PROCUREMENT & SUPPLY CHAIN

The Value of the Lakehouse

HOW T-MOBILE ARTICULATED THE BENEFIT OF A MODERN DATA PLATFORM

Why You're Here

To learn how we convinced non-technical leadership that the Lakehouse architecture was the right path forward for T-Mobile data.

Agenda

Intro

Pain Points to Strategy

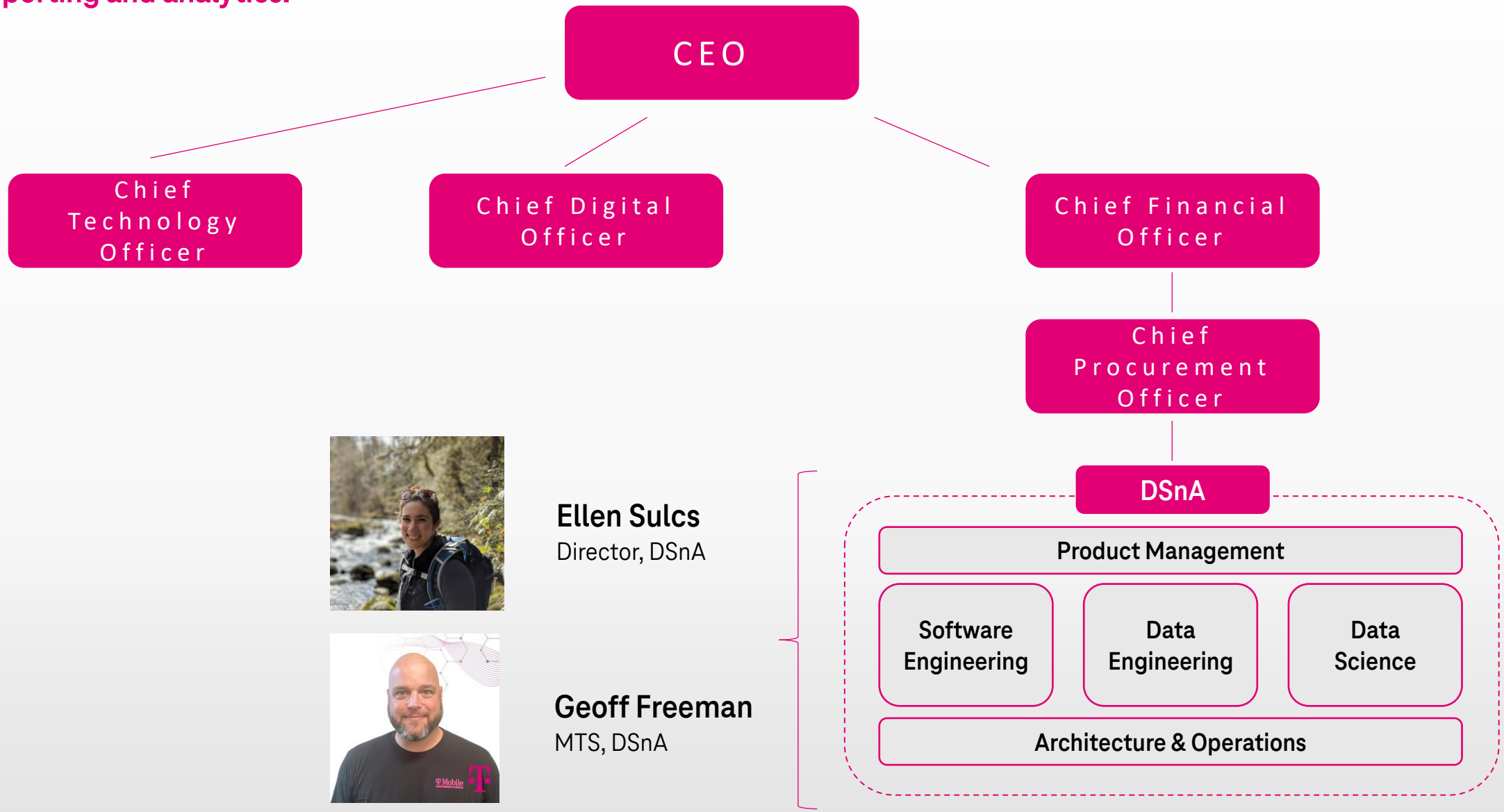
Strategy to Architecture

Managing Concerns and Expectations

Lessons Learned

Who We Are

Technologists in a business focused organization with a mandate to manage data for Procurement & Supply Chain reporting and analytics.



Ellen Sulcs
Director, DSnA

Geoff Freeman
MTS, DSnA

Pain Points to Strategy

The Common Pain Points – Analyst to Exec

I just onboarded a new vendor and I need data from their system

My reports are out of date and won't refresh

Your data looks different than what I see in the system

I need to join my spend data with data from the network build team

Our cloud spend is getting out of hand

I use this data for operational decisions and it's not up to date

I need to share the output of my algorithm with my 3PL

My query keeps failing

Organizing the Pain Points into Platform Strategy

- Your data looks different than what I see in the system
- I use this data for operational decisions and it's not up to date

Data Latency

- I just onboarded a new vendor and I need the data from their system
- I need to join my spend data with data from the network build team
- I need to share the output of my algorithm with my 3PL

Data Mobility

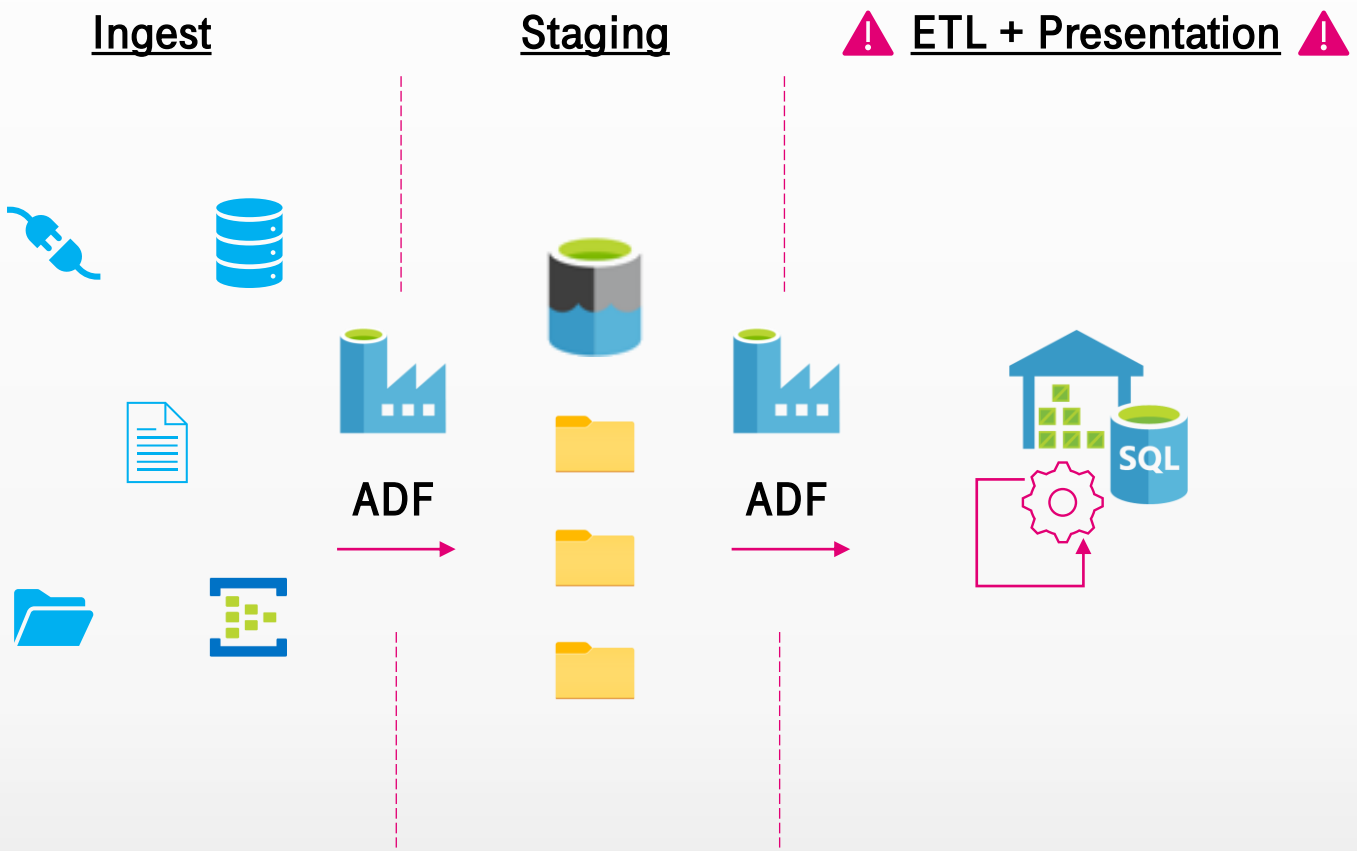
- My reports are out of date and won't refresh
- My query keeps failing

Predictable Performance

- Our Azure spend is getting out of hand

Economies of Scale

The OG Architecture – Why the Pain Points Hurt



DW appliance is the problem:

Predictable performance impacted by shared and constrained storage/compute resources create noisy neighbor issues

Latency caused by performance issues blocking ETLs

Mobility is directly limited by the siloed nature of the DW; sharing requires redundancy.

Economies of scale limited by ADW hardware – bigger hardware is exponentially more expensive

The Data Strategy

Platform Vision: Scalable data management that sets the enterprise standard

Low Data Latency

The **most up to date raw and transformed data** is made available for all consumers and use cases.

- Refresh Cadence
- ETL Processing Time

High Data Mobility

Data is easily **consumed where it resides** and **democratized.** Various organizations and solutions can consume and **share data seamlessly.**

- # of Data Connections (in and out)
- Time to Onboard Data

Predictable Performance

Workload isolation through serverless workloads ensures **consistent query experience,** eliminating “noisy neighbor” problem.

- Data Availability
- Query Failures

Economies of Scale

Ensure a **cost-effective management** of Lakehouse architecture and realize **enterprise data management efficiencies.**

- Infrastructure Spend
- Resource Efficiency

The Data Strategy

Platform Vision: Scalable data management that sets the enterprise standard

Low Data Latency

The **most up to date raw and transformed data** is made available for all consumers and use cases.

- Refresh Cadence
  - ⬇ **75%** *Average refresh cadence from 6hrs to 1.5hrs*
- ETL Processing Time
  - ⬇ **88%** *Complex build times reduced from 4hrs to 30mins*

High Data Mobility

Data is easily **consumed where it resides** and **democratized.** Various organizations and solutions can consume and **share data seamlessly.**

- # of Data Connections (in and out)
  - ⬆ **30%** *55 connection points (41 in / 14 out)*
- Time to Onboard Data
  - ⬇ **80%** *Data onboarding time from 5 days to 1 day*

Predictable Performance

Workload isolation through serverless workloads ensures **consistent query experience,** eliminating “noisy neighbor” problem.

- Data Availability
  - ⬆ **1%** Data availability increase from 98% to 99%
- Query Failures
  - ⬇ **60%** Reduction in query failures from 1.5% to 0.6%

Economies of Scale

Ensure a **cost-effective management** of Lakehouse architecture and realize **enterprise data management efficiencies.**

- Infrastructure Spend
  - ⬇ **30%** $120k Azure spend saved per month
- Resource Efficiency
  - ⬆ **30%** 10 sources added with equivalent staff

Strategy to Architecture

Mapping the Architecture to the Strategy

Lakehouse principle of true separation of compute and storage addresses all pillars of our strategy.

High speed storage – noisy neighbors – data movement = faster data

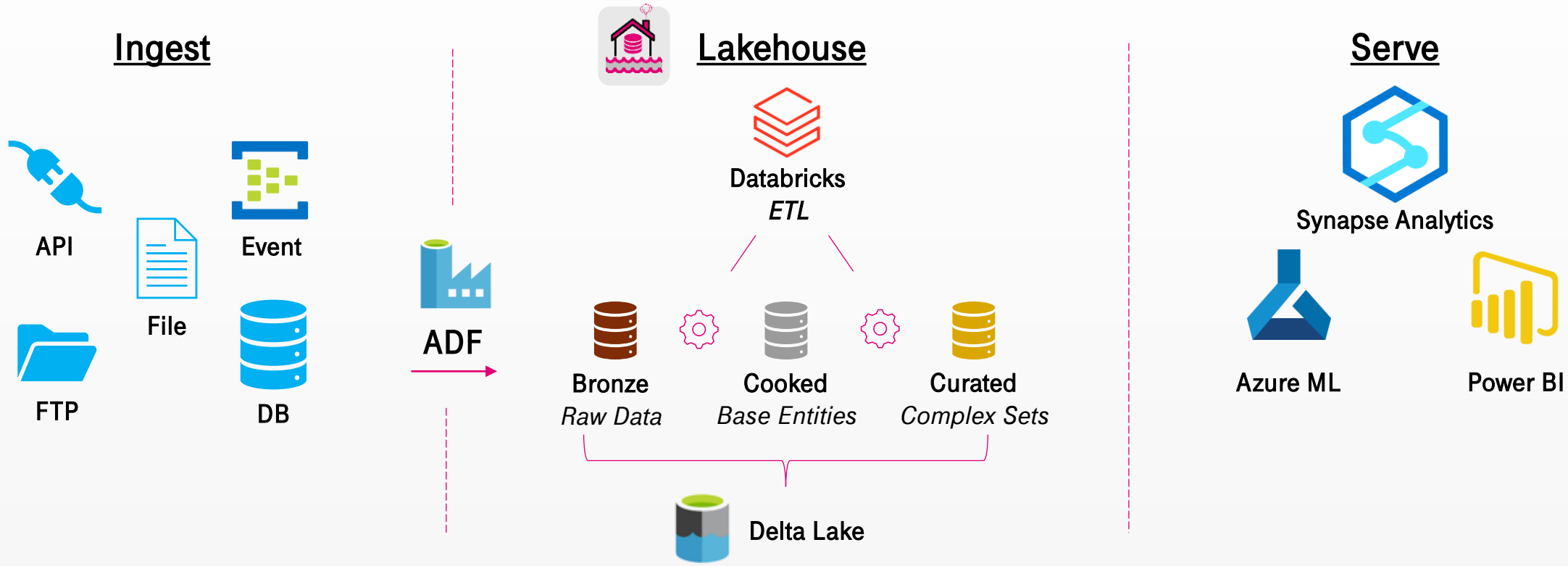
Query data where it resides – if it's in the data lake, it can be queried. No more silos.

No more noisy neighbors hogging resources, no more waiting on locked objects.

Commodity hardware is cheap to add, reduction in data duplication reduces TCO of data.

Lakehouse Deep Dive

Lakehouse provides data ingest, exchange, centralization and transformation for both operational and analytics needs.



Generic ingest simplifies onboarding

Open source (delta) format makes data exchange simple and extensible

Elimination of contention reduces latency

Increasing capacity scales linearly

Separation of storage & compute ensures predictable performance

Business partners can bring their own compute

Managing Concerns and Expectations

Feedback on the Strategy

How much time?

Not a relevant concern
Every ETL migrated reduces cost and improves performance. Even if we never finish, we'll see benefits.

We gave ourselves 1yr to migrate 30+ systems (2k+ tables) of data to Lakehouse. Primarily lift and shift.

How much money?

No resource costs/additions
Rolled into the reserved 30% capacity for tech debt reduction

Minor, initial infrastructure increase
Planned for small spike (~15%) prior to migration of first ETL chunk

We never saw an initial spike in cost.

What are the user impacts?

A change in the database connection used
Serverless endpoints ensure compatibility across existing suite of reporting tools (SSMS, Alteryx, PBI, …)

What are the security implications?

Improved security with process driven access barriers
Use of AAD groups (or IAM roles in AWS, etc) is industry best practice. Using the same groups across all fabrics makes it easy to understand, audit, and maintain.

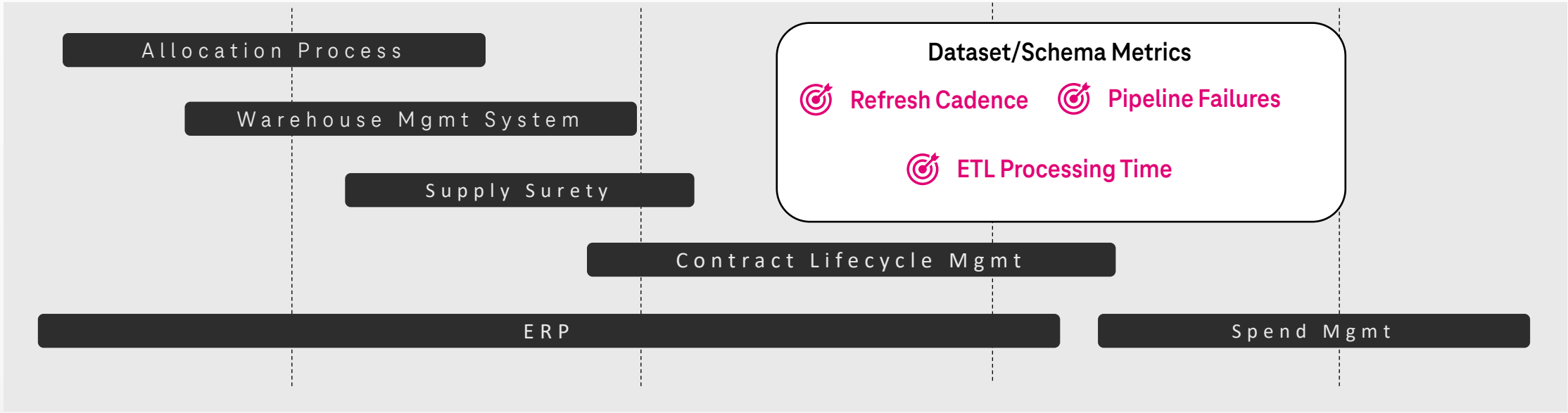
Make barriers to access process driven, instead of technology driven

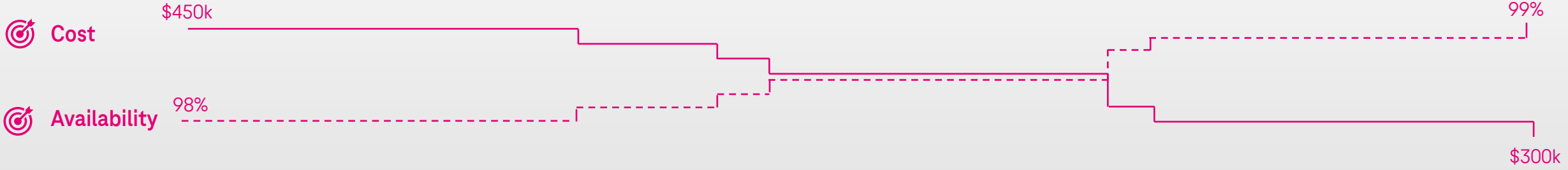
Supporting the Migration

Prioritize based on metrics and business value to demonstrate immediate incremental value.

Top Priority: 5G Build Allocation Process

$20M loss per allocation miss

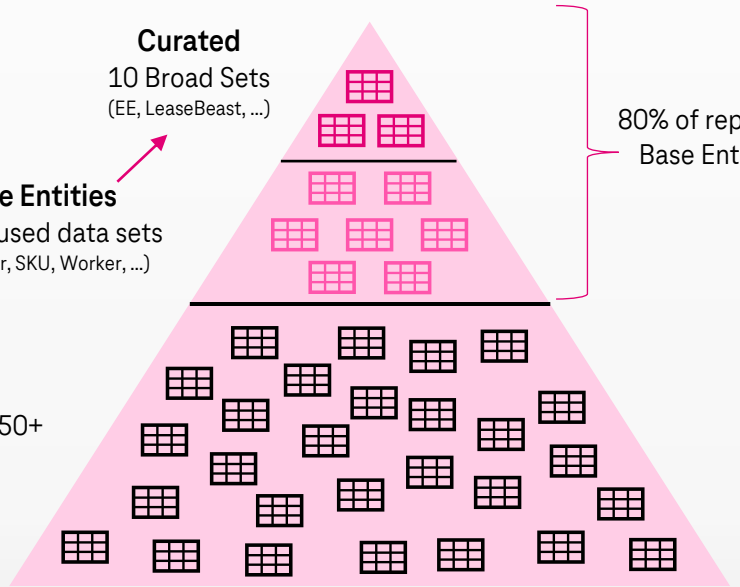




Where We Are Today

In 2022, we used the Lakehouse to reliably contain the chaos. In 2023, we seek to simplify what exists. Here's what we're doing internally:

Focus on Data Modeling

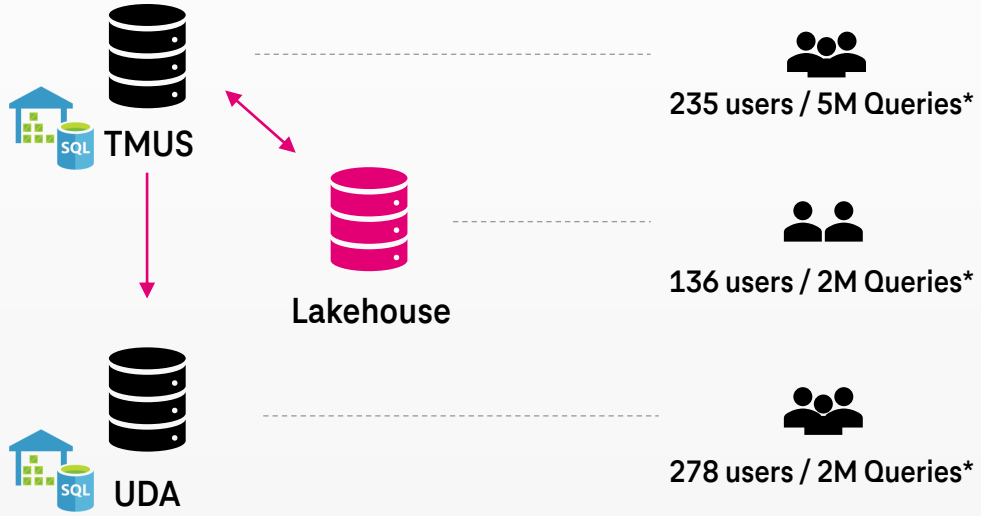


Goal:
80% of reporting comes from Base Entities and Curated Sets.

Current State:
80% of reporting comes from Raw Data.
Curated Data sets are built directly from Raw Data.

Unify Compute Experience

Goal: Push it all to the Lakehouse



Tech Benefit:

- ☑ Improved security posture
- ☑ Enhanced data transparency
- ☑ Increased data observability

User Benefit:

- ☑ Minimal user upskilling required
- ☑ Consistent handling of any query type
- ☑ Compatible with existing toolset

**Users & Query Volume from Q1 '23*

Lessons Learned

Quantifying the Lakehouse Success

The Lakehouse exceeded our expectations.

Low Data Latency

The **most up to date raw and transformed data** is made available for all consumers and use cases.

- Refresh Cadence
  - ⬇ **75%** *Average refresh cadence from 6hrs to 1.5hrs*
- ETL Processing Time
  - ⬇ **88%** *Complex build times reduced from 4hrs to 30mins*

High Data Mobility

Data is easily **consumed where it resides** and **democratized.** Various organizations and solutions can consume and **share data seamlessly.**

- # of Data Connections (in and out)
  - ⬆ **30%** *55 connection points (41 in / 14 out)*
- Time to Onboard Data
  - ⬇ **80%** *Data onboarding time from 5 days to 1 day*

Predictable Performance

Workload isolation through serverless workloads ensures **consistent query experience,** eliminating “noisy neighbor” problem.

- Data Availability
  - ⬆ **1%** Data availability increase from 98% to 99%
- Query Failures
  - ⬇ **60%** Reduction in query failures from 1.5% to 0.6%

Economies of Scale

Ensure a **cost-effective management** of Lakehouse architecture and realize **enterprise data management efficiencies.**

- Infrastructure Spend
  - ⬇ **30%** $120k Azure spend saved per month
- Resource Efficiency
  - ⬆ **30%** 10 sources added with equivalent staff

Lessons Learned

Enterprise tools and applications are the limiting factor in data latency and cost management.
Business platforms struggle to exchange large data with NRT capability. Change data capture is not the norm.

A data model is still valuable.
Data models aren't just for performance improvements. We can reliably contain an infinite amount of data, but a model that directs users to tried and true sources and curated sets is invaluable.

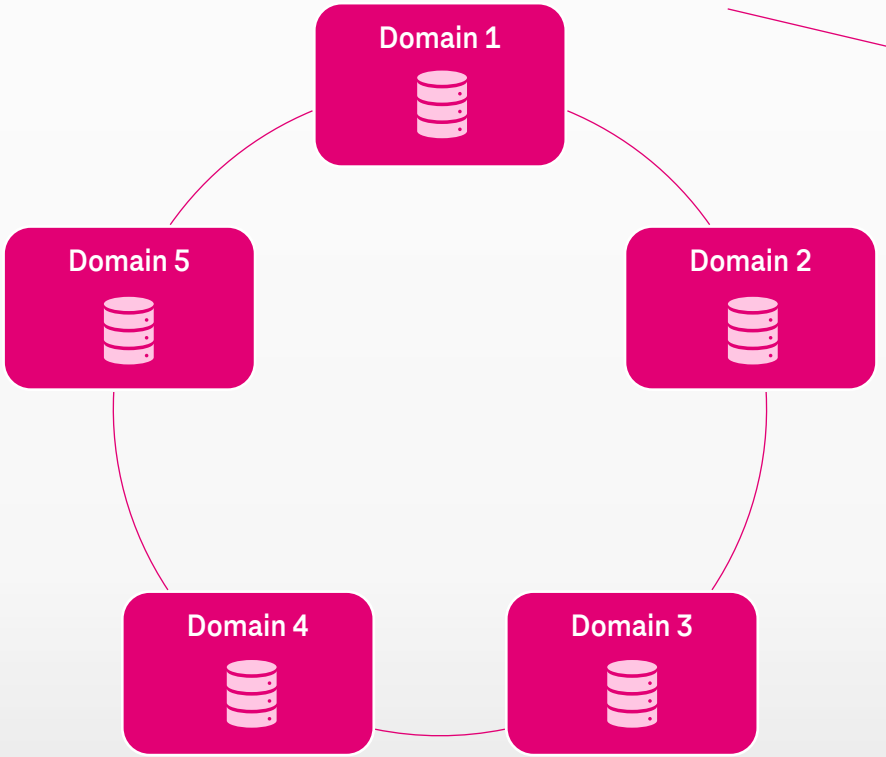
Old habits die hard.
Don't underestimate the culture investment and the eventual need for a top-down push to migrate users to the Lakehouse.

Now What?

Taking the Lakehouse architecture enterprise wide: networked Lakehouses with shared governance platforms.

Networked Lakehouses

Domain data managed by BU-embedded data teams



Shared Governance Toolset

Simple data discovery and access experience

Automated Access Management
Custom service over AAD API

MDM/RDM Capabilities
Through standard SaaS solutions

Data Catalog
User-friendly data discovery and context

Data Quality
Centralized health monitoring

End to End Lineage
Leveraging openlineage.io and Unity Catalog

Interested in the Specifics?

Check out some of our publications:

Articles/Publications:

Why a Delta Lakehouse?
https://medium.com/tmobile-dsna/why-delta-lakehouse-27d325c18c42

What is a Lakehouse?
https://medium.com/tmobile-dsna/what-is-a-data-lakehouse-8bdad0f7a67a

Case Study – How Databricks Photon Saves us 25% on Compute
https://medium.com/tmobile-dsna/how-databricks-photon-saves-us-25-on-compute-9941f178d7b7

Case Study – Connecting Millions of Customers with the Largest 5G Network
https://www.databricks.com/customers/t-mobile

Case Study – T-Mobile Supports 5G Rollout with Azure Synapse Analytics and Power BI
https://customers.microsoft.com/en-us/story/1581506762107713013-t-mobile-telecommunications-power-bi

Recordings:

D3L2: Implementing a Data Lakehouse for Improved Data Science and Analytics
https://www.youtube.com/watch?v=IJlvlWa_X1o

D3L2: Why Did We Migrate to a Data Lakehouse on Delta Lake
https://www.youtube.com/watch?v=1V_u9hboI-Q

Q&A