

Statement Hive: Solving Multi-Bank Statement Reconciliation at Scale

The Problem We Faced

In reconciliation process, vouchers were already standardized and generated internally in a fixed, reconciliation-ready format. However, bank statements remained the biggest challenge, as each bank and CMS partner followed its own structure, delivery method, and file format. This diversity made reconciliation complex, manual-heavy, and difficult to scale, while also impacting accuracy and audit compliance.

Problem Statement and Challenges

Challenge Area	Details
Number of Banks	31 different banks
Multiple CMS partners	Multiple CMS partners
Statement Variability	Different statement formats for every bank-CMS combination
Ingestion Channels	Email, APIs, SFTP
File Formats	XLS, XLSX, XLSB, XML, PDF, CSV password-protected ZIP files
Processing Method	XLS, XLSX, XLSB, XML, PDF, CSV password-protected ZIP files
Operational Impact	XLS, XLSX, XLSB, XML, PDF, CSV password-protected ZIP files

What Is Statement Hive?

Statement Hive is a centralized platform designed to bring order to fragmented bank and CMS statements. It ingests data from multiple sources, standardizes it into a single format, validates its integrity, and prepares it for seamless reconciliation with vouchers.

The platform ensures that no matter where the statement comes from or how it is formatted, it follows one consistent, recon-ready flow.

Statement Hive supports active participation from multiple banking partners. The participating banks include ICICI Bank, Standard Chartered Bank, SCB Central Bank of India (CBI), Federal Bank, HDFC Bank, IDBI Bank, Kotak Mahindra Bank, IndusInd Bank, Indian Overseas Bank (IOB), Punjab National Bank (PNB), Punjab & Sind Bank (PSB), and UCO Bank.

What Statement Hive Does

- Ingests bank and CMS statements
- Normalizes all formats into a single structure
- Performs validations and integrity checks
- Prepares clean data for reconciliation

Technology Overview

Statement Hive is built with scalability, performance, and auditability in mind.

The architecture separates ingestion, transformation, and reconciliation responsibilities across dedicated services.

Technology Stack

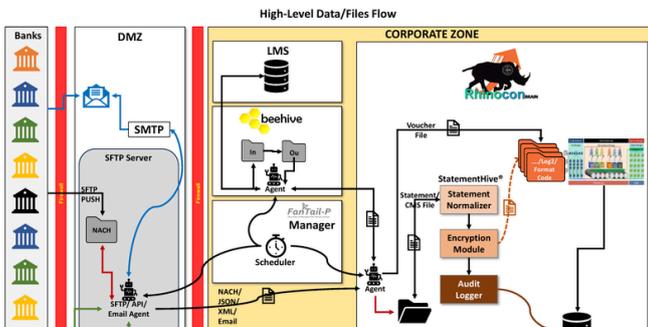
- Python (multithreaded utilities)
 - File ingestion
 - Format conversion and transformation
- Java Spring Boot (multithreaded services)
 - Parsing
 - Validation

- *Staging and reconciliation Staged database architecture Ensures data accuracy Maintains traceability at every step*

How Statement Hive Works (End-to-End Flow)

Step 1: Statement Ingestion

Statements enter Statement Hive through multiple channels. Regardless of the source, every file is brought into a common ingestion flow to maintain consistency.



Ingestion Channels

Email

- Statements arrive as email attachments
- Our ingestion utility, Fantail:
 - Downloads emails
 - Extracts attachments
 - Renames files using a standard naming convention filename#senderMailId#mailSubject#date
 - Transfers files to the Statement Hive input directory

APIs

- Some banks provide statement APIs
- A dedicated Python utility (API Banking):
 - Fetches data from bank APIs
 - Converts responses into files
 - Pushes them into the same Fantail-driven flow

SFTP

- Files placed in configured SFTP folders
- Automatically picked up by Statement Hive listeners

It uses three '#' values to fetch metadata mapped to the subject line from the database. The subject line provides the format code, module ID, and bank ID, which are required for the RhinoCon file-naming convention

XLS, XLSX, XLSB, CSV, XML, PDF → CSV

Extracts ZIP files

Unlocks password-protected statements

Normalizes column structures.

Format Code Framework

Each bank statement is mapped to a Format Code, which defines:

- Column mapping
- Date formats
- Debit and credit logic
- Balance positioning rules

Outcome: Every statement is transformed into a single standardized CSV format.

Step 3: File Parsing & Staging

After transformation, the standardized CSV files are processed by a Java Spring Boot File Processing Service. This service ensures high throughput while safely loading data into the system.

Responsibilities

- Parse CSV records
- Load data into staging tables
- Use controlled multithreading for performance and stability

Step 4: Statement Validation & Integrity Checks

Before reconciliation, every bank statement undergoes strict validation checks. This ensures only clean and consistent data enters the reconciliation layer.

Validation Checks

- Opening balance integrity validation
- File-level transaction integrity
- Duplicate transaction detection
- Last processing date validation

Step 2: Universal Format Transformation

Once files reach Statement Hive, a multithreaded Python transformation engine takes over. Its primary role is to remove format diversity and bring every statement into a single standardized structure.

Transformation Capabilities

Converts all formats to CSV:

Data Flow

Validated data moves through the following stages:

Stage 1 → Stage 2 → Stage 3 → Stage 4

Step 5: CMS Partner Processing

CMS partner files follow the same ingestion and transformation pipeline to maintain consistency. However, since CMS files do not contain bank balances, certain validations are skipped.

CMS Processing Rules

- Balance and file integrity checks are skipped
- CMS data is promoted directly to Stage 2

This approach keeps CMS processing efficient without compromising standardization.

Step 6: Matching & Reconciliation Readiness

At Stage 2, Statement Hive performs transaction matching using predefined reconciliation rules. This is where bank statements and CMS transactions come together.

Matching Outcomes

- Matched CMS transactions move to Stage 3
- Stage 3 data becomes fully reconciliation-ready with vouchers
- Unmatched or suspicious transactions remain available for investigation and audit

Why Statement Hive Works

Statement Hive succeeds because it combines automation, standardization, and scalability.

USE CASES

Statement Hive is primarily used by L&T Finance. It supports 31 banks, including HDFC, ICICI, Kotak, SCB, and IndusInd Bank, and 47 CMS partners that participate in the Statement Hive ecosystem.



Poulami Mondal
Business Analyst