

# API Extraction Workflow

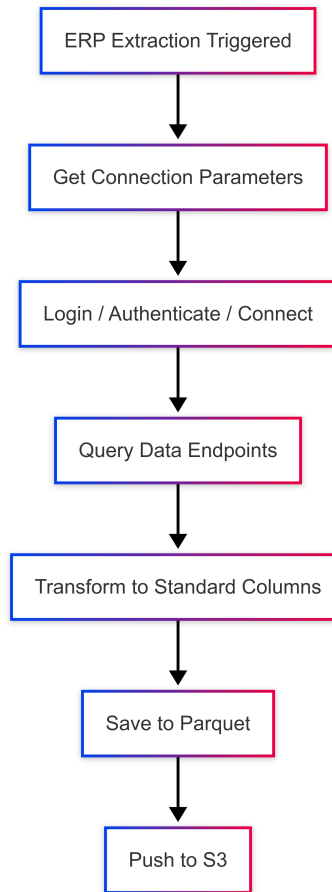
Nathan Lunceford

2025-03-05

## Table of contents

<b>1</b>	<b>Workflow Overview</b>	<b>2</b>
<b>2</b>	<b>Process Steps</b>	<b>2</b>
2.1	ERP Extraction Triggered . . . . .	2
2.2	Get Connection Parameters . . . . .	3
2.3	Query Data Endpoints . . . . .	3
2.4	Transform to Standard Columns . . . . .	3
2.5	Save to Parquet . . . . .	3
2.6	Push to S3 . . . . .	4

# 1 Workflow Overview



## 2 Process Steps

### 2.1 ERP Extraction Triggered

The workflow begins when an extraction request is initiated. This could happen:

- On a scheduled basis (e.g., nightly extractions)
- On-demand through user-initiated requests
- As part of a larger data processing workflow

## 2.2 Get Connection Parameters

The system retrieves the necessary connection parameters for the specified ERP system and client:

- API endpoint URLs
- Connection settings
- Timeout configurations

The system establishes a secure connection to the ERP API by:

- Retrieving secure credentials from the vault
- Applying the appropriate authentication method (API key, OAuth token, etc.)
- Establishing the connection with the ERP API
- Verifying successful authentication

## 2.3 Query Data Endpoints

With a successful connection established, the system queries the appropriate endpoints to extract the required data:

- Sales orders
- Sales order details
- Inventory spec fields
- Customer details
- Vendor details
- Product spec fields

## 2.4 Transform to Standard Columns

The raw data from the ERP system is transformed to follow our standardized data model:

- Column names are mapped to our standard naming conventions
- Basic data quality checks are performed

## 2.5 Save to Parquet

The transformed data is converted to Parquet format:

- Efficient columnar storage format
- Compressed to save space
- Optimized for analytical queries
- Schema is preserved with appropriate data types

## **2.6 Push to S3**

The Parquet files are uploaded to our S3 storage: - Organized by client and data type - Stored with appropriate metadata - Secured with proper access controls - Made available for downstream processing