

# ORCHESTRATING ORDER

How Apache HOP Powers Data Integration  
Across a Fast-Growing Business Ecosystem



## 2025

WHITE PAPER BY

know**.bi**

## EXECUTIVE SUMMARY

Fast growth often comes with technological growing pains. In industries where technology maturity lags, businesses frequently end up with a patchwork of tools, platforms, and data formats that make integration a massive challenge. At the center of our 3PL-focused family of companies, Apache HOP plays a critical role in bridging these gaps—connecting systems, automating processes, and enabling data to move freely and reliably across the organization...and keep our customers' goods and payments moving efficiently.

## THE CHALLENGE

Our business ecosystem spans several rapidly evolving vertical markets from telecom to industrial supplies. As we scaled, we adopted tools to meet immediate needs. Each served a purpose but created complexity when combined:

- Salesforce as our CRM
- On-premise ERP system for core operations
- On-premise data warehouse
- Power BI for analytics and reporting
- An eCommerce website with custom integrations

Our primary customer adds additional complexity with platforms through their own third-party partners, namely:

- A custom web-hosted retail POS software platform, developed specifically for our primary customer's interests – no pre-built integrations.
- A large financial payment processing platform with limited reporting capability

On top of that, we work with our own partners who introduce more data points:

- A gamification platform to drive employee engagement, requiring details sales metrics down to the individual front line staffer's daily transaction details.
- A rewards program system to drive end-user engagement with our platforms
- A trade-in program platform to simplify upgrades for end-users who want them, and bring cash to stores in the program.

The majority of data transfer historically relied on flat files (CSV) transfers by SFTP in the best cases, and hand-scraping Excel or PowerPoint files in the worst ones. Two exceptions are UPS and Salesforce, which integrate via API. The rest created a fragile, manual web of dependencies, with almost no standardization or automation.

## ECOSYSTEM OVERVIEW

The diversity of tools and platforms meant data silos were rampant. Each team and department had their own practices. Data also moved on unpredictable schedules between teams or team members. Errors introduced in one system could cascade downstream, going undetected for days, and sometimes weeks. Business leaders struggled to access timely, accurate data. New integrations took weeks or months, and each required custom logic and manual oversight. Every integration was a new integration, starting from scratch.

Apache HOP at the Core Apache HOP became our answer to this fragmentation. It sits at the center of our data ingestion process, providing:

- **Orchestration:** Apache HOP, in conjunction with Cronicle, coordinates the movement of data between systems, triggering workflows based on time or event.
- **Transformation:** Using reusable pipelines and metadata injection, we standardize disparate file formats, naming conventions, and data models.
- **Transport:** Whether the endpoints are pulling files from FTP, reading from APIs, or consuming local folders, Apache HOP handles diverse methods of data movement.
- **Monitoring and Logging:** Each workflow logs success/failure, processing times, and data volumes, giving us visibility into system health.
- **Error Handling:** Apache HOP workflows include conditional logic, notifications, and retries to gracefully handle failures.
- **Scheduling:** Workflows are triggered on demand, by schedule, or downstream from other events, creating a cohesive flow of information.
- **Scale:** We are presently executing approximately 180 workflows and pipelines daily, ranging from seconds to minutes of run time each. All of this is achieved using docker and orchestrated on a single virtual server which interfaces with our SFTP, API, and DB servers in our on-premise cloud. We are essentially future-proof at the current time due to the efficiency of Hop's microservice architecture.

## REAL-WORLD EXAMPLES

- Data from the customer's POS arrives via CSV over FTP. Apache HOP retrieves, verifies, standardizes, and loads it into our data warehouse, making it available for reporting. Having HOP pre-build summary tables each day makes reporting a breeze.
- Our primary customer receives daily summaries of warehouse activity including on-hand quantities and shipment details by customer, compiled by Hop and emailed automatically, streamlining inventory forecasting.
- Stakeholders receive timely reports of inventory and current orders from our ERP based on logic requested by the business and executed in Hop.
- The finance team compiles upwards of \$1.5MM in micropayments into customer reimbursements, processing millions of rows in seconds to minutes, allowing quicker customer payment turn time.
- New product data arrives over FTP or API. Apache HOP standardizes and loads it into our data warehouse, making it available for reporting, as well as loading it to our ERP for website product updates.
- Our gamification vendor requires up-to-date frontline staff information for each retail store (2000+) with whom we partner. Using HOP allows us to compile and provide consistent and correct information even in a high-turnover retail environment. HOP lets us detect new users or terminated ones and highlight them to the vendor as the users come and go.
- UPS is polled daily by a HOP workflow to retrieve shipment tracking details, matched to customer orders in ERP for predictive ordering shipment calculations and for use in ERP and Ecommerce site.
- HOP is one of the few open ETL platforms to seamlessly and relatively painlessly interface with Salesforce, allowing movement of data both in and out of our CRM to reporting and ERP platforms.

## OUTCOMES AND BENEFITS

- Reduce manual effort by approximately 3-4 FTE equivalent.
- Improved data accuracy by applying validation and transformation rules at ingestion
- Enabled daily executive dashboards with near-real-time data from previously siloed systems
- Onboarded new partners and systems in weeks rather than months
- Centralized data management and auditability with consistent workflows and logging

## LESSONS LEARNED & BEST PRACTICES

- Reusable Pipelines: Designing modular workflows speeds up new integrations.
- Utilize Project and Environment variables: HOP's use of the project and environment configuration files allows a large amount of pre-configuration and allows development of workflows to be deployed to a heterogeneous server environment without re-configuring each item in a workflow or pipeline. Develop on one platform (i.e. Mac) and deploy on Windows or Linux by selecting the appropriate environment. Switch to a development environment to safely test ingestion, and then deploy to production.
- Increased security: HOP's ability to store encrypted user info and passwords in the environment files means fewer places for sensitive information to leak.
- Standard Naming Conventions: A shared taxonomy across files and fields reduces confusion.
- Centralized Logging: Helps with monitoring, debugging, and compliance.
- Version Control: HOP's compatibility and integration with git means change tracking and deployment are not just desirable, they're literally part of the workflow development process.

## NEXT STEPS AND FUTURE PLANS

- Transition more flat file processes to APIs and event-driven architecture
- Implementing real-time pipeline execution with queue-based triggers
- Expand our HOP repository of framework jobs for faster onboarding of new systems and partners
- Continue to take advantage of Metadata Injection and reusable code to simplify future development
- Implement full unit testing for validation of core data ingestion workflows

## CONCLUSIONS

Apache HOP has transformed how we think about data movement. From chaotic, manual processes to streamlined, automated orchestration, HOP has become the backbone of our data infrastructure—allowing us to scale, integrate, and innovate with confidence.

## **MARCECO: A 30-YEAR EVOLUTION FROM HUSTLE TO HIGH-TECH**

Marceco's story began humbly—with little more than a vision, a credit card, and the determination to be a participant in the for-profit economy. Registered at the local courthouse and armed with a shoulder bag, we set out to build something meaningful. It was a vision rooted in hard work, grit, and above all, the commitment to operate with honesty and integrity.

That vision has carried us far. From our early days of door-to-door hustle, Marceco has grown into a national-scale sales organization. Today, we support a retail channel of over 1,500 locations and operate as a value-added contractor for one of the four major U.S. carriers. In 2025, we proudly celebrate our 30th anniversary—an uncommon milestone in any industry, especially in one as fast-moving as mobile communications.

One thing that has always set Marceco apart is our relentless desire to create value and solve challenging problems—not just for ourselves, but for our partners and customers. We've never been content with off-the-shelf answers. Instead, we've built our own tools, processes, and platforms—driven by a mindset we call “clock building” rather than just “telling time.” That philosophy has fueled innovation, differentiation, and growth.

But it also introduced a new kind of challenge: how to scale and support that complexity. As we expanded, so did the technology burden—volumes of third-party data, no standardized processes, and an ecosystem of partners with custom platforms and limited integration paths.

To stay ahead, we needed more than just tools—we needed orchestration, automation, and confidence in our data. That need led us to Apache HOP.

**For more information,** visit <https://www.marceco.net/>





## ABOUT KNOW BI

At know.bi, we specialize in building robust, scalable, and modern data integration solutions. Apache Hop is at the core of everything we do.

As one of the main contributors to the Apache Hop project, we've been deeply involved since its early days. Our team not only helps shape the platform through code, documentation, and community support, we also bring that expertise directly to our clients.

Our goal is simple: to help organizations take full control of their data projects. Integrate, according to industry best practices and powered by Apache Hop

**For more information,** visit <https://www.know.bi/>