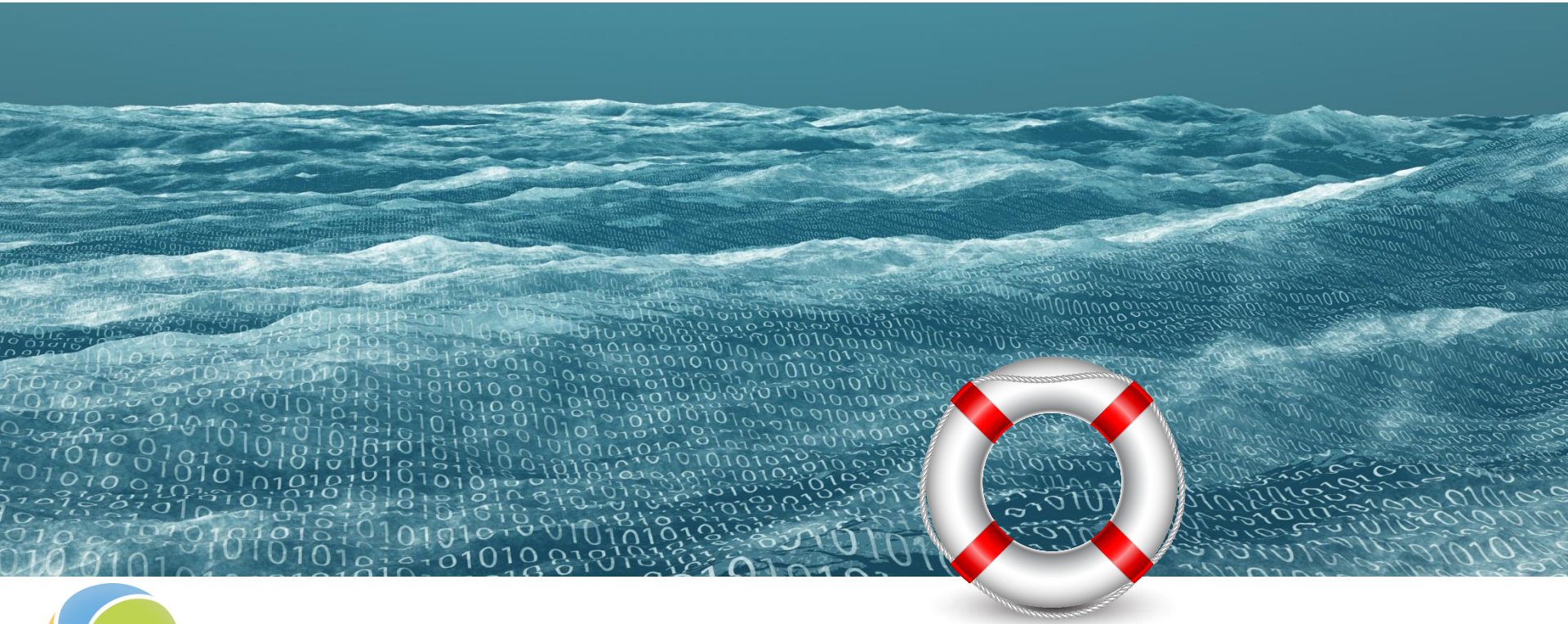
BARCLAYS | Data Optimisation and Simplification



MDE

Metadata Patterns Decide Who Lives and Who Dies

Delivered by TERADATA

Powered by Ab Initio

Objectives & Priorities

1. Accurately analyse data lineage and usage/queries end to end across ALL PLATFORMS

2. Use insights from (1) to prioritise
   a. Which DBs are addressed first
   b. Which new platform the DB goes onto

3. Automate the production of Info Packs to support discussions with DB business owners – persuade them to agree new platform decision

Two Parallel Programmes: Objectives

New Estate Build

Data Access

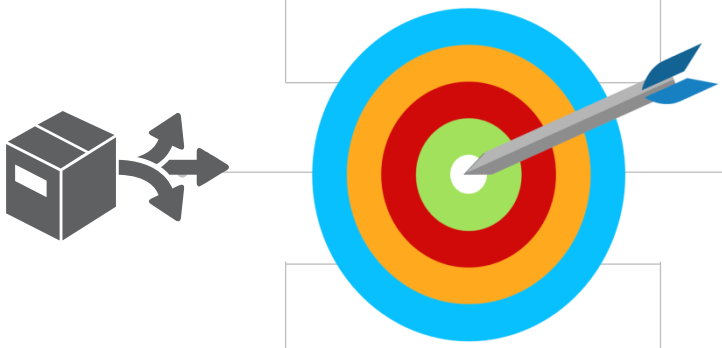
Single Integrated Layer

Data Loads

Cleaner patterns with clear lineage

Data Complexity - Managed

Access layers geared towards Business areas



Current Estate Cleanup

Data Duplication

Identify and remove duplicate data. Optimise investments

Data Quality

Profile & Report on Data Quality issues to enhance confidence in BIW 2.0

Data History

Identify Hot & Cold data items and store appropriately. (Archive, Near line, Real time)

Challenges

- At Barclays, we have >2 Petabyte estate in our BI area. Like the ocean, the data evolves and changes daily

Usage patterns change – sometimes predictably, sometimes not

- With over 50 different business units using the environments – and a recent reorganisation – the need to distill and categorise quickly and succintly was critical
- But what do you do with an ocean of data ?

Question: How do you drink a 2 Petabyte ocean?
Answer: Segment by segment

- How to get it into the buckets and quantify quickly so that we could make decisions ?
- Pattern based analysis - how does data flow and usage change over time, and how does that result in carving data up into different buckets
- We used Teradata MDE* for lineage, usage analysis and segmentation
- Barclays in-house knowledge gave context, then we agreed policy decisions for each segment
- Finally we began the process of moving to a new generation of BI at Barclays

*powered by Ab Initio

BARCLAYS

So what did we do . . .

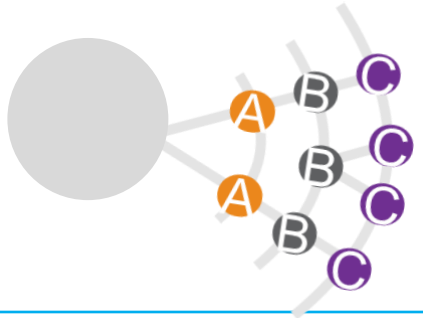
. . . And how did we start?

Phase 1: Pilot

EXPECTATION: Take slice of ocean, ask Teradata MDE team to determine patterns and recommend segments for future action i.e. what lives (goes to BIW2.0 or Hadoop), or Queryable archive, or dies (quarantined then deleted)

Pilot area chosen because well understood & Simple. **Only 5 Databases** …

REALITY: MDE produced end to end data lineage to show how data fragmented across 74 Databases not 5. Some on Teradata, most on MS SQL Server, then data fed back to Teradata

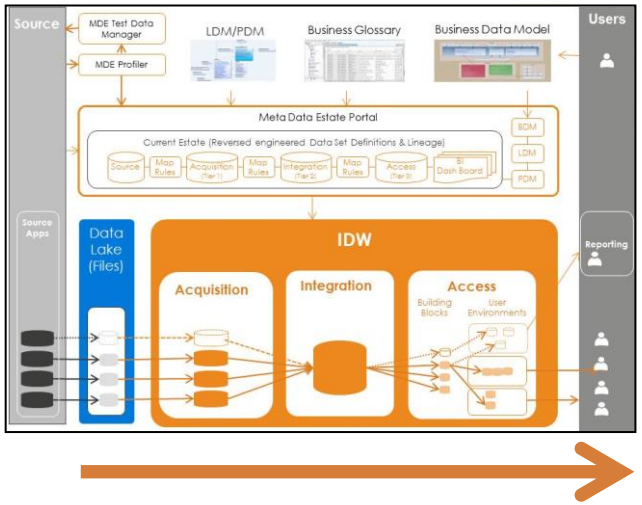


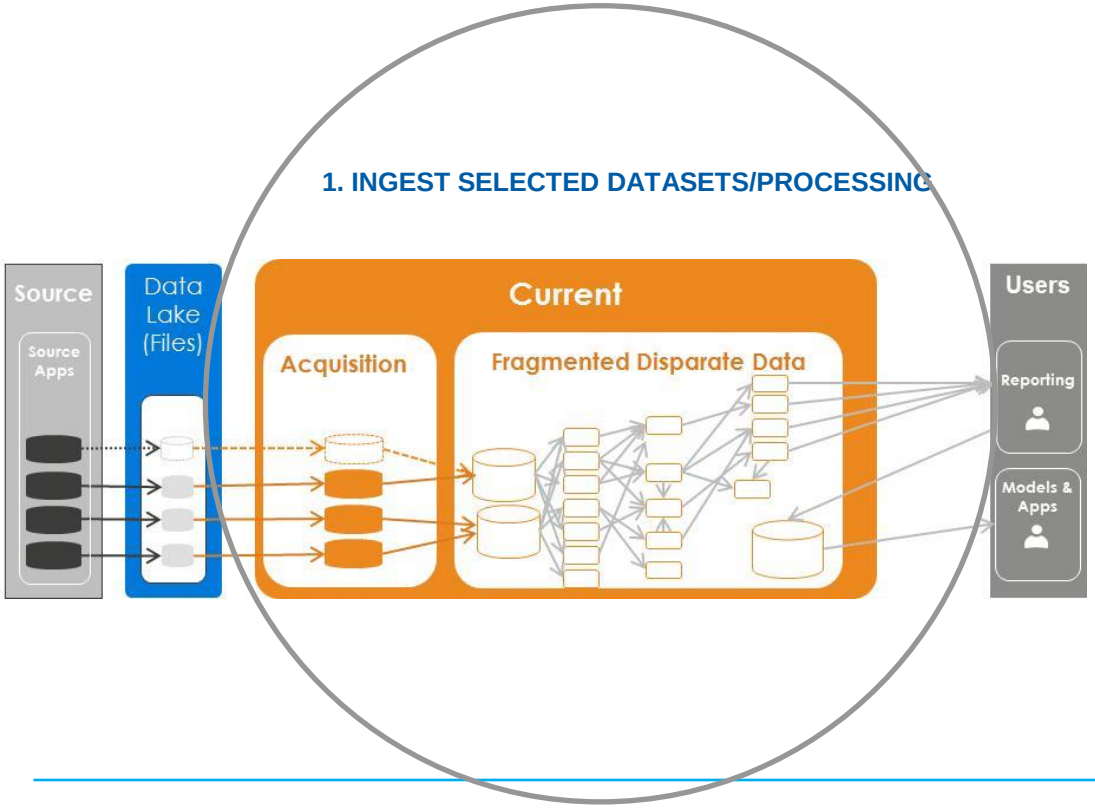
Reality: 74 Databases fragmented across different platforms

Systematic Approach:

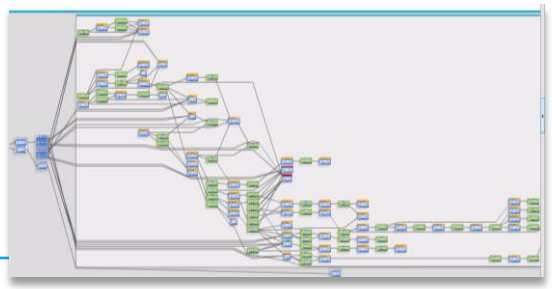
Step A: Left To Right Ingest

- Ingest metadata of current estate
- Derive Lineage - reveal complexity
- Identify breaks in Lineage
- Iterate





Example Lineage

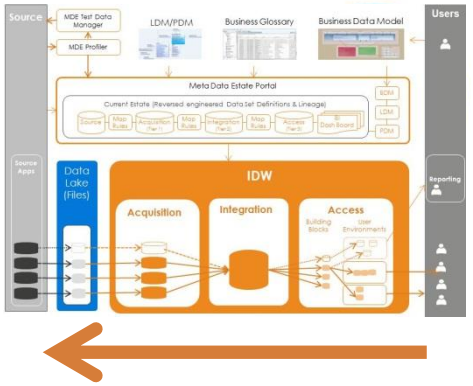


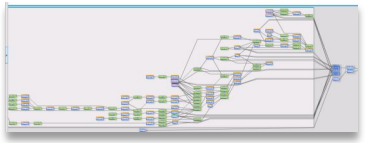
BARCLAYS

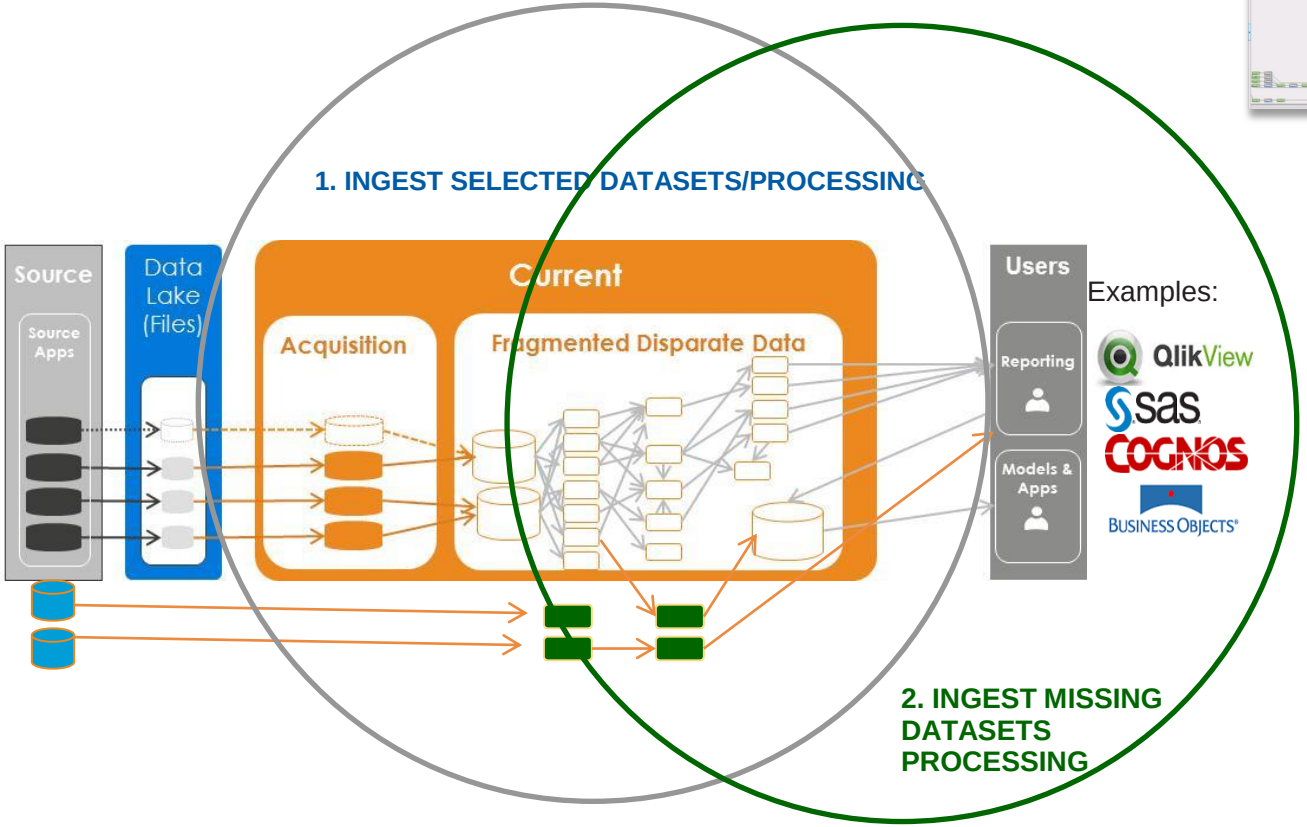
Systematic Approach:

Step B: Right to left Analysis

- **Formal Reports to Sources – needle in a haystack**
- One business entity populated from multiple places.
- Multiple business entities populated from the same place



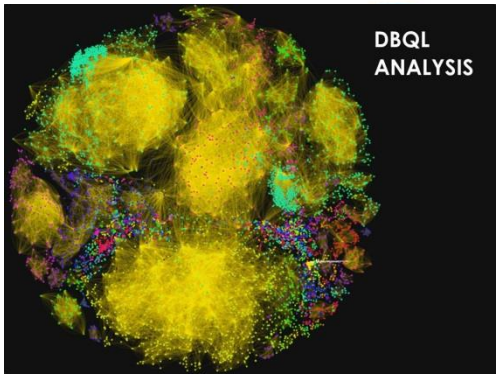




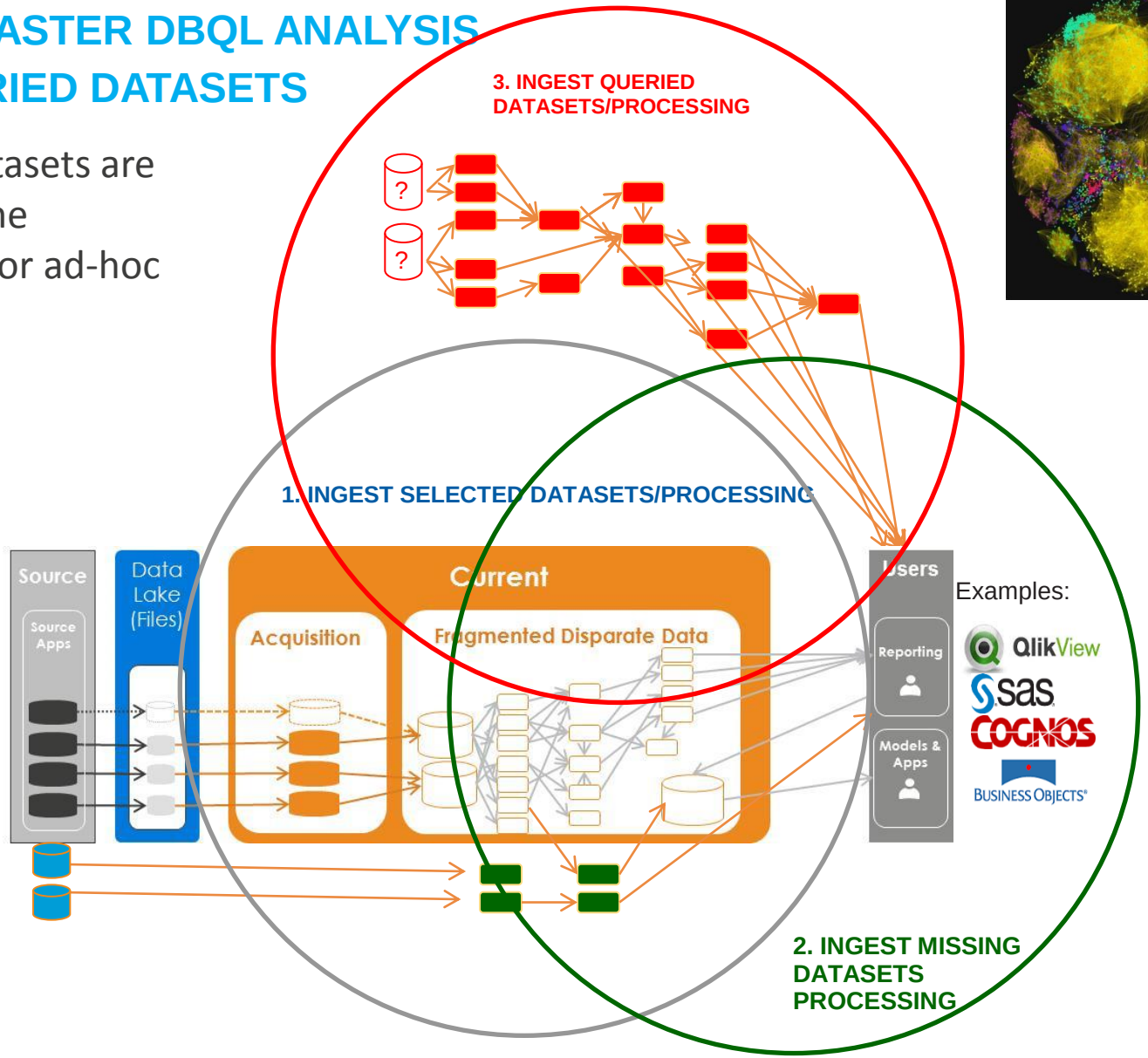
BARCLAYS

STEP C: ASTER DBQL ANALYSIS OF QUERIED DATASETS

Which datasets are used by the business for ad-hoc queries?



Example ASTER DBQL analysis



STEP D: PROFILE/DQ ANALYSIS

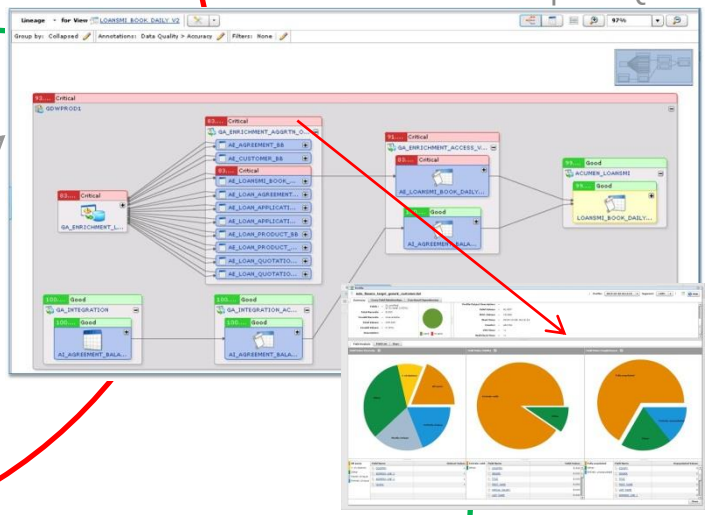
Only executed on datasets that are used and have uncertain DQ

Steps A, B, C: incremental ingest to populate integrated end to end Metadata

Step C: ASTER DBQL analysis of queried datasets

3. INGEST QUERIED DATASETS/PROCESSING

Example DQ



STEP D: PROFILE/DQ ANALYSIS

(ON SELECTED DATASETS

Step A: Left to right ingest

1. INGEST SELECTED DATASETS/PROCESSING

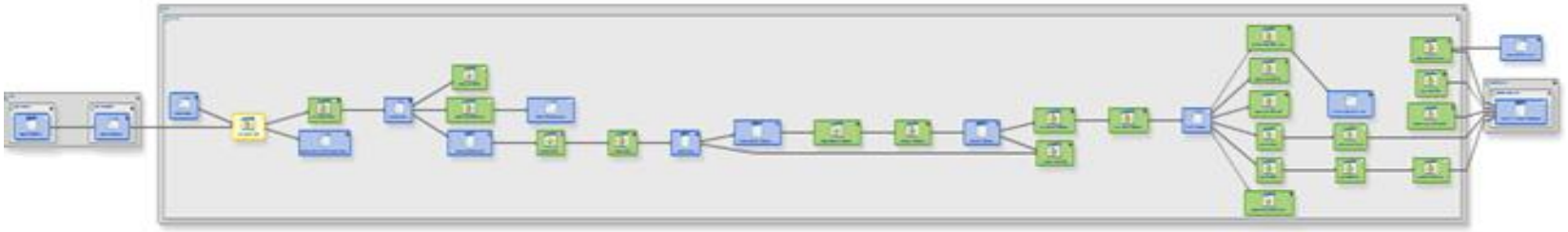
Step B: Right to left analysis

2. INGEST MISSING DATASETS/PROCESSING

BARCLAYS

Results: Source to Target (Data Lineage Samples)

Teradata → SQL Server → Teradata



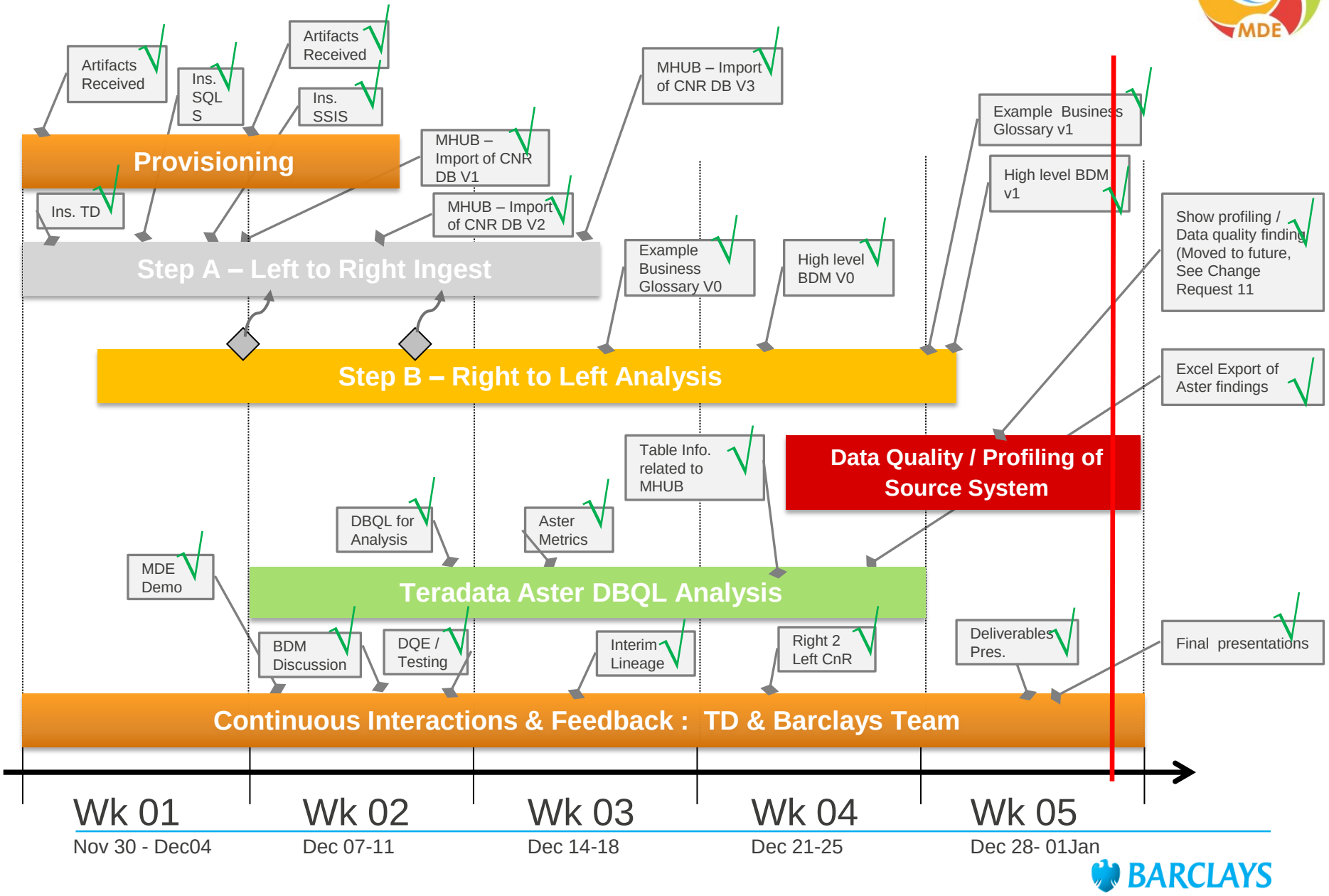
Data Lineage	Estate Clean Up	New Estate Build
Helps to identify end to end source and targets	✓	✓
Acts as a trusted source of truth for data		✓
Capable of showing transformations & business logic details.		✓
Helps perform impact analysis of upstream/ downstream Sys.	✓	✓
Data quality overlay possible		✓

Oracle → Teradata

Source is a file System feed into SQL Server

IBM Mainframe → Oracle

Phase I - All done in a 5 Week Timeline

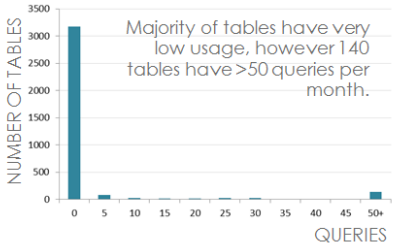


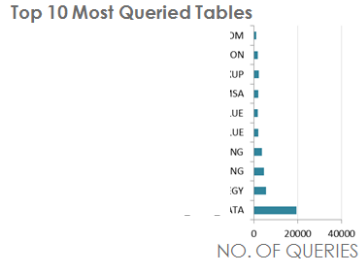
Phase 1 – some example deliverables

5 DB's: DB AxY example

13.4TB SIZE | 3,531 TABLES | 282,880 TOTAL QUERIES | 329 USERS

QUERIES (monthly)	53,400 DQL QUERIES	43,716 DDL QUERIES	185,595 DML QUERIES





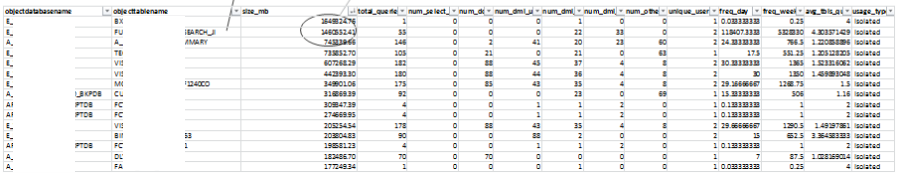
Table's with No Selects

Description	Category	PPT slide	Excel name
Tables with no usage in 3 mths	Data Duplication – Optimise Investments	Ref: Q 2	Tables with No Selects – no usage

Ex: Tables with No Selects

2 unique users
No selects

1.4 TB



Data Lineage – different areas

Teradata → SQL Server → Teradata (Feed – Dialler)

Data Lineage	Estate Clean Up	New Estate Build
Helps to identify end to end source and targets	✓	✓
Acts as a trusted source of truth for data		✓
Capable of showing transformations & business logic details.		✓
Helps perform impact analysis of upstream/ downstream Sys.	✓	✓
Data quality overlay possible		✓

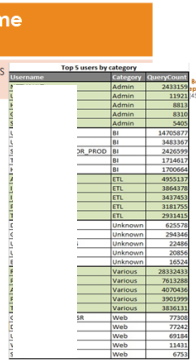
SQL Server → Teradata (Feed - Total View)

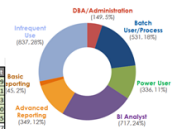
Source is a file System feed into SQL Server

SQL Server → Teradata (Feed - BBC)

User Metrics

Description	Category	PPT slide	Excel name
All user metrics by categories	User categories	Ref: Q 3	All User Metrics



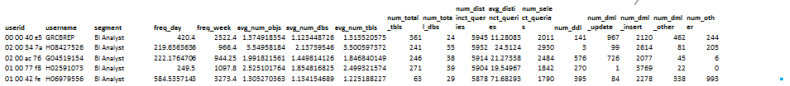


Ex: Advanced Reporting

E01909598
158 queries a day?

All types of queries

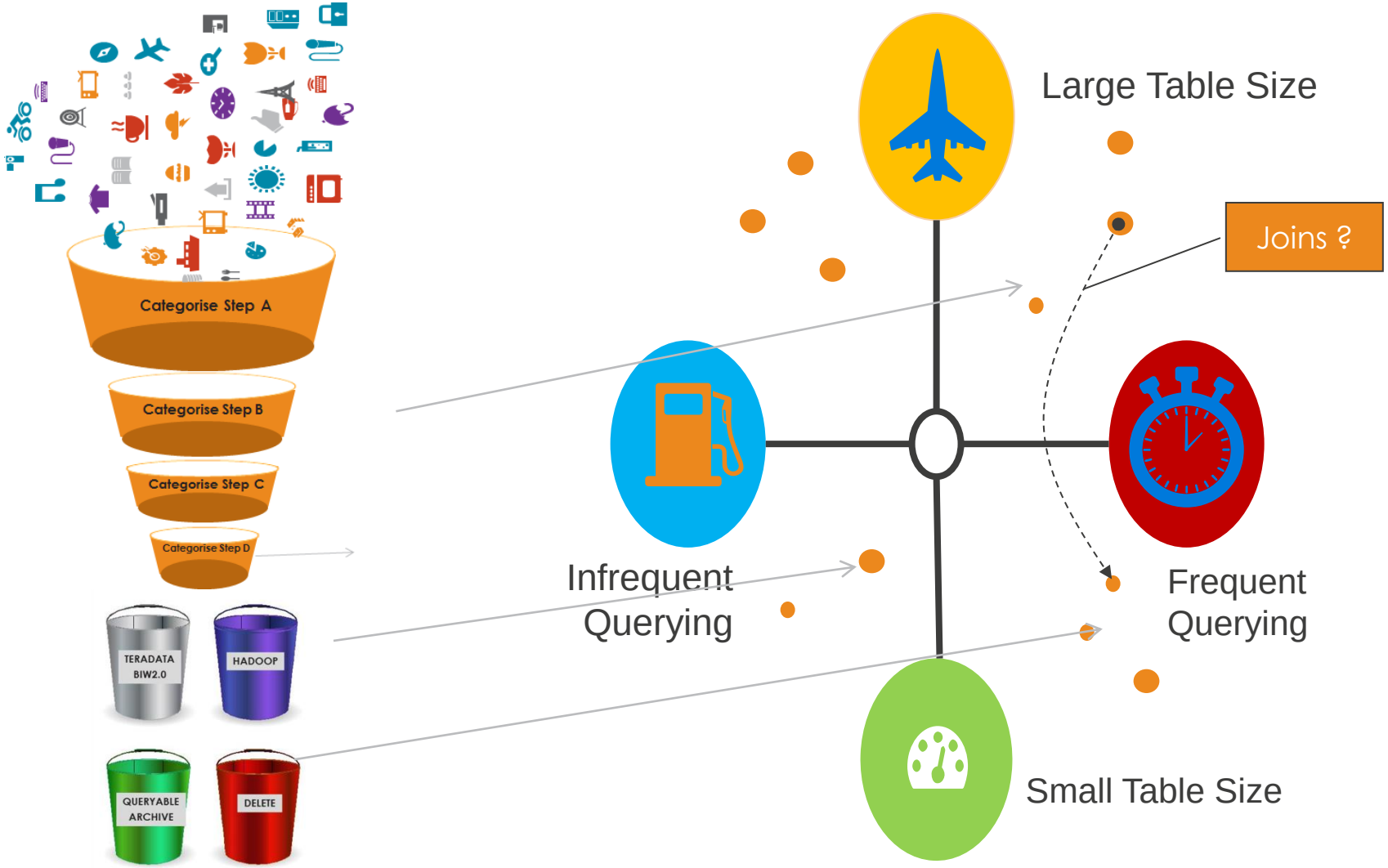
Ex: BI Analysts



Phase 2: Remaining 2 Petabytes in 5 Week Timeline

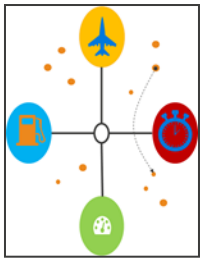
REPEATABLE PROCESS TO APPLY POLICY

CATEGORISATION / SEGMENTATION
by Area to support User Meetings / Adoption

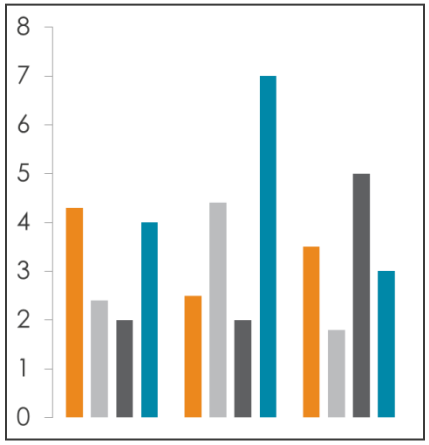


ACTION STEPS:
Apply rules, Document, Review, Decide

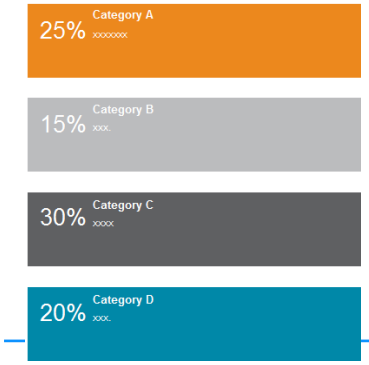
Category

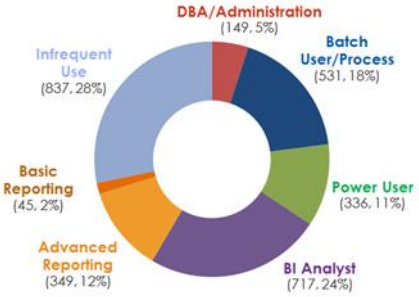


Segment



Rules by Segment





Decision Tree

- **Apply Rules**
  Type of Use
  Type of User
  Criticality
  Business Area

- **Document**
  Informed Packages to Review with Business

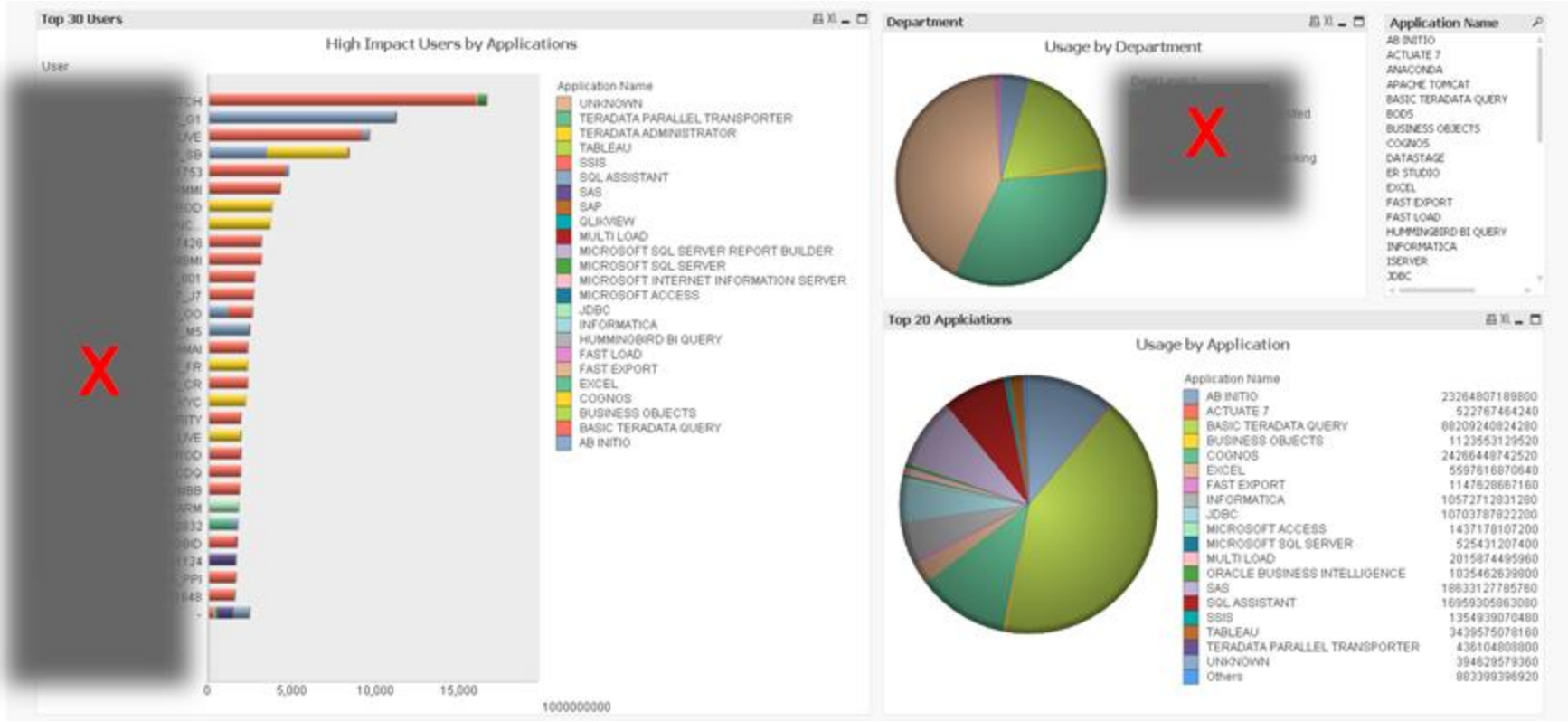
- **Review**
  Capture Feedback to Close the Loop

- **Decide**
  Take action

Decisioning Loop By Subject Area

BARCLAYS

Phase 2 – some example deliverables



So why does the business care?

- Accurate datasets with end to end data lineage baked-in (as required by regulator e.g. BCBS)
- Understand business usage/value of data assets
- Maintaining a “general ledger” of data assets
- “Non-productive” footprint removed to make space for new datasets
- Overnight batch reduced (not populating “non-productive” footprint)

Thank you for your time

BARCLAYS

For more information on Metadata Driven Estate visit Teradata and Ab Initio at the Ab Initio stand, or attend the Innovation Hub:

Presentation Date & Time:	Tuesday 19 April, 16:30-16:55
Presentation Zone:	Zone D
Presentation Title:	**Using Metadata to Accelerate Delivery**
Speakers	Elaine Fletcher, Partner, Teradata UK Professional Services
	Damian Worsdall, Technical Account Manager, Ab Initio Software