Beyond the hype

Practical steps to establishing and scaling your data & ML team

Ellissa Verseput
LeadDev London, June 2024

sympower

sympower.net

GENAI HAS WITNESSED UNPRECEDENTED LEVELS OF ADOPTION BY CONSUMERS

LeadDev
Berlin

LLM IS A NEW PARADIGM FOR BUILDING AI SYSTEMS

AT THE PEAK OF INFLATED EXPECTATIONS OF **GENAI-AUGMENTED** SOFTWARE ENGINEERING

COMBINING **LLMS** WITH OTHER TOOLS AND DATA UNLOCKS MORE POTENTIAL

Why is nobody talking about

"Normal"

Machine learning and Analytics anymore?

Talk title *

Beyond the Hype: Practical Steps to Establishing and Scaling Your Data & ML Team

Talk description *

Join me for a deep dive into a down-to-earth data story that has been overshadowed on the stage. While discussions around cutting-edge technologies like genAI, ChatGPT, and deep learning, this talk shifts the spotlight to what truly ma companies—practical data analytics and real-world machine learning applications.

Learn why enabling data-driven decision-making and automating tasks with machine learning are the keystones that difference, particularly for companies transitioning from start-up to scale-up. We'll explore the essential steps in establi proving the product-market fit of a new, small data team in a scale-up environment, essentially a data start-up within yo company.

This talk explains how to kickstart a data team, establish a robust data platform, and implement the first data analytics learning use cases. With the new data team proving its value, it's crucial to avoid getting stuck when experiencing a su Drawing parallels with the growth pains often experienced by expanding engineering teams, we'll use concepts from T to navigate the challenges and scale your data initiatives effectively.

Ellissa Verseput

Data Team Lead @Sympower

Hi!

- ML Engineer by craft
  - 2x scale-up
  - 2x corporate innovation
- Now nearly 3 years in data leadership in green scale-up
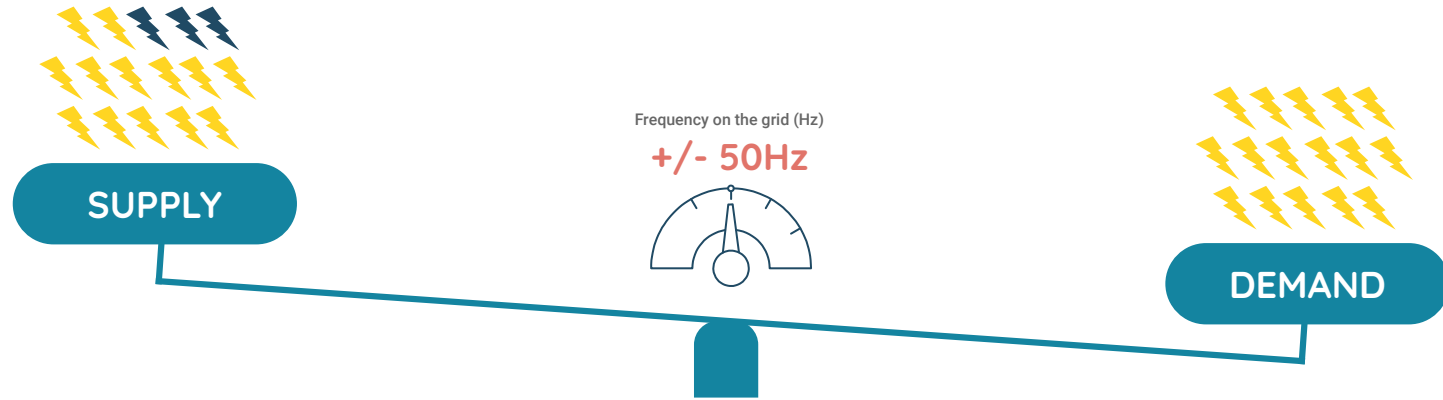- Sustainability enthusiast

Journey
of today

- Evolving data for 3 years in a scale-up
- Each year a new story: new phase, new challenge
- Actionable takeaways how to progress
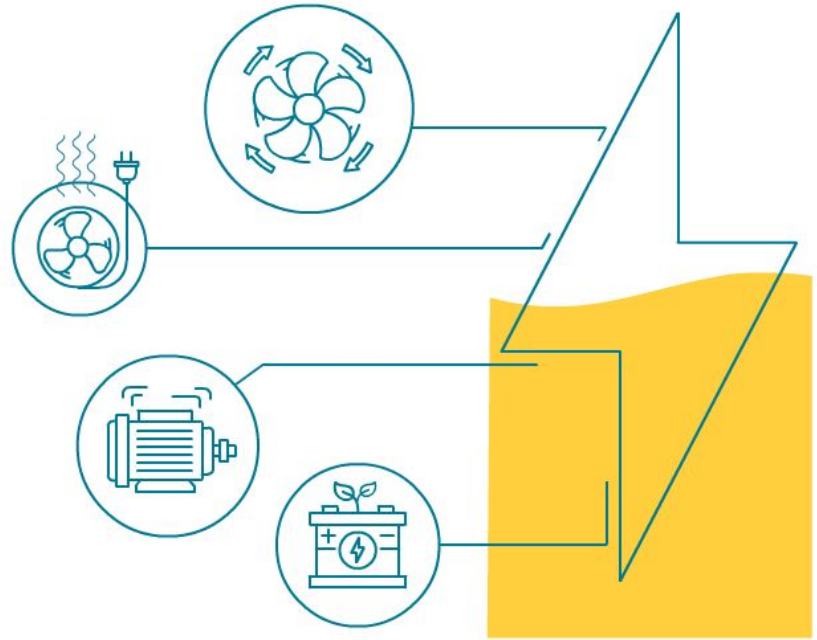
All with 'normal' ML and Analytics

2022
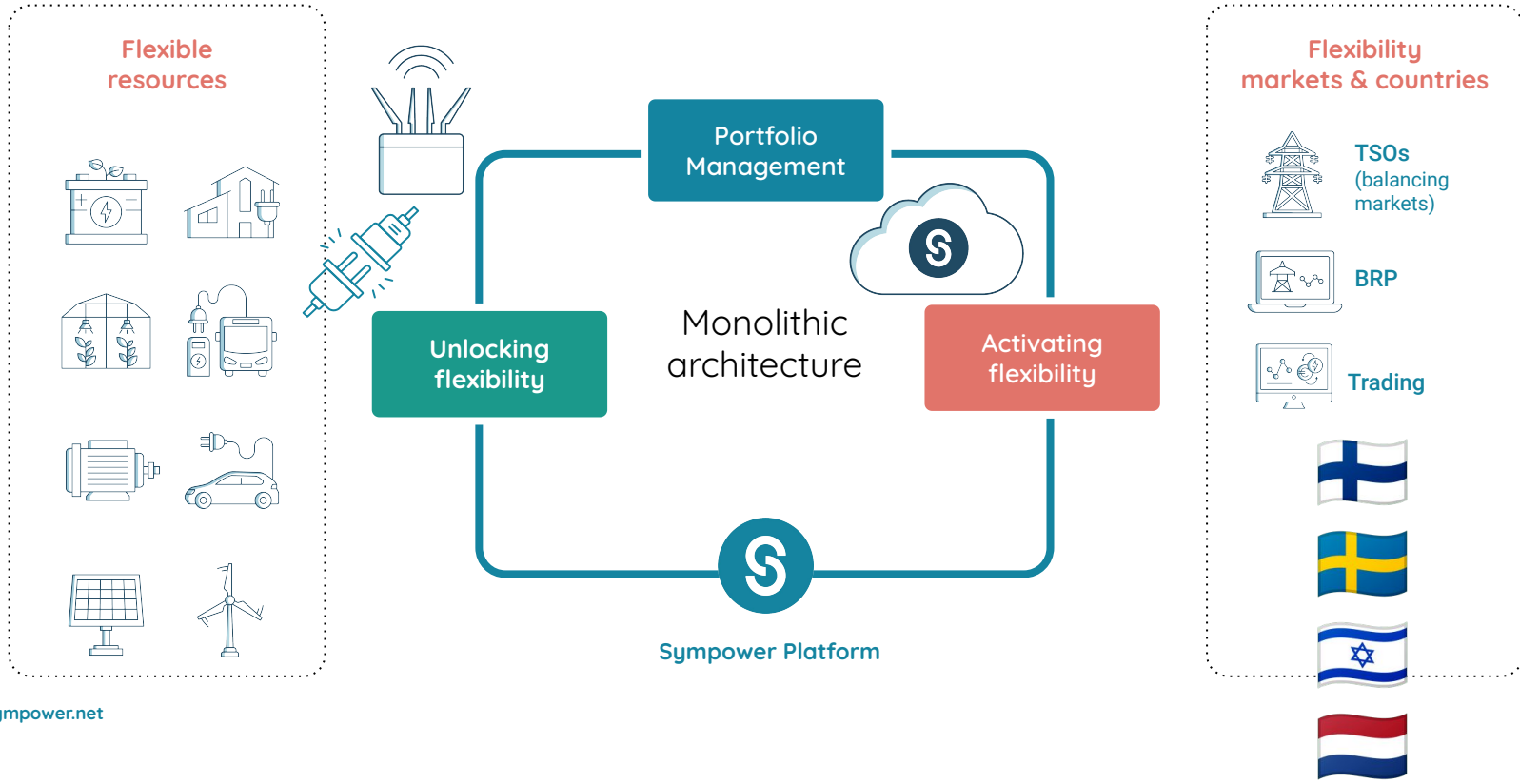
Joining a green energy tech scale-up
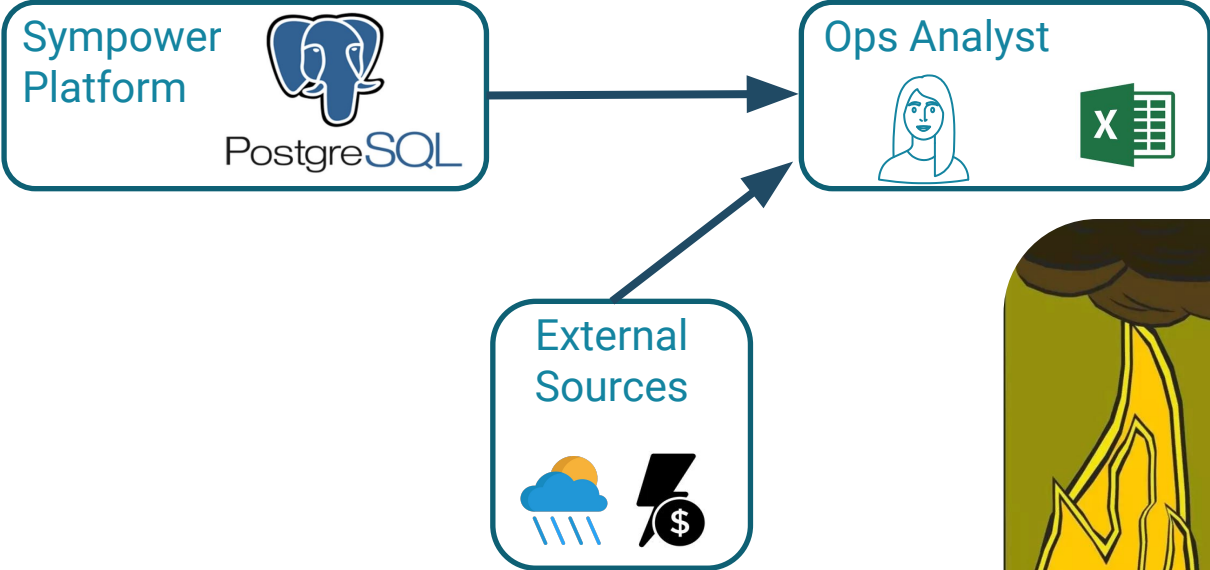
Keeping the energy grid in balance…

SUPPLY

Frequency on the grid (Hz)

+/- 50Hz

DEMAND

| LONG TERM | DAY-AHEAD | INTRA-DAY | → | BALANCING SERVICES |

… by turning off/on resources of industries

Sympower's startup platform

Flexible resources

Monolithic architecture

Portfolio Management

Unlocking flexibility

Activating flexibility

Sympower Platform

Flexibility markets & countries

TSOs (balancing markets)

BRP

Trading

Critical daily forecasting processes in hands of a colleague and Excel

Sympower Platform — PostgreSQL

Ops Analyst

External Sources

THIS IS FINE

sympower.net

#data-requests

Slack channel for making ad-hoc data requests by business, in need of operational platform data...

Shuneel Kumar 12:46 PM
Hi,
I need frequency data from bx2 for 21june 2022, for 24 hour or at least 1 hour with 100mHz sample. The

Kristian ⛔ 11:02 AM
Yearly and hourly market, offered and accepted capacities and prices for them. FCR-D down & up, FCR-N and FFR
Instance 🇫🇮
Urgency: Today preferably
Time: 2022-02-10 23:00:00 - latest values

Hi @Siimo , would you have the time to run the same query as I requested the last time? I would need updated values for the rest of February.

✅ 1   😃+

Irene Almenar 🤓 2:42 PM
DATA REQUEST: Invoicing FCR-D
Hi Data team!
It was just right after your message Ellissa that I have a request of data 😉
I am trying to download controllable capacity per active resource owner for SE1 and the month of May
I am using a query (see thread) that it has worked for downloading data for SE2, SE3 and SE4.
Can you help me with this?

🖼️🖼️🖼️ **25 replies** Last reply 3 months ago

Rosario 🌴 2:49 PM
Power meter data in 100ms request
Instance: SE3 🇸🇪
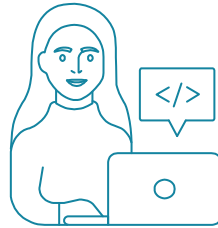**Controller ID: 115 & 116**
RO: Holmen Hallstavik
Time:
- start 2022-08-05 10:00:00 (UTC)
- end 2022-08-05 11:00:00 (UTC)
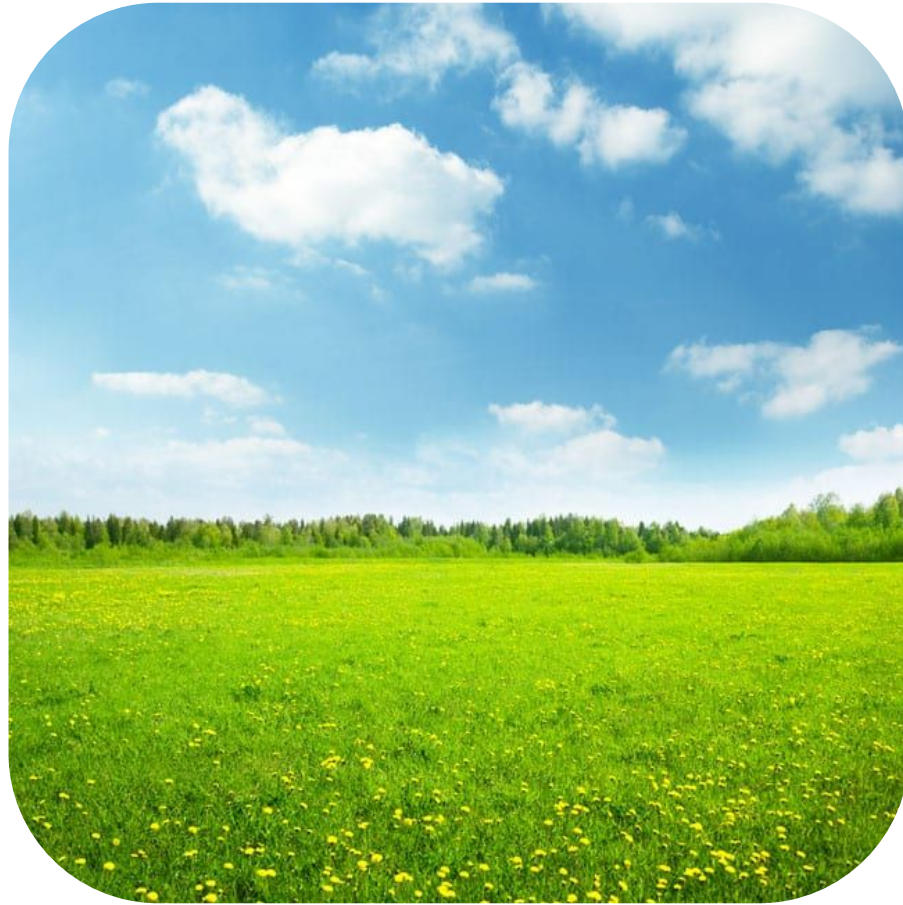
Comment: This request is similar to the earlier

🖼️🖼️ **6 replies** Last reply 1 month ago

The data team

Hands-on Data Team Lead (me)

The Data Platform

⚡ Challenge 2022

- Mini team
- No data infrastructure
- Lots of data "buzz", but no data value

Establishing Data Team's Value Proposition

Forecasting

Responsibilities
Provide all live countries and market programs with **automated, reliable, usable forecasts.**

Machine Learning!

Insights

Responsibilities
Decrease the time spent on **manual data gathering** and processing.

Automate data insights that enable **data-informed decision making**

Data Analytics!

Sympower Platform

PULSAR · Microservices · Amazon Timestream · Prometheus · Amazon S3 · PostgreSQL

Sympower Teams

.TXT · N · X

Sympower's Data Platform

Service Layer
- BI reports
- Dashboards
- Output Stream — PULSAR

EL

Lakehouse — Transform → → →

Production ML models

Research ML models

Track model performance

Notebooks/ SQL Analytics

External Sources

Goal architecture for our future Data Platform

sympower.net

17

Sympower's Data Platform

EL

Service Layer

Output Stream

PULSAR

databricks Lakehouse Platform

SIMPLE ○ OPEN ○ COLLABORATIVE

| Data Engineering | BI & SQL Analytics | Real-time Data Applications | Data Science & Machine Learning |

Data Management & Governance

DELTA LAKE

Open Data Lake

Structured   Semi-structured   Unstructured   Streaming

aws   Microsoft Azure   Google Cloud

Let's go team!

Hands-on Data Team Lead (me)

Analytics Engineer
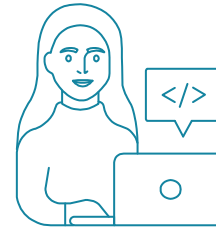
Analytics Engineer

Let's start with what's there

Full

Incremental
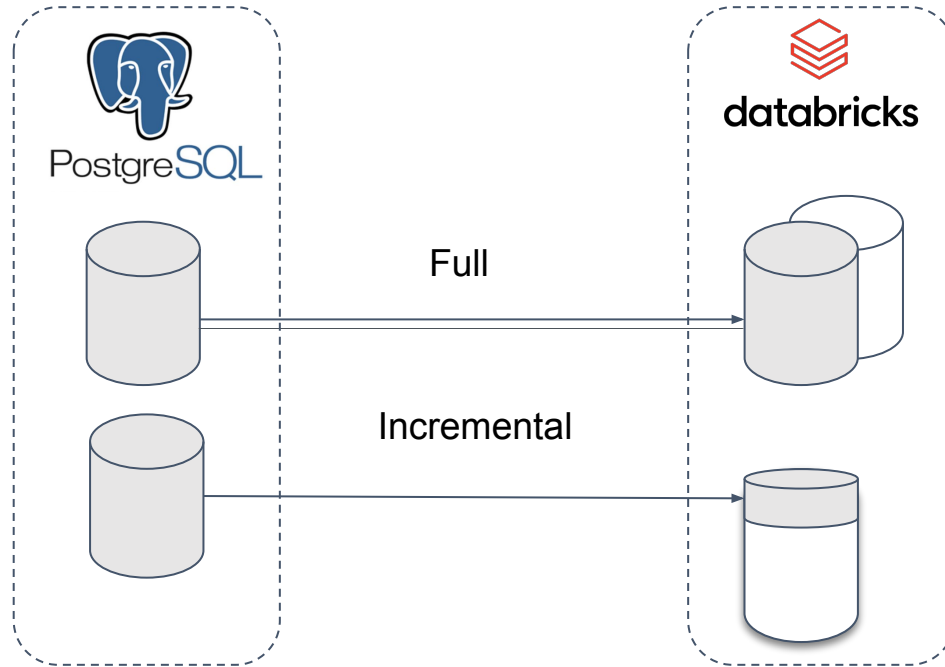
Sympower Platform — SCRAPING monolith DBS! — PostgreSQL — Monolith — Monolith

Sympower Teams

Sympower's Data Platform

Service Layer
- BI reports
- Self-service dashboards
- Output Stream — JSON

EL

Lakehouse — Transform → → →

1 forecasting pipeline
2 dimensional star models

External Sources

1 prod ML model

Research ML models

Track model performance

Notebooks/ SQL Analytics

sympower.net

MVP Data Platform

21

💡 Takeaways 2022

- Hire for crucial missing skills
- Buy, not build, data-infrastructure
- Pick 1-2 burning use cases and proof!

2023

More demand for the Data Team
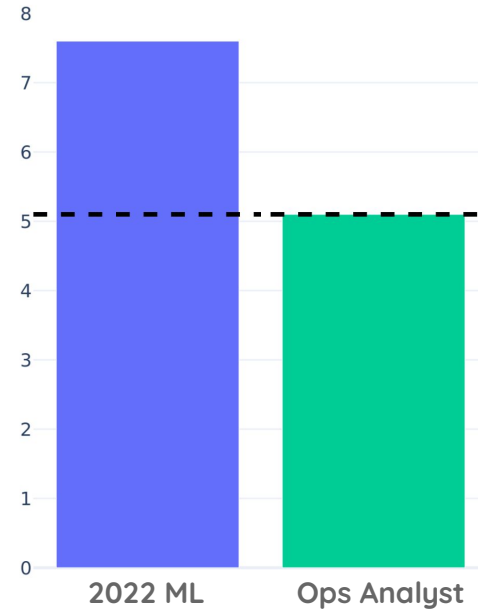
Insights: what about top-down demand?

But steering the business still happens based on gut-feeling
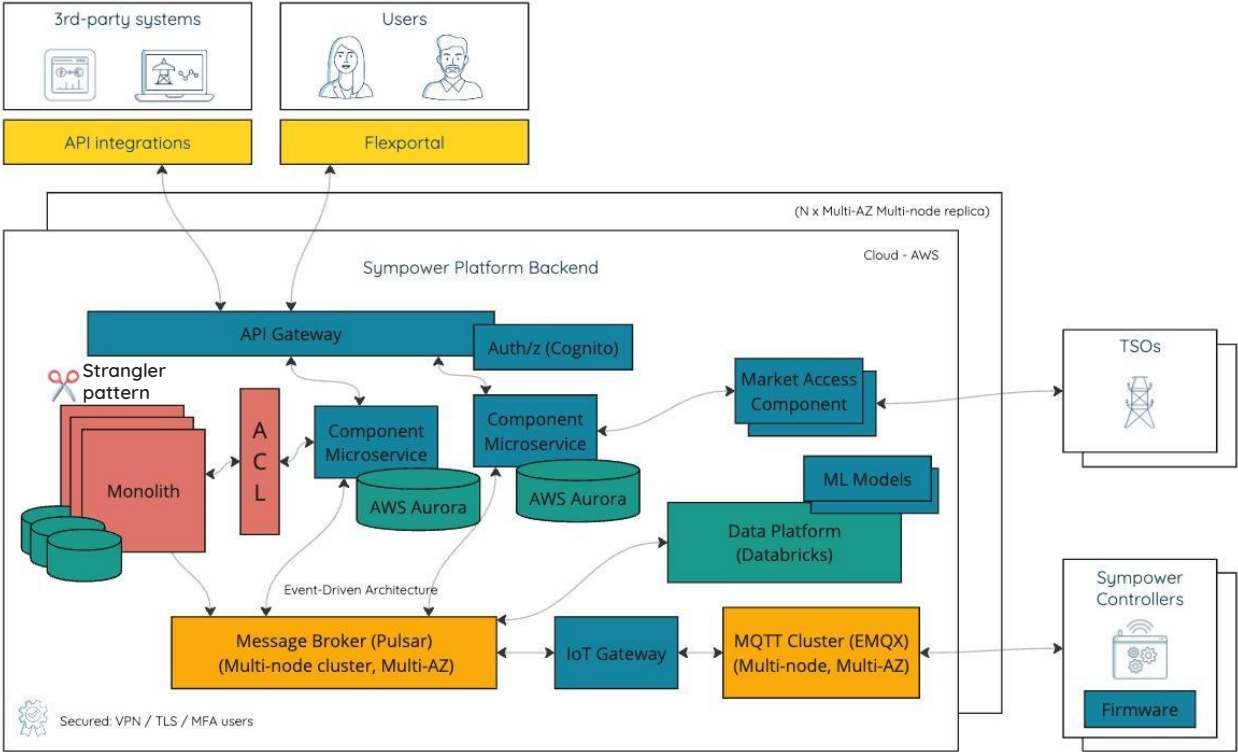
Forecasting: We Want More! And better accuracy!

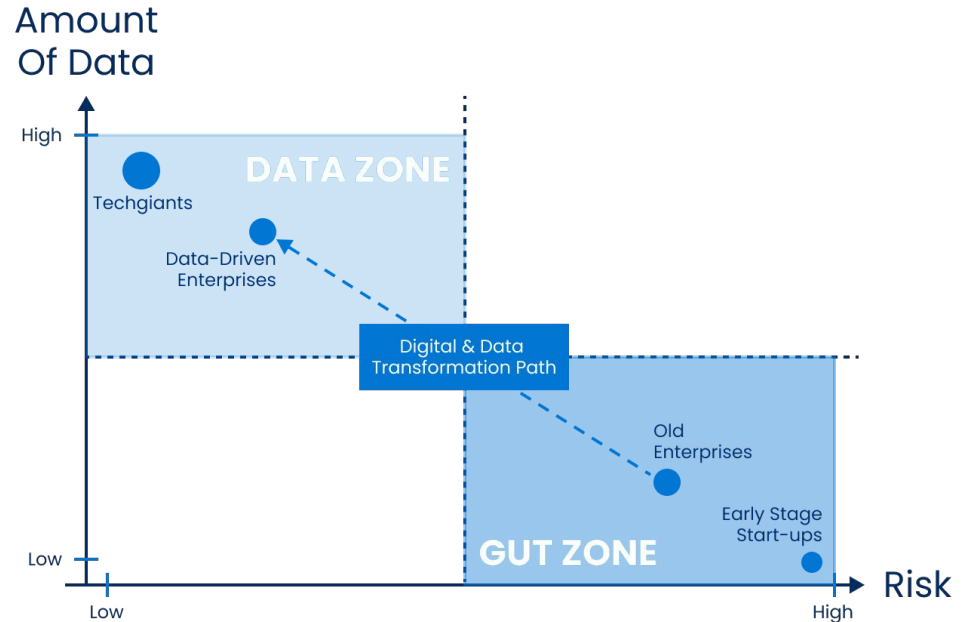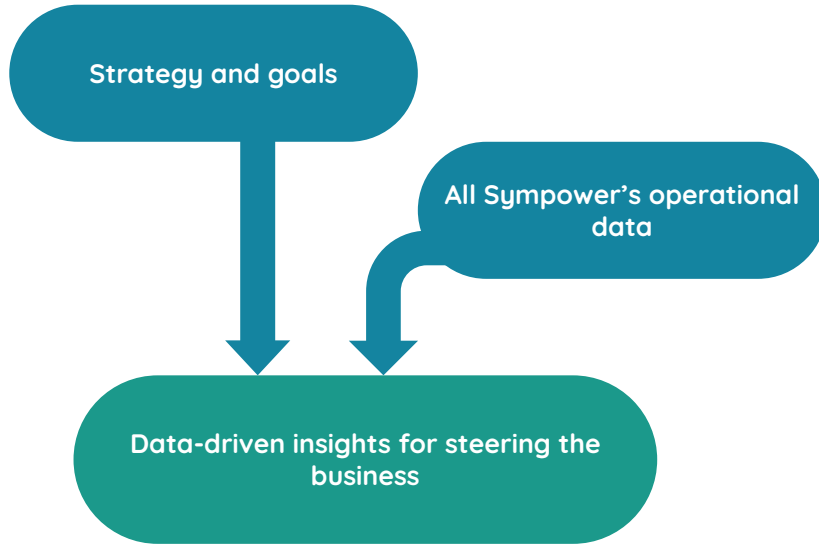

Average forecasting error

Data infrastructure: from scraping monolith dbs to…

⚡ Challenge 2023

- How to get all business users onboard
- Help, I can't work on all ML projects and lead the team
- Maturing data platform architecture

Incentivising top-down demand for data-driven insights

Strategy and goals

All Sympower's operational data

Data-driven insights for steering the business

Amount Of Data

High

DATA ZONE

Techgiants

Data-Driven Enterprises

Digital & Data Transformation Path

Old Enterprises

Early Stage Start-ups

Low

GUT ZONE

Low

High

Risk

Low

High

Sympower's Data Platform

Service Layer

databricks Lakehouse Platform

SIMPLE ○ OPEN ○ COLLABORATIVE

Data Engineering

BI & SQL Analytics

Real-time Data Applications

Data Science & Machine Learning

Data Management & Governance

DELTA LAKE

Open Data Lake
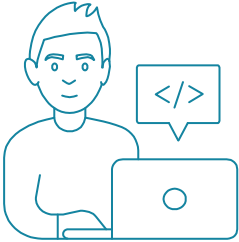
Structured

Semi-structured

Unstructured

Streaming

aws
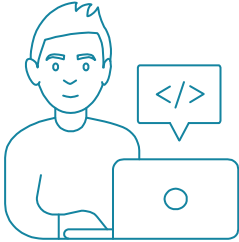
Microsoft Azure

Google Cloud

Power BI

Track model

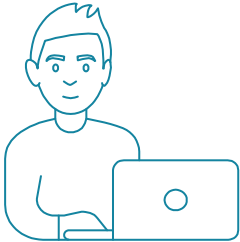# Growing the team

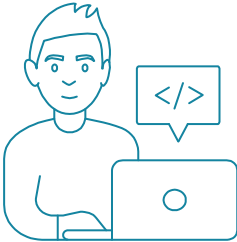ML Engineer

ML Engineer

Analytics Engineer

Analytics Engineer

Data Team Lead
Hands-on-*ish*
(me)

Product Manager Data

Data Engineer

Sympower Platform | PULSAR | Microservices | Amazon Timestream | Monolith | PostgreSQL | Monolith ⚠️

Sympower Teams | .TXT | N | X

Sympower's Data Platform

Service Layer

5 BI reports

15 Self-service

Output Stream | PULSAR

EL

Lakehouse

Transform → → →

3 forecasting pipelines
6 dimensional star models

3 prod ML model

Research ML models

Track model performance

Notebooks/ SQL Analytics

External Sources | bambooHR | pipedrive | recruitee

sympower.net

The Data Platform is doing more and more!

31

💡 Takeaways 2023

- Manage your top-down data demand
- *First-time data manager:* delegate!
- Spread cognitive load
  - What (Product Manager Data) VS How (Data Team)
  - Different specialists working together in the same team
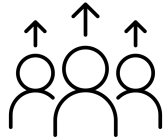
2024

Growth pain

Data team start 2024

Data team has established product market fit within Sympower
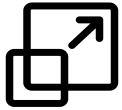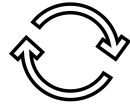
Our products are end-to-end solutions

Full roadmap & budget to grow the team 2x during 2024

However… growth bottlenecks for analytics

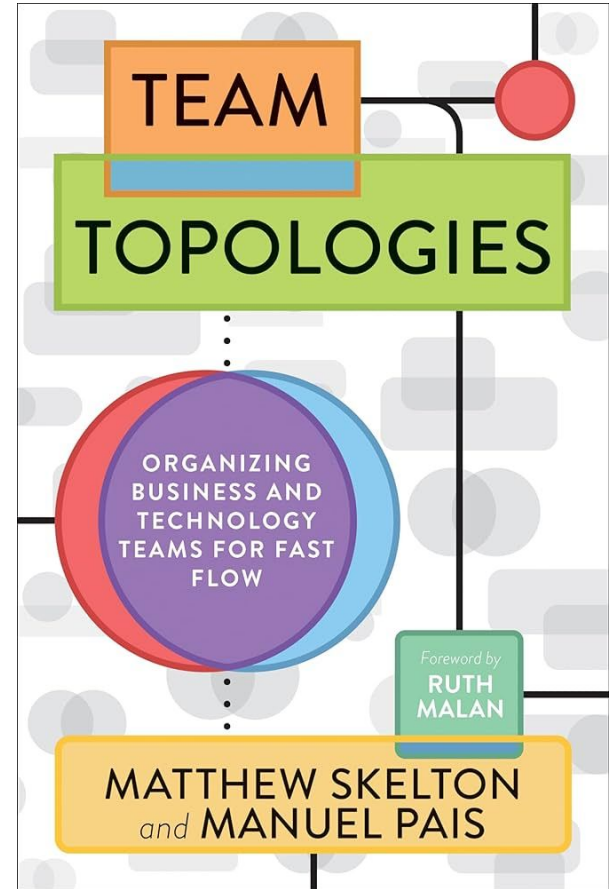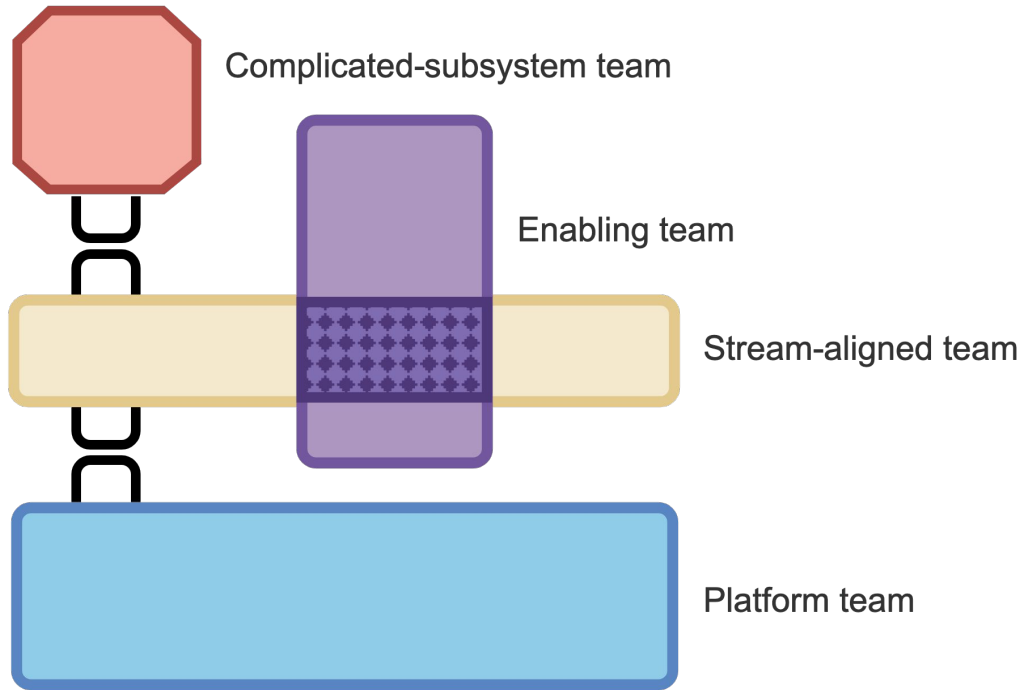| Not scaling E2E approach | Slow feedback cycle | Increasing bus factor | Incomplete data coverage | Accumulating Tech debt |

sympower.net

⚡ Challenge 2024

- Growing from 7 to 14 people
- How to best structure the data team(s)
- Scaling analytics

Team Topologies to the rescue

- Seven-to-nine rule
- Cognitive load and domains in team
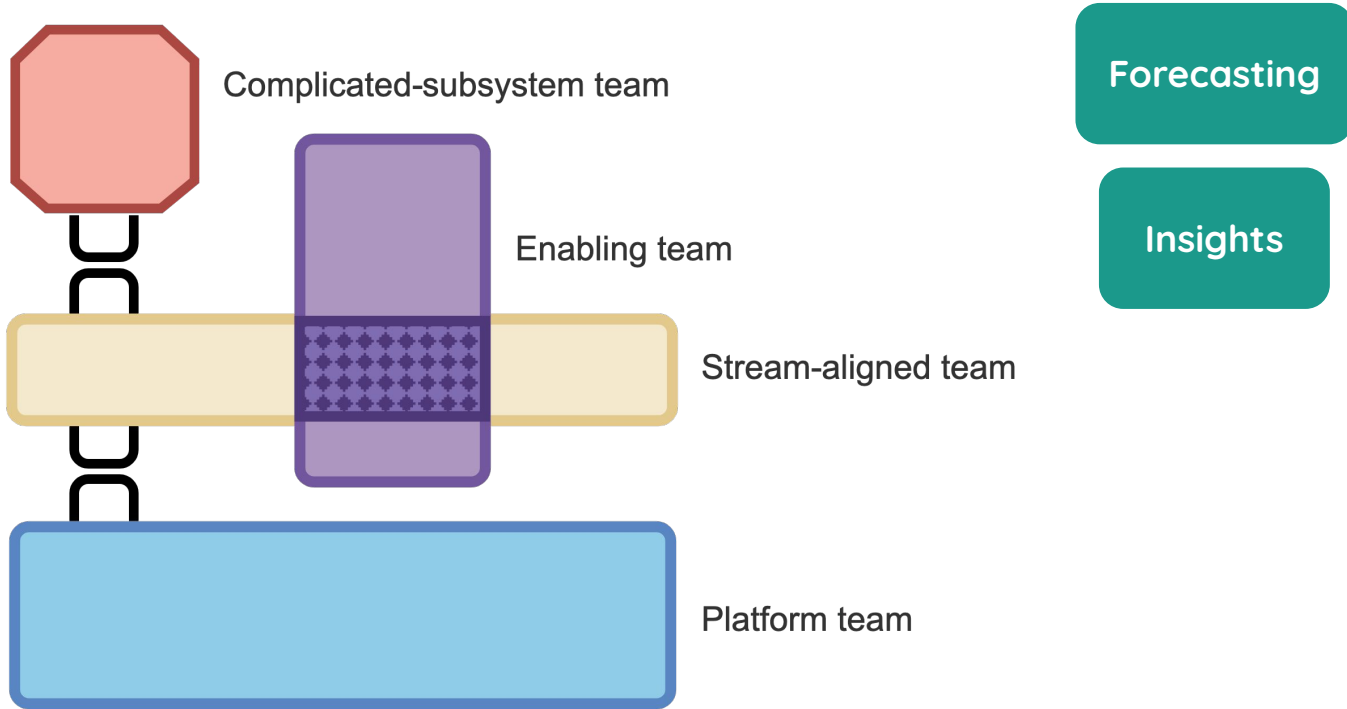- Other team purpose options than E2E delivery

What we do now…

Complicated-subsystem team

Enabling team

Stream-aligned team

Platform team

Insights

Forecasting

© Matthew Skelton and Manuel Pais from *Team Topologies*

… to ideas how we could start map these responsibilities

Complicated-subsystem team

Enabling team

Stream-aligned team

Platform team

Forecasting

Insights

© Matthew Skelton and Manuel Pais from *Team Topologies*

Keep E2E VS enabling approach for Analytics

We keep our E2E solution provider approach as it is right now

We start moving towards an enabling approach, where data-savvy business users users create their own insights

Upgrades to the way of working

Data-savvy business users → value stream aligned activities

Analytics engineers → enabling team activities

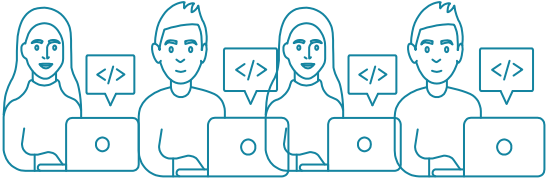**Machine learning engineers → comlex sub-system team activities**

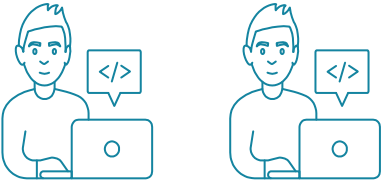**Data engineers → platform team activities**

💡 Takeaways 2024

- In hindsight, pivot faster from E2E approach
- You are not the first one, educate yourself
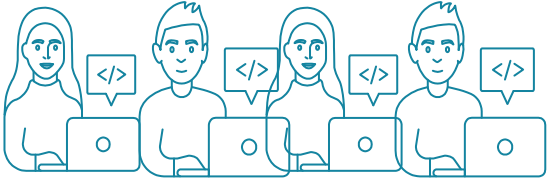- Try and learn!

Conclusion, but not the end! Work in progress!

4x ML Engineer

2x Data Engineer

4x Analytics Engineer

Data Team Lead Hands-on-*ish* (me)

Product Manager Data

Data Team Lead Hands-on-*ish*

Product Manager Data

symplower.net

Thank you!

Want to know more?

Ellissa Verseput
Data Team Lead
ellissa.verseput@sympower.net
+31 6 133 688 21

Want to join Sympower?

Interested people can keep an eye on our career page
https://sympower.net/careers/

Hiring for data roles and other engineering roles!