

# Accounting Firm: Data, Automation & Governance Assessment

---

## Independent Assessment for Multi-Client Practices (50–500 Clients)

---

### Overview

A multi-client accounting firm services between 50 and 500 SME clients across tax compliance, bookkeeping, payroll, and statutory reporting. Day-to-day processing is handled by junior accountants and administrative staff, while a small number of Chartered Accountants (CAs) focus on complex technical reviews and final sign-off.

The firm's operating model works, but is under increasing strain from scale, document volume, and data fragmentation:

- Hundreds of invoices, receipts, and bank transactions per client
- Client documents arriving via email, portals, and ad hoc uploads
- Manual invoice reading, duplication checks, and bank matching
- Heavy reliance on junior staff for first-pass processing
- Senior staff spending time validating work instead of applying judgment

The owner does not want to hire more people. The objective is to **improve throughput, accuracy, and control using existing processes**, while strengthening data security, auditability, and confidence in outputs.

---

### The Scenario

#### Current Operating Reality

- Interns or junior accountants email clients (SMTP hosted on cPanel) requesting:
  - Tax information
  - Payslips
  - Medical payments
  - Supplier invoices
  - Bank statements
- Documents are returned via:
  - Email attachments
  - PDFs
  - Scanned images
  - Re-sent or duplicated files
- Processing involves:
  - Manual invoice reading and coding
  - Visual duplication checks
  - Spreadsheet-based tracking
  - Manual bank statement matching
  - Posting into Sage, Xero, SAP, Oracle, or similar systems
- Chartered Accountants:

- Re-check classifications
  - Validate reconciliations
  - Resolve inconsistencies
  - Carry professional and regulatory risk
- 

## Core Risks & Constraints

### Operational Risk

- Repeated manual effort on low-value work
- Inconsistent processing quality across juniors
- Scaling constrained by human throughput

### Professional Risk

- Duplicate invoices posted unknowingly
- Missed or misclassified expenses
- Weak defensibility of decisions under review

### Data & Security Risk

- Sensitive client data moving via email
  - No unified view of document lineage
  - Limited visibility into who accessed or modified data
- 

## Assessment Objective

This assessment evaluates how data enters, moves through, and exits the firm's accounting processes—and how automation and data governance can reduce workload **without replacing existing accounting systems**.

The assessment aims to:

- Absorb higher volumes with the same team
  - Reduce duplication and reconciliation errors
  - Improve confidence in junior-level processing
  - Free CAs to focus on judgment and complex matters
  - Strengthen security, audit trails, and regulatory defensibility
- 

## Scope of the Assessment

### 1. Data Sources & Intake Mapping

Review how information enters the firm:

- Client emails and attachments (SMTP / cPanel)
- Document formats (PDF, image, scan)
- Bank statements (PDF, CSV, OFX)
- Repeated submissions and versioning issues

#### Key Questions:

- Where does duplication first occur?

- How often are documents reprocessed unknowingly?
  - Are documents uniquely identifiable today?
- 

## 2. Invoice Duplication & Integrity Review

Independent analysis of:

- How invoices are identified as new vs existing
- Whether the same invoice can be posted multiple times
- How modified or reissued invoices are treated

### Diagnostic Questions:

- How many invoices are duplicates with different filenames?
  - Are MD5 or content hashes used to fingerprint documents?
  - Can visually different PDFs represent the same invoice?
- 

## 3. Invoice & Receipt Reading Accuracy

Assessment of:

- Manual invoice interpretation by junior staff
- Consistency of expense classification across clients
- Treatment of partial invoices and bundled receipts

### Questions Answered:

- Which invoice fields are most commonly misread?
  - How often do CAs override junior coding?
  - Are similar invoices treated consistently?
- 

## 4. Bank Matching & Reconciliation Analysis

Review of:

- Manual bank-to-invoice matching processes
- Handling of unmatched or ambiguous transactions
- Time spent on obvious vs true exception items

### Key Questions:

- Which transaction types cause the most friction?
  - How much CA time is spent resolving basic matches?
  - Where does reconciliation rely on memory instead of data?
- 

## Role of Databricks (As an Intelligence Layer)

This assessment explicitly considers how Databricks can operate **alongside** existing accounting systems.

Databricks is **not a replacement** for Sage, Xero, SAP, or Oracle. It performs the work those systems are not designed to do.

### What Databricks Enables

#### AI-Based Invoice & Receipt Reading

- Reads invoices and receipts (PDF, image, scan)
- Extracts supplier, date, VAT, totals, and line items
- Classifies expenses consistently across clients
- Assigns a confidence score to every extraction

Junior staff only review low- and medium-confidence items.

---

### **Duplication Detection Using Hashing**

- Generates document fingerprints (MD5 / content hashes)
- Detects exact and near-duplicate invoices
- Flags re-sent or modified documents automatically

Prevents silent overstatement of expenses and VAT errors.

---

### **Intelligent Bank Matching**

- Probabilistic matching of bank transactions to invoices
  - Scoring based on amount, date, supplier, and description
  - Clear separation of:
    - High-confidence auto-matches
    - Items requiring junior confirmation
    - True exceptions for CA review
- 

### **Email & Intake Control**

- Tracks inbound emails and attachments securely
  - Links documents to client and reporting period
  - Identifies repeated submissions automatically
  - Maintains full document lineage and audit trail
- 

## **Data Governance & Security Review**

The assessment evaluates:

- Encryption of data at rest and in transit
- Role-based access controls
- Client data segregation
- Full logging of access and changes
- Human-in-the-loop validation of AI outputs

### **Governance Questions:**

- Who can see which client data?
  - How are AI decisions reviewed and overridden?
  - Are adjustments traceable to a person and timestamp?
  - Can decisions be defended to auditors and regulators?
- 

## **Human-in-the-Loop by Design**

This model does not remove people.

- AI performs reading, matching, and duplication checks
- Every output includes a confidence score
- Junior staff confirm or correct flagged items
- Chartered Accountants focus on judgment, not data cleanup

This preserves professional oversight, compliance, and staff development.

---

## Deliverables

- End-to-end data flow maps (email → document → ledger)
  - Duplication risk analysis
  - Invoice and receipt error patterns
  - Bank reconciliation friction points
  - Data governance and security findings
  - Prioritised recommendations ranked by:
    - Risk reduction
    - Time saved
    - Professional exposure
    - Ease of adoption
- 

## What This Is — and Is Not

### **This is:**

- An independent operational assessment
- Automation-aware, not vendor-led
- Security- and governance-focused
- Designed for firms that already function well

### **This is not:**

- A system replacement project
- A headcount reduction exercise
- A software sales pitch
- An experiment with professional risk