Case Study:
Migrating Legacy ETL to Cloud
About Bitwise

Experienced Global Player
- 1200+ strong team of elite data specialists
- Global delivery centers in Pune, India with corporate headquarters in Chicago, USA
- 21 years in business and a certified Minority Business Enterprise

Technology Leadership that Challenges the Status Quo

Bitwise Innovation
Our customers are never satisfied, and neither are we. Which is why we constantly challenge conventional wisdom to develop cutting-edge technology solutions that drive key business results.

Bitwise Expertise
Our elite team of data and digital experts partner with our customers to solve their most complex problems by pushing the boundaries of innovation and delivery.

Bitwise Difference
We take a different approach to providing a customer-centric experience built on our passionate people, technology innovations and agile, collaborative processes.
Overview

About the Customer

**Fortune 500 Financial Services Company**

- **Diversified bank** that offers a broad array of financial products and services to consumers, small businesses and commercial clients
- Identifies as a **Fintech Company**
- Embraces **open source tools** and fosters environment of internal innovation at an accelerated pace

**Present**

**2013**

- Feature and capabilities comparison of Talend, Ab Initio and Informatica
- Convert complex Ab Initio graphs to target ETL platform
- Develop an open source ELT/ETL platform: Hydrograph
- Hydrograph used to productionize Recovery LOB on Hadoop cluster
- Ab Initio to Hydrograph Conversion evolution
- Hydrograph metrics on Spark cluster proven to be better than any other existing data integration platform
- Hydrograph passes the governance standards for all things data
- Retail; Commercial; Small Business Cards; Auto and Mortgage Loans adopt Hydrograph as their standard data integration product

**Our Journey with the Customer**
Business Case for Migrating to Cloud

Reduce reliance on vendor resources
Reduce product IT costs by migrating all legacy applications from data centers and building all new apps in Cloud
Minimize infrastructure costs by exiting data centers and mainframes

In other words...create a lean team by leveraging Cloud technology.

USE CASE: Build a Centralized HR Analytical Environment in support of organizational strategy to move to Cloud

• Required a process for quickly and efficiently converting legacy Ab Initio graphs to an ETL tool that is compatible with the AWS Cloud platform.
• The target ETL tool needed to provide an easy-to-use development environment that required minimal retraining for developers
• Deliver robust capabilities for matching the performance of Ab Initio.
Bitwise partnered with the customer’s team to implement Hydrograph.

**Hydrograph Implementation**

- Bitwise configured Hydrograph for Cloud environment.
- The solution was developed to run on AWS EC2/EMR (Spark) with Redshift as the target Data Warehouse.

**ETL Conversion (Ab Initio to Hydrograph)**

- Bitwise used its ETL Converter solution which helps to accurately scope the conversion and automate up to 70% of the overall process, including data validation.

Hydrograph is an open source ETL development tool designed for the big data ecosystem that provides a GUI environment with Drag-and-Drop functionality that developers are familiar with.

- Drag-and-Drop GUI
- Open source and license-free
- Easily supports new / custom component development
Hydrograph Implementation

Data Flow

- Workday
- S3
- EC2 **Hydrograph** (Spark)
- Load into Redshift

Methodology

- Bitwise used CI/CD to work within the customer’s development framework
- CI/CD Pipeline details
  - Stargate - onboarding
  - GitHub - code repository
- Maven - manage code dependencies and build
- Jenkins - automated deployment

Infrastructure
Ab Initio to Hydrograph Conversion

Phase 1 – Assessment
Phase 2 – Conversion
Phase 3 – Validation

Assessment

- Understanding compatibility of source & target ETL Technologies
- Current environment Assessment
- Inventory Assessment

Conversion

Export XML

Automated Conversion

Import XML to Hydrograph

Conversion Automation

~70% Effort Savings

Manual Fix

Testing / Data Validation

Test Automation

~40% Effort Savings

Assessment Automation

~60% Effort Savings

Ab Initio to Hydrograph Conversion

Manual Fix

Automated Conversion

Import XML to Hydrograph

Conversion Automation

~70% Effort Savings

Manual Fix

Testing / Data Validation

Test Automation

~40% Effort Savings

Assessment Automation

~60% Effort Savings

Ab Initio to Hydrograph Conversion

Manual Fix

Automated Conversion

Import XML to Hydrograph

Conversion Automation

~70% Effort Savings

Manual Fix

Testing / Data Validation

Test Automation

~40% Effort Savings

Assessment Automation

~60% Effort Savings
ETL Development Environment

Developer Interface
## Addressing Functional Requirements

<table>
<thead>
<tr>
<th>Required Features</th>
<th>Internal ETL Tool</th>
<th>Hydrograph</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUI</td>
<td>No</td>
<td>Intuitive GUI</td>
</tr>
<tr>
<td>Workflow</td>
<td>JSON</td>
<td>XML</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Hard</td>
<td>Easy</td>
</tr>
<tr>
<td>CI/CD</td>
<td>No</td>
<td>Company standards</td>
</tr>
<tr>
<td>Reusable code</td>
<td>No</td>
<td>Subjobs</td>
</tr>
<tr>
<td>Filetypes</td>
<td>CSV, Parquet, JSON</td>
<td>CSV, Fixed Width, Parquet, Hive, XML</td>
</tr>
<tr>
<td>Custom Transformation Development</td>
<td>Difficult</td>
<td>Easy</td>
</tr>
<tr>
<td>Parallel Flows</td>
<td>Not supported</td>
<td>Supported using phasing</td>
</tr>
<tr>
<td>Job tracking</td>
<td>Not supported</td>
<td>Execution tracking</td>
</tr>
<tr>
<td>Data Watchers</td>
<td>Not supported</td>
<td>View data</td>
</tr>
</tbody>
</table>

**Hydrograph met all Functional requirements**
Addressing Data Governance

Customer’s Architectural Governance Requirements

<table>
<thead>
<tr>
<th>Ability to orchestrate job flow across different shops</th>
<th>Ability to support cross technology payload integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to ensure business continuity</td>
<td>Ability to configure to CI/CD pipelines</td>
</tr>
<tr>
<td>Ability to maintain Chain of Trust, perform Impact Analysis and record Quality Stats</td>
<td>Hydrograph met all Governance requirements</td>
</tr>
</tbody>
</table>
Results

Migration of Legacy ETL to AWS Cloud using Hydrograph

One-Time Conversion
No continuous implementation program means no ongoing support costs

No Ongoing Licensing Fee
Hydrograph is open source and license free

Processing Power of Hadoop
Hydrograph harnesses the processing power of Hadoop *(up to 33% faster execution time in Spark compared to native)*

Flexibility of Operating System
Hydrograph runs on Windows and MacOS

Meet or Exceed All KPIs
All KPIs met or exceeded in Cloud environment / *Jobs running in production with 100% accuracy of data loads*
Thank You

Inquiries?
Call us at **847.969.1500**
Email **sales@bitwiseglobal.com**