

PhUSE SDE Ridgfield
The Conformed Data Model
An Innovation Step for the Aspiring Clinical Data Scientist

Commodity Prices Table:

	Sale	Buy	Grow
Gold	\$647.00	\$694.51	39.80%
Platinum	\$941.00	\$1,061.93	37.20%
Silver	\$774.00	\$1,061.93	37.20%
Copper	\$516.00	\$537.76	36.00%
Steel	\$443.00	\$537.50	19.80%
Beryllium	\$743.00	\$754.39	1.50%
Manganese	\$598.00	\$754.39	33.50%
Aluminum	\$299.00	\$384.21	18.00%
Chrome	\$686.00	\$727.17	9.30%
Nickel	\$421.00	\$453.84	7.80%
Bauxite	\$340.00	\$453.84	32.00%

Projected sales of main products in 2013:

Product	Q1	Q2	Q3	Q4
Product A	701	719	786	760
Product B	429	583	403	558
Product C	305	413	294	558
Product D	295	240	295	558

Percentage Impact on the Securities Markets:

Division	Impact
TV division	18%
FRT division	21%
SEC - BUYER	15%
SEC - FUNDS	15%

Retail sales levels:

City	Q1	Q2	Q3	Q4
Birmingham	1460.0	1460.0	1460.0	1460.0
Leeds	1460.0	1460.0	1460.0	1460.0
London	1460.0	1460.0	1460.0	1460.0

Abstract

Clinical development has made huge efficiencies over recent decades, (eCRF and CDSIC standards to name a few).

However, just as we're reaching the summit of traditional database management techniques we're facing another step-change in our journey – the big data mountain.

The methodologies we're currently using cannot cater for these vast and varied data sources, without a huge increase in resource - which is something most BioPharma companies haven't provisioned for. So how can we bridge the gap between current database management techniques and faster more innovative digital technologies?

This presentation discusses a practical step-by-step approach for Data Analyst to begin to transform themselves into Data Scientists; bringing with them cost-saving efficiencies for clinical development and career-enhancing skills in the realm of digital innovation.



Guy Garrett CITP MBCS
Managing Director
Achieve Intelligence

Guy is Managing Director of Achieve Intelligence, a consultancy offering Enterprise Architecture and Information Strategy solutions to the BioPharma, Finance & Retail industries.

He is a guest lecturer at several UK universities, teaches the SAS EMEA: Data Warehouse Architecture & Design Principles course, and is co-leading the PhUSE Data Engineering Project.

The Conformed Data Model



[5 mins]

Past



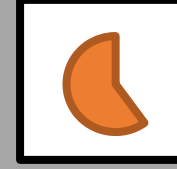
[5 mins]

Present



[5 mins]

Future



[5 mins]

Q&A

The Conformed Data Model



[5 mins]

Past



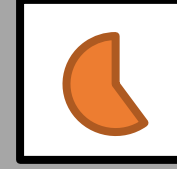
[5 mins]

Present



[5 mins]

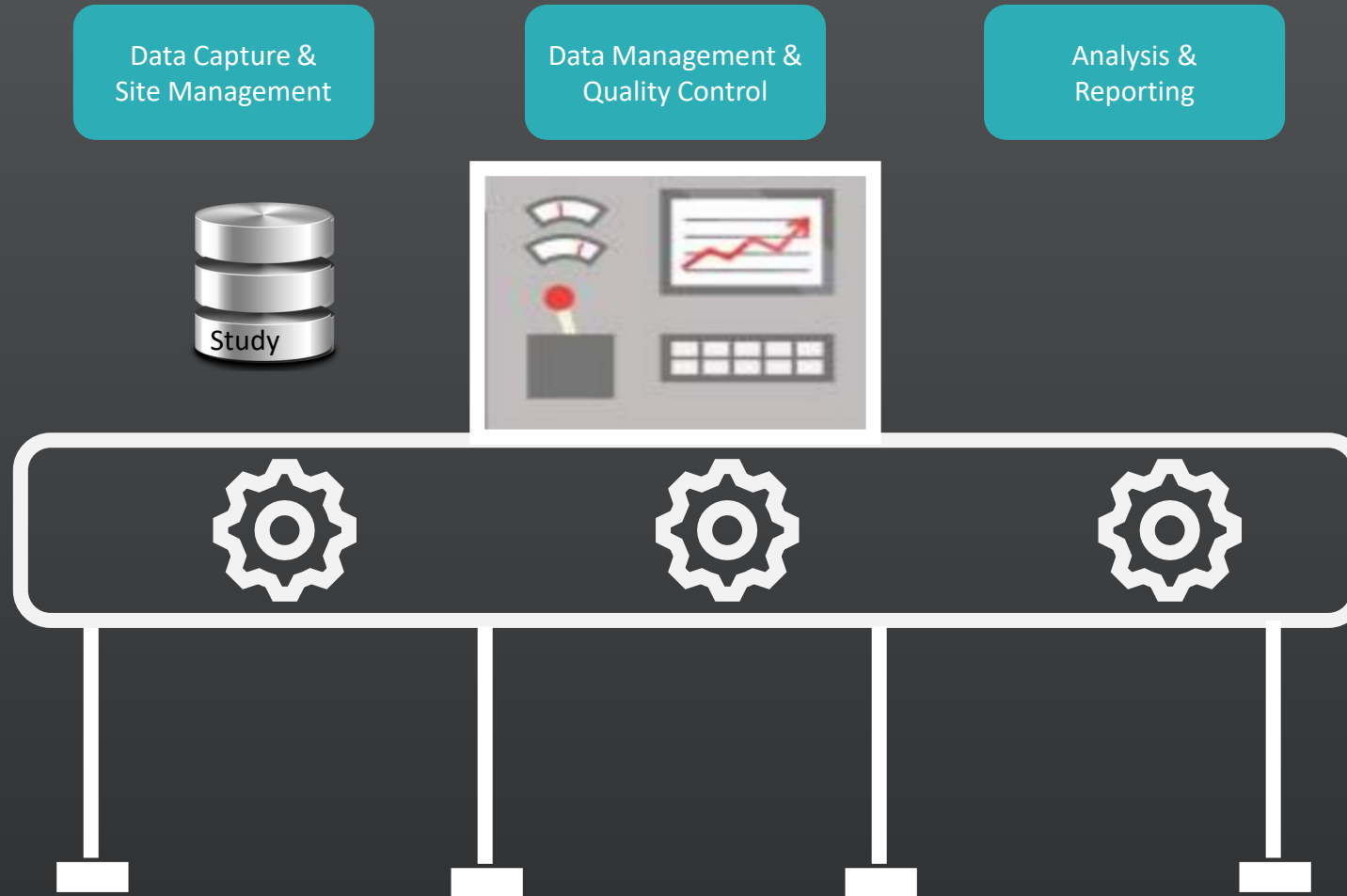
Future



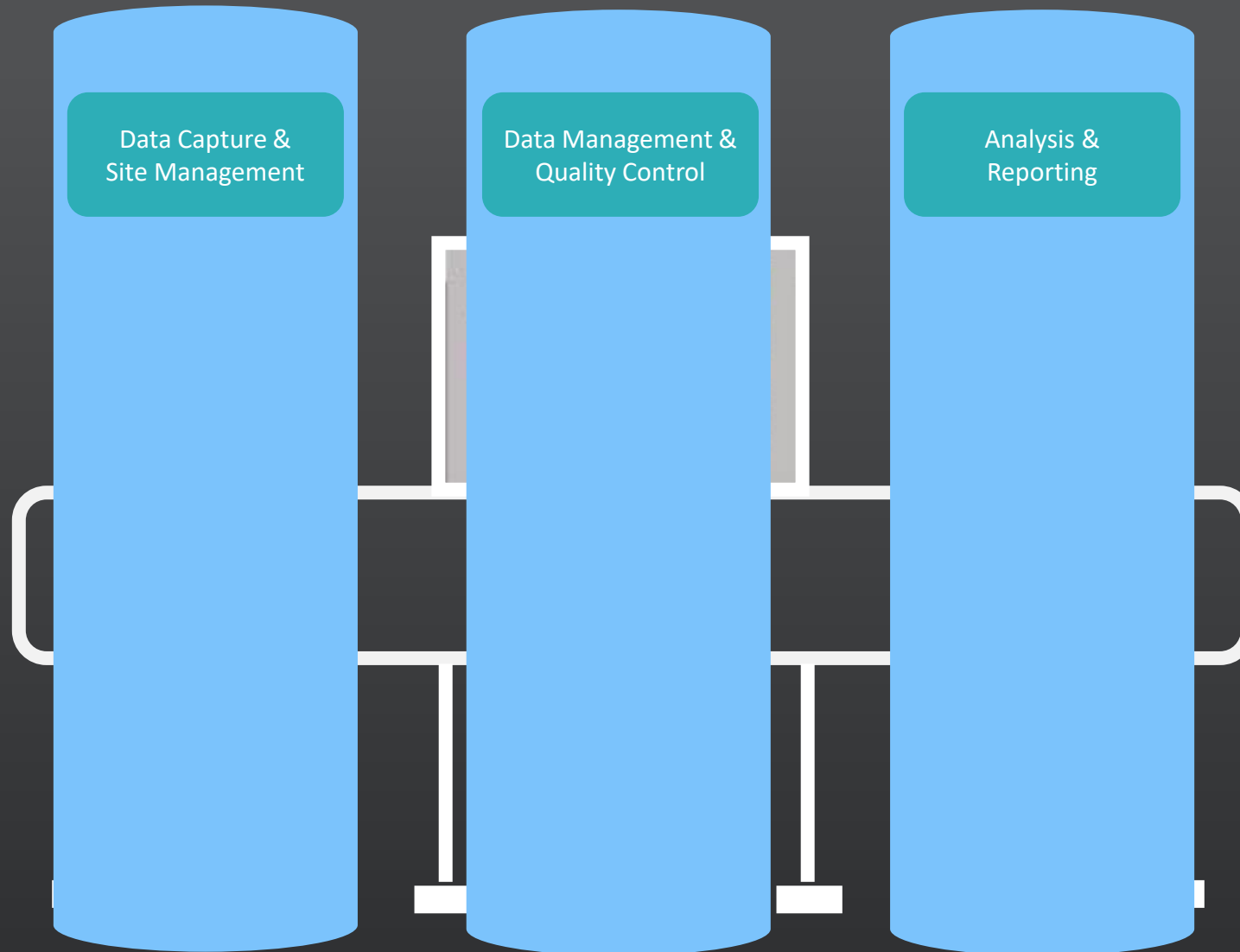
[5 mins]

Q&A

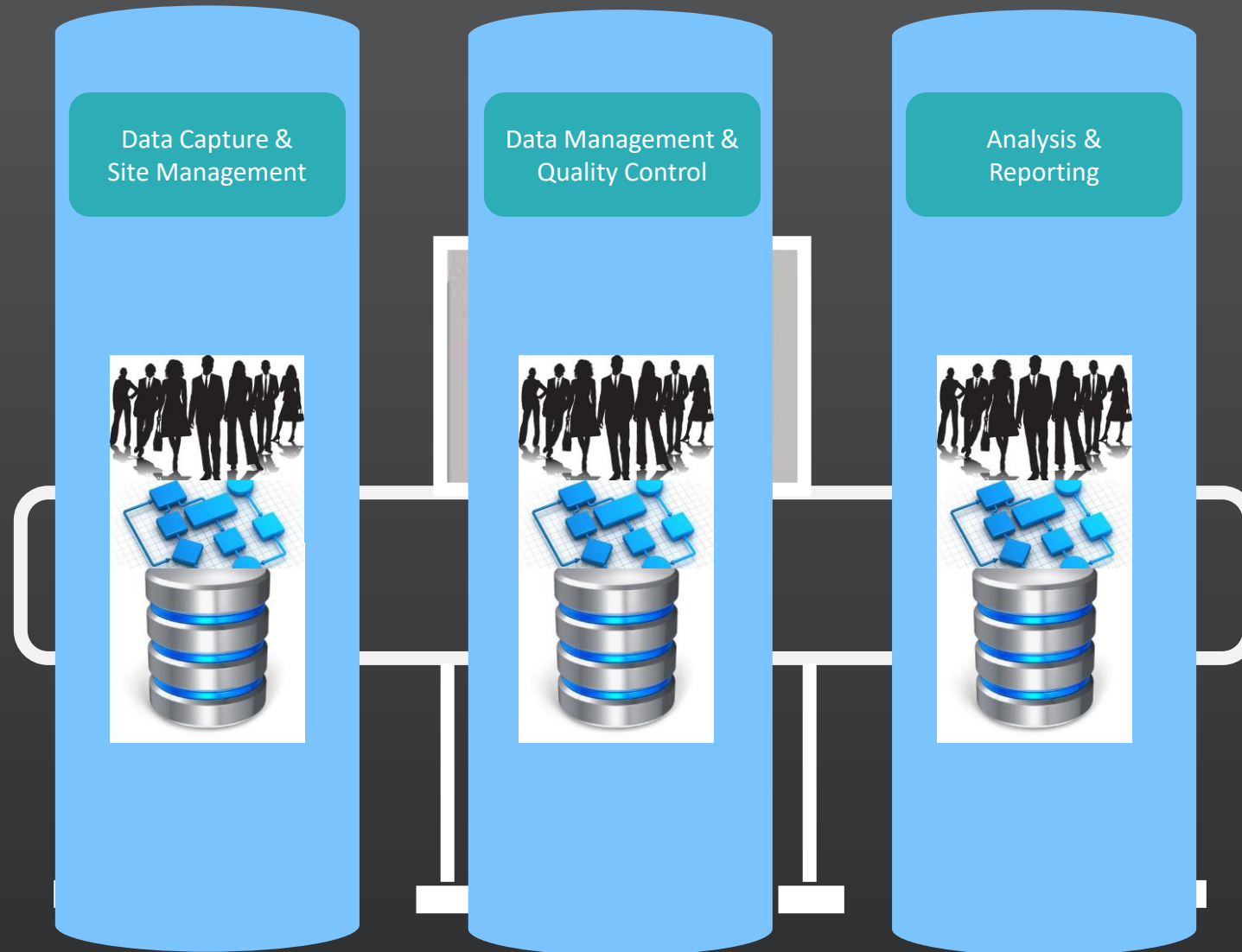
Assembly line / functional silo



Assembly line / functional silo



Assembly line / functional silo



Multiple Studies = Multiple Databases



Multiple Studies = Multiple Databases



Multiple Studies = Multiple Databases



The Conformed Data Model



[5 mins]

Past



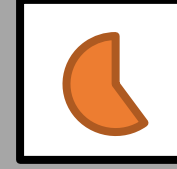
[5 mins]

Present



[5 mins]

Future



[5 mins]

Q&A

A New Approach



Data Engineering in other industries



Finance



Retail



Telecoms



Insurance

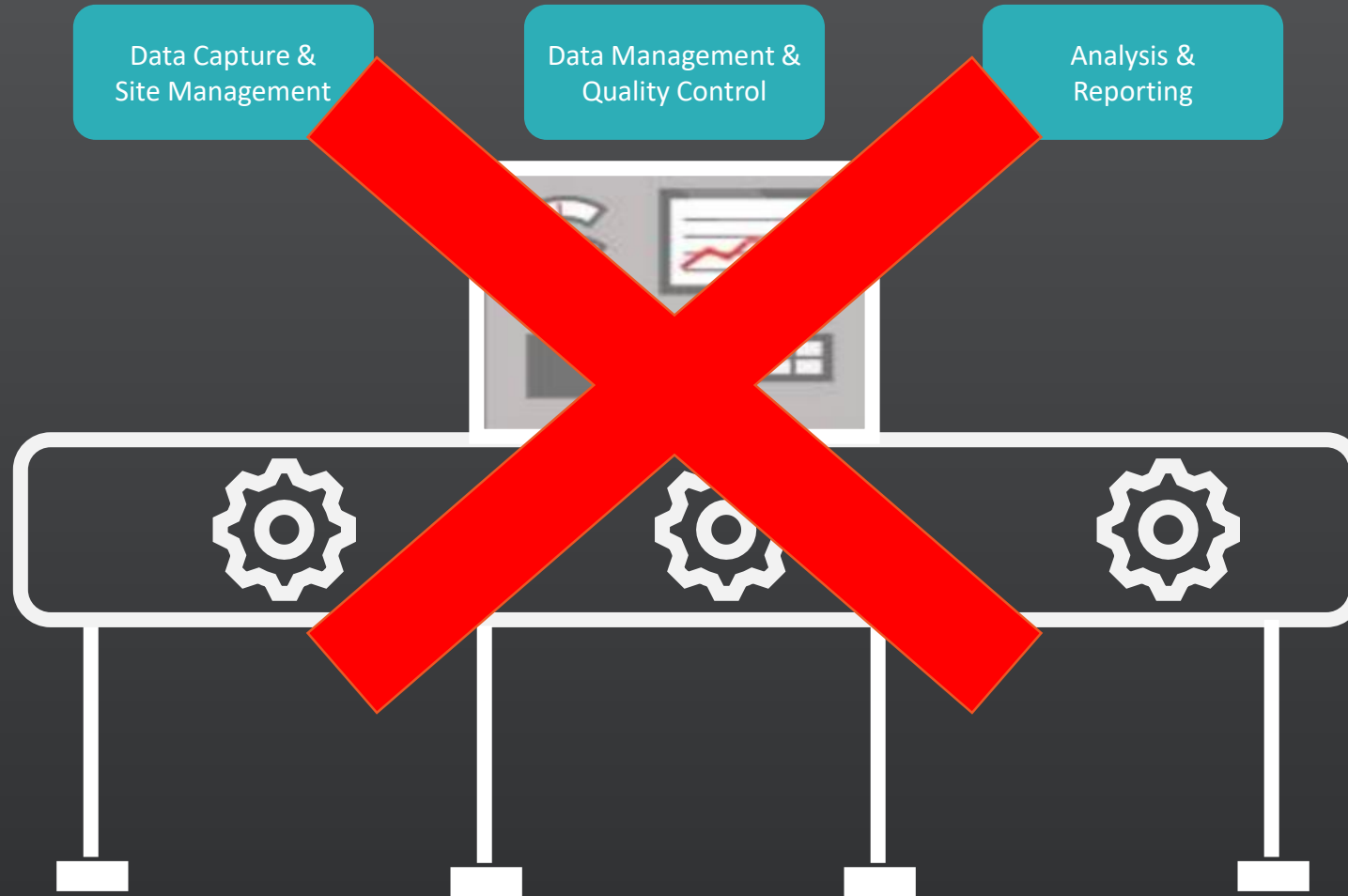
Why isn't BioPharma utilizing Data Engineering?

1. We have focused on other improvements (CDISC and communication with health authorities)
2. Fragmentation prevents holistic decisions
3. Budget pressures have been lower than other industries (until recently)
4. Resistance to change / Inertia of perceived cost of change

or

5. Data Engineering **just cannot work** for the BioPharma industry

REMOVE Assembly line



REMOVE functional silos



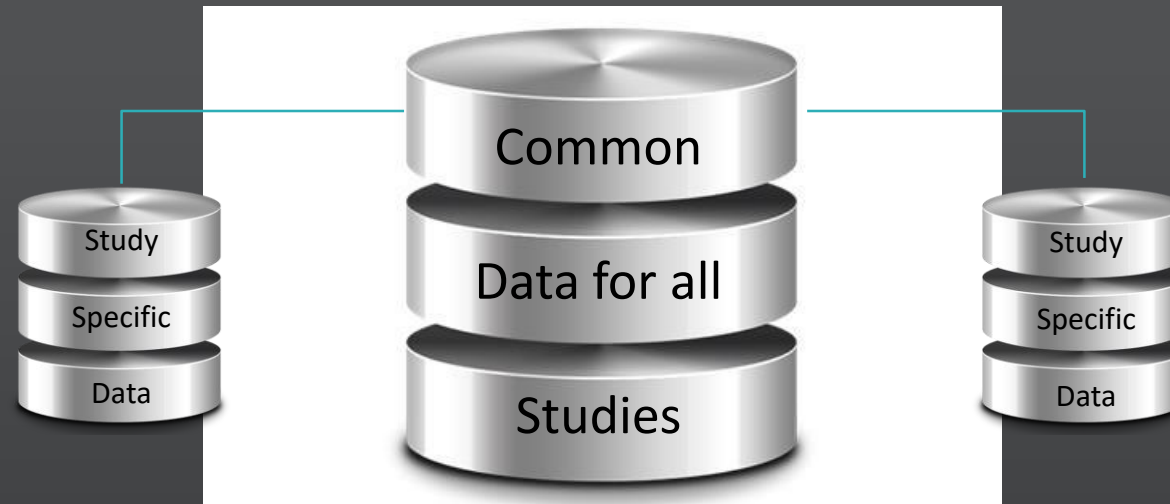
Multiple Studies = Multiple Databases



Conformed Data Model



Conformed Data Model

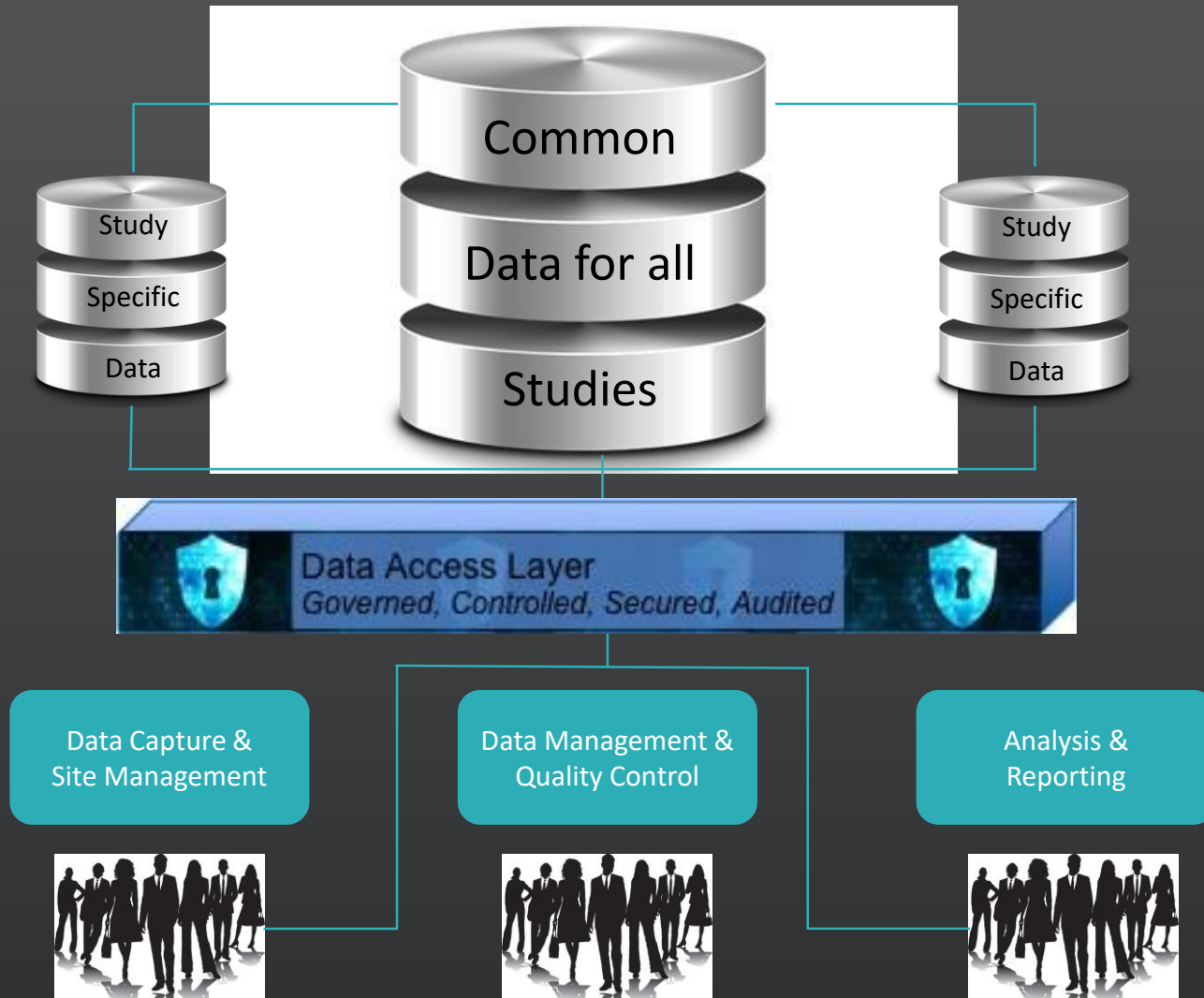


One Central database
build once what you use repeatedly

Conformed Data Model



Conformed Data Model



The Conformed Data Model



[5 mins]

Past



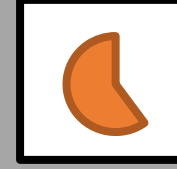
[5 mins]

Present



[5 mins]

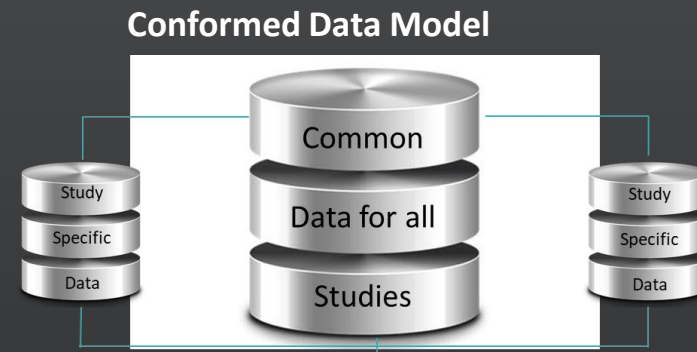
Future



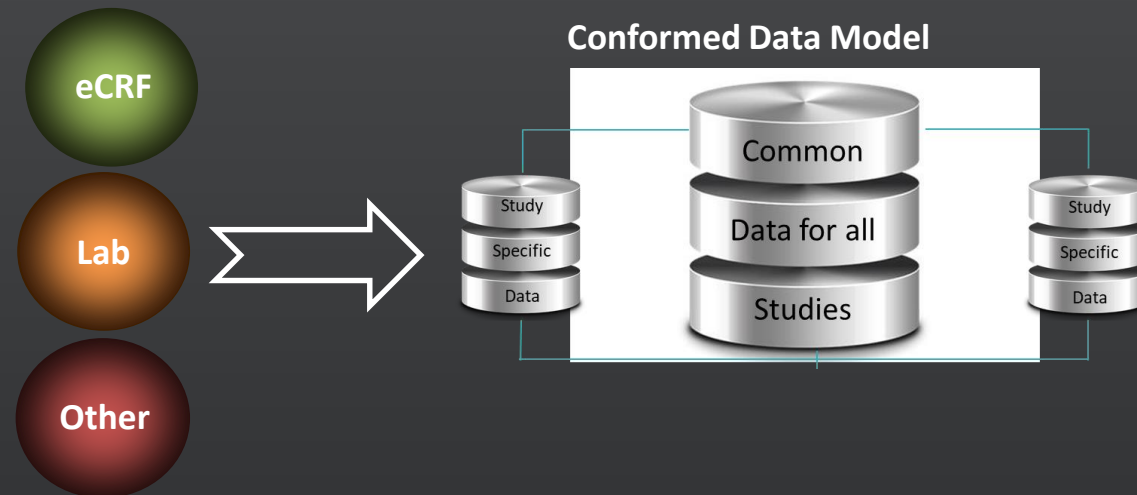
[5 mins]

Q&A

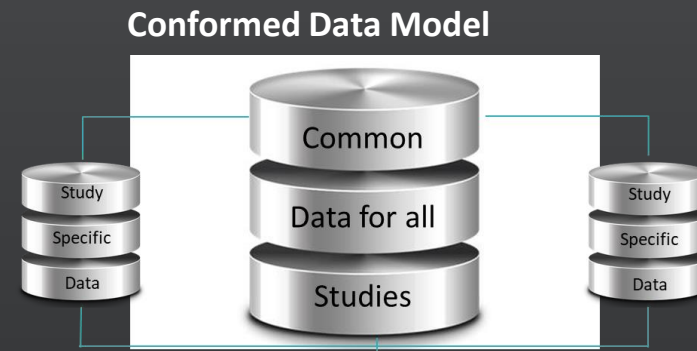
Data Science Pipelines



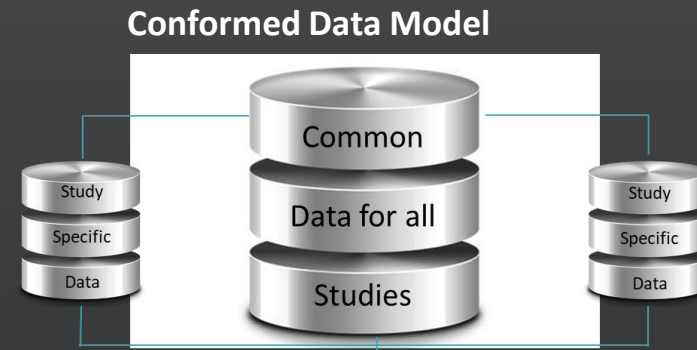
Data Science Pipelines



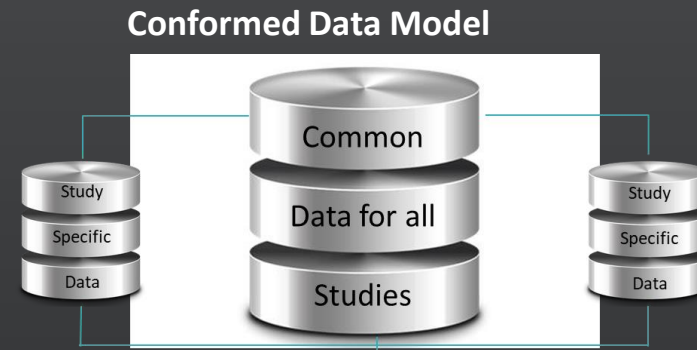
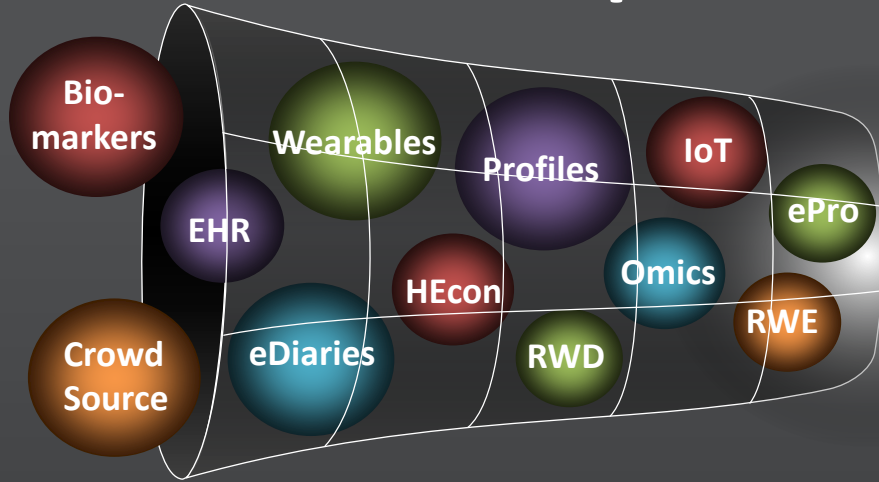
Data Science Pipelines



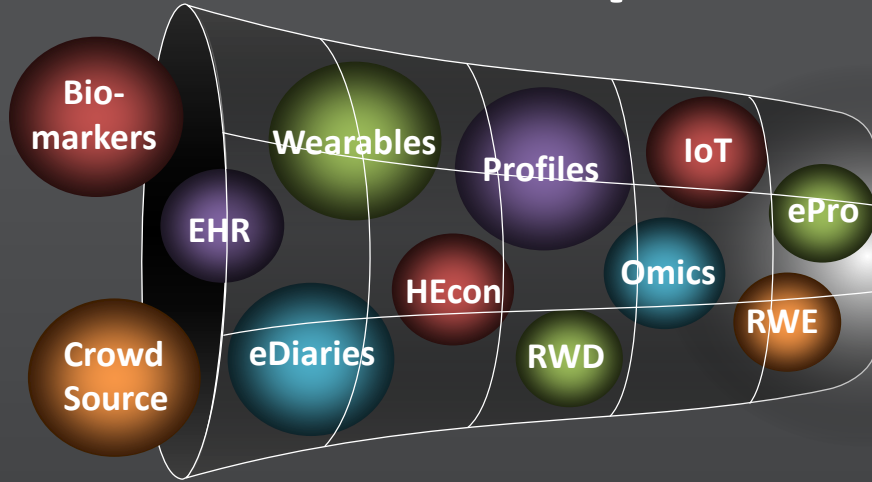
Data Science Pipelines



Data Science Pipelines



Data Science Pipelines



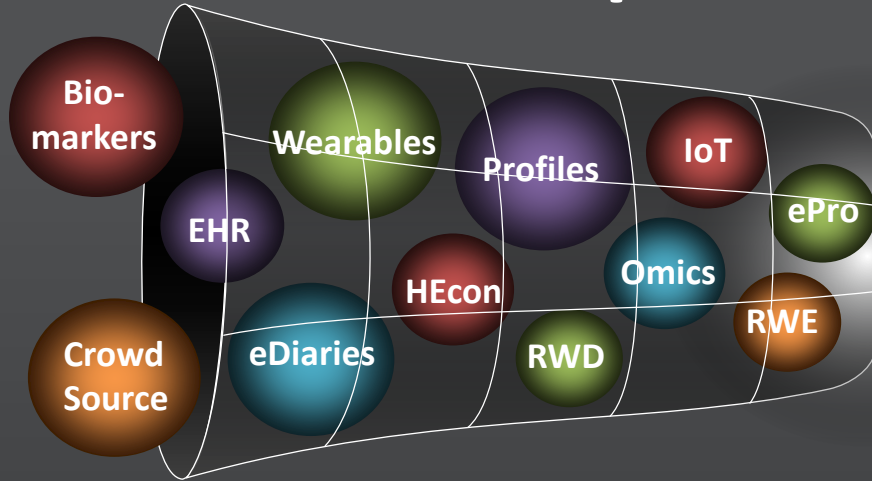
Unstructured Data Lake



Conformed Data Model



Data Science Pipelines



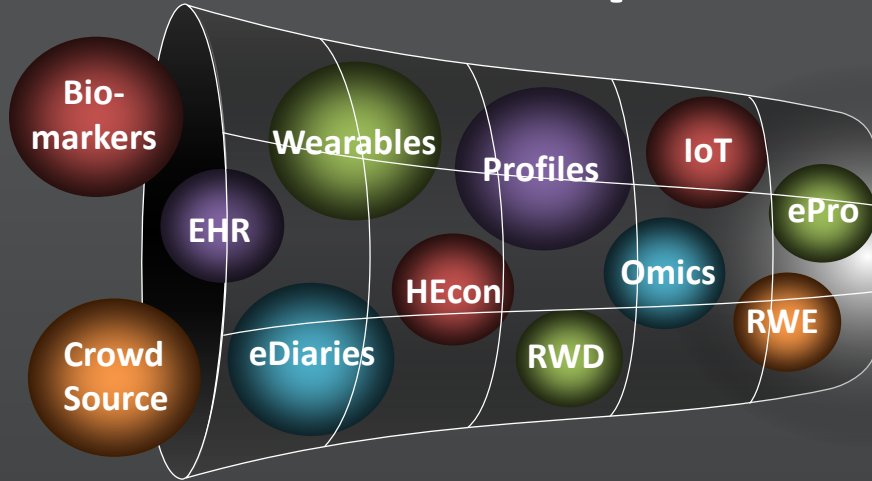
Unstructured Data Lake



Conformed Data Model



Data Science Pipelines



Unstructured Data Lake



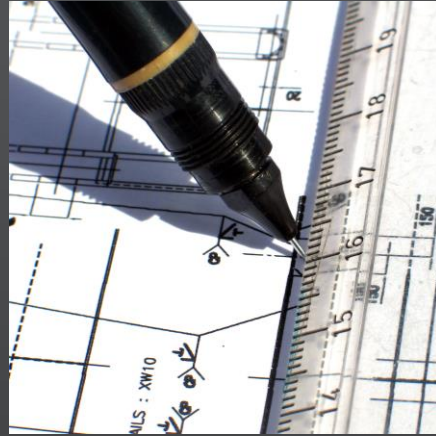
Conformed Data Model



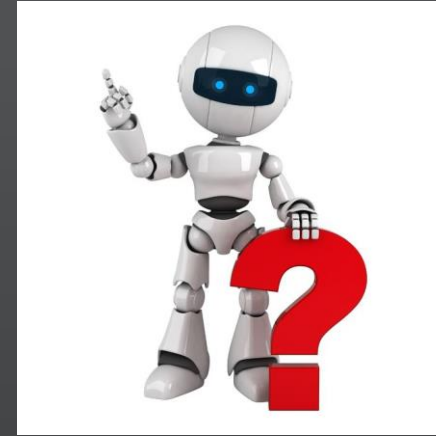
Benefits



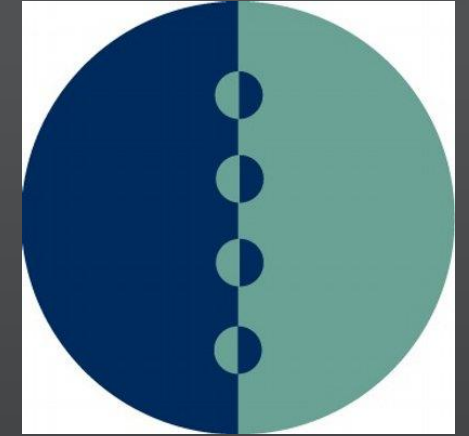
Stats Programmers
-> Data Scientists



Data Managers
-> Data Engineers



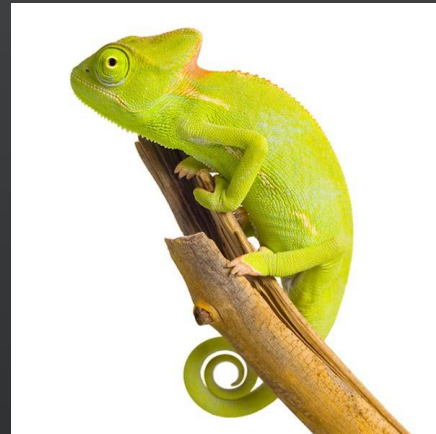
Automatic
Query Identification



Better CDISC
Management



KPI Capability



Adaptive Trial
Capability



Serendipitous
Results



Risk-Based
Monitoring

The Conformed Data Model



[5 mins]

Past



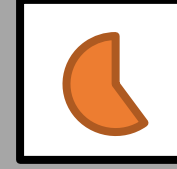
[5 mins]

Present



[5 mins]

Future



[5 mins]

Q&A