Journey to a Modern Data Architecture

Leo Dong

Agenda

- Motivations for improving data architecture

- Where are we now?

- From A to B

  - How do we manage change in organisations

  - Common pitfalls and risks

- Trends in architecture

  - What's popular

  - My takes on the future

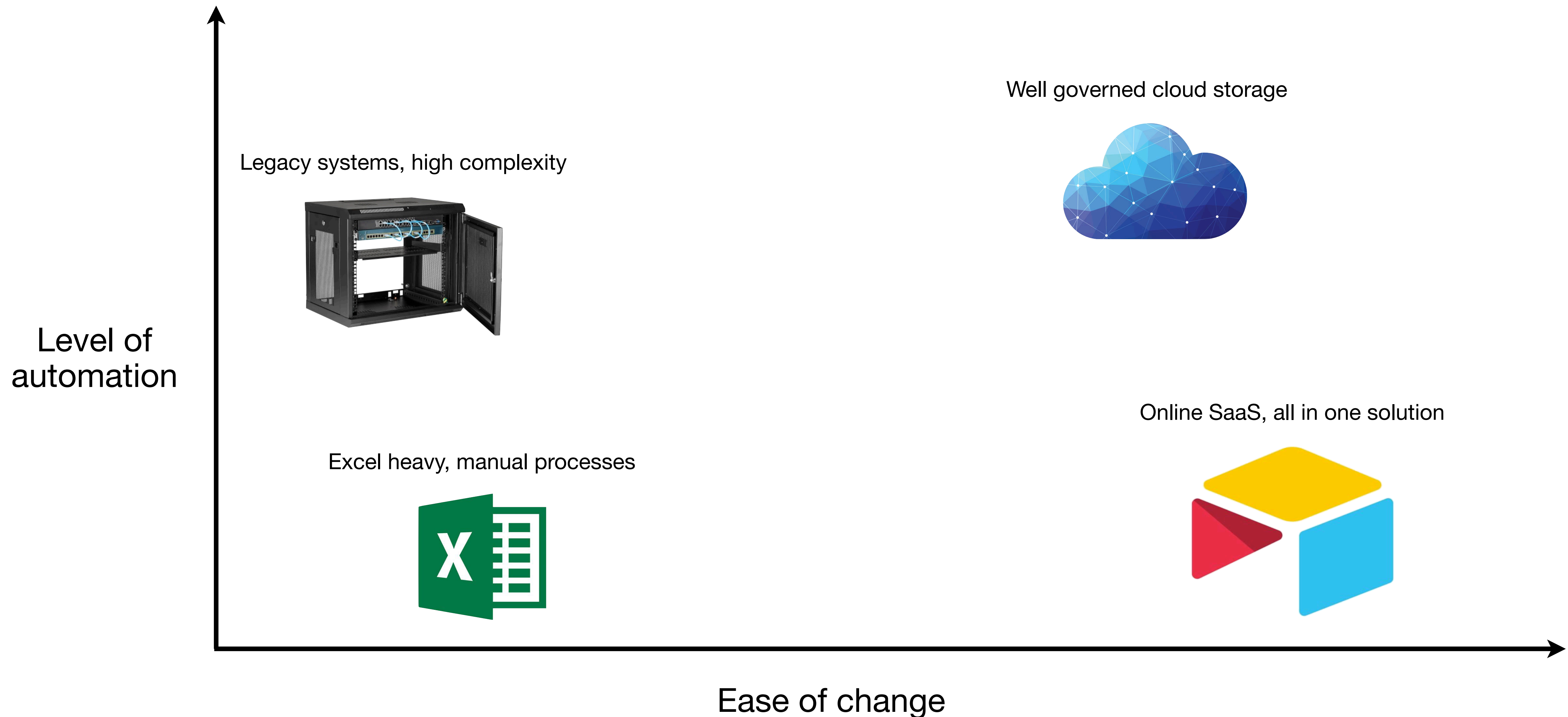# The importance of data architecture
A summary

- Innovation and competitive advantage

- Speed of deployment

- Cost-effectiveness

- Scalability

- Regulatory adherence
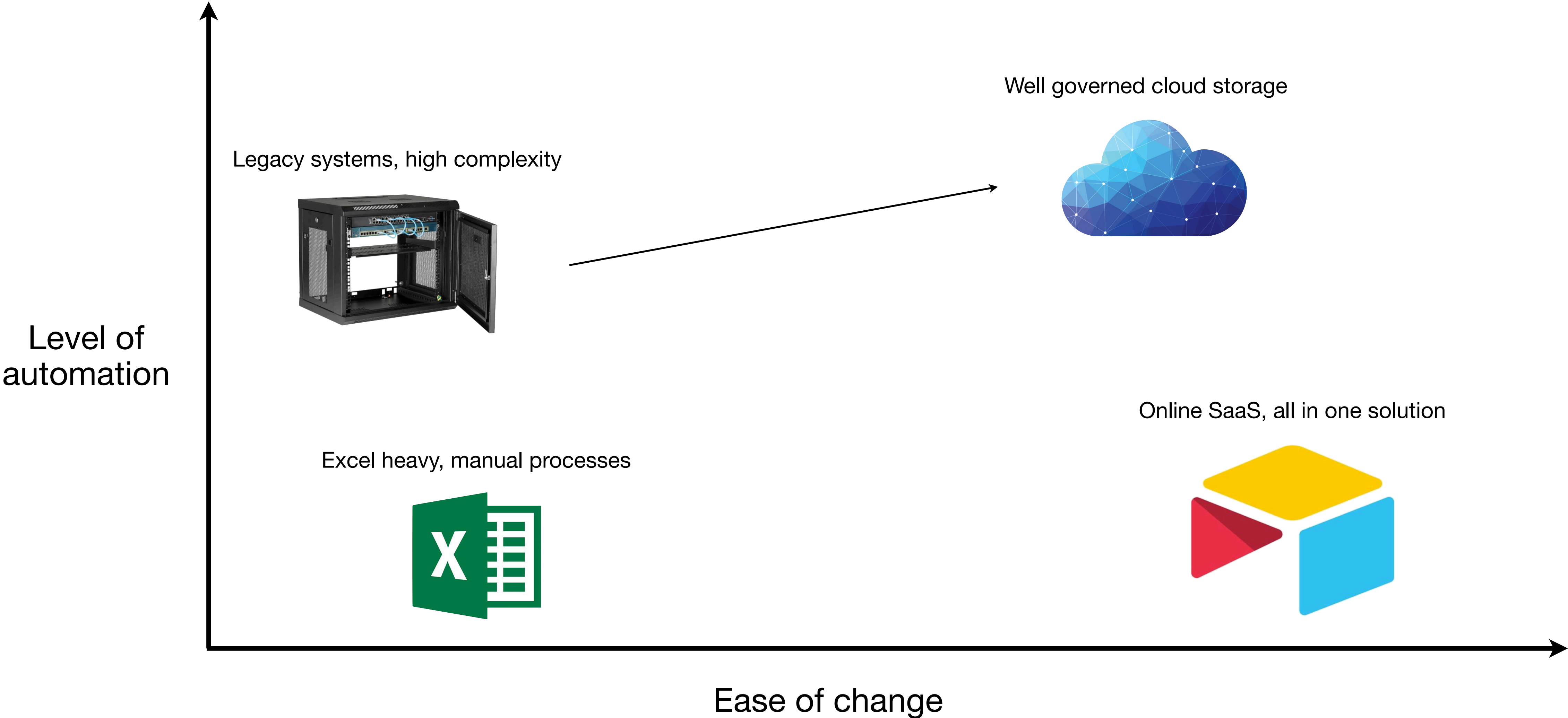
The importance of data architecture
What do we gain from investing in improving our architecture?

- Innovation and competitive advantage

  - You don't want your data architecture to slow you down

- Speed of deployment

  - Imagine if you wanted wanted to follow-fast a competitor; this is only possible if your foundations are the same.

- Cost-effectiveness

  - Costs can skyrocket for storage costs, processing, etc.

  - There is always a trade-off between employing people vs SaaS

    - It makes more sense paying for SaaS the further your USP or core competency lies away from managing data.

- Scalability

  - If your organisation is growing, its important to visualise a path forwards

- Regulatory adherence

  - Data governance sounds like red tape, but done right it can propel you forward

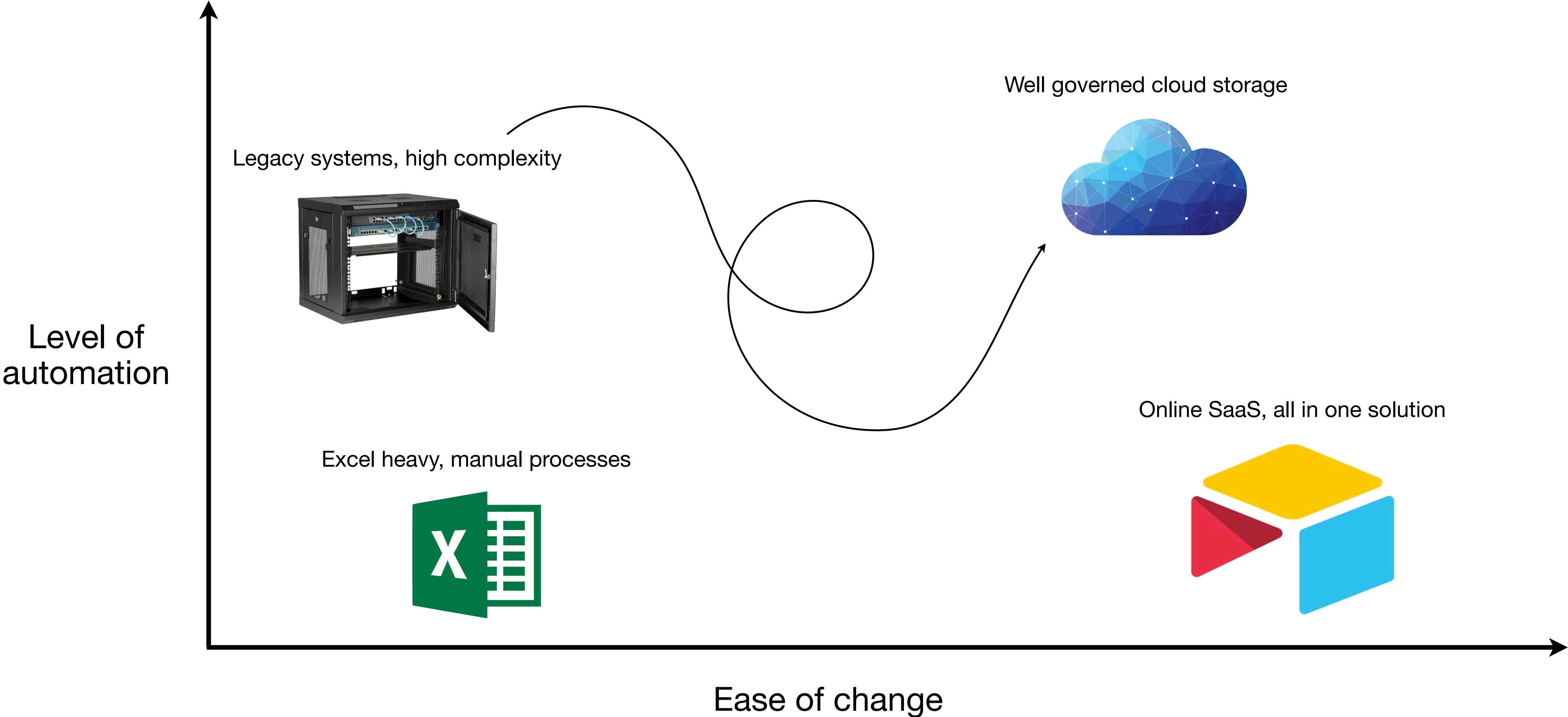 - Meta's $1.3bn fine a year ago (is this the tip of the iceberg?

Where does your organisation sit?

Level of
automation

Ease of change

Well governed cloud storage

Legacy systems, high complexity

Online SaaS, all in one solution

Excel heavy, manual processes

Getting to where we want to be

Well governed cloud storage

Legacy systems, high complexity

Level of automation

Online SaaS, all in one solution

Excel heavy, manual processes

Ease of change

Real pathway

Well governed cloud storage

Legacy systems, high complexity

Level of automation

Online SaaS, all in one solution

Excel heavy, manual processes

Ease of change

Case Study
Moving to the cloud

- Company A is a telecommunications provider that sells phone plans directly to consumer

- They have weighed the pros and cons of moving to a more modern data architecture:

  - Competitors have all moved to the cloud, and have

    - Reported additional profits from giving personalised product recommendations to customers

    - Saved on costs on internal platform teams

  - A new initiative to white-label their phone plans would be possible with this new infrastructure - would have been too difficult to implement on the old

  - The new architecture would allow them to use SaaS products to manage their data catalogue and help with monitoring

Legacy systems, high complexity

Well governed cloud storage

Sounds like a great idea right?

Case Study
It all falls apart

- Existing staff are excited initially about the changes as the old system was already struggling to handle the load, especially with more customers purchasing e-sims through their platform

- The excitement falls apart when they are tasked with maintaining both the existing system and the new one in parallel

- The new system functions the same as the old, but the expertise of how to handle the data is lost as staff turnover increases

- The white-labelled product is delayed for now since the migration has gone over budget and time.

Legacy systems, high complexity

Well governed cloud storage

Case Study
What went wrong?

- Be aware of good-will evaporating with change

- The best time to make a change is before you have to

- Demonstrate the advantages of change

  - Build a PoC of the white labelled product first before running two systems in parallel

- Documentation and a culture of data governance is important

  - Documenting everything after a migration means you don't know what you don't know

Legacy systems, high complexity

Well governed cloud storage

Looking ahead
Trending topics

- Handling streaming data

  - Now that you've got your batch processing jobs down, whats the next step?

  - First, its best to decide whether you have a business case for streaming/real time APIs

    - You likely don't need this for reporting purposes as it is costly in terms of resources and complexity

    - If you want a responsive, data driven customer experience though, this may be required

- Data Lakehouses

  - Seem to be the popular way to scale up a company that has not had the investment into high quality data warehouses, allowing users to self-serve to some extent off raw data.

- Data Mesh

  - Microservices for data governance, ownership and management.

  - Benefits organisations with unmanageably large data teams in a monolithic structure

  - Case study: *Zalando* has moved to the Data Mesh.

- Graph databases

Looking ahead
My take

- Automated monitoring and summary of tables/databases

  - Data governance is tedious, but tools such as data catalogues are still worth their weight in gold.

  - It is still hard to trust the data

    - If I do work, I will still need to make sure there are no duplicates of a presumed primary key, etc.

- Lineage is important

  - With more scrutiny and regulation, tracking data lineage will be more important; and I haven't seen it done well yet

- Agile will hit business teams

  - A lot of businesses work in sprints and believe that they are agile

  - Perhaps the tech teams are, but the underlying business they support is holding onto legacy processes

Any questions?