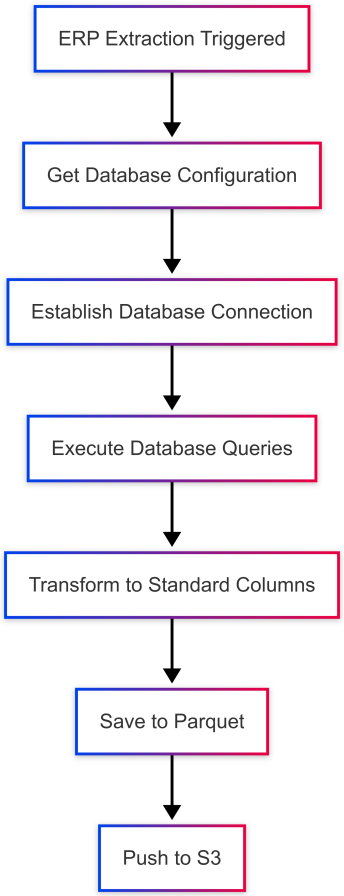
Database Extraction Workflow

Nathan Lunceford

2025-03-05

Table of contents

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 Database Configuration

The system retrieves the necessary database configuration for the specified ERP system and client:

- Database connection strings
- Schema information
- Table structures
- Connection pool settings
- Query timeout configurations

The system securely accesses the ERP database by:

- Retrieving secure credentials from the vault
- Establishing a secure database connection
- Verifying connection success
- Configuring connection parameters (transaction isolation, timeouts)

2.3 Execute Database Queries

With a successful connection established, the system executes the appropriate queries 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 database 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