# MASTER CAREER KNOWLEDGE DOCUMENT (FINAL)

Subject: Zaur Tskhvaradze

Primary Role: Senior Data Quality Engineer

This document is the single source of truth for a career-focused chatbot.

## 1. Professional Identity & Narrative

Zaur Tskhvaradze is a Senior Data Quality Engineer with a non-linear but highly intentional career path. After more than a decade working in public and private sectors (including finance and procurement leadership roles), he deliberately transitioned into tech to pursue continuous growth, complex problem-solving, and long-term relevance.

His entry into data engineering through EPAM's intensive training program became the turning point where he discovered a strong aptitude for data accuracy, reliability, and validation at scale. He views clean, trustworthy data as the foundation of sound decision-making and has built his career around making data issues visible early, measurable, and fixable through automation.

## 2. Core Expertise Areas

- Data Quality Engineering (DQE)
- Automated Data Testing Frameworks
- ETL / ELT Validation
- Enterprise Data Migration & Integration QA
- Regression & Reconciliation Testing
- Large-scale Data Warehouses
- Stakeholder-facing Quality Reporting

## 3. Technical Stack & Tools (Verified Usage)

Primary
- Databricks: notebooks, test automation jobs, result persistence
- PySpark: re-implementing ETL logic for comparison and validation
- SQL: complex joins, window functions, reconciliation logic
- Snowflake: analytical data warehouse
- Stonebranch: orchestration
- Python: automation, scripting, validation logic

Secondary
- Azure Data Factory
- Azure Data Lake

Data Formats
- Delta
- Parquet
- CSV
- JSON

## 4. Scale & Complexity

- Pipelines: ~100–1000 depending on project
- Tables: from small reference tables to billion-row fact tables

- Source systems: up to 10 per project
- Batch frequency: hourly, daily, near-real-time

Zaur's work focuses on system-wide correctness, not isolated pipelines.

## 5. Databricks & Spark Responsibilities

Zaur actively:
- Builds Databricks notebooks for automated DQ checks
- Configures scheduled jobs to test ETL pipelines
- Stores validation results as queryable tables

He frequently re-created developer Spark logic to ensure exact source–target equivalence.
Performance tuning was not the main objective, but correctness and consistency were.

## 6. Data Quality Frameworks (Key Differentiator)

Zaur has designed multiple DQ frameworks from scratch, adapting architecture to project constraints and maturity.

Framework Characteristics
- Rule-based and config-driven (YAML/JSON)
- Designed for multi-source environments
- Built for regression-heavy testing

Checks Implemented
- Source-to-target row & column validation
- Primary key validation
- Null & duplicate detection
- Leading/trailing space checks
- Reconciliation
- Schema / DDL validation

Results & Reporting
- Persisted in Databricks tables
- Documented and tracked in Confluence
- Used in go/no-go release decisions

Business Value
- ~70% reduction in manual testing
- Enabled fast, repeatable regression testing
- Made large-scale comparisons feasible

## 7. Migration & Integration QA

Zaur led QA for:
- On-prem → cloud migrations
- Legacy → new system migrations

Approach
- Systematic reconciliation
- Business rule validation
- Schema alignment checks

Issues were investigated and logged in Jira with clear expected outcomes.
Zaur participated in and signed off go-live readiness.

## 8. Leadership & Seniority

- Mentored up to 10 junior engineers
- Mentoring included onboarding, reviews, and learning plans
- Conducted technical interviews (theoretical + practical)
- Acted as internal assessment expert:
- evaluated promotion candidates
- assessed technical depth and soft skills

## 9. Measurable Impact

- ~70% reduction in manual testing effort
- ~95% of defects caught before production
- ~50% reduction in testing cycle time
- Consistently positive client feedback
- Multiple long-term contract extensions
- EPAM CEO Award – IMPACT (2024)

10. Work Philosophy & Behavior
Data Quality Philosophy
“Data quality is about trust and predictability. My goal is to make data issues visible early, measurable, and repeatable through automation.”

Automation vs Manual Testing
Automation is preferred for all repeatable and regression scenarios.
Manual testing is reserved for early project stages and exploratory validation of complex logic.

Communication Style
Zaur communicates issues early, clearly, and without blame, focusing on impact and next steps for both technical and non-technical stakeholders.

Under Pressure
He remains calm, structured, and solution-oriented, prioritizing critical risks over reactive fixes.

11. Learning Mindset & Community
Zaur actively:
- Participates in local data communities
- Attends webinars and training
- Continuously upskills through courses and self-study
- Uses AI tools to research, learn, and prototype ideas faster

He strongly believes in growth mindset and knowledge sharing, values he developed early during EPAM Campus training.

12. Professional Boundaries (Important for Chatbot Accuracy)
Zaur does not position himself as:
- ML model developer
- Backend application engineer
- BI / dashboard specialist

His core value is data correctness, reliability, and testability at scale.

13. Personal Dimension (Light, Human Context)
Outside of work, Zaur:
- Enjoys preparing healthy meals and experimenting in the kitchen
- Swims and exercises regularly
- Enjoys camping and time in nature

These activities help him stay balanced, focused, and creative.