# Deciphering Data Architectures: Choosing Between a Modern Data Warehouse, Data Fabric, Data Lakehouse, and Data Mesh

**James Serra**

## Table of Contents

Preface ..... 6
Part I. Foundation ..... 7
1. Big Data ..... 7
1.1. What is big data and how can it help you? ..... 8
1.2. Data Maturity ..... 11
1.3. Self-service Business Intelligence ..... 13
1.4. Summary ..... 14
2. Types of data architectures ..... 15
2.1. Evolution of data architectures ..... 15
2.1.1. Relational data warehouse ..... 17
2.1.2. Data Lake ..... 19
2.1.3. Modern Data Warehouse ..... 21
2.1.4. Data Fabric ..... 21
2.1.5. Data Lakehouse ..... 22
2.1.6. Data Mesh ..... 22
2.2. Summary ..... 23
3. Architecture design session (ADS) ..... 24
3.1. What is an ADS and why have one ..... 24
3.2. Preparing and participants ..... 25
3.3. Conducting an ADS ..... 28
3.4. Summary ..... 33
Part II. Common Data Architecture Concepts ..... 33
4. Relational Data Warehouse ..... 33
4.1. What is a data warehouse? ..... 33
4.2. What a data warehouse is not ..... 35
4.2.1. DW Prefix ..... 35
4.2.2. Views with Unions ..... 36
4.2.3. Dumping Ground ..... 36
4.3. Top-down approach ..... 36

4.4. Why use a relational data warehouse? ..... 38
4.4.1. Reduce stress on the production system ..... 38
4.4.2. Optimize for read access ..... 38
4.4.3. Integrate multiple sources of data ..... 39
4.4.4. Run accurate historical reports ..... 39
4.4.5. Restructure/Rename tables ..... 39
4.4.6. Protection against application upgrades ..... 39
4.4.7. Reduced security concerns ..... 40
4.4.8. Keep historical data ..... 40
4.4.9. Master Data Management (MDM) ..... 40
4.4.10. Improve data quality by plugging holes in source systems ..... 40
4.4.11. No IT involvement needed to create reports ..... 40
4.4.12. Easy to create BI solutions ..... 41
4.4.13. Better reporting ..... 41
4.5. Populating a data warehouse ..... 41
4.5.1. How often to extract the data ..... 42
4.5.2. What extract method to use ..... 42
4.5.3. How to extract the data ..... 43
4.5.4. How to determine the data that has changed since the last extraction ..... 43
4.6. The death of the relational data warehouse has been greatly exaggerated ..... 44
4.7. Summary ..... 45
5. Data Lake ..... 45
5.1. What is a data lake? ..... 45
5.2. Why use one? ..... 46
5.3. Bottoms-up approach ..... 47
5.4. Data lake design ..... 48
5.5. Multiple data lakes ..... 50
5.6. Summary ..... 52
6. Approaches to Data Stores ..... 52
6.1. Data marts ..... 52
6.2. Operational data store (ODS) ..... 53
6.3. Master Data Management (MDM) ..... 55
6.4. Data virtualization ..... 55
6.4.1. Replacement for data warehouse ..... 56
6.4.2. Replacement for data movement ..... 57

6.4.3. Use cases ..... 58
6.5. Data Catalog and Data Marketplace ..... 59
6.6. Data Hub ..... 61
6.7. Summary ..... 63
7. Approaches to Design ..... 64
7.1. OLTP vs OLAP ..... 64
7.2. Operational vs analytical data ..... 65
7.3. SMP vs MPP ..... 65
7.4. Lambda Architecture ..... 66
7.5. Kappa Architecture ..... 68
7.6. Polyglot Persistence ..... 69
7.7. Summary ..... 70
8. Approaches to Data Modeling ..... 70
8.1. What is Data Modeling? ..... 70
8.2. Relational Modeling ..... 71
8.3. Dimensional Modeling ..... 73
8.4. Kimball vs Inmon ..... 76
8.4.1. Inmon ..... 77
8.4.2. Kimball ..... 78
8.4.3. Hybrid ..... 79
8.4.4. Myths ..... 81
8.5. Common Data Model (CDM) ..... 83
8.6. Data Vault ..... 84
8.7. Summary ..... 86
9. Approaches to Data Ingestion ..... 86
9.1. ETL vs ELT ..... 86
9.2. Reverse ETL ..... 86
9.3. Batch vs real-time processing ..... 87
9.4. Data governance ..... 88
9.5. Summary ..... 89
Part III. Data Architectures ..... 89
10. Modern Data Warehouse (MDW) ..... 89
10.1. Pre-cursor architectures to the MDW ..... 89
10.2. Combine relational data warehouse and data lake ..... 92
10.3. Five MDW stages ..... 94

10.4. Use cases ..... 96
10.5. Summary ..... 97
11. Data Fabric ..... 97
11.1. Definition ..... 97
11.2. Use cases ..... 100
11.3. Summary ..... 100
12. Data Lakehouse ..... 100
12.1. Storage software layer ..... 100
12.2. Concerns skipping relational data warehouse ..... 104
12.3. Relational serving layer ..... 105
12.4. Use cases ..... 106
12.5. Summary ..... 106
13. Data Mesh Foundation ..... 107
13.1. Overview: Centralization vs decentralization ..... 107
13.2. Data mesh hype ..... 109
13.3. Four principles of data mesh ..... 110
13.4. Reasons to move to a mesh ..... 114
13.5. Minimum viable mesh ..... 115
13.6. Data domains ..... 116
13.7. Data products ..... 117
13.8. Example domains ..... 118
13.9. Summary ..... 121
14. Data Mesh Adoption ..... 122
14.1. Data mesh logical architecture ..... 122
14.2. Concerns and challenges ..... 123
14.3. Myths ..... 130
14.4. Should you adopt a data mesh? ..... 131
14.5. Keys for a successful data mesh ..... 133
14.6. Compare data mesh vs data fabric ..... 133
14.7. Different topologies ..... 134
14.8. Use cases ..... 135
14.9. Real-world implementations ..... 135
14.10. Data mesh future ..... 136
14.11. Summary ..... 137
Appendix ..... 137

15. People and process ..... 137
15.1. Why projects succeed ..... 137
15.2. Why projects fail ..... 138
15.3. Team organization ..... 138
16. Technologies ..... 138
16.1. Open Source ..... 139
16.2. Hadoop and data lakes ..... 139
16.3. Benefits of the cloud ..... 139
16.4. Major cloud providers ..... 145
16.5. Multi-cloud ..... 145
16.6. Databricks ..... 147
16.7. Snowflake ..... 147
17. Data architectures on Microsoft Azure ..... 147
18. Data architectures on AWS ..... 147
19. Data architectures on GCP ..... 147