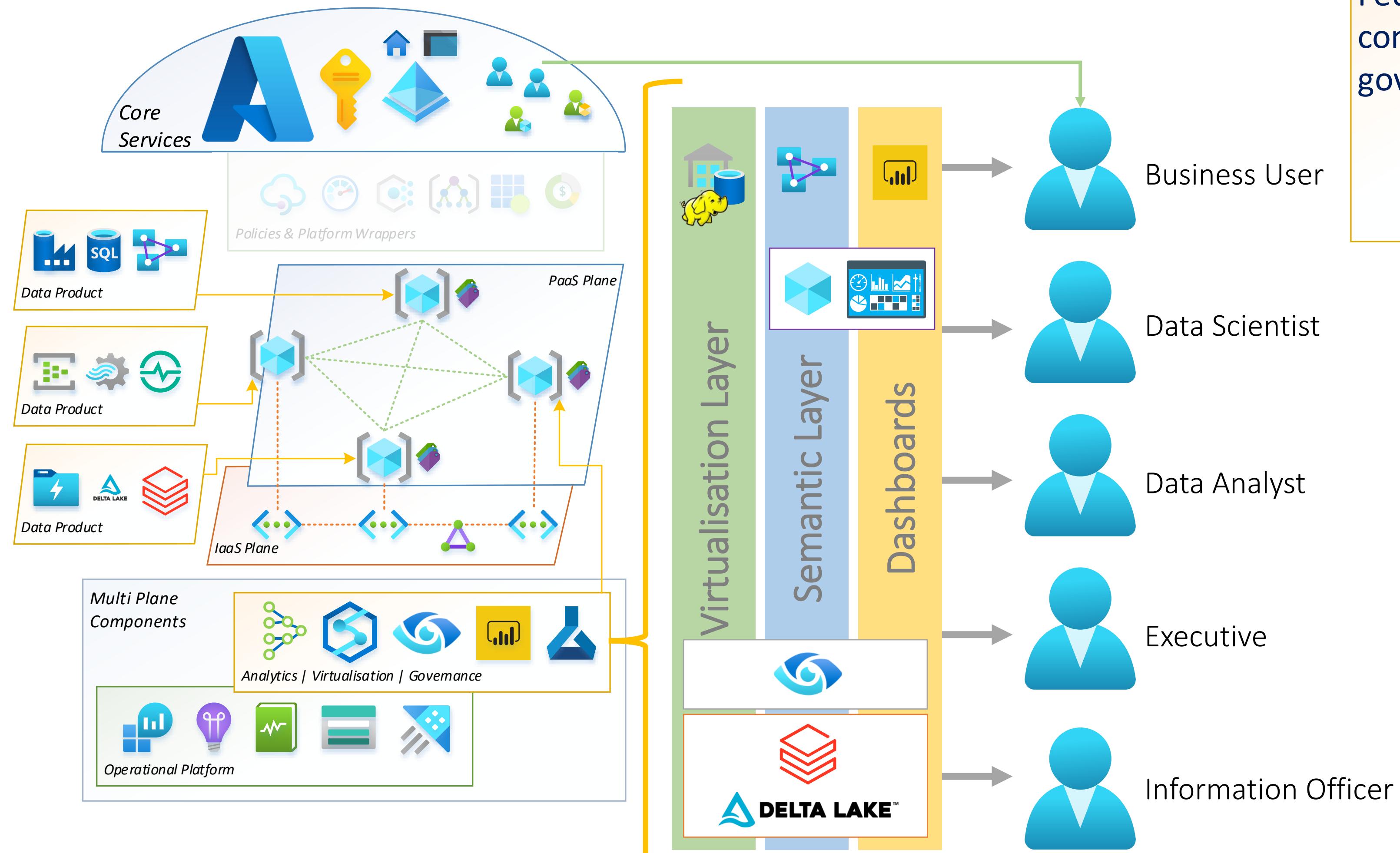




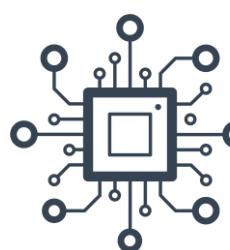
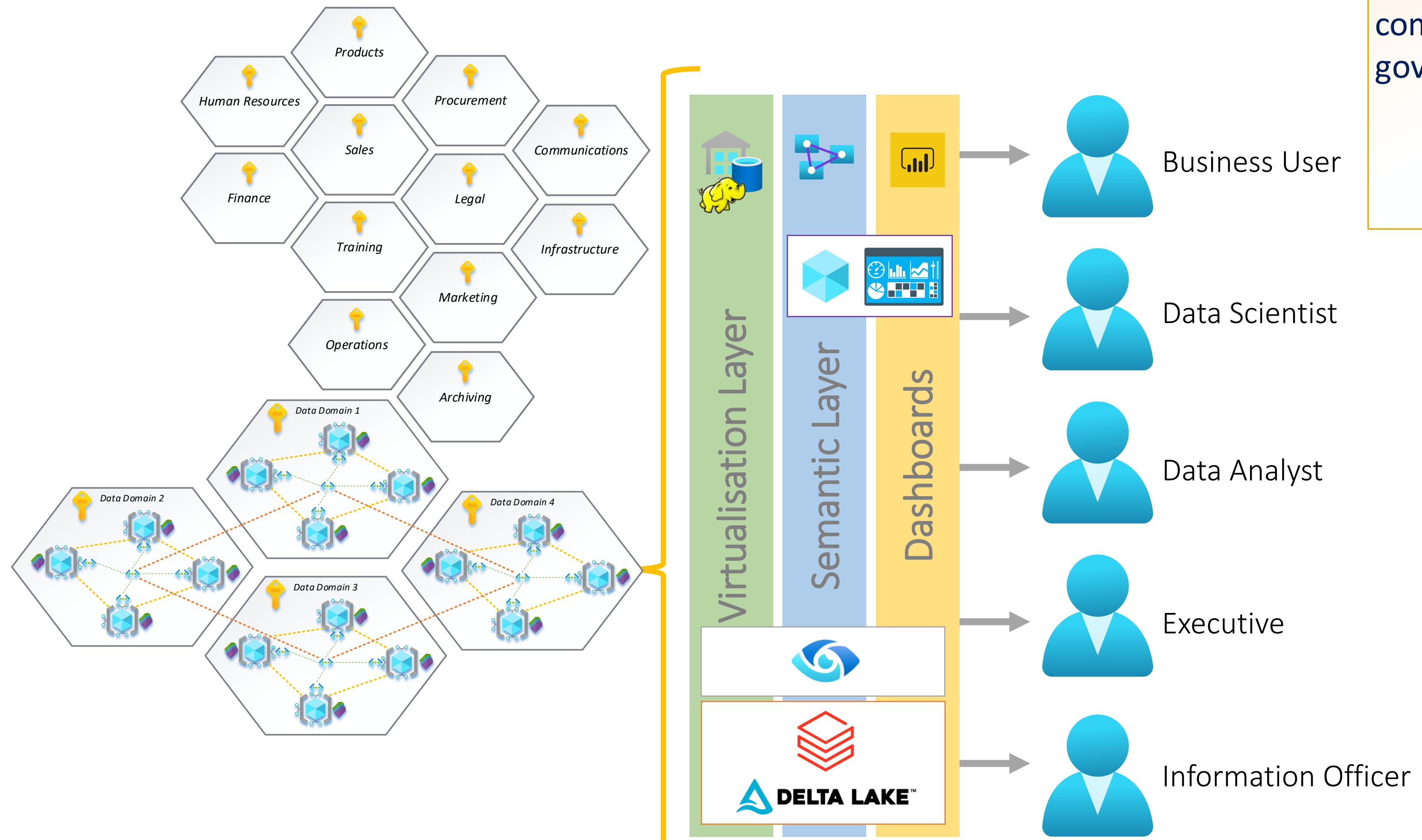
Federated computational governance.

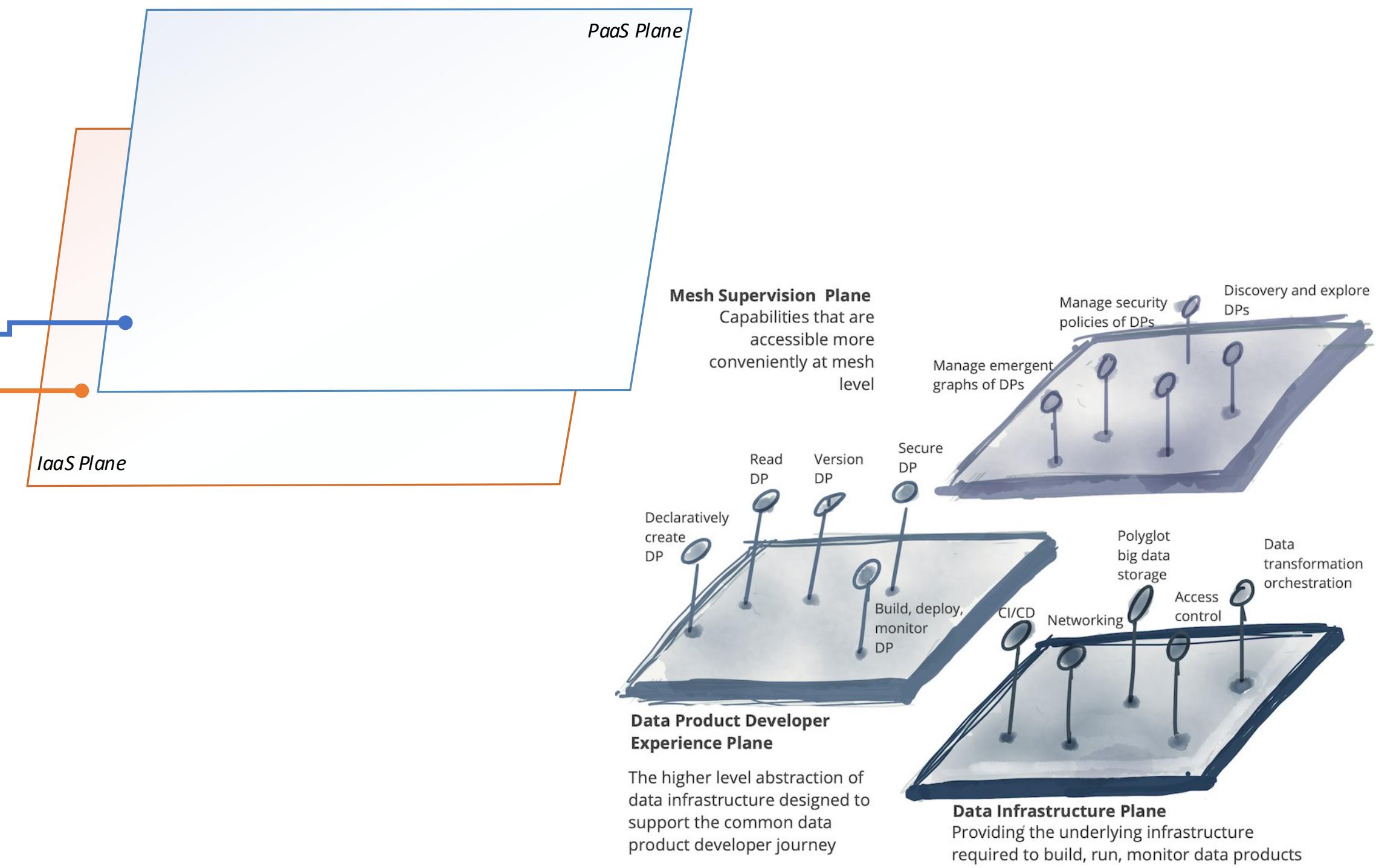
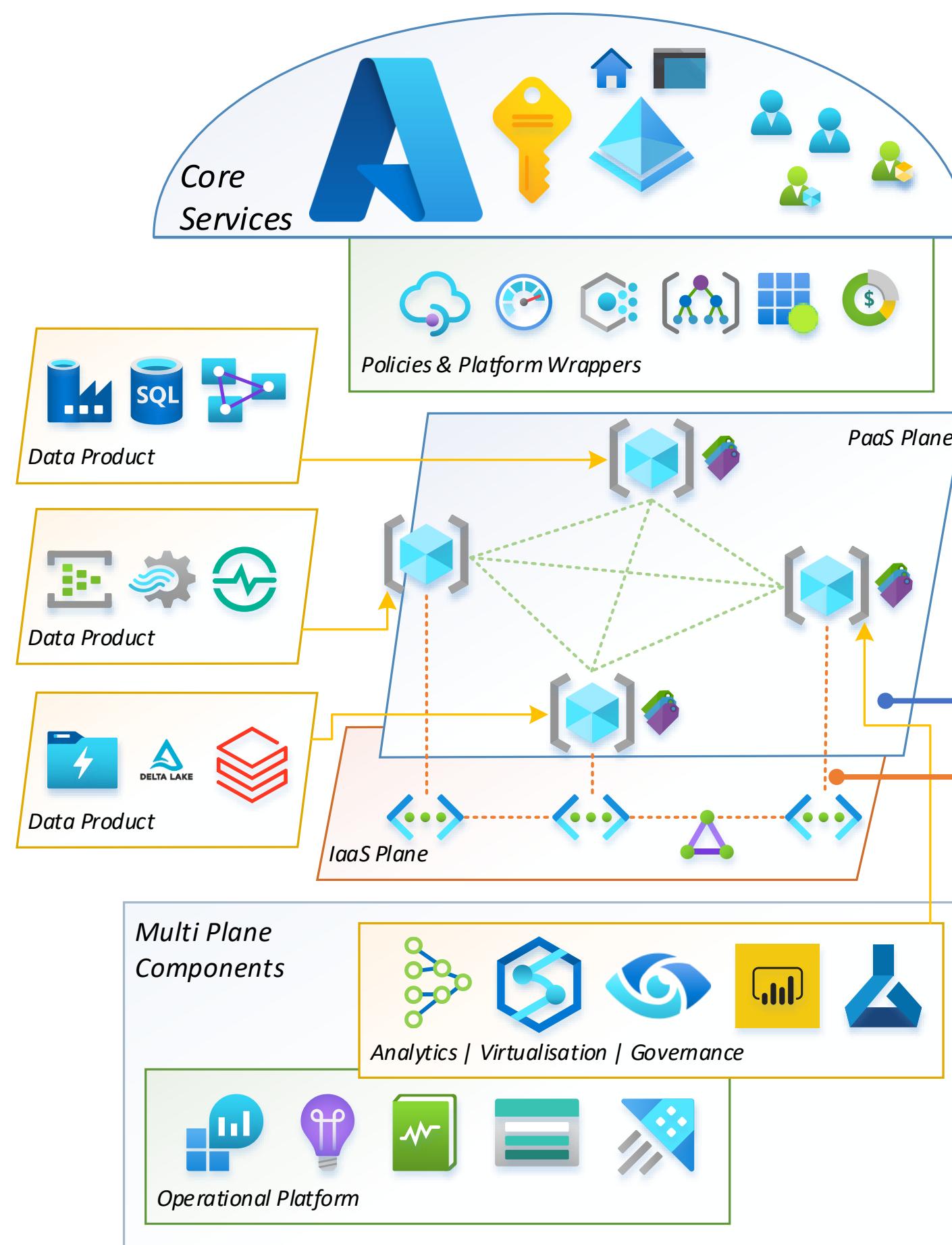
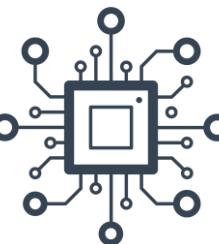


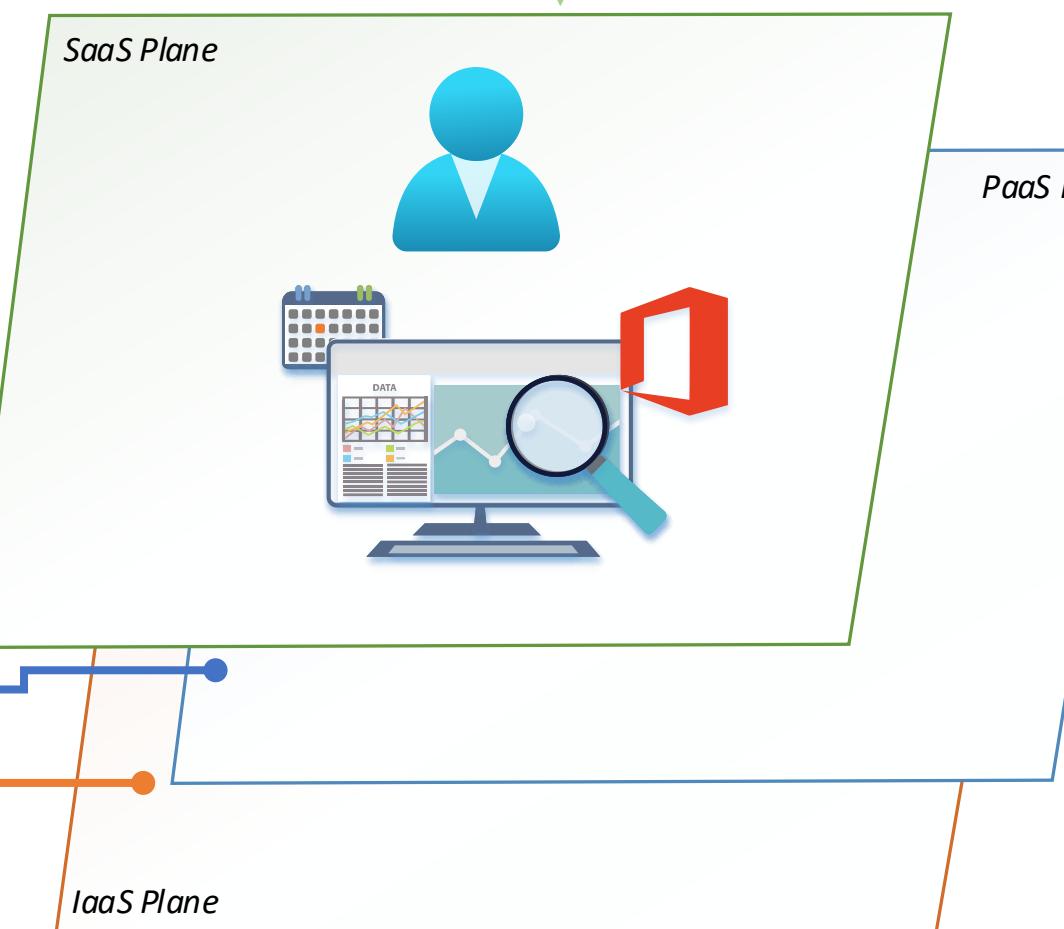
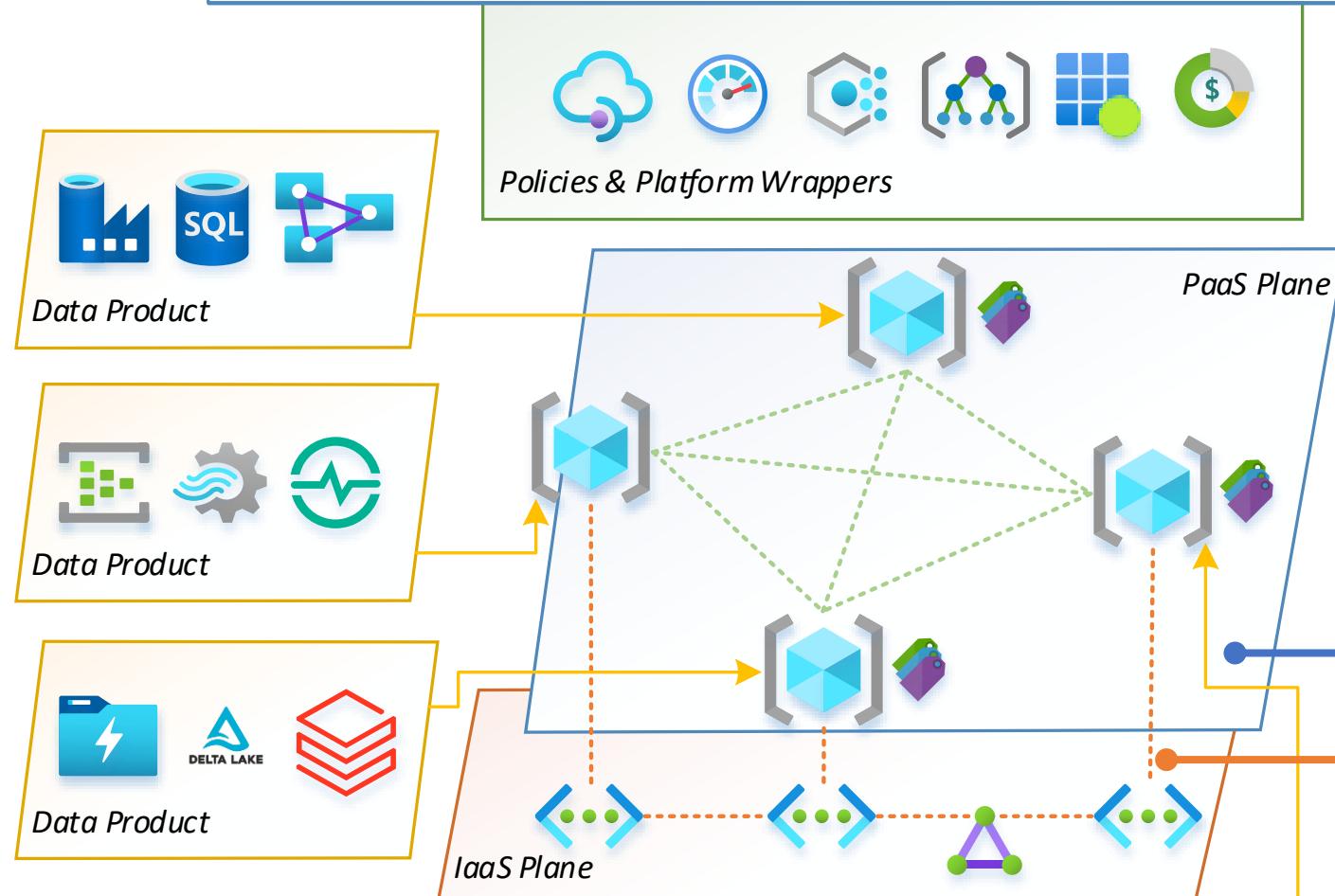
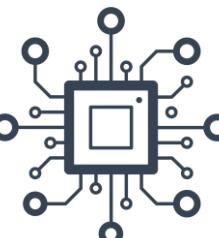
Federated computational governance.



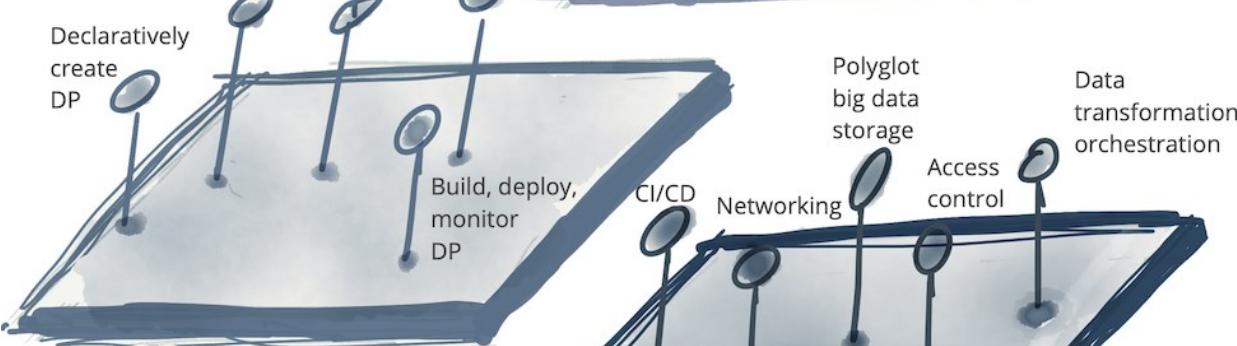
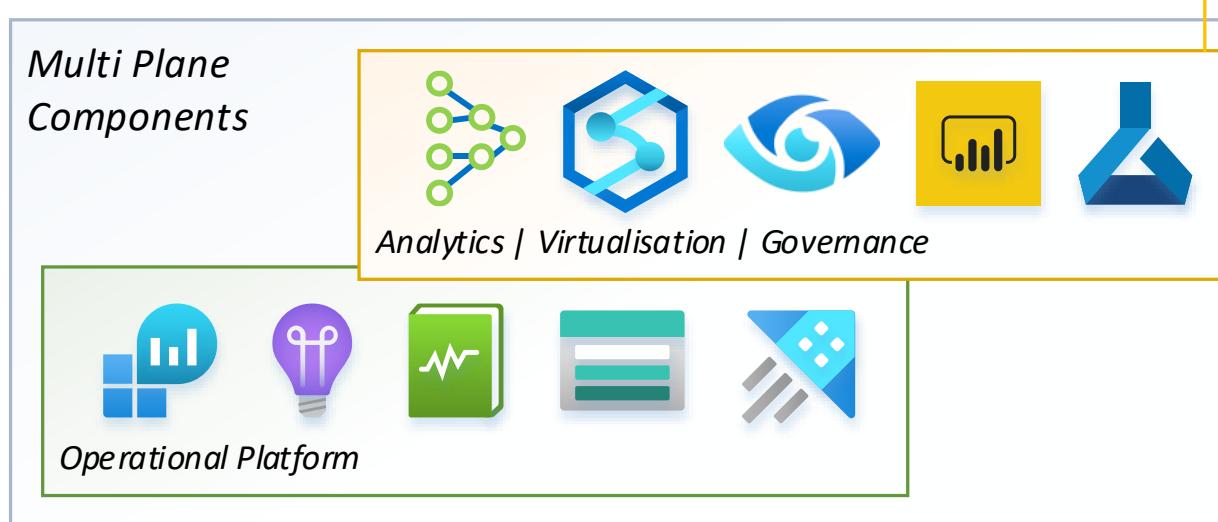
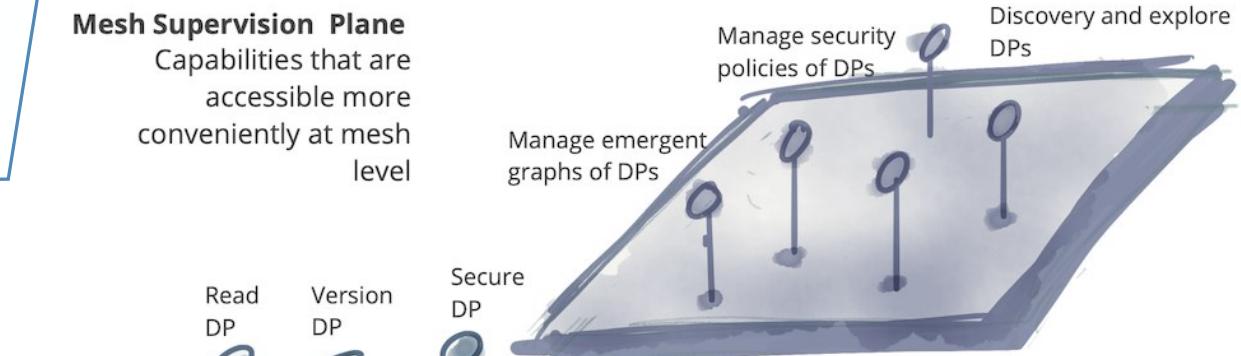
Federated computational governance.







Mesh Supervision Plane
Capabilities that are accessible more conveniently at mesh level



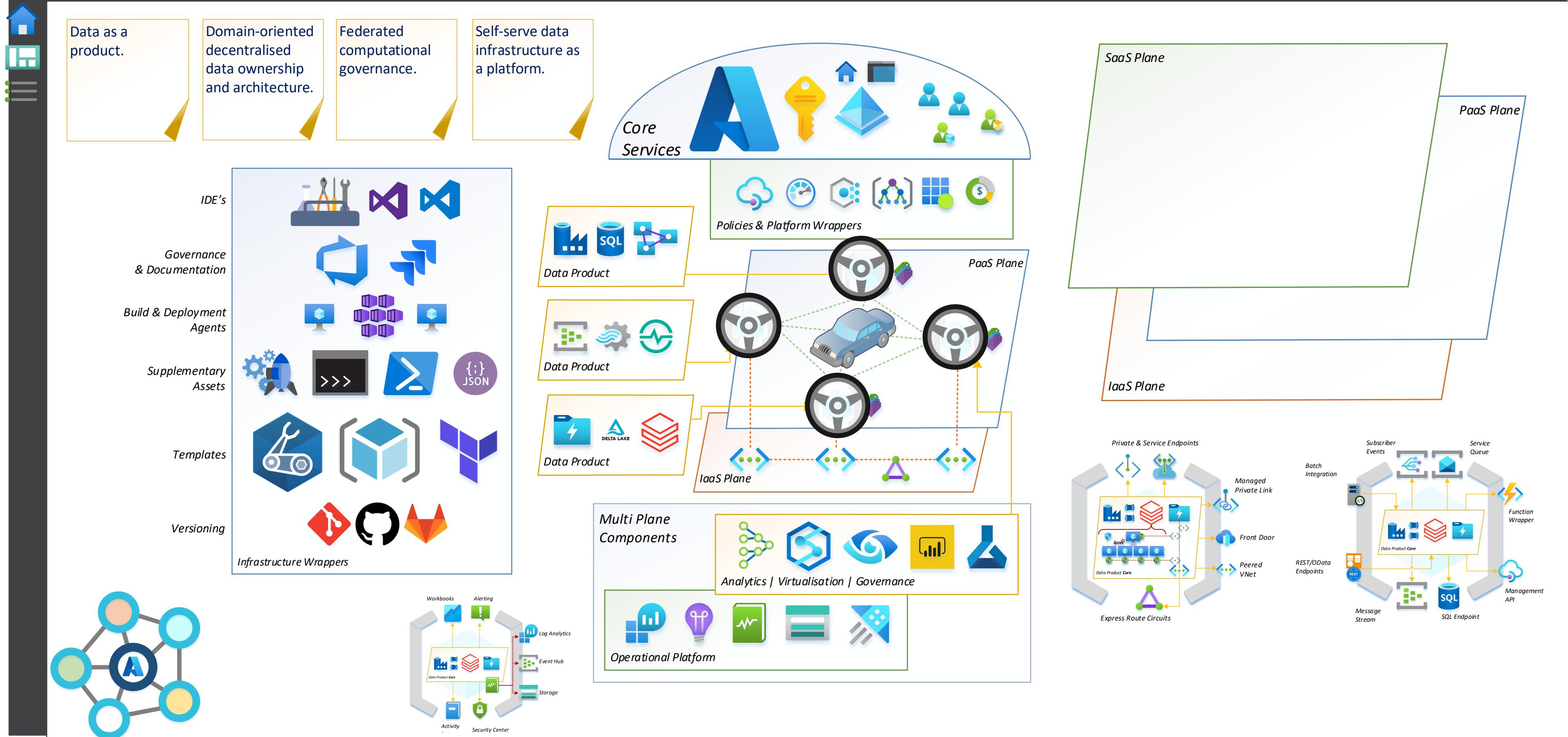
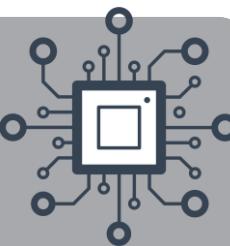
Data Product Developer Experience Plane

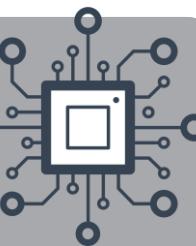
The higher level abstraction of data infrastructure designed to support the common data product developer journey

Data Infrastructure Plane
Providing the underlying infrastructure required to build, run, monitor data products



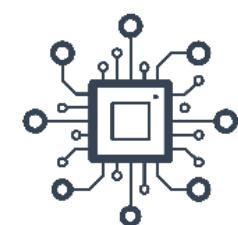
What does our minimum viable Data Mesh contain?





Thank you for listening...

Paul Andrew



Mr Paul Andrew
Consulting Ltd

Blog: mrpaulandrew.com

YouTube: [c/mrpaulandrew](https://www.youtube.com/c/mrpaulandrew)

Email: paul@mrpaulandrew.com

Twitter: [@mrpaulandrew](https://twitter.com/mrpaulandrew)

LinkedIn: [In/mrpaulandrew](https://www.linkedin.com/in/mrpaulandrew)

Github: github.com/mrpaulandrew

