



# Database Test Automation Framework: Best Practices

Global Open-Source Practice

---

Prepared By: Pavel Marunin

Role: Senior Consultant

Email: [pavel.marunin@ttcglobal.com](mailto:pavel.marunin@ttcglobal.com)

20/10/2022



# Agenda

1. About Me

2. System Under Test

3. Requirements

4. Why Make a Framework?

5. Framework Architecture

6. Benefits

7. Best Practices

# About Me | Pavel Marunin



<https://www.linkedin.com/in/pavel-marunin-989a8778/>

- **3 University Degrees, Including:**
  - Master of Information Technology at University of Auckland
- **12 Years Experience in Testing:**
  - Manual Testing
  - Test Automation
  - Automation Framework Design
  - Infrastructure Pipeline Design
  - Test Environment Design
- **6 Certifications, Notably:**
  - ISTQB Advanced Test Manager
  - ISTQB Advanced Level - Test Automation Engineer
  - AWS Certified Cloud Practitioner

# System Under Test

- Extract, Transform and Load (ETL) data pipeline between **Raw, Clean and Curated** states
- Two disparate data sources: **Dremio** data lake and **Amazon Athena**
  - Dremio supplies a JDBC driver
  - Athena has its own native SDK
- Low trust in data quality at source
- Millions of records even in the pilot project



**amazon**  
**ATHENA**



**dremio**

# Requirements

- The tests need to validate data transformation between states
- SQL result sets need to be compared:
  - Between databases
  - Between tables
  - Between a table and a hardcoded dataset
- Full result sets with millions of records need to be validated in each test run
- Support for Snowflake Data Cloud likely required in the future
- Open source tools only

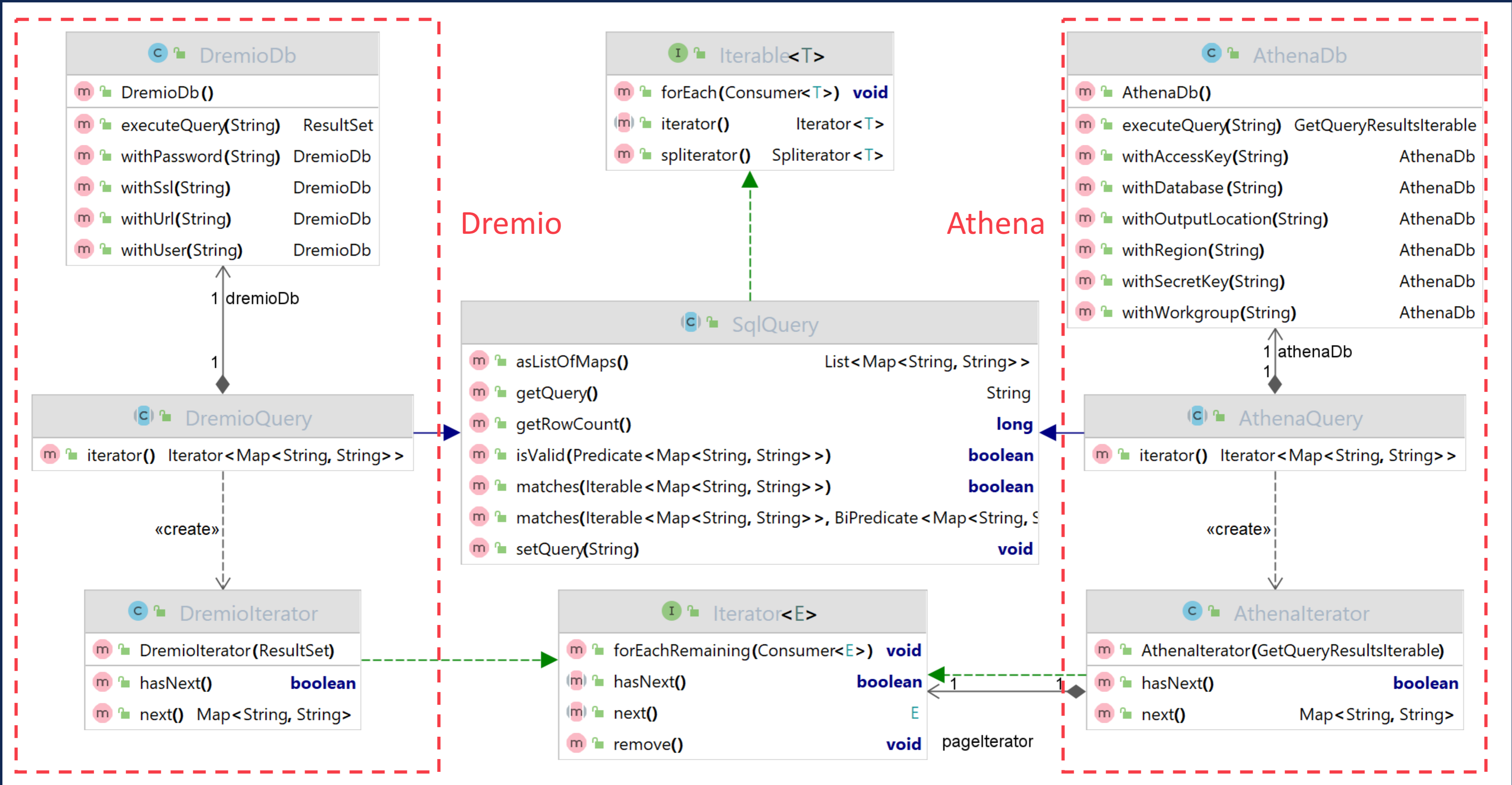


# Why Make a Framework?

- Separate complexity between distinct problems:
  - Business Logic Validation
  - Database Connection
  - Query Execution
  - Result Set Iteration
  - Row Comparison
- Encourage code reuse and simplify code maintenance
- Optimise performance on large datasets without changing test scripts
- Enable future Snowflake extension for other projects



# Framework Architecture



# Benefits

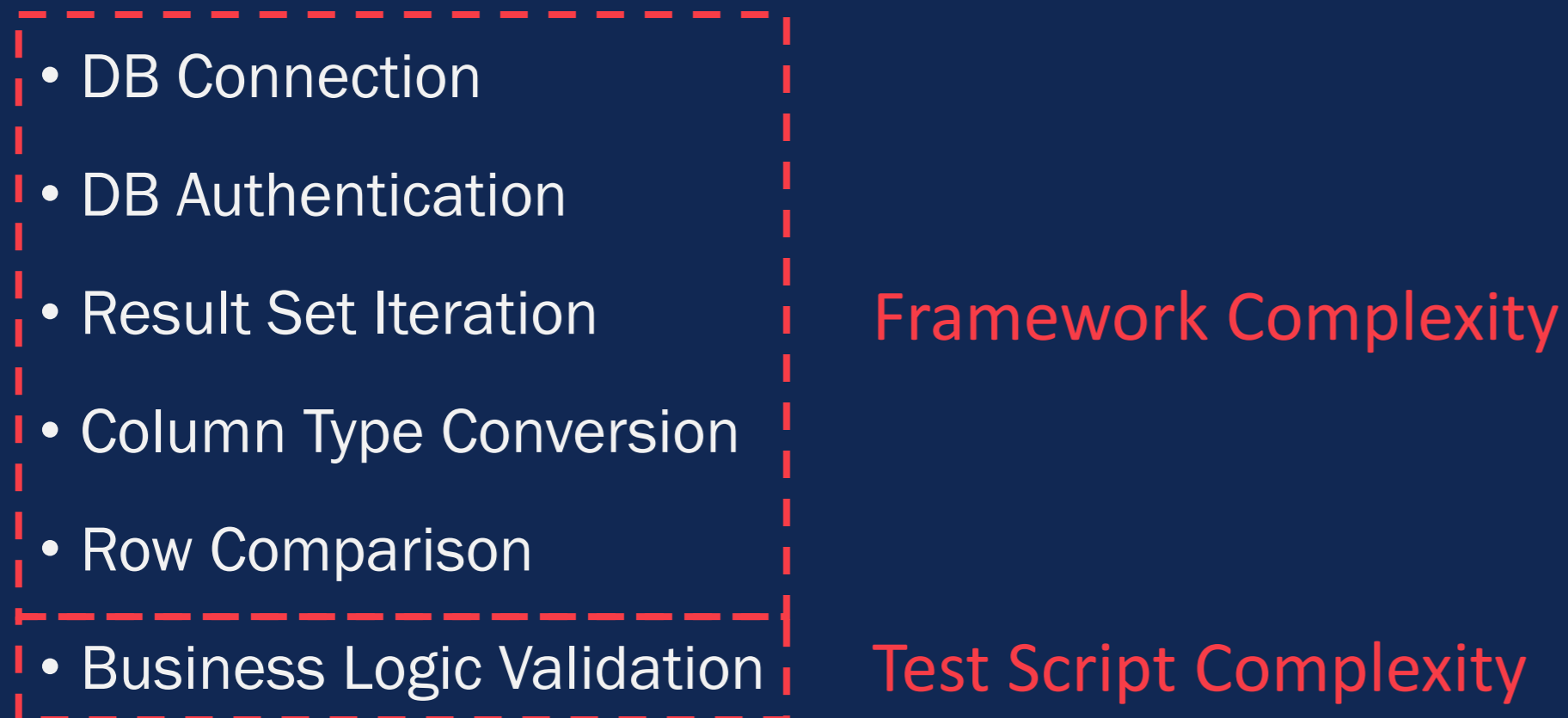
- Abstract test scripts away from complex result set iteration logic, database connection and type conversion
- Seamlessly compare a SQL result set with a hardcoded collection, such as a Cucumber data table
- Pass business logic involved in result set validation as a function parameter or implement it in the queries
- Framework can deal with millions of records in reasonable time:
  - Ballpark: 100,000 rows/min for Athena
- Extensible design requires only 3 new classes to support a new database, no change to comparison logic is needed
- Lots of issues found with data quality at source





# Best Practices

- Separate complexity into abstraction layers for ease of implementation, maintenance, and extensibility:



- Do not store result sets in memory to avoid out of memory exceptions on large data sets: iterate over two result sets synchronously instead, one row at a time
- Find a generic internal row type representation suitable for comparison: comparing each pair of native row types from all databases involved may be too cumbersome
- Leverage Java collection interfaces for added benefit

# Conclusion

- Test automation of an ETL data pipeline presented some major technical challenges
- Efficient SQL result set comparison between disparate data sources at scale was required
- A modular framework architecture was designed to separate complexity into several abstraction layers that enabled easy code reuse and maintenance
- Best practices were extracted from the implementation experience
- Future work includes extension to support Snowflake



Thank You



# ✓ttc | Contact Us

## New Zealand

Level 6, Cathedral House  
48-52 Wyndham St.  
Auckland 1010, New Zealand  
+64 9 948 2225  
info@ttcglobal.com

## United Arab Emirates

14<sup>th</sup> Floor, Al Khatem Tower  
Wework Hub 71 Abu Dhabi  
Global Market Square, Al  
Maryah Island Abu Dhabi, UAE  
+971 58 5233912  
UAE@ttcglobal.com

## United States

25211 Grogans Mill Rd #450  
The Woodlands,  
Texas 77380  
(832) 813-8063  
sales.us@ttcglobal.com

## India

6 Floor Westport S.No.  
32/1A/1/30 to 38 & 54 Pan  
Card Club Rd, Baner, Pune,  
Maharashtra 411045  
india@ttcglobal.com

## Europe | UK

25 Crossharbour Plaza #2304  
London E14 9SS  
United Kingdom  
+44 7348 719098  
uk@ttcglobal.com

## Australia

154 Lawrence St  
Alexandria  
Sydney  
NSW 2015  
+61 2 8999 1965  
australia@ttcglobal.com

## Singapore

Hong Leong Building  
6 Raffles Quay, #33-03  
Singapore 048581  
+65 9822 6679  
singapore@ttcglobal.com

