

Data Democratisation

Distributing the Workflow

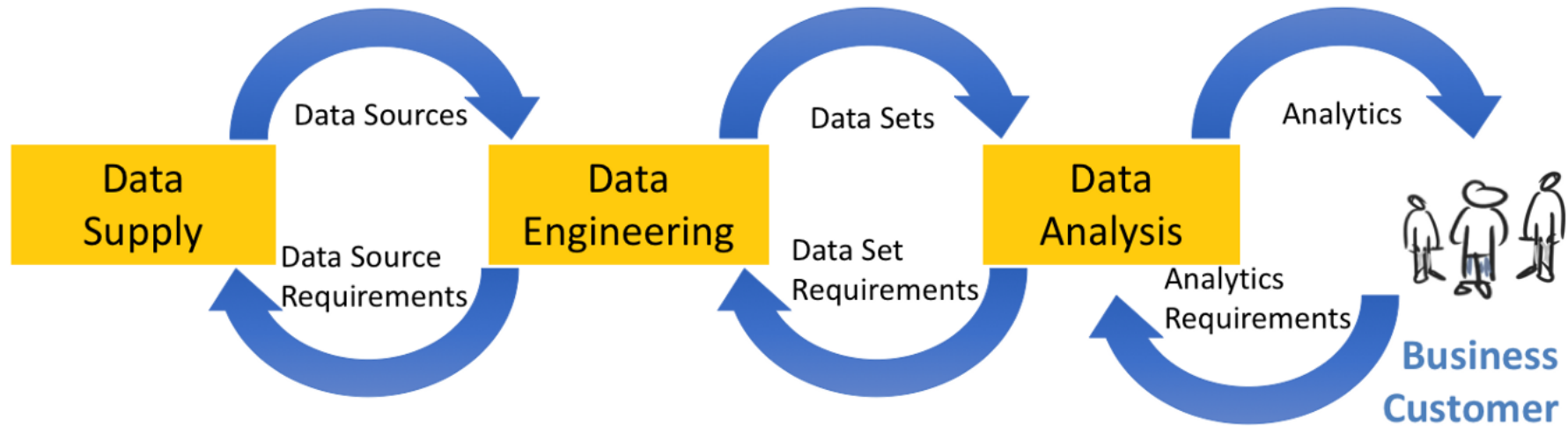


Filipa Peleja

Lead Data Scientist, Levi Strauss & Co.

The struggle

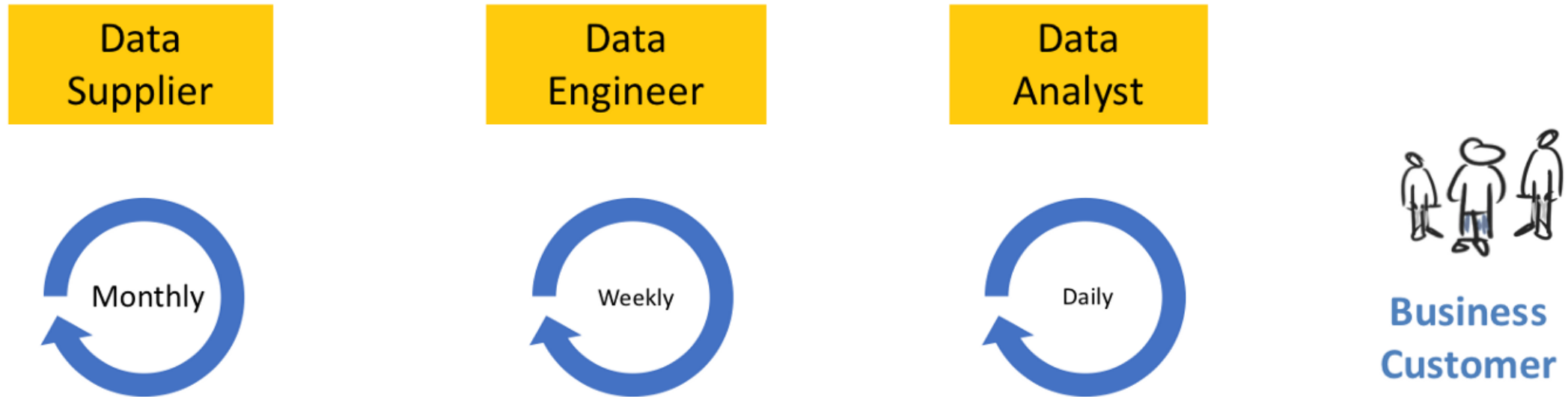
Centralisation vs "freedom"



The struggle

Centralisation vs "freedom"

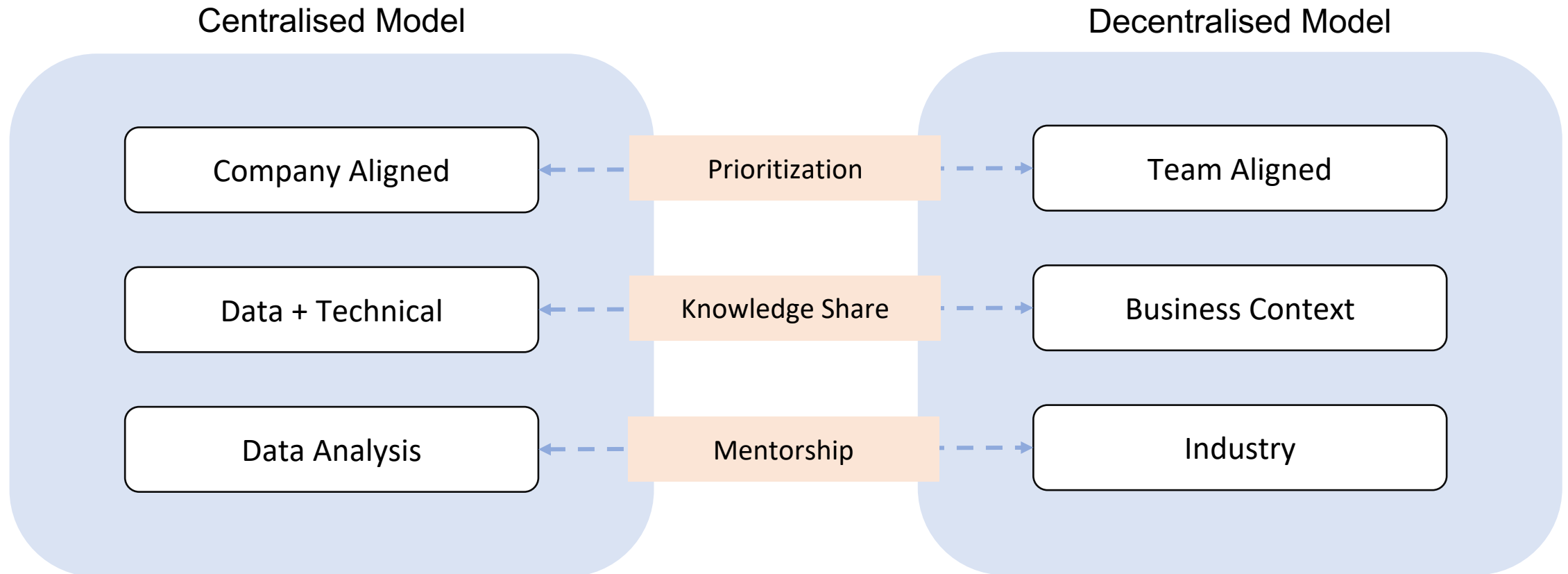
Timing matters..



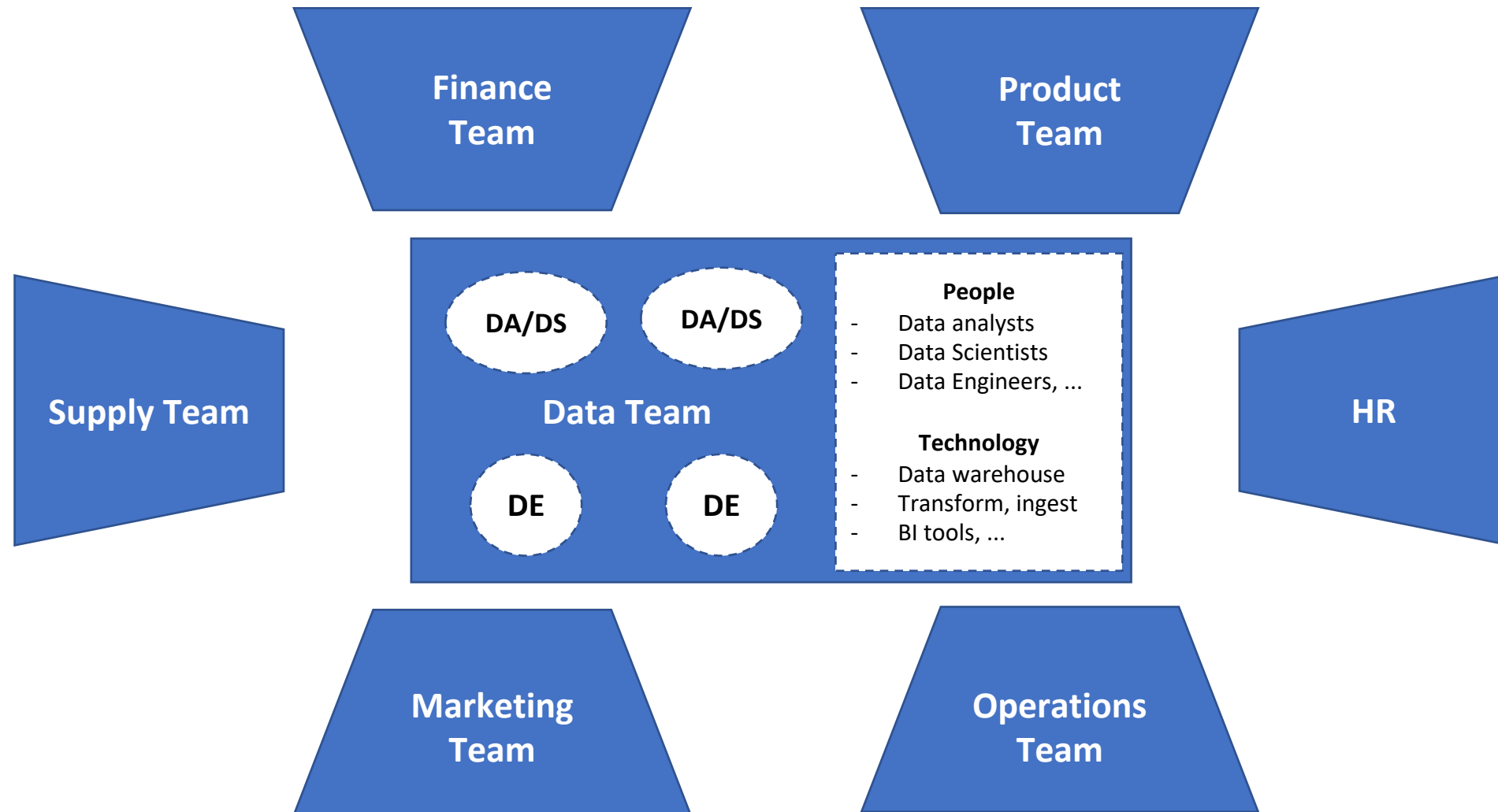
DataKitchen, 2022

<https://medium.com/data-ops/dataops-resolves-the-struggle-between-centralization-and-freedom-in-analytics-7f746fbc856b>

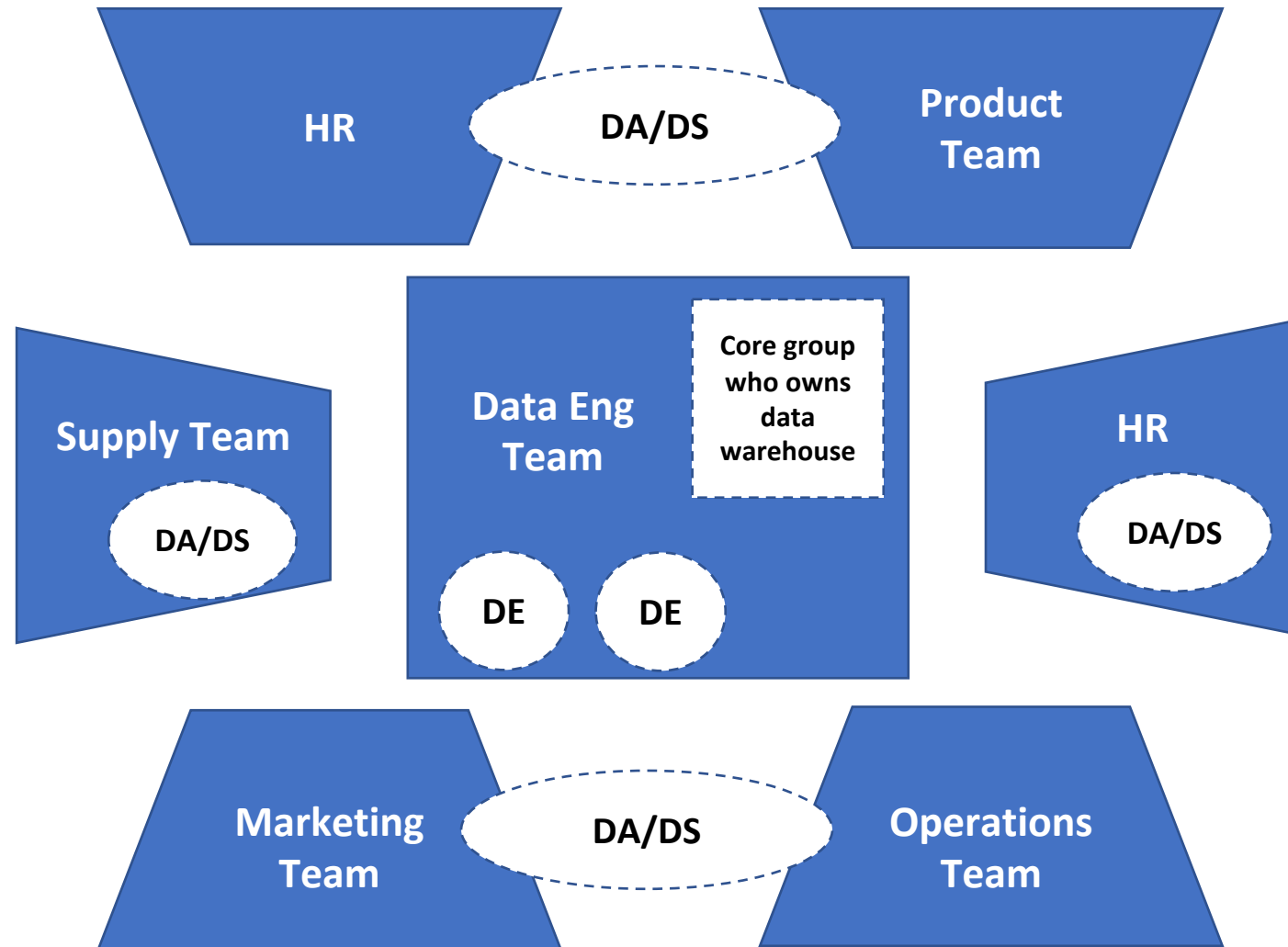
Centralised vs Decentralised



Centralised Model



Decentralised Model



Data Mesh: a data platform architecture highly decentralised distributed data architecture

"Data mesh envisions a world where central IT provides and maintains data infrastructure and is in control of data governance. Everything else is delegated to the business domains, including data ingest, modelling and transformation among other tasks is on the analysis side.

In the world of Data Mesh, the business domains have cross-functional responsibility, and not merely self-service autonomy."

By Andrew Brust, May 2022, "Centralized or Decentralized Analytics"

Data Mesh *pros/cons*

Pros

- + Scalability
- + Ownership – data business are responsible for it
- + Quality – data business teams know the data well

Cons

- No standard definition of a data mesh
- Investment in organizational change and technical implementation
- Duplication of data and ingestion
- Difficult to see the "big picture" to combine data
- Performance of combining multiple domains/business units

Can Data Analysts & Data Scientists

*support business without
constant need of close
support of
Engineering teams*



*for tasks related to data
ingestion and automation
tasks?*

Democratisation for ingesting data

Data Build Tool (dbt)

- dbt is an example of a tool that aims to adopt software engineering best practices
- Allows to ingest data in its raw form, and there will be a process with version control that tests its quality and pre-prepared data dictionaries to manage the data that goes live

***Does not solve all data ingestions needs but helps business move faster
and
DE have more time to support/improve the process with different timings***

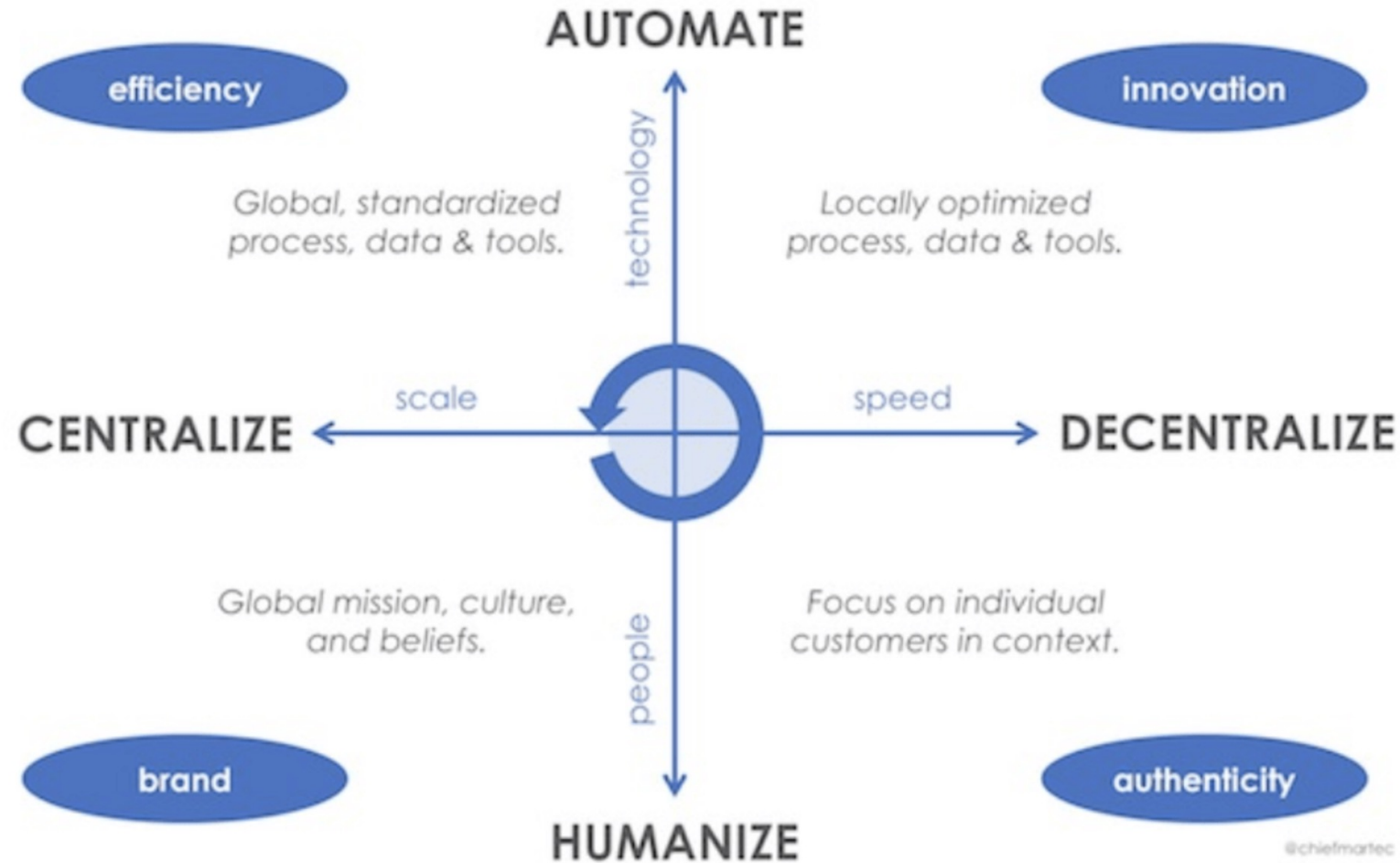
Self-service analytics

Demand for advanced analysis & reporting

- **Higher demand for companies to democratize analytics** and, consequently, push more for decentralization models
- Advance analyses and reporting is growing fast
- Why decentralised? Business units are closer to the market and know best the required analysis that drive best value and performance

Self-service analytic tools are growing in sophistication as well in user-friendliness

Summary: Centralised vs Decentralised



The 4 Forces of Marketing Operations & Technology

Diagram courtesy of Scott Brinker/chiefmartec.com

Summary: Centralised vs Decentralised

Centralised

- + **Alignment of data resources** to company-need
- + **Knowledge sharing** – easier to build cultural data norms
- + **Mentorship** – learn from more senior analysts / data engineers

Decentralised

- + **Speed** to address business needs

Summary: Centralised vs Decentralised

Centralised

- + **Alignment of data resources** to company-need
- + **Knowledge sharing** – easier to build cultural data norms
- + **Mentorship** – learn from more senior analysts / data engineers
- Speed is the biggest challenge
- Harmonise/standardise definitions e.g., revenue definition is "not the same" in all departments
- Ownership over data sources if pipelines or models break

Decentralised

- + **Speed** to address business needs

Summary: Centralised vs Decentralised

Centralised

- + **Alignment of data resources** to company-need
- + **Knowledge sharing** – easier to build cultural data norms
- + **Mentorship** – learn from more senior analysts / data engineers
- Speed is the biggest challenge
- Harmonise/standardise definitions e.g., revenue definition is "not the same" in all departments
- Ownership over data sources if pipelines or models break

Decentralised

- + **Speed** to address business needs
- Difficult to keep analysts working closely together
- Not easy to scale and share knowledge of data analytics and organization data
- Highly dependent on how empowered analysts are e.g., own their engineering workflow
- Disconnect between data ingestion done by DE and analysis done by DA/DS

Join us!

levistrauss.com/work-with-us/data-science/

- Product Recommendations
- Search & Browse
- Loyalty
- Consumer Experience

Computer Vision

NLP & Information Retrieval

Recommendation Systems

and more

The future of fashion, apparel and retail is digital — powered by data, computer vision and machine learning.

*We are harnessing petabytes of images and a treasure trove of data from the last 167 years to **solve exciting problems in fashion, design, manufacturing, supply chain, business and society** — all of it while leading with our values. Join us to make your mark on our digital transformation.*

Katia Walsh, Ph.D.

Senior Vice President & Chief Strategy and AI Officer

Thank you!



Filipa Peleja

Lead Data Scientist, Levi Strauss & Co.