

Business Insights as a Service Cultivating Data Citizenship

Introduction

26 February 2020



About the presenter

Certified Data Management Professional (CDMP) from DAMA International with over 25 years experience in Data Management (Data modeling, data architecture, master data management, data quality, metadata management, business intelligence and Data security and Operations).

Industry experience in the following industries:

- Retail – Vendor Master and Rebates Systems - (Massmart group as Data modeler and Solutions architect)
- Local Government – Land Information Systems, Property Value chain (Johannesburg Metro, City of Cape Town and Nelson Mandela Bay as Data Architect)
- Telecomms – Converged Billing, Prepaid and Postpaid billing and CRM (MTN as Enterprise Data Architect)
- Petro/Chemical – Owner of Analytics Ecosystem, architecture and data engineering (Sasol as Snr. Manager of Data Architecture in Data & Analytics)
- Mining – Data Architecture and Building out Analytics platform (DeBeers as Data Architect in Data & Analytics)

- Have spoken twice at DAMA International conferences in the states, published article in Position IT on Why GIS is a tool for Enterprise Information Management in 2010
- Won the Teradata Best Innovation for Cloud Analytics award in 2018.

And...

In my spare time I help people through horses (Equine Assisted Thereapy) and have started an Equine Theatre practice at my stables.



Contents



INTRODUCTION: THE NEED
FOR DATA ANALYTICS



CHANGING WAYS OF
WORK: DATA COE AND
DATA CITIZENS



ENABLING ARCHITECTURE
– END STATE



Q & A

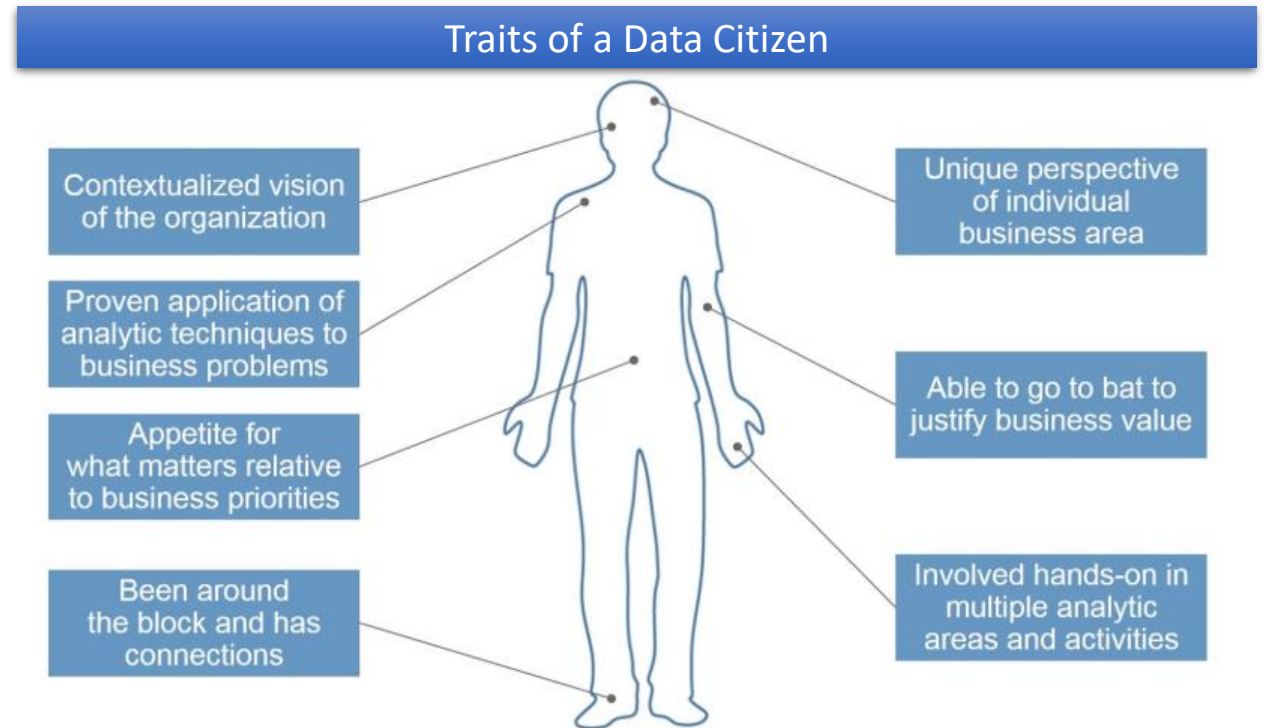
Empowering our business with information to do things better

The traditional way of doing Business Intelligence with a Centralised delivery model that only delivers on a specification is no longer good enough to respond to the increased business demand

Whilst there will always be a need for Management Reports and dashboards, business users require a more “Insights as a Service” model and the ability to discover and apply their extensive business knowledge on trusted data

What we need to do:

- Create a Data and Analytics Services capability
- Create analytical data platforms that enable business easy access to information and easy methods enabling them to build their own insights
- Cultivate “Data Citizenship” in our business



Data Analytics



Why Data Analytics?

- Having large volumes of information accessible results in nearly **unlimited potential for data discovery and business/operational insight.**
- Currently, data lives in disparate legacy systems of record. To enable data discovery and operational insight, a new **analytic ecosystem architecture** is required where data is aligned and structured into a well governed enterprise data model. The enterprise can then explore this data and convert it into data products. These range from calculated insights to graphs and algorithms. **This context can then be used by decision-makers to minimize operational risks, take action with the aim of enhancing business gain and/or exponentially change the way we do business.**

A journey to value

The purpose of the Data Analytics Team



1. To build a scalable and sustainable data platform that is **cost effective** and **easily maintainable** to support the vast varieties and formats of data.



2. Prepare this data for analytical consumption to **enable business with a consistent and reliable path to faster and accurate insights.**

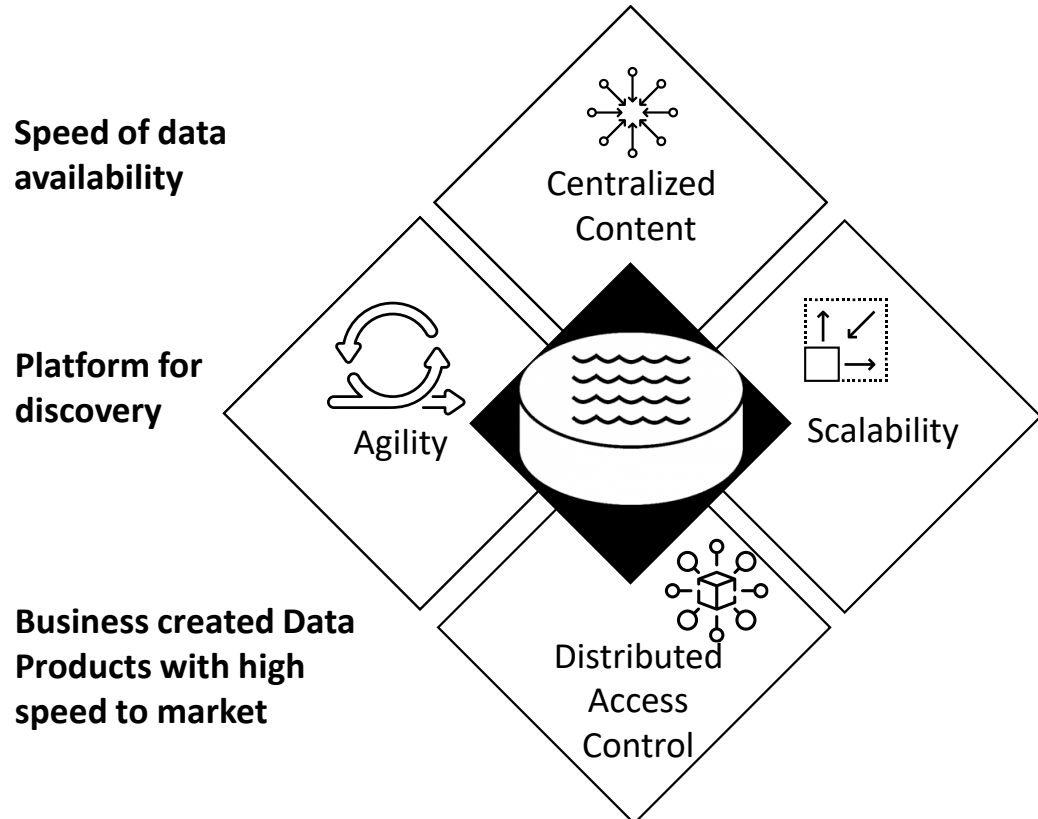
The role of the data lake and the data warehouse

Data Lake

The data lake is a **multipurpose storage layer**.

The data lakes economically store large amounts of structured or unstructured data (images, audio, video)

- Data scientists and data citizens use the lake for **discovery and ideation**
- Highly **agile environment** for data generalists / programmers that tap the stream data for real time analytics
- The lake can serve as a staging area for the data warehouse

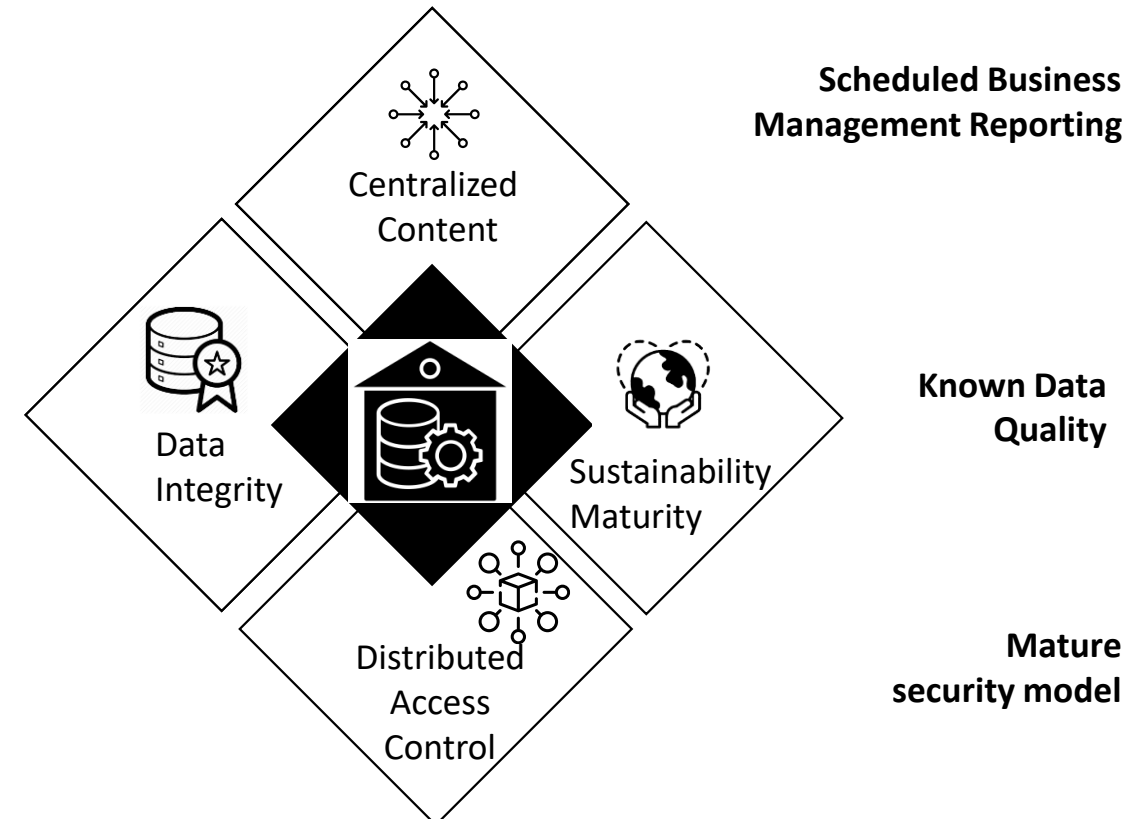


Data Warehouse

The data warehouse is a **trusted storage layer from where all scheduled reporting is done out of**.

The data warehouse can store large amounts of structured data enabling invaluable business/ operations intelligence with the following characteristics:

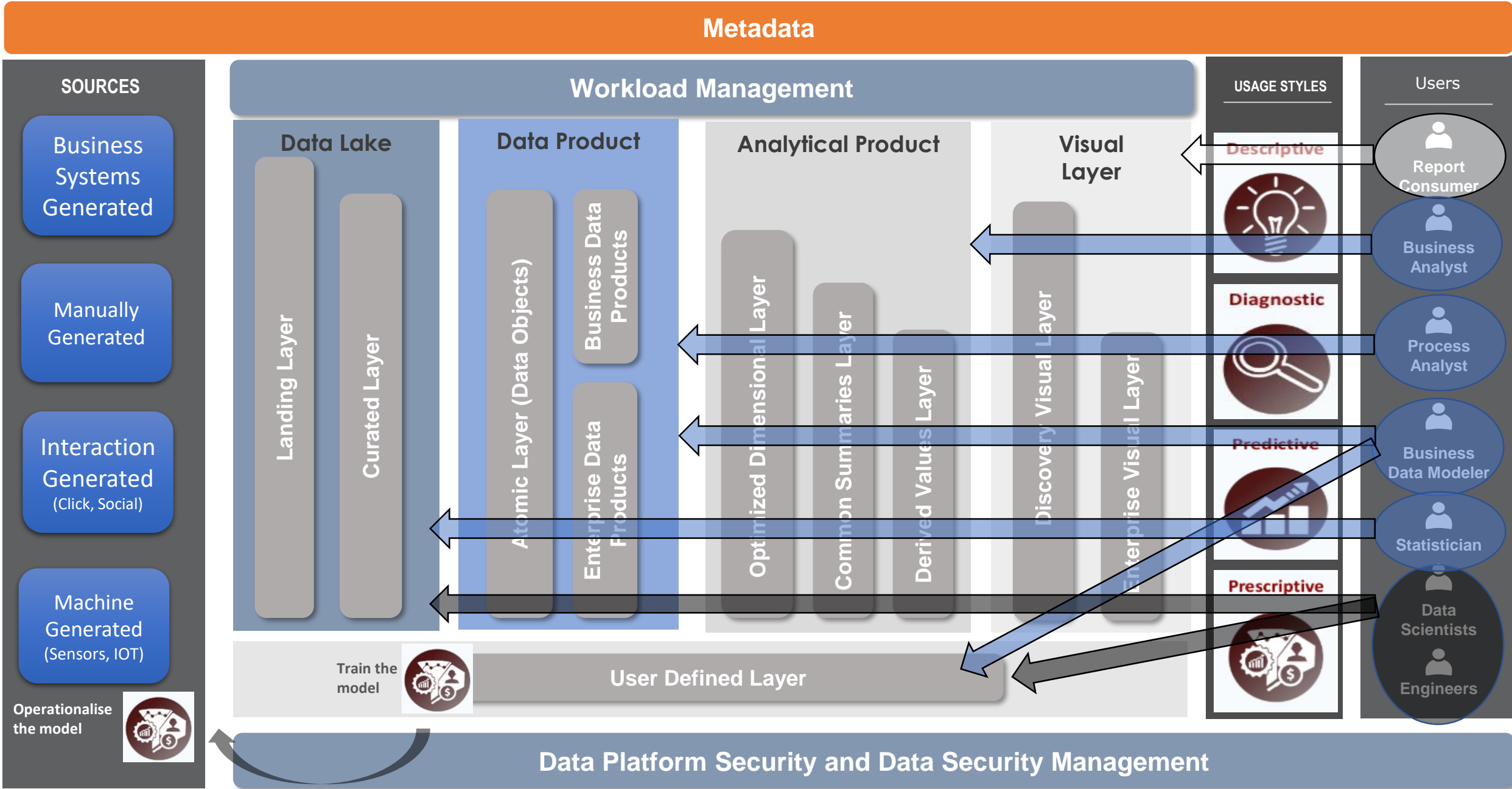
- Caters for business professionals providing **management, statutory and regulatory reporting**
- It has a fixed configuration resulting in lower speeds for change management, but in turn assuring data quality and data integrity
- It is the location of more carefully “treated” data for reporting and analysis in batch mode



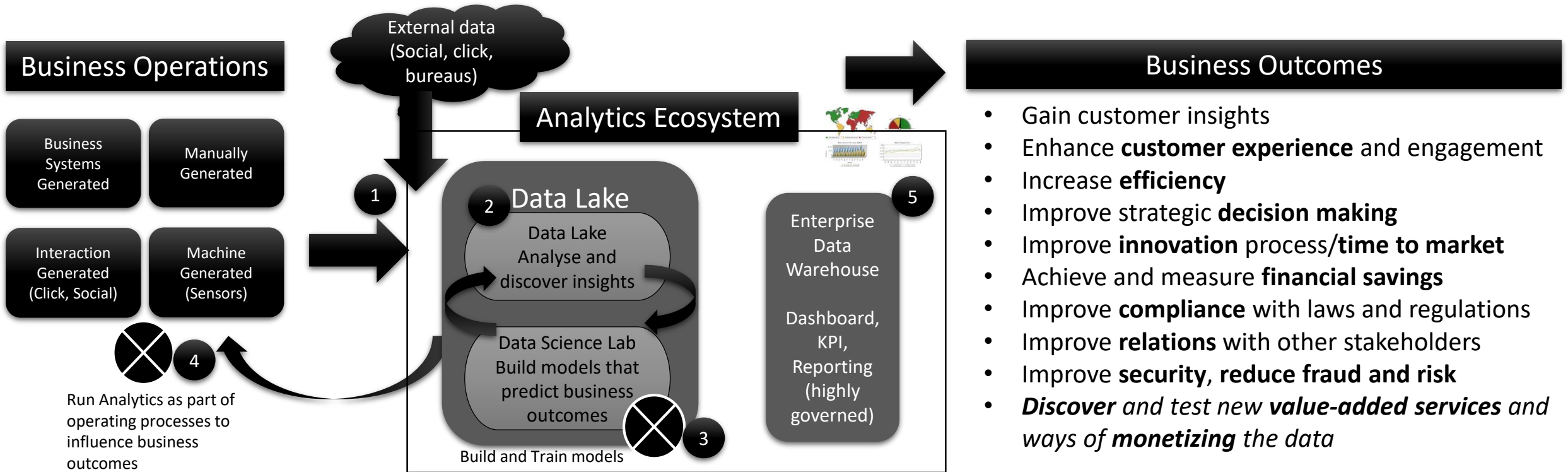
Enabling and End-state architecture



Conceptual Data Architecture for Analytics – Catering for various users



From data sets / assets to data products and insights



No.	Key steps to build end state
1	Choose the right data – source data creatively, quickly, external data with IT enablement providing self provisioning and secure accessibility to data from ops systems
2	Environment and tools to enable business to discover their data and easily build insights.
3	Co-creation of models that predict and optimize business outcomes in an easy to use environment enabling quick to market deployment processes
4	Embedment of analytical and predictive models into the business operational processes
5	Provision of environment that still provides the regulatory and management reporting consistently with integrity

End State Analytical Architecture ...Where we want to go...

Space/Time Analytics data preparation framework

Data sources

External Weather data

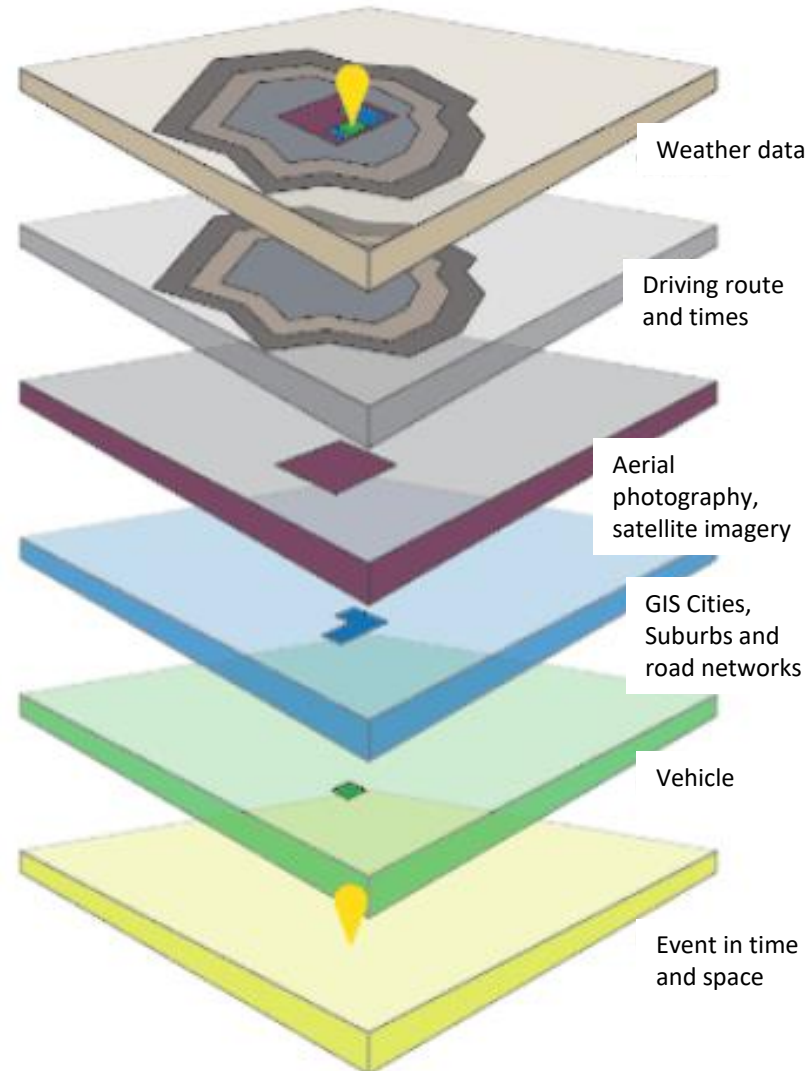
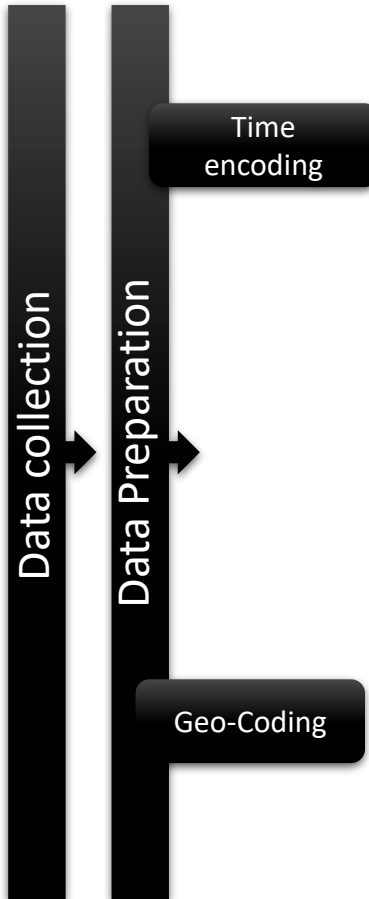
Trucks, drive telematics data

Photo, Video and Audio media files

Cities, suburbs and road networks

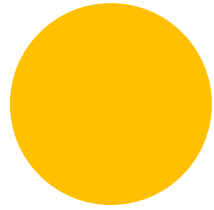
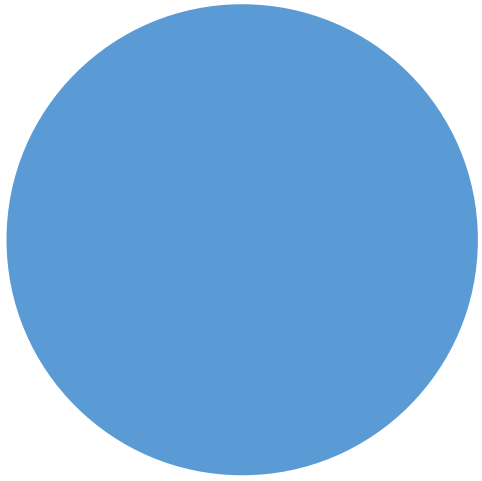
Vehicle Information

Telematics Data, signal or call data



Benefits Enablement

- Time and Location dimensions are enabled in a Spatial / Time context
- Users use time and space relationships to derive insights
- Users load their unstructured data (audio, video) which are all space and time referenced as layers in a Self-Service environment
- AI and machine learning models can be built to find trends
- Little / or no data wrangling required as data is loaded through data preparation framework and is time and geo-referenced for analytics



Data COE and Data Citizen Roles

New ways of work

Data citizen roles vs Data COE roles

Data Citizen role

Data COE role

Report Consumer

Requires trusted reports, dashboards, KPIs defined periodically

Business Analyst

Requires ability to slice and dice and look for information in a known and defined Semantic layer

Process Analyst

An analyst (including process and data analysts) which helps the enterprise understand and improve its processes and operations. Process Analysts aim to build/propose designs to improve quality, reduce errors and increase business value

Business Data modeler

Requires ability discover new insights without having a requirement, needs to be able load own data and create new data products using user friendly tools like Excel or PowerBI and quickly deploy new Data Products for wider business use

Statistician

Focuses on modelling and sampling of data in order to develop and validate mathematical and statistical theories and probabilities leveraging a wide array of data using user friendly tools that provide packaged methods and visualisation without needing to write code

Data Scientist

An analyst with a mathematics and programming background who combines data from across the analytic ecosystem with other interesting "external" data. Combined data is then used to develop models, (learning) algorithms and other advanced visualisations in order to optimise business processes, uncover relationships and establish datasets to be utilised via other forms of analysis.

Data Engineer

Creator of data products, beginning with an ingest process to source data into the Data Lake, will take project requirements and turn them into data products, with the required levels of provenance as per specification

Data Owner

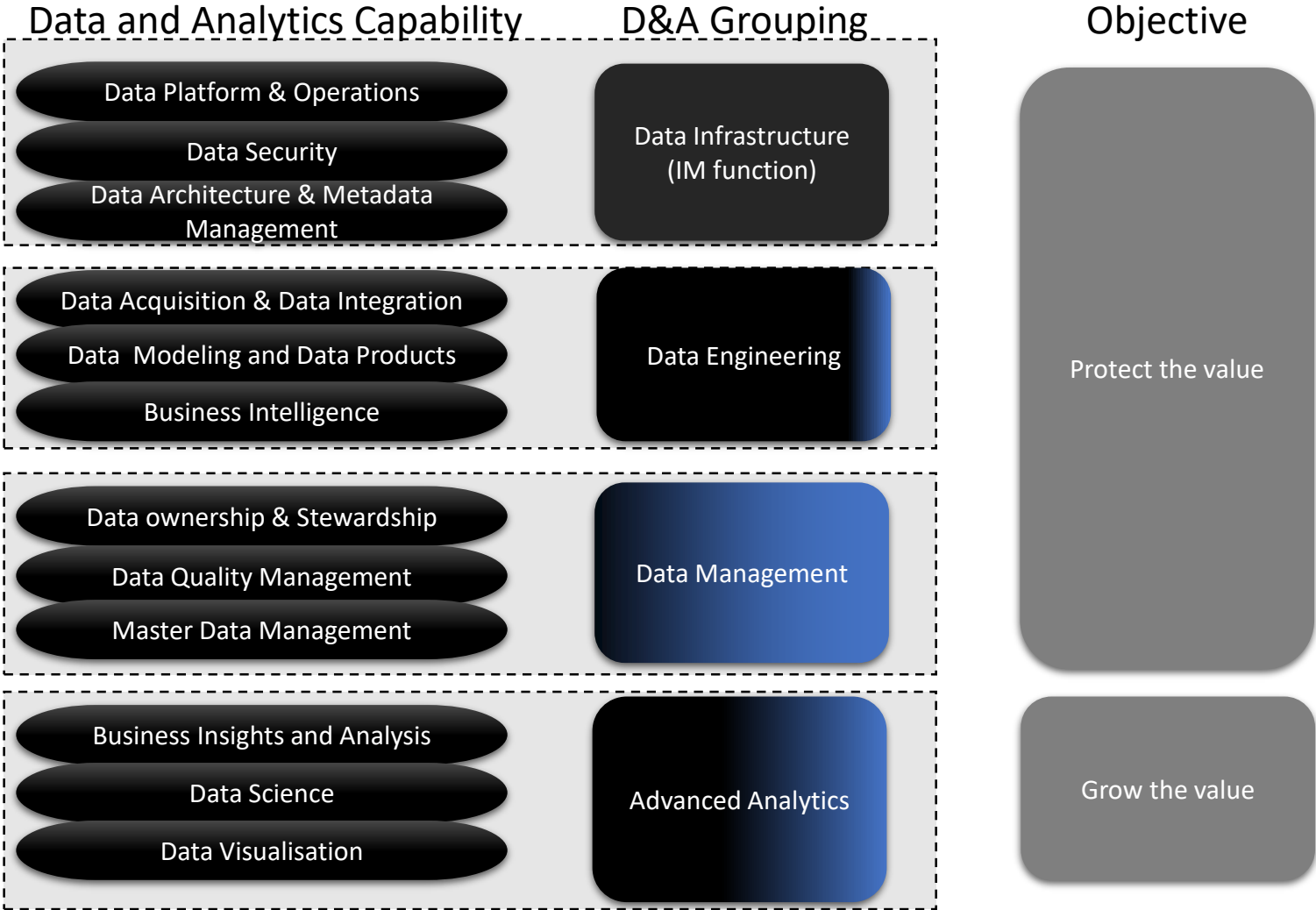
Owns the business process and the data created by these processes. Has the mandate to change the business processes and the data rules. Is accountable for the data management of the data assets within the subject area allocated (this includes data retention, regulatory and data privacy compliance)

Data Stewards

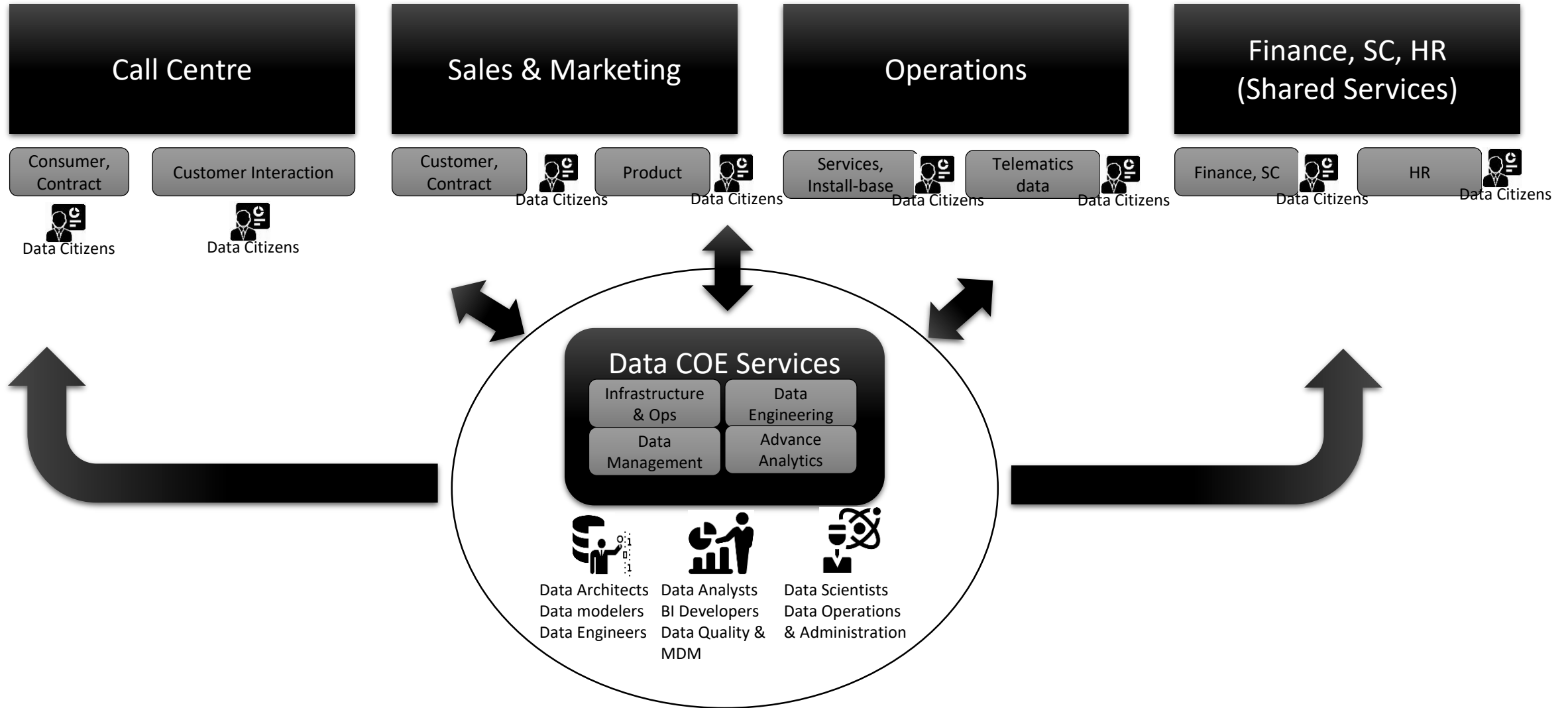
Is responsible for the management and implementation of the data rules defined by the data owner. Is responsible for the data quality management and master data and defining of how the data should be used through out the organisation under direction of the data owner

Data COE Services – List of Data Capabilities

- In order to build a healthy Data services capability, and referencing the Data Management Body of Knowledge we can define the following 12 key data capabilities
- These capabilities shown are classified into 4 main areas as they possibly fit into different areas of the organisation and have different objectives
- There are some of the Groupings which require business involvement and co-creation, this is highlighted using a blue gradient – the blue gradient indicates more business co-creation/collaboration required
- Initially the Data COE will drive most of the data capability services, but as business maturity improves more business participation will be implemented



Channel and Hub approach to deliver Data Services model approach



Q & A



- An implementation of such enabling systems and practices require executive sponsorship and involvement
- It will impact on current ways of work
- It is a program rather than a Project which will require business involvement to support these changes both from the enabling architectures as well as to embed the practices and new processes / practices in business