

# **Modern Data Engineering with SQL in 2025: Architect, Transform, and Tune Big Data Workflows Using SELECT, JOIN, GROUP BY, and Advanced Techniques**

**Newman Chandler**

# TABLE OF CONTENTS

PREFACE .....	III
<b>INTRODUCTION.....</b>	<b>1</b>
<b>CHAPTER 1 FOUNDATIONS REFRESHER &amp; MODERN DIALECTS</b> .....	<b>3</b>
1.1 EXPLICIT COLUMN SELECTION & OPTIMAL DATA TYPES .....	3
1.2 COMPARING DIALECTS: POSTGRESQL 16, SNOWFLAKE, BIGQUERY .....	7
1.3 VERSIONED SQL SCRIPTS & DEPLOYMENT PIPELINES.....	11
<b>CHAPTER 2 MODULAR QUERY DESIGN WITH CTES &amp;</b> <b>LATERAL .....</b>	<b>16</b>
2.1 BREAKING DOWN COMPLEXITY WITH CTES.....	16
2.2 FLATTENING JSON & SEMI-STRUCTURED DATA VIA LATERAL .....	20
2.3 REFACTORING NESTED SUBQUERIES .....	23
<b>CHAPTER 3 FILTERING STRATEGIES: WHERE VS. HAVING .....</b>	<b>28</b>
3.1 PREDICATE PUSHDOWN & STORAGE PRUNING .....	28
3.2 POST-AGGREGATION FILTERS WITH HAVING .....	31
3.3 DYNAMIC FILTERS & PARAMETERIZED QUERIES .....	34
<b>CHAPTER 4 ADVANCED AGGREGATIONS &amp; WINDOW</b> <b>FUNCTIONS.....</b>	<b>40</b>
4.1 MULTI-LEVEL AGGREGATES: ROLLUP, CUBE, GROUPING SETS ..	40
4.2 CUMULATIVE & SLIDING WINDOWS.....	44
4.3 PERCENTILES & STATISTICAL FUNCTIONS .....	46
<b>CHAPTER 5 SCALABLE JOINS &amp; SEMI-JOINS.....</b>	<b>50</b>
5.1 INNER/OUTER JOINS: WHEN & HOW .....	50
5.2 SEMI-JOINS (EXISTS/IN) & ANTI-JOINS (NOT EXISTS).....	54
5.3 JOIN ORDERING & BROADCAST VS. SHUFFLE STRATEGIES.....	57
<b>CHAPTER 6 PERFORMANCE TUNING: INDEXES, PARTITIONING</b> <b>&amp; CLUSTERING.....</b>	<b>61</b>
6.1 DESIGNING B-TREE & HASH INDEXES .....	61
6.2 TABLE PARTITIONING FOR TIME & RANGE .....	64
6.3 CLUSTERING KEYS & MATERIALIZED VIEWS.....	67

6.4 READING & INTERPRETING EXECUTION PLANS.....	71
<b>CHAPTER 7 OPEN TABLE FORMATS &amp; CHANGE DATA CAPTURE .....</b>	<b>75</b>
7.1 APACHE ICEBERG & DELTA LAKE ESSENTIALS .....	75
7.2 SCHEMA EVOLUTION & TIME TRAVEL .....	78
7.3 REAL-TIME INGESTION WITH SNOWPIPE & KAFKA CONNECT .....	82
<b>CHAPTER 8 SQL-FIRST ORCHESTRATION &amp; CI/CD .....</b>	<b>87</b>
8.1 EMBEDDING SQL IN AIRFLOW & PREFECT TASKS .....	87
8.2 VERSION CONTROL OF SQL MODELS (DBT).....	90
8.3 AUTOMATED TESTING & DATA QUALITY CHECKS .....	94
<b>CHAPTER 9 GOVERNANCE, SECURITY &amp; LINEAGE.....</b>	<b>98</b>
9.1 ROW-LEVEL SECURITY & DYNAMIC MASKING.....	98
9.2 AUDIT LOGGING & DATA LINEAGE TOOLS.....	102
9.3 GDPR/CCPA-READY PIPELINES .....	106
<b>CHAPTER 10 EXTENDING SQL: UDFS &amp; AI-DRIVEN OPTIMIZATION.....</b>	<b>110</b>
10.1 PYTHON & JAVASCRIPT UDFS IN MODERN WAREHOUSES.....	110
10.2 LEVERAGING BUILT-IN TIME-SERIES FUNCTIONS .....	113
10.3 AI-POWERED QUERY OPTIMIZERS & AUTO-TUNING .....	117
<b>CHAPTER 11 END-TO-END CASE STUDIES .....</b>	<b>123</b>
11.1 RETAIL SALES PIPELINE ON SNOWFLAKE .....	123
11.2 STREAMING CLICKSTREAM ANALYTICS IN BIGQUERY .....	127
11.3 MULTI-CLOUD HYBRID ARCHITECTURE.....	132
<b>CHAPTER 12 FUTURE TRENDS &amp; NEXT STEPS.....</b>	<b>139</b>
12.1 SERVERLESS SQL ENGINES.....	139
12.2 DATA MESH & FEDERATED ANALYTICS .....	141
12.3 CONTINUOUS LEARNING RESOURCES .....	142
<b>APPENDIX A SQL STYLE GUIDE &amp; NAMING CONVENTIONS ..</b>	<b>145</b>
<b>APPENDIX B CHEAT SHEET: COMMON CLAUSES &amp; SYNTAX</b>	<b>148</b>