McKinsey & Company

Future of Enterprise Data Management: Data Products and Knowledge Graphs

Knowledge Graph Conference 2021

Presenter: Mohammed Aaser, Chief Data Officer, McKinsey and Company

Key points

Organizations have seen considerable value from central data teams and data lakes, however **challenges remain in scaling use of data**

Data products represent the next wave of progress for the enterprise – defining explicit shared meaning and designed with scale and value in mind

Knowledge Graphs are an approach to develop these products and evidence is amassing on their benefits

Your organization can take steps to get started, namely identifying potential data products, building a business case, and an operating model

Where are we today?

- Data Lakes / Lakehouse

- ETL to ELT

- Virtualization and connectivity across clouds/premise

- Data governance efforts to improve quality

Have unlocked a number of use cases and has been very valuable

However, organizations still face challenges in scaling

>70%
of organizations have been either unable to scale or capture full potential of digital transformation efforts[1]

>80%
cite a lack of a data-driven culture[2]

>2/3
Identify complex IT environments[2]

1. How to restart your digital transformation, McKinsey.com, March 2020
2. McKinsey's sixth Global Data Survey, in the field from February to September 2019. >70 responses.

What are the key unlocks of this next wave

Transparency

What data is available for use?

Knowledge

How do we define and agree on definitions for data elements, including their construction and use?
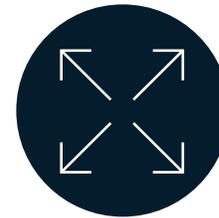
Integration

How can I better manage integration of data to improve efficiency and unlock new applications?

Quality

What is the provenance, lineage, and issues with the data?

Scalability

How do we drive at-scale adoption, access, and use of the data in a governed way?

Example of setting up data for customer retention at financial services organization

Define hypotheses

Review Hypotheses with busines leaders

Identify data available

Capture and learn about data available

Integrate data and test effectiveness

ELT data into Lake for use

Automate production of dataset for modeling
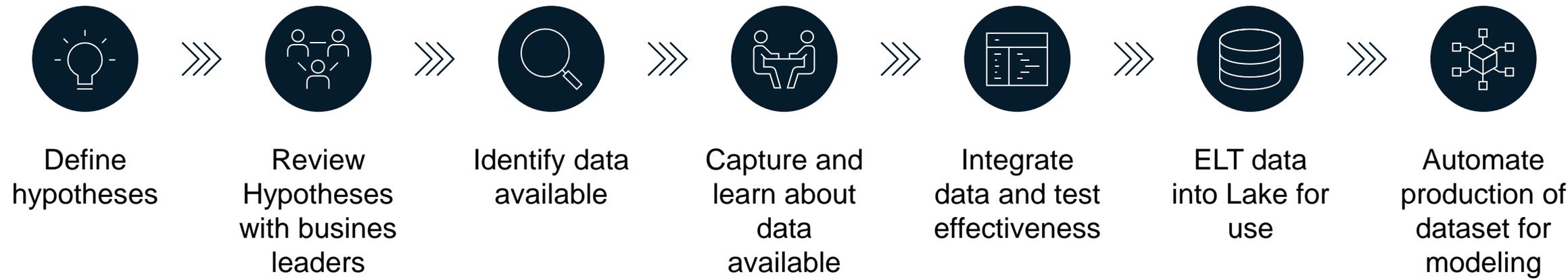
Example: Financial services company and customer retention

Digital data
- Website activity data
- App activity data
- Email opens

Contact Data
- Meetings and appointments
- Emails
- Contact signals (e.g., new account opened, advisor trades, advisor contact to service centers)

Portfolio and Account data
- Product Mix
- Fees
- Inflows/Outflows
- Performance
- Number of accounts
- Type of accounts
- Account history

Advisor Data
- Advisor history
- Ratings and satisfactions scores
- Distance from advisor

Service center data
- Client contacts
- Type of problem, number of transfers
- Client satisfaction scores

Client profile
- Demographics
- Suitability
- Risk profile
- Householding
- Type of account

Products
- Credit card usage
- P&C insurance offerings
- Life insurance offerings

External data
- Micro-market data
- Competitive intensity
- Life events/moves
- Est. Investible assets/share of wallet

Hundreds of questions

What is an active customer?

Which channels are included in our analysis?

What defines churned customer?

How does a Trust account work?

Why did everyone get rid of multiple accounts in 2016?

Are advisor contacts required when opening a new account? Does it classify as a "contact signal"

Huge impact – big celebration!

However, that data we developed was designed just for this use case

Detailed **knowledge of data in the heads of experts**, not well documented or exposed

Everything purpose built for use case, **not reusable** for other analyses. New integrations needed to conduct additional analysis

Data was directionally correct and sufficient to make decisions on, but **finance would often have different definitions** on certain data elements

Great impact – but it wasn't designed to scale across the organization

Current state: Bespoke data create to power use cases

Use-cases

Customer Retention   Customer deepening

Data Domains

(Client)   (Advisor)   (Service)   (Products)   (Digital)

Data assets

Customer suitability	Reviews	Contacts	Products	Email opens
Customer trades	Performance	Transfers	Products to customers	App usage
Channel	Client mapping	Satisfaction score	Product value	Website visitation

Future view: Data products developed to power multiple use cases, enabling organizations to scale impact of data

Use-cases

Customer Retention   Customer Segmentation   Advisor Retention   Advisor growth

Products

Client 360   Advisor 360

Data Domains

Client   Advisor   Service   Products   Digital

Data assets

Customer suitability	Reviews	Contacts	Products	Email opens
Customer trades	Performance	Transfers	Products to customers	App usage
Channel	Client mapping	Satisfaction score	Product value	Website visitation

Data products are a potential solution to scaling and capturing value from data

Details to follow

Data products have following characteristics

Ownership – There is an owner and multi-disciplinary team responsible or producing, managing, exposing and giving access to the data

Explicit shared meaning – Common terms, such as "active customer" should be defined and easily accessible, without needing to understand underlying systems it originates from. Enables data to follow FAIR principles

Quality management – teams are responsible for maintaining quality of the data, including provenance, audibility, traceability and quality metrics
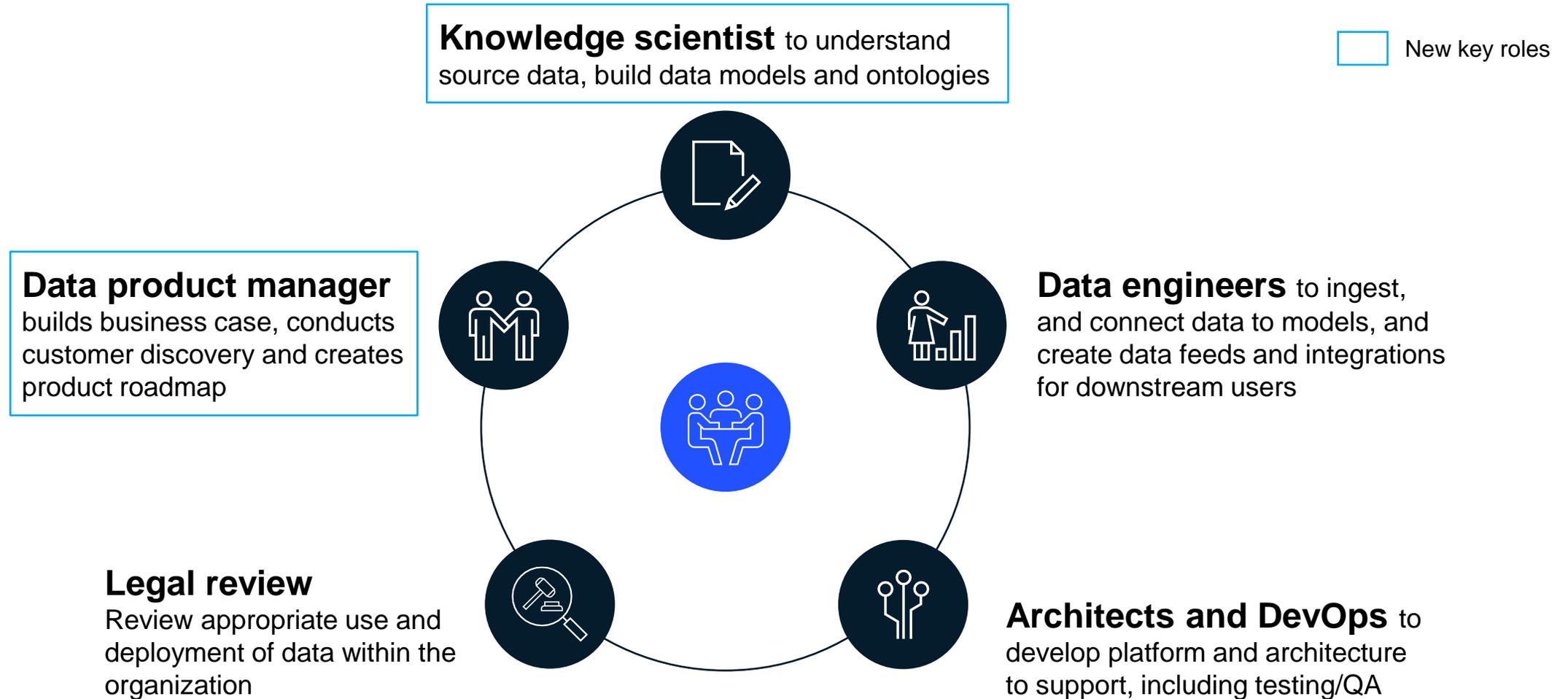
Scalability – Data is produced to serve multiple customers and teams and architecture to deliver the data should support various downstream use cases

Value generation – The data should generate business value across the number of customers that are using it. Further, the data product should have a revenue model to ensure its sustainability

Data product team – key roles in place

Knowledge scientist to understand source data, build data models and ontologies

New key roles

Data product manager builds business case, conducts customer discovery and creates product roadmap

Data engineers to ingest, and connect data to models, and create data feeds and integrations for downstream users

Legal review Review appropriate use and deployment of data within the organization

Architects and DevOps to develop platform and architecture to support, including testing/QA

Knowledge Graphs support development of a "dynamic" data product enabling rapid integration of data for multiple use cases

How they work

Layer of abstraction

- **Create data objects** which have a shared meaning and business value **(e.g., active customers)**
- **You can map these objects to the underlying data structures** (e.g., active customers = transaction done in the last year and have $X in assets)
- **These data objects can be linked easily to each other to represent real-world relationships –** for example active advisors can have many active customers
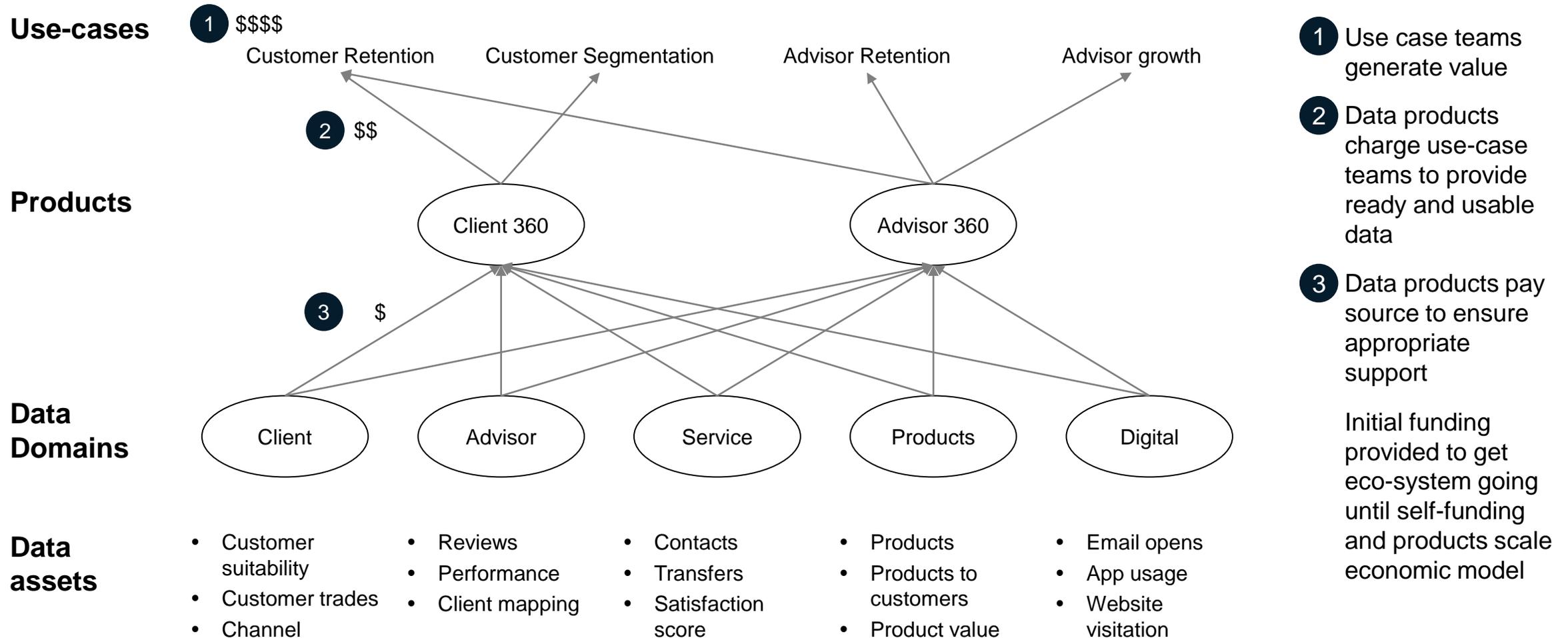
Benefits

- **Offers flexibility to adapt data models** without having to modify underlying data structures
- **Enables integration of data rapidly** to serve high variety of downstream use cases using business logic integrated into the data model
- **Provides standards for all users**, and can be easy to understand for business users
- **Ideal for products with high number of sources to integrate**, and high degree of diversity in outputs needed (e.g., Customer 360, Employee 360, Digital Twins)

Considerations

- **Time investment to develop and maintain model**
- **Tooling still nascent** especially deployment and versioning KGs, but hold significant potential
- Data with limited integration needs do not warrant KG implementation

Data products should be treated like a revenue generating entity, with initial support to get eco-system off the ground

Use-cases

① $$$$
Customer Retention   Customer Segmentation   Advisor Retention   Advisor growth

② $$

Products

Client 360   Advisor 360

③ $

Data Domains

Client   Advisor   Service   Products   Digital

Data assets

- Customer suitability
- Customer trades
- Channel

- Reviews
- Performance
- Client mapping

- Contacts
- Transfers
- Satisfaction score

- Products
- Products to customers
- Product value

- Email opens
- App usage
- Website visitation

① Use case teams generate value

② Data products charge use-case teams to provide ready and usable data

③ Data products pay source to ensure appropriate support

Initial funding provided to get eco-system going until self-funding and products scale economic model

Although early days, evidence of commercial value is amassing

BBC
Most **content management** done with aid of knowledge graph
Coordinates journalist's work and powers article recommendations

EA
"Intelligent Content Ecosystem" (videos, games, articles, ...)
Meaningful product and article recommendations
age ratings across multiple jurisdictions and languages, along with justifications

WUNDERMAN THOMPSON
Supports $2bn marketing agency for all **people based data product**
Integration of 100s of millions of datapoints from 1000s of sources
Development efficiency: "do in 12 hours what we couldn't in 6 months"

reonomy
"Unprecedented Products": **ownership portfolio of commercial real-estate**
Development efficiency: new dataset live in 5 days with junior dev

amazon
Graph holds hundreds of millions of **products**, millions of types
Allows natural queries like 'gluten-free yoghurt' or 'best toys for baby'
Population and curation highly automated with machine learning

UHG
Member profiles from single system, near real-time, including contact timeline and care-path recommendations
Claimed $150 million per year savings

DIFFBOT
Multiple data products based on unstructured web-based data **– news, organizations, people**

enigma
Gathers data from many sources to create a **private/SMB data product**

Steps to get started

Identify and prioritize data products to be built
- Conduct discovery to identify use cases to be powered
- Determine those with greatest demand and value to start with

Align executives on value capture approach
- Develop a business case and business model
  — Benefits: use cases powered and downstream impact
  — Costs: team, technology, data
  — Longer term revenue model

Build an initial operating model for one product with multiple use cases supported
- Product manager to build business case and generate demand for data
- Knowledge scientist to build context-aware data model
- Data engineers integrate and link data to this model

Familiarize yourself with tooling and tech
- Look at knowledge graph approaches to organizing and linking data together

LED Lights

Please connect with me

Mohammed_Aaser@mckinsey.com

https://www.linkedin.com/in/mohammedaaser/