Internal Reference – Unified CDD Schema Challenges & BA Action Plan

# Background:

As part of the Quantexa integration project, we initially began mapping CDD source systems individually into the landing layer. However, Quantexa's latest requirement involves designing a unified schema that consolidates all CDD data into a single structure. This introduces several complexities, especially due to fragmented data and inconsistent field representations across systems.

# Challenges and Corresponding BA Action Items:

## 1. Fragmented Data Across Tables

🔹 Challenge: Information related to a single customer is spread across multiple tables in the same system (T1 level).

* ✅ Action Items:
* - Identify primary and supporting tables in each source system.
* - Map relationships and dependencies between tables.
* - Request ERDs or source documentation from data owners.

## 2. Column Name Misalignment Across Systems

🔹 Challenge: Same data element is named differently in different systems.

* ✅ Action Items:
* - Create a synonym dictionary or crosswalk sheet.
* - Align with Quantexa team on naming convention to be followed in unified schema.

## 3. Data Semantics Confusion

🔹 Challenge: Similar fields might have different business logic or scales (e.g., different risk score ranges).

* ✅ Action Items:
* - Capture field definitions and logic from each source.
* - Confirm expected standard format with Quantexa (e.g., unified risk score scale).

## 4. Duplicate or Conflicting Records

🔹 Challenge: Merging sources may introduce duplicate customer records or conflicting details.

* ✅ Action Items:
* - Check Quantexa's preferred deduplication logic.
* - Ensure traceability by storing source and timestamp.

## 5. Missing or Optional Fields

🔹 Challenge: Some systems may lack certain fields expected in the unified schema.

* ✅ Action Items:
* - Use MoSCoW prioritization per field.
* - Seek clarification from Quantexa whether null/defaults are allowed.

## 6. Transformation Overhead

🔹 Challenge: Fields may require standardization or data enrichment.

* ✅ Action Items:
* - Prepare transformation logic for each required field.
* - Validate with Quantexa and internal dev teams.

## 7. Scalability of the Unified Schema

🔹 Challenge: Schema needs to handle future additions without full redesign.

* ✅ Action Items:
* - Check if schema should be versioned or extensible.
* - Recommend semi-flexible formats (e.g., JSON with optional fields).

## 8. Tight Timelines vs Requirement Clarity

🔹 Challenge: Risk of misdevelopment due to partial clarity.

* ✅ Action Items:
* - Split requirement delivery in phases (Must-have vs. Nice-to-have).
* - Regularly update field-level closure tracker with decisions from Quantexa.

# Suggested Talking Points for Quantexa Meeting:

Hi team, as part of our current scope, we are now focusing on consolidating multiple CDD sources into a unified schema for your contextual scoring engine. We’ve identified a few challenges that we believe need collaborative resolution.

1. Many systems store information across multiple tables. We’ll need help validating which tables are primary vs. supporting.  
2. Field names and logic vary across systems. We'd like to propose a field crosswalk and confirm preferred standards.  
3. In case of duplicate values across systems, can you help us finalize source priority rules?  
4. What transformation rules does Quantexa expect for inputs like risk score, customer type, and watchlist flags?  
5. Lastly, how flexible should the schema be in terms of future additions—strict schema or schema-on-read model?

Once these questions are aligned, we’ll prepare a clean requirement document and pass it to developers for implementation.