

Publishers Clearing House Martech Data Ecosystem

Data Collection from Multiple Sources

The process begins with gathering raw data from a variety of platforms and systems.

These include SQL Server and DB2 (for orders and sweepstakes), Google Analytics and other web analytics tools.

Third-party demographic and enrichment data, such as Experian, are also integrated.

All ingestions are orchestrated by Airflow and containerized Python pipelines running on Kubernetes.

Identity Resolution (User Unification)

Once data is collected, the system resolves user identities across disparate sources.

It matches records based on unique identifiers such as customer ID, email address, device fingerprint, or hashed email.

Advanced matching logic with survivorship rules ensures that the most reliable attributes are retained, producing a single view of the customer.

Data Cleaning and Standardization

Data from different systems often varies in format and quality.

This stage focuses on normalization—cleaning inconsistencies, deduplicating records, validating formats, and imputing missing values.

Business rules are applied to enforce opt-ins, consent preferences, and channel restrictions.

Automated quality checks and audit trails ensure data integrity before loading into the core warehouse.

Establishing a Common Data Model (Core Warehouse)

All cleansed and standardized data is loaded into a centralized data warehouse—typically BigQuery or Snowflake.

The model follows a dimensional design with conformed dimensions such as Customer, Device, Channel, Product, and Time.

Data Mart Creation

Specialized data marts are built from the warehouse to support distinct business areas—Marketing, Web Analytics, Sales, etc.

Each mart aggregates relevant metrics and KPIs while providing privacy-controlled, role-based access for analysts.

SAS-Driven Segmentation

SAS runs segmentation logic over the data marts, applying RFM (Recency, Frequency, Monetary) scoring, behavioral clustering, and predictive modeling.

This process identifies high-value customer clusters, campaign audiences, and engagement segments, with real-time updates.

Personalization in SQL Server Applications

Segmented data is exported back into SQL Server to drive real-time personalization within customer-facing systems.

Each user receives a defined path and tailored ad or content experience based on their segment attributes and behavior.

Campaign Activation via ESP (Email Service Provider)

The same segmented datasets are pushed to an ESP for batch or triggered email campaigns.

Exports respect all consent and suppression rules and include frequency caps to prevent over-messaging.

Lead Distribution to Partners

Finally, enriched and segmented customer leads are packaged and securely distributed to external partners.

Each record includes contact, intent, and consent information.

SLAs and delivery monitoring ensure compliance, while feedback loops from partners feed into future segmentation.

