



Addressing the Gap in Data Standards: Introducing the Aggregation Layer

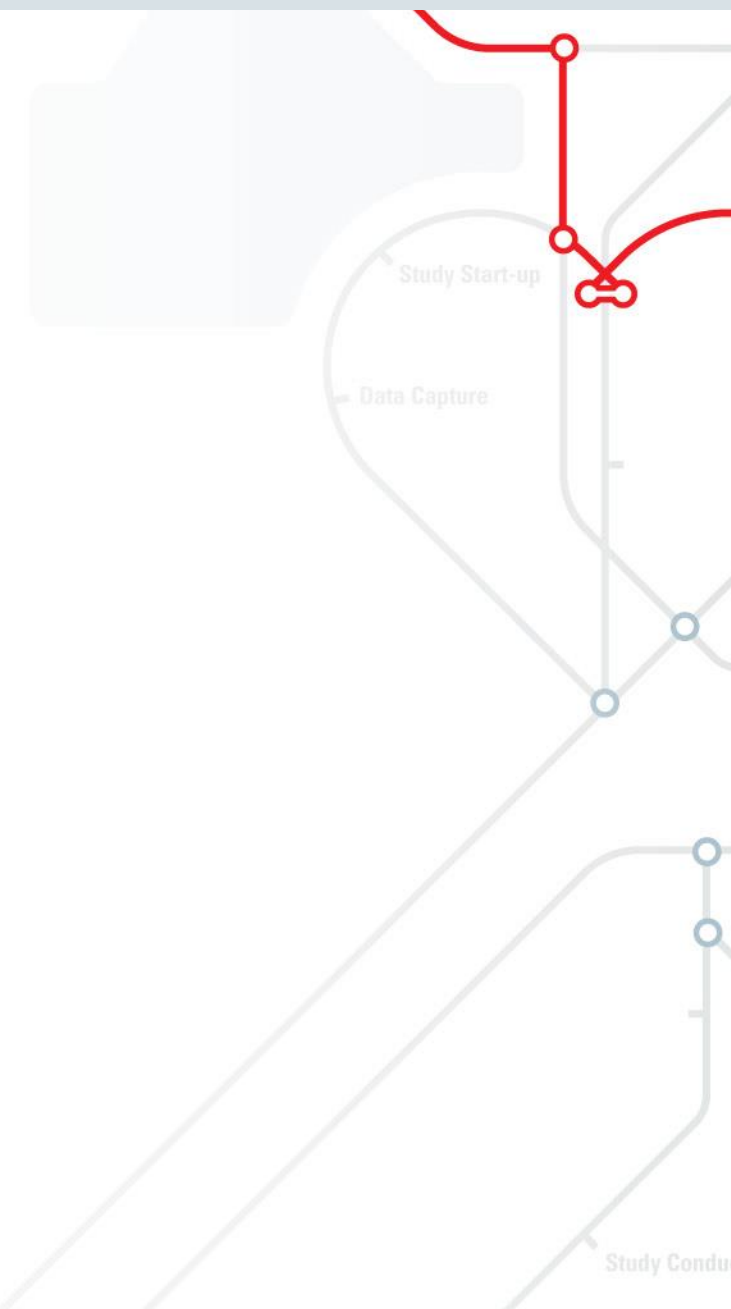
DH11 – Data Handling stream
PHUSE EU Connect 2019

Claudio Garutti, Principal Solutions Consultant
Marko Bomyer, Senior Consulting Practice Manager
Amsterdam, 12 November 2019

Outline

- 1 Introduction
- 2 Identifying the Gap
- 3 Case Study
- 4 Best Practices
- 5 Conclusion

Introduction



Identifying the Gap

Data Models

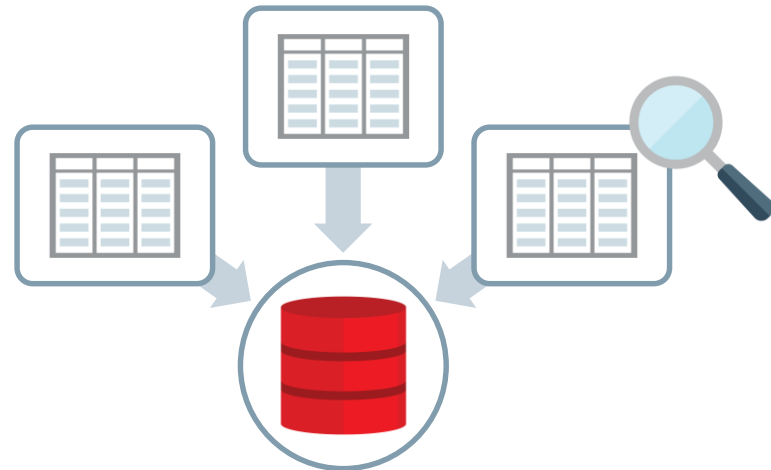
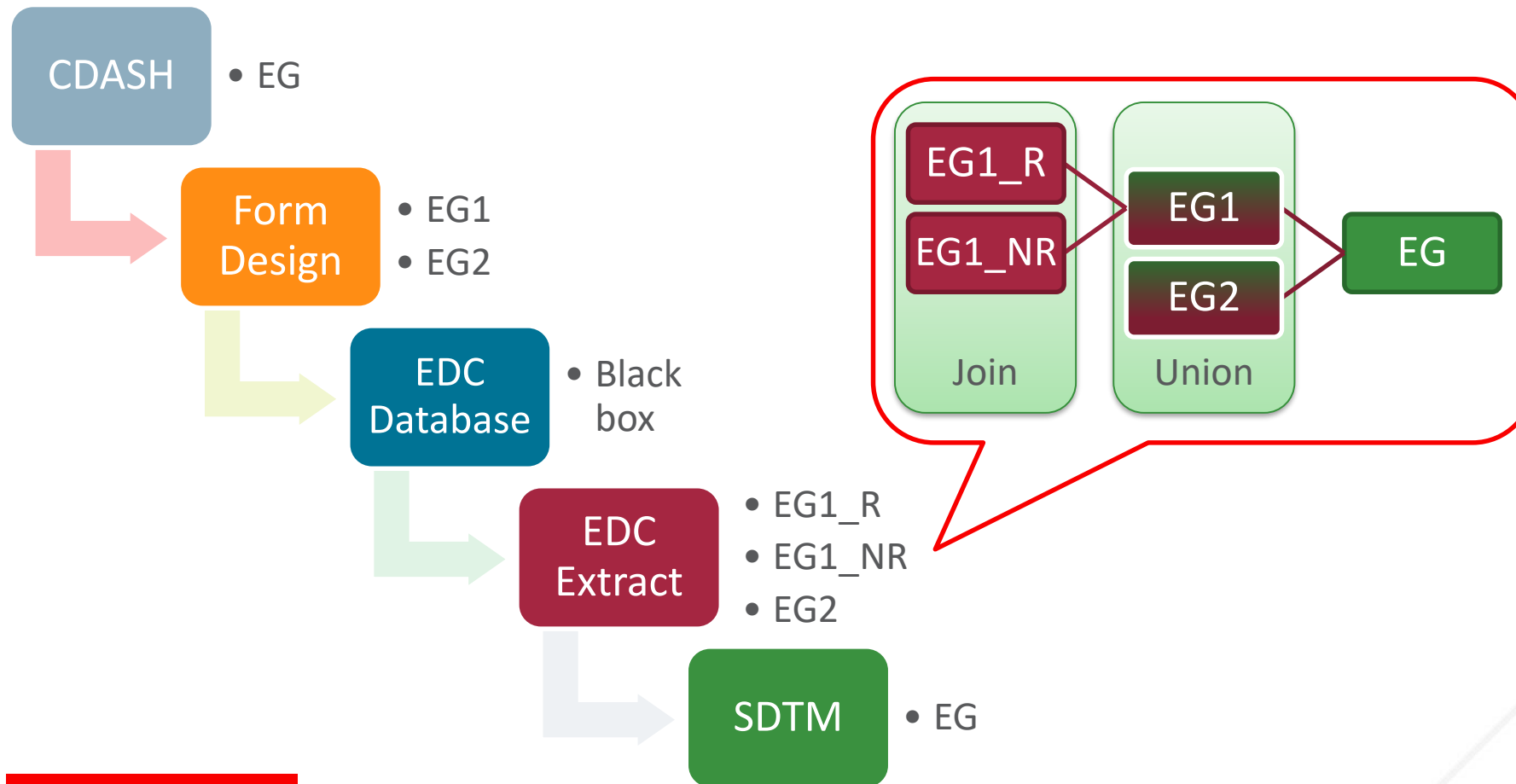


table names	column names	column formats
constraints (e.g. key)	code lists	allowed ranges

Data Collection	EDC raw data	Lab data	SDTM
☐	☐	☐	☐
☐ MH (CDASH)	☐ MHIST	☐	☐ MH (SDTM)
☐ VS (CDASH)	☐ VS_1, VS_2, ...	☐	☐ VS (SDTM)
☐	☐	☐	☐

Identifying the Gap – Gap #1

CDASH is the standard for how data is to be collected, not for raw data output



Identifying the Gap (#2)

Lack of a standard for data cleaning



DB designer
uses
CDASH



Stats
uses
SDTM (ADaM)



Data Manager
uses
something?

Identifying the Gap (#2)

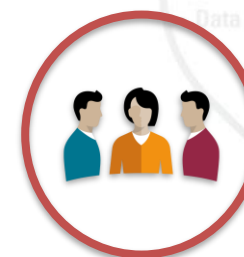
Lack of a standard for data cleaning

Normalized

STY1A	CTR1N	SBJ1D	VIS1N	ETSTNM	ETSTRES
ACB123D456	25067	26557	1	HR	87
ACB123D456	25067	26571	1	HR	58
ACB123D456	25067	26585	1	HR	98
ACB123D456	25067	26599	1	HR	56
ACB123D456	25067	26613	1	HR	78
ACB123D456	25067	26627	1	HR	75
ACB123D456	25067	26641	1	HR	76
ACB123D456	25067	26655	1	HR	68
ACB123D456	25067	26669	1	HR	98
ACB123D456	25067	25914	1	SYSBP	120
ACB123D456	25067	25997	1	SYSBP	112
ACB123D456	25067	26499	1	SYSBP	115
ACB123D456	25067	26513	1	SYSBP	124
ACB123D456	25067	26528	1	SYSBP	145
ACB123D456	25067	26543	1	SYSBP	123
ACB123D456	25067	26557	1	SYSBP	154
ACB123D456	25067	26571	1	SYSBP	133

De-normalized

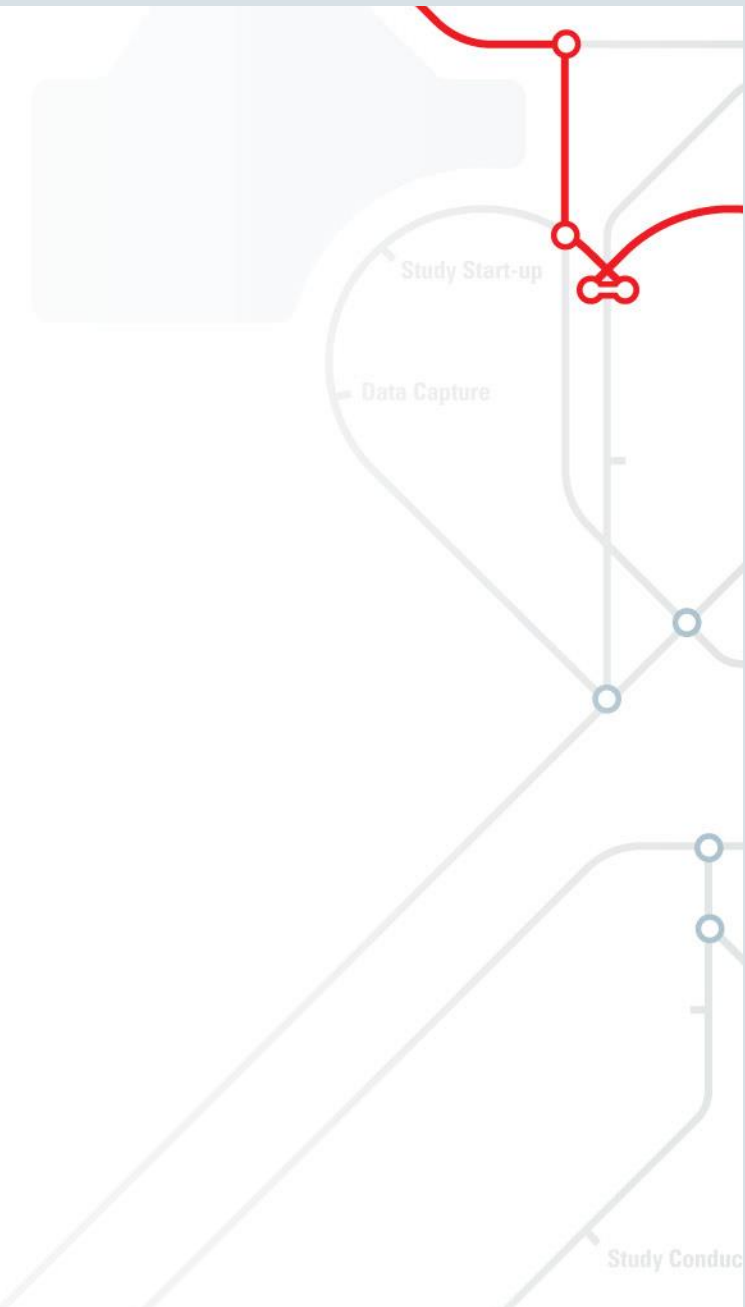
STY1A	CTR1N	SBJID	VIS1N	HR	SYSBP	DIABP	HT	WT
ACB123D456	25067	26585	1	98	154	98	174	78
ACB123D456	25067	26513	1		124	67	177	56
ACB123D456	25067	26599	1	56	122	56	182	91
ACB123D456	25067	26641	1	76	130	76	192	46
ACB123D456	25067	26655	1	68	111	68	177	87
ACB123D456	25067	26669	1	98	143	98	188	84
ACB123D456	25067	26499	1		115	55	166	55
ACB123D456	25067	26627	1	75	127	75	188	56
ACB123D456	25067	26683	1	56	132	56	184	92
ACB123D456	25067	26697	1	67	128	67	156	91
ACB123D456	25067	26711	1	89	137	89	182	83
ACB123D456	25067	25914	1		120	87	185	87
ACB123D456	25067	25997	1		112	93	156	63
ACB123D456	25067	26528	1		145	87	189	82
ACB123D456	25067	26557	1	87	154	87	167	78
ACB123D456	25067	26613	1	78	114	78	166	64
ACB123D456	25067	26571	1	58	133	58	177	66



Data Manager
uses
something?

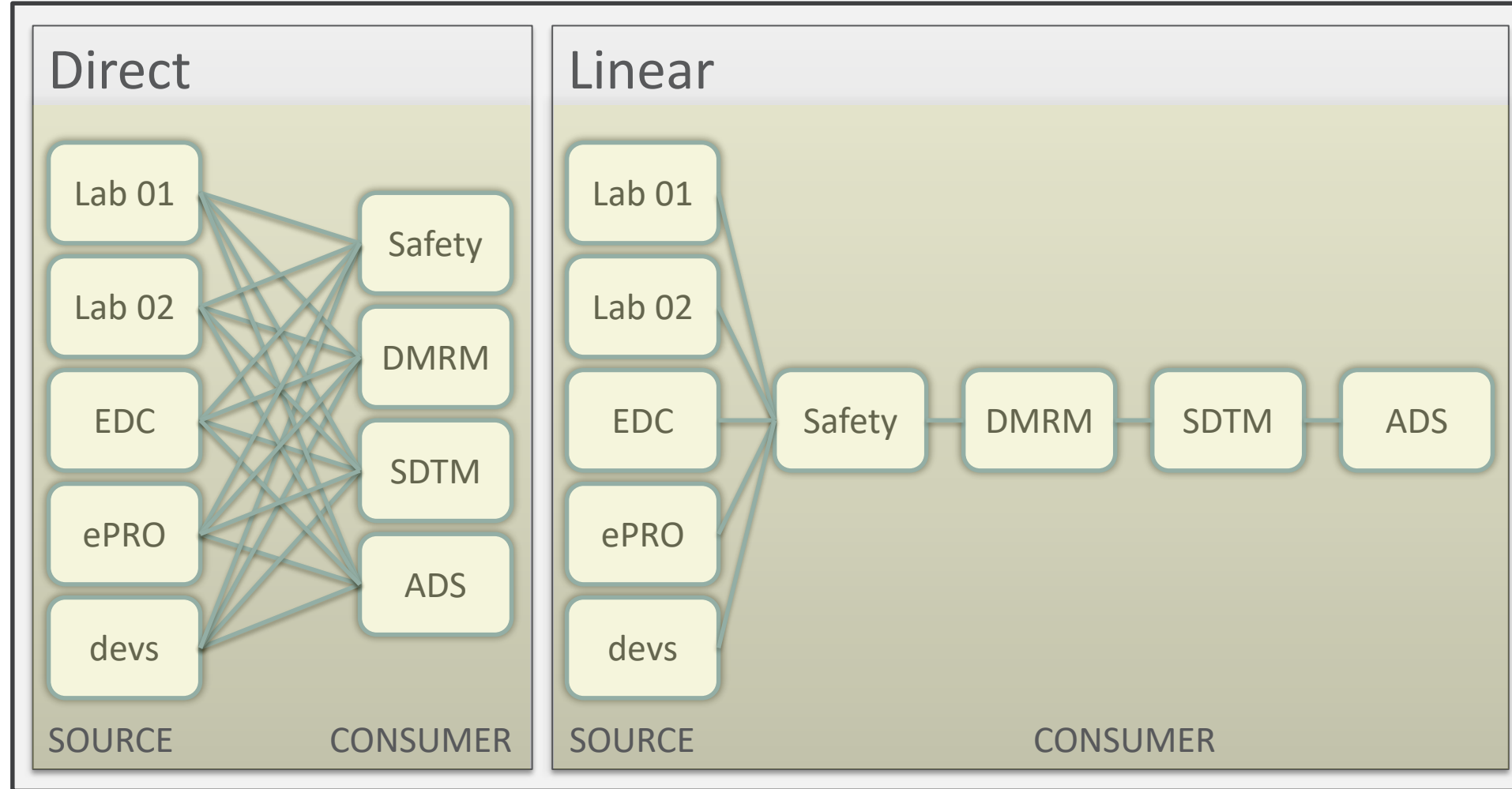
Identifying the Gap (#3)

Lack of a deliberate data architecture



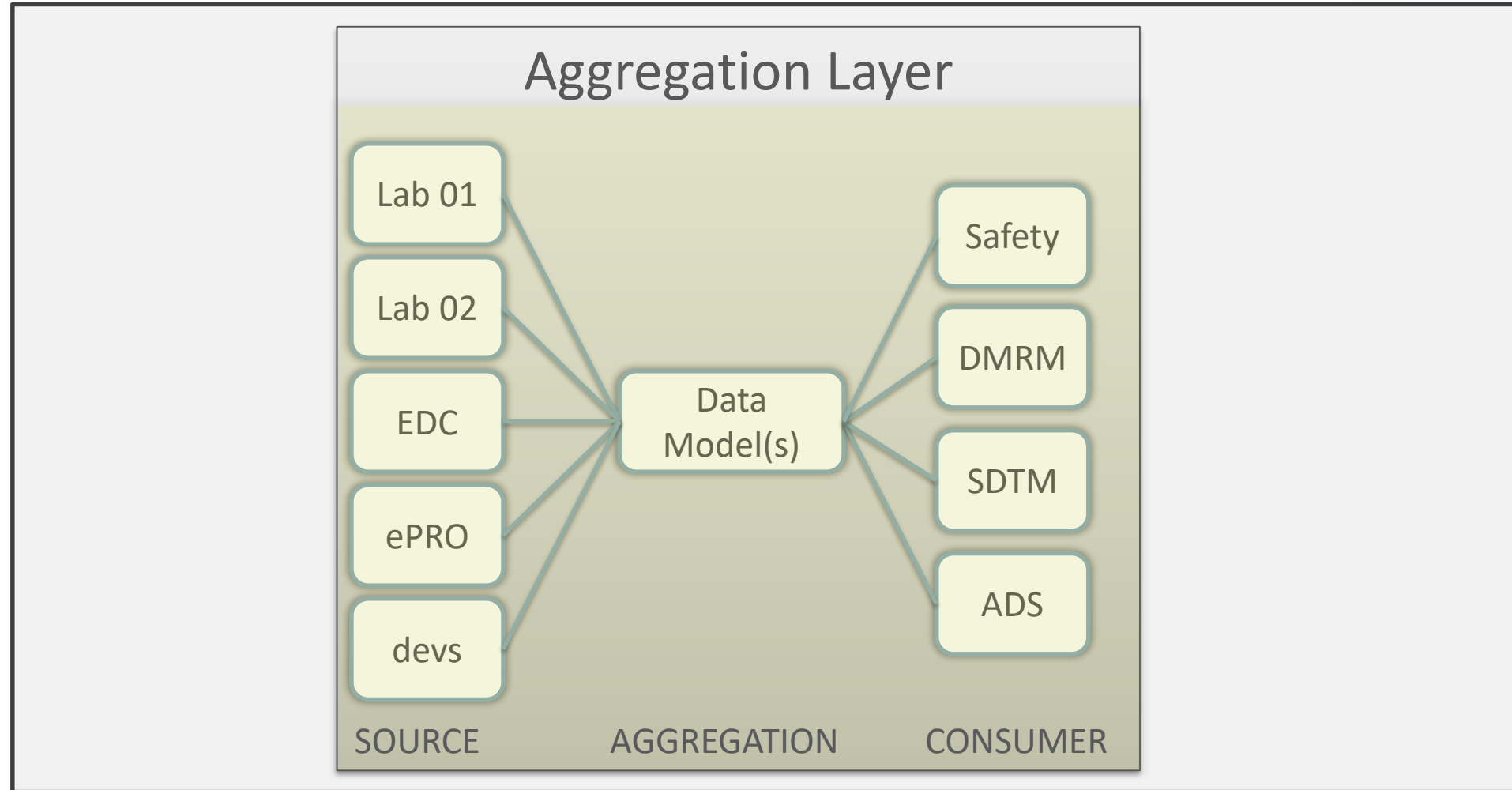
Identifying the Gap (#3)

Lack of a deliberate data architecture



Identifying the Gap (#3)

Lack of a deliberate data architecture

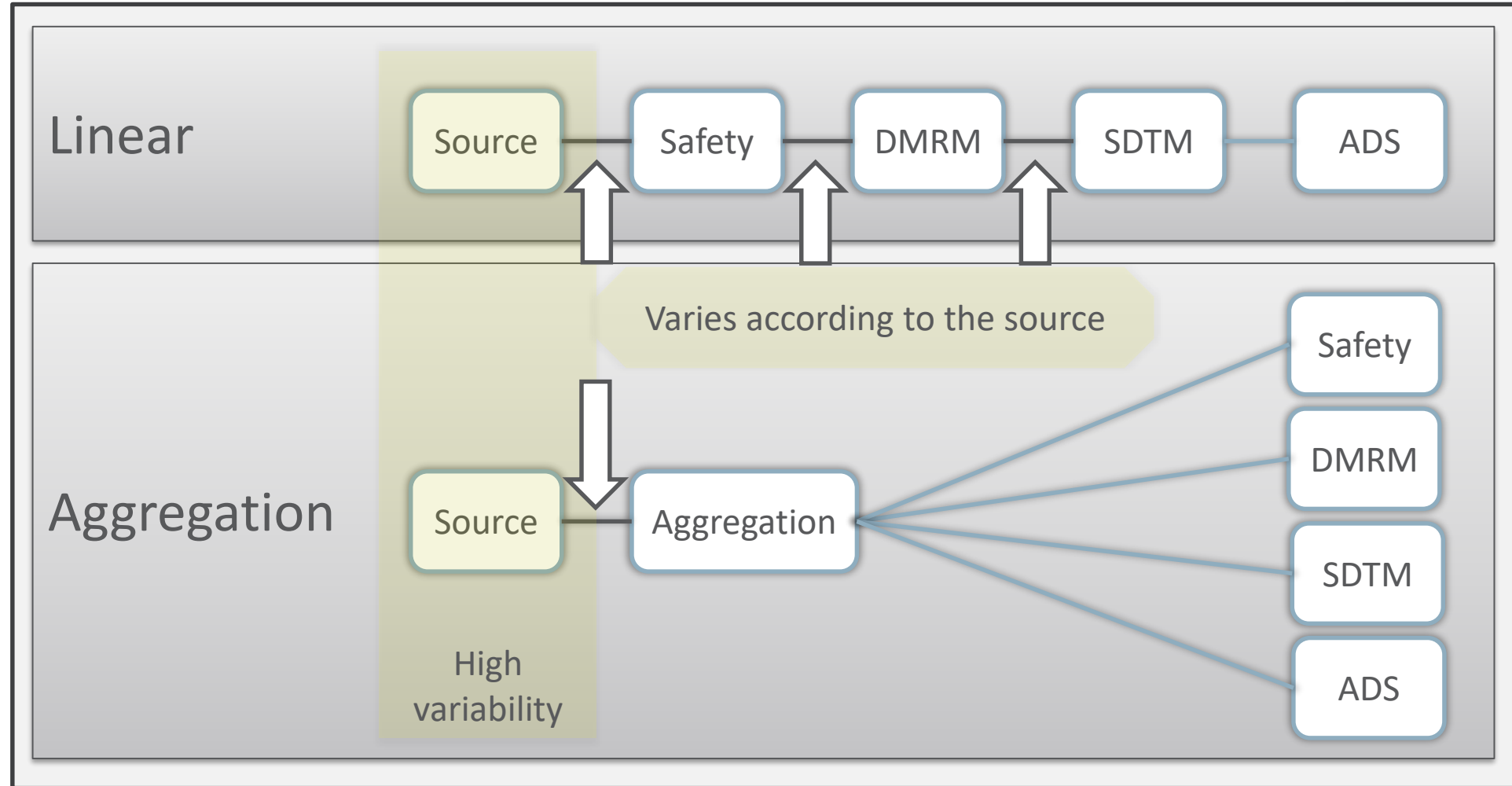


Outline

- 1 Introduction
- 2 Identifying the Gap
- 3 Case Study**
- 4 Best Practices
- 5 Conclusion

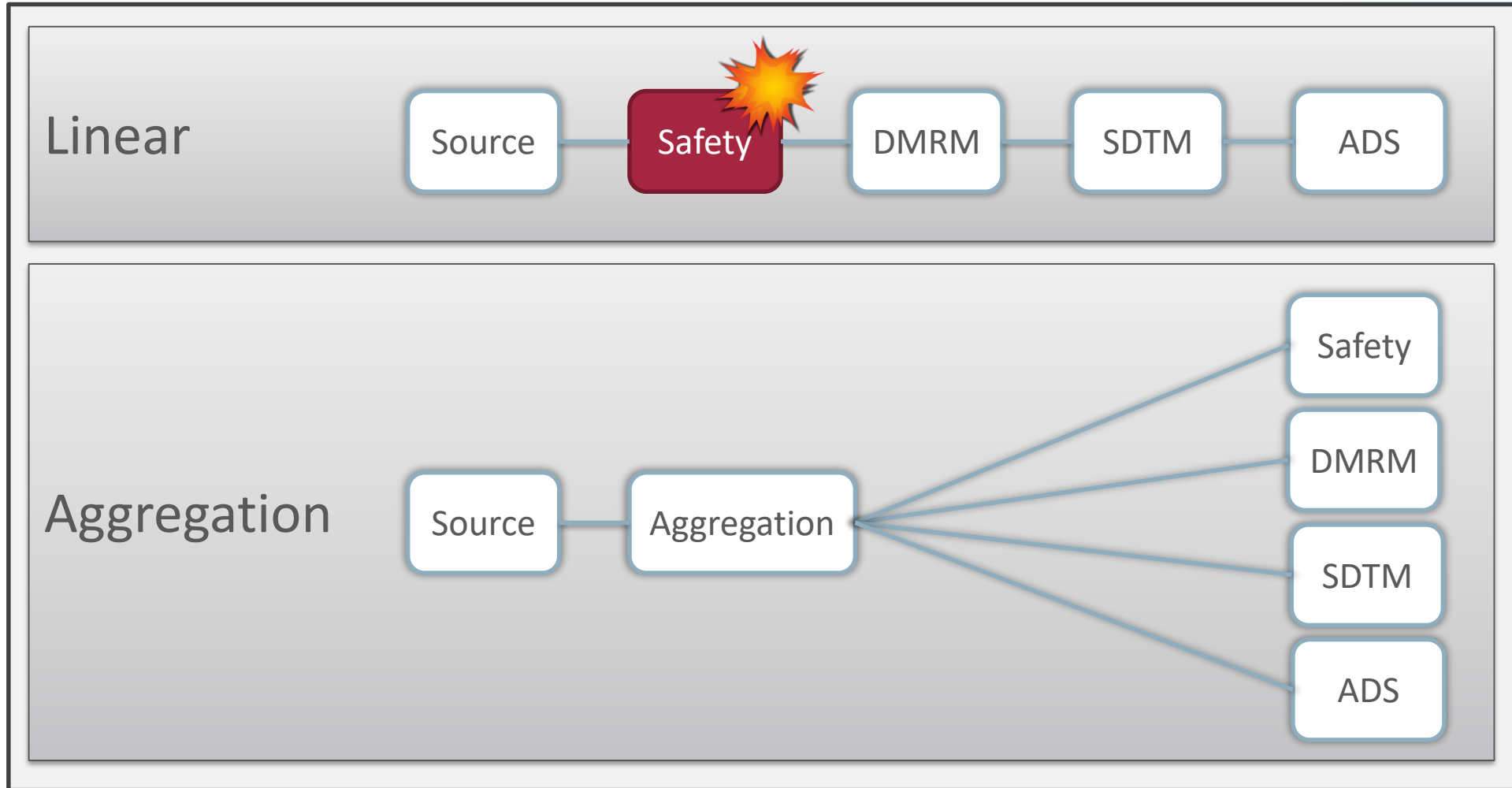
Case Study

Linear architecture vs Aggregation Layer architecture



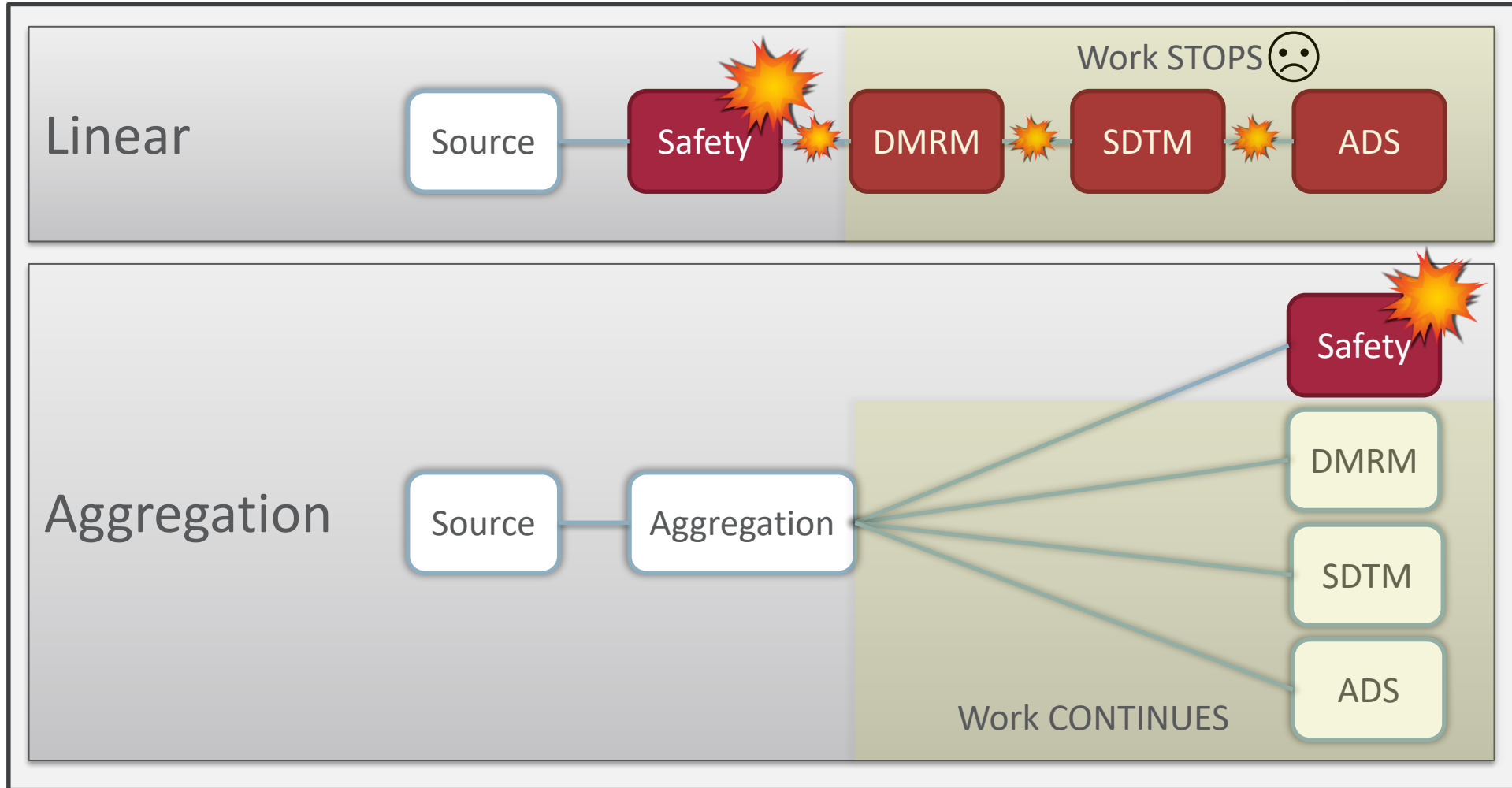
Case Study

Limiting the impact of errors



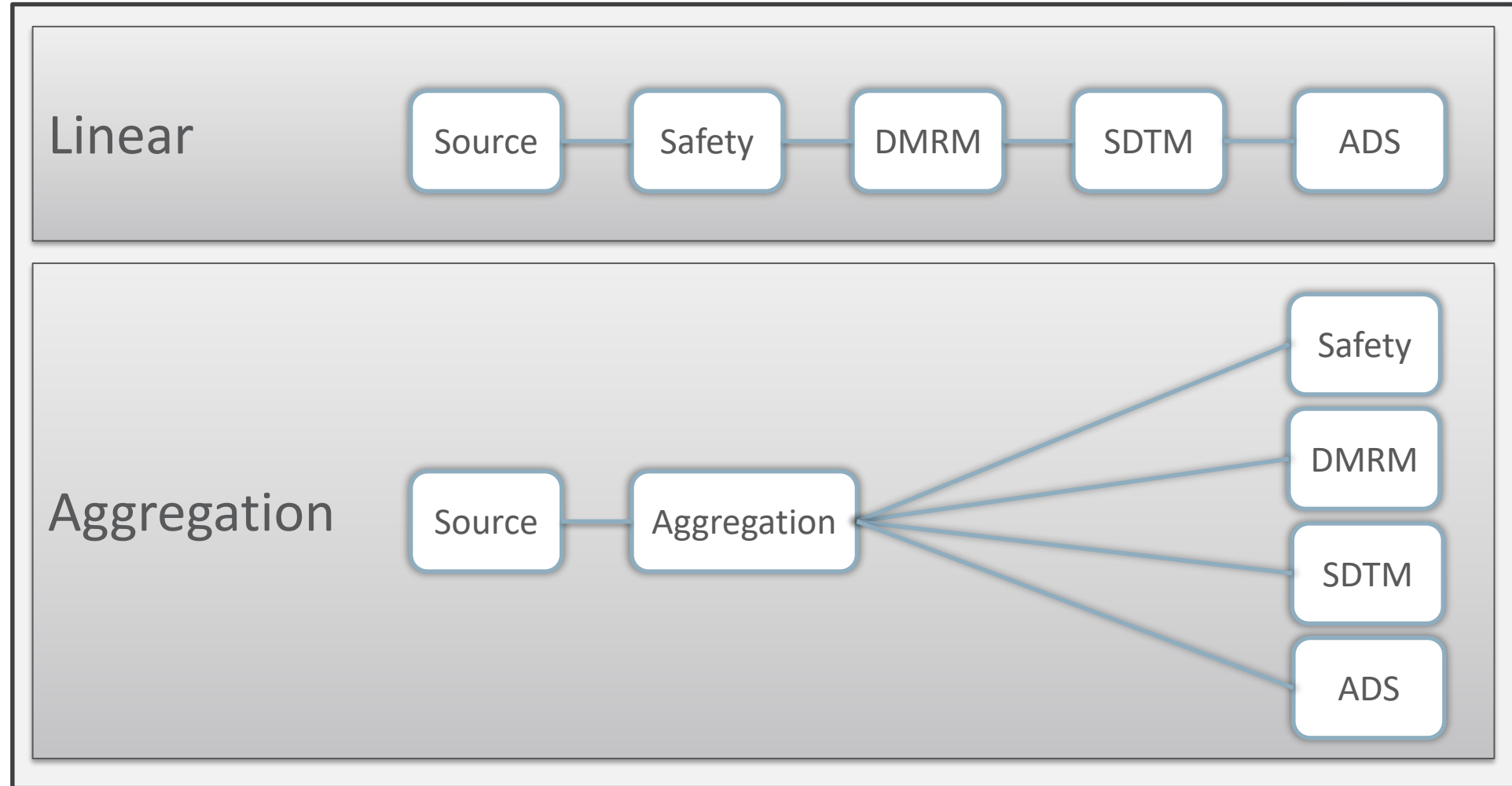
Case Study

Limiting the impact of errors



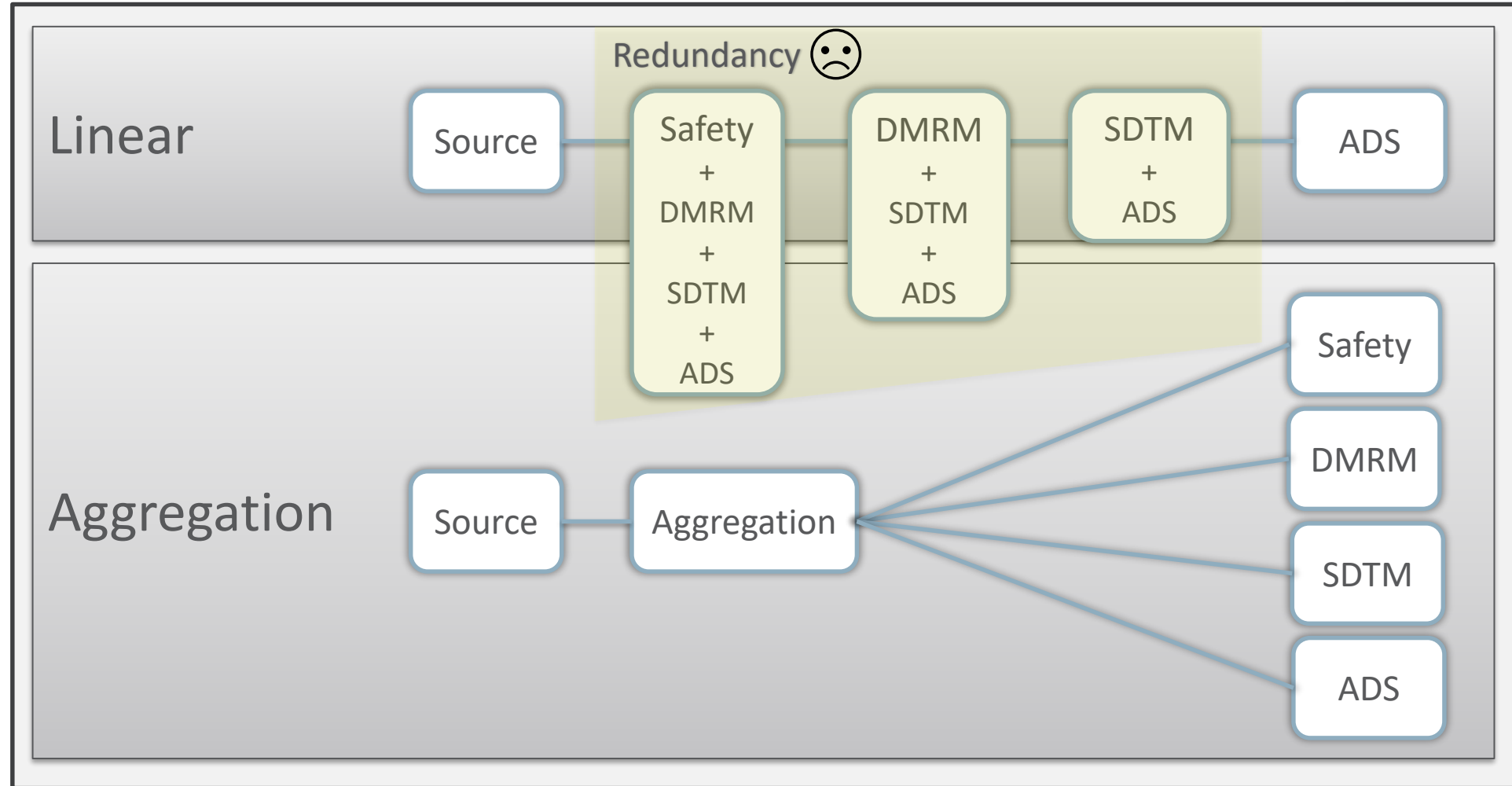
Case Study

Reducing Redundancy To Simplify Maintenance



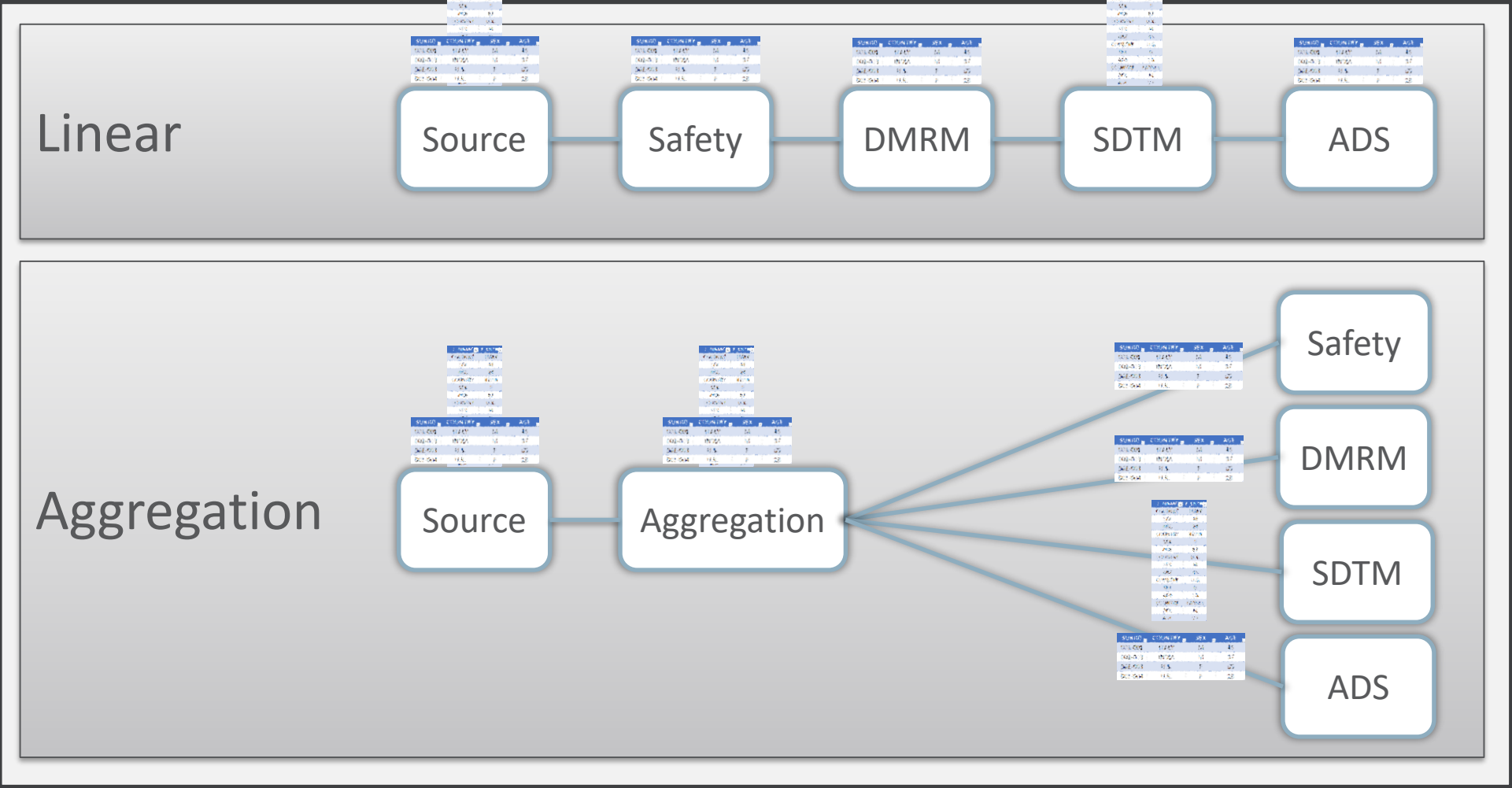
Case Study

Reducing Redundancy To Simplify Maintenance



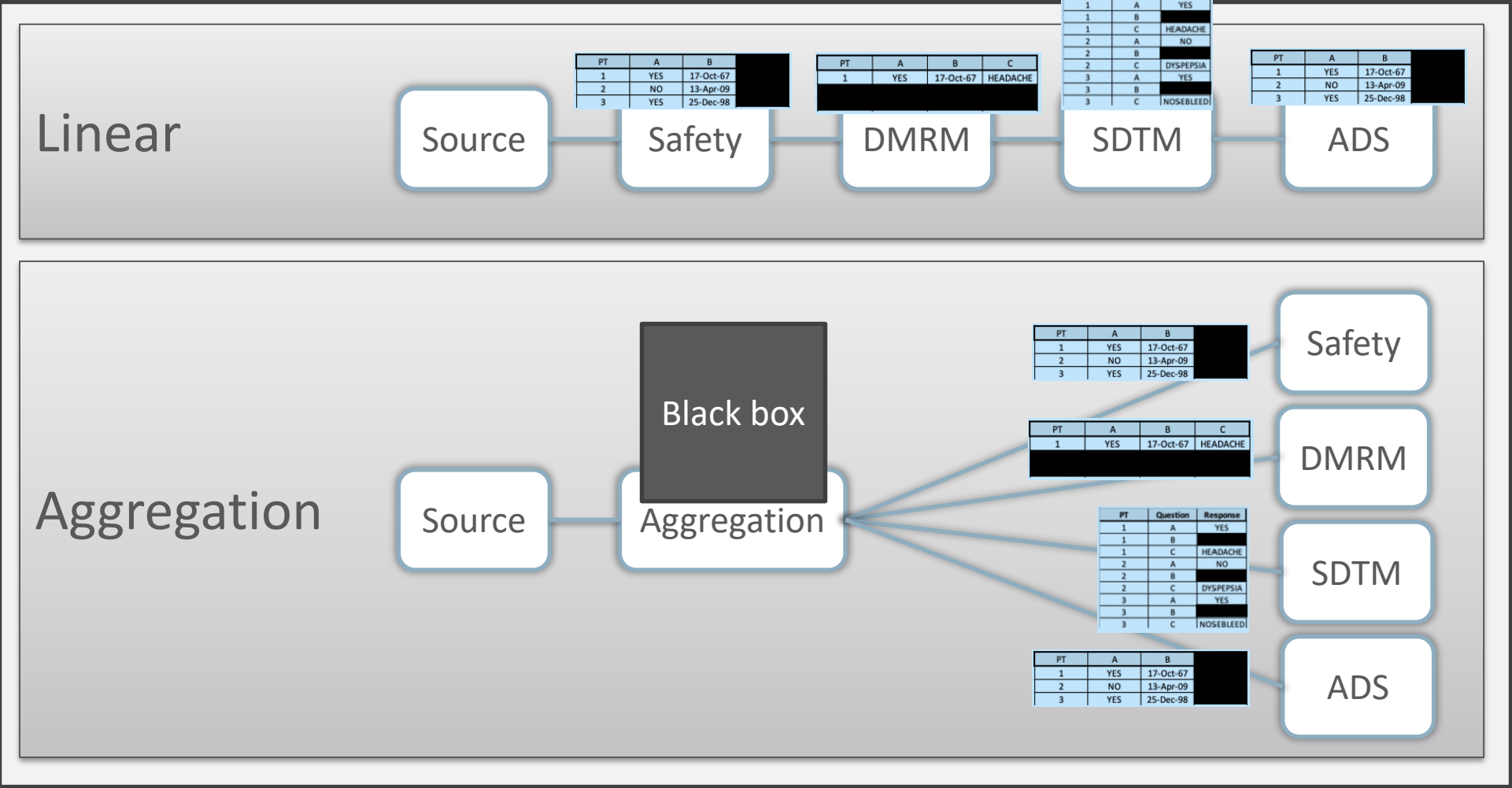
Case Study

Avoiding Unnecessary Transformations



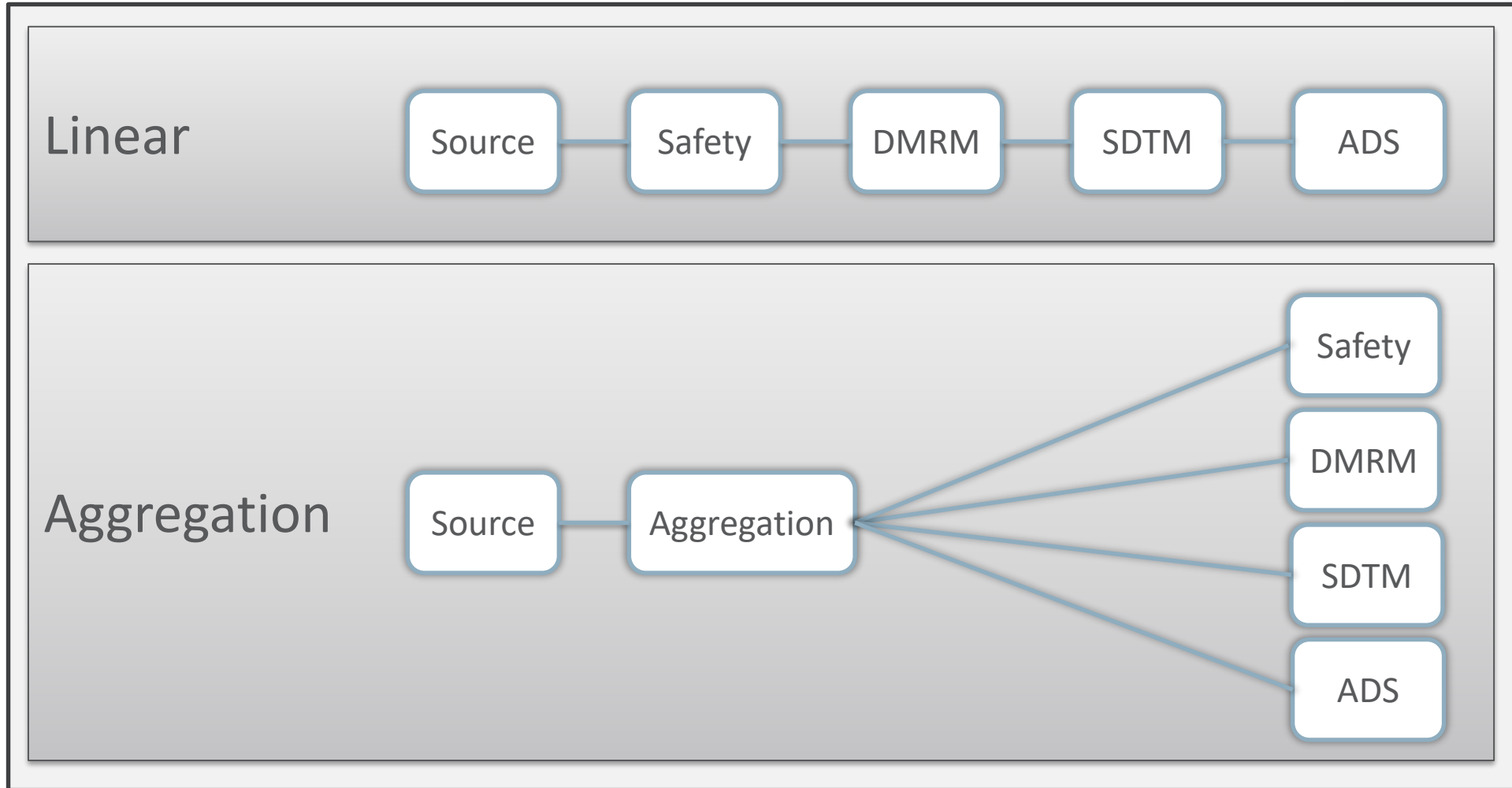
Case Study

Keeping Control Of Blinding



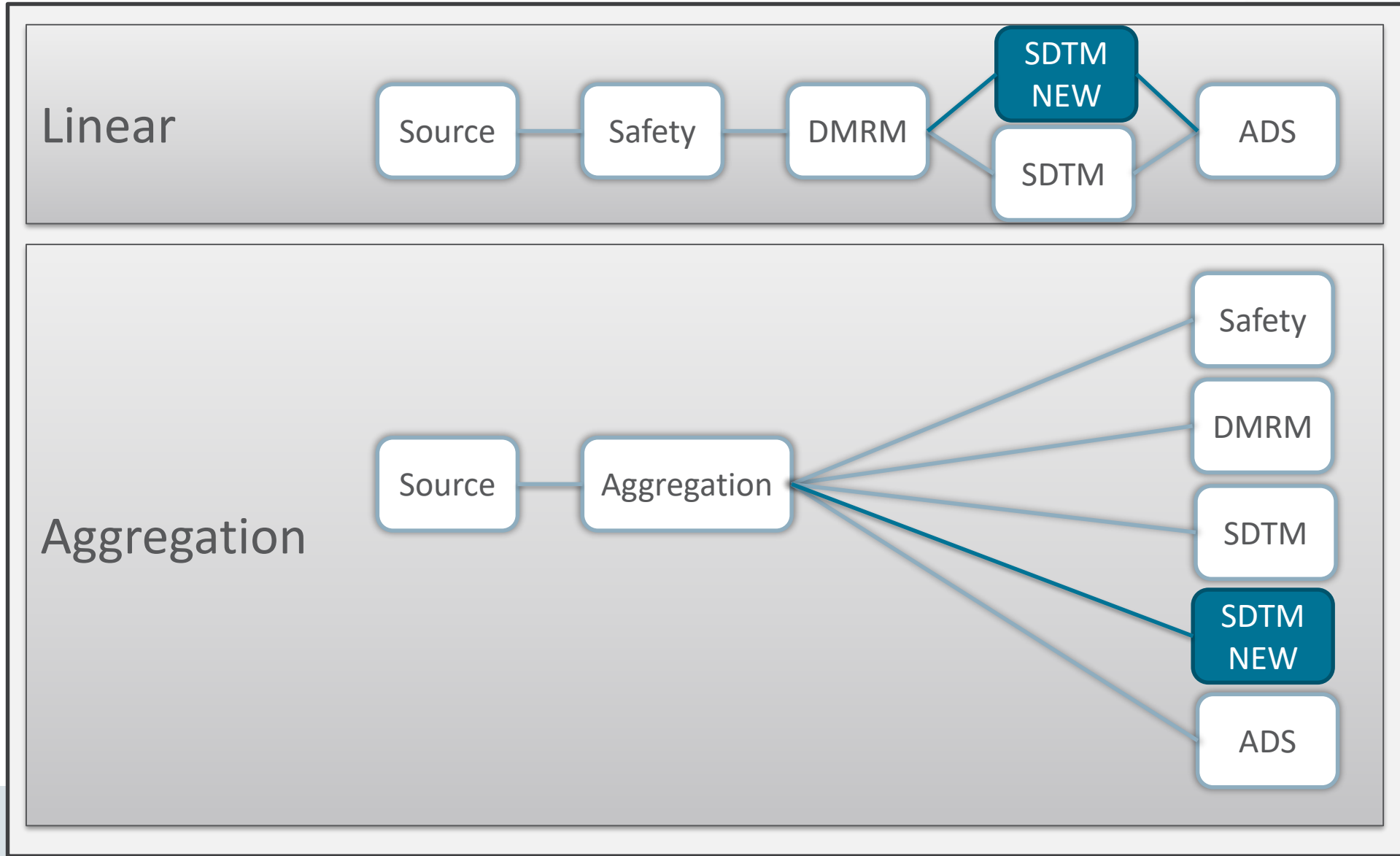
Case Study

SDTM "plug-n-play"



Case Study

SDTM "plug-n-play"



Best Practices

Aggregation Layer



Internal

Hidden to
end-users



Contains All
Derivations

(e.g. calculate BMI once,
serve multiple models)



Standardize
Structure

for re-usability
across studies

Outline

- 1 Introduction
- 2 Identifying the Gap
- 3 Case Study
- 4 Best Practices
- 5 Conclusion**

Conclusion

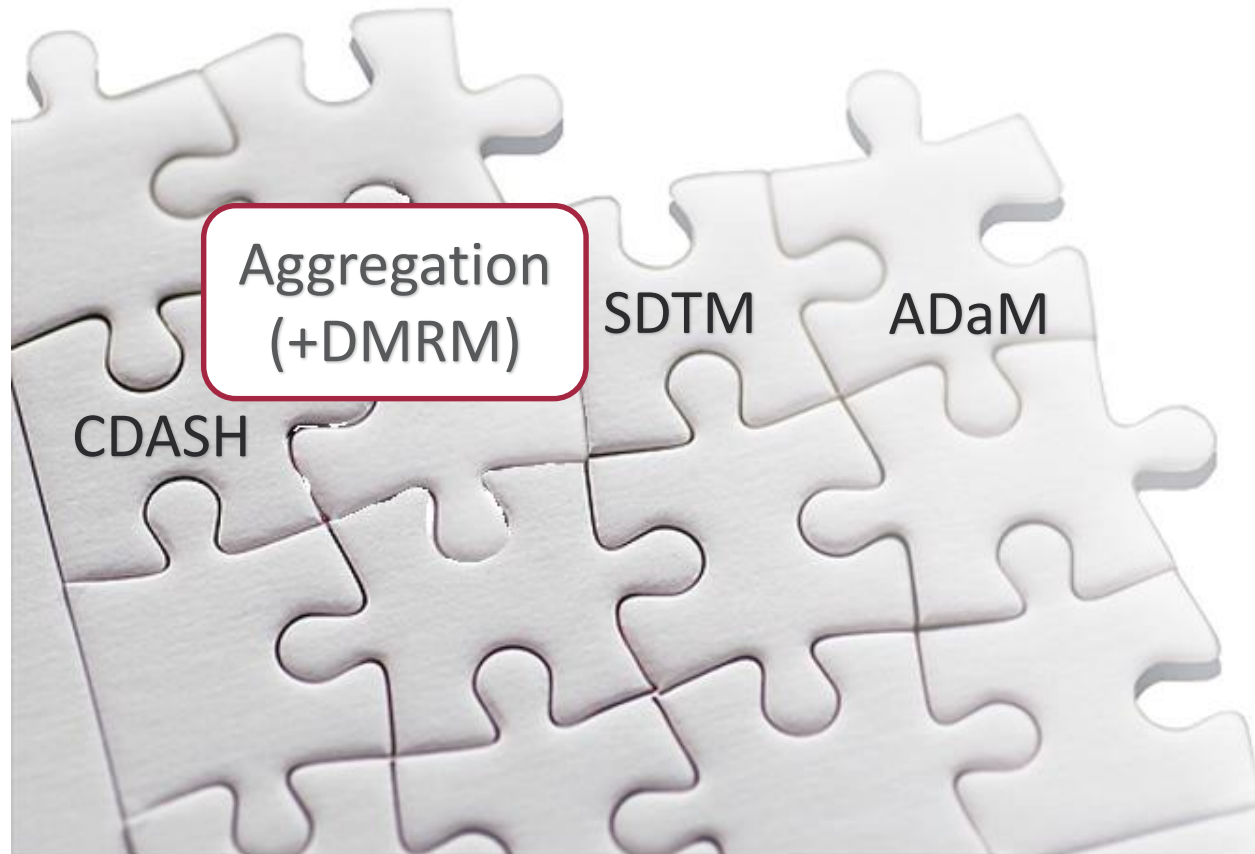


(1) CDASH $\neq$ EDC extract

(2) Data managers need to use *something*

(3) Deliberate data architecture

Conclusion



(1) CDASH $\neq$ EDC extract

(2) Data managers need to use *something*

(3) Deliberate data architecture

Thank you! Questions?



Claudio Garutti

claudio.garutti@oracle.com

Marko Bomyer

marko.bomyer@oracle.com