Runbooks
HANDOVER PACKAGE

A complete operational guide delivered at project close, empowering your team to run, govern, and grow your data stack.

PREPARED FOR	PREPARED BY	DELIVERY DATE	VERSION
Client ABC	**One Big Table Consulting**	**May 2026**	**1.0 — Final**

STACK: Snowflake · dbt · Fivetran · Power BI

TABLE OF CONTENTS

01 OPERATIONAL RUNBOOKS

This section provides step-by-step operating procedures for Apex Retail Group's data platform. These runbooks are designed to be used by your internal data or IT team for day-to-day operations, without requiring One Big Table's ongoing involvement.

1.1 Pipeline Maintenance

All ELT pipelines are orchestrated via Fivetran (data ingestion) and dbt Cloud (transformation and testing). The procedures below cover routine monitoring, manual triggering, and restart protocols.

Triggering a Pipeline Run Manually

1. **Log in to Fivetran** at app.fivetran.com using your admin credentials (see Credentials Register in SharePoint).
2. **Navigate to Connectors** and locate the relevant source system (e.g., Shopify, Salesforce, Google Ads).
3. **Click Sync Now** to trigger an immediate incremental sync. For a full historical re-sync, select Re-sync from the connector menu.
4. **For dbt transformations,** log in to dbt Cloud, navigate to Deploy > Jobs, and click Run Now on the appropriate job (e.g., "Daily Production Build").
5. **Monitor the run log in real time.** A green checkmark confirms success. A red X indicates a failure — refer to Section 1.2 for resolution steps.

Pipeline Schedule Reference

Pipeline	Tool	Trigger / Frequency	Primary Owner
Shopify Orders & Customers	Fivetran	Every 6 hours	Data Engineering
Salesforce CRM Accounts & Leads	Fivetran	Daily — 02:00 UTC	Data Engineering
Google Ads Campaign Performance	Fivetran	Daily — 03:00 UTC	Marketing Ops
Meta Ads Performance	Fivetran	Daily — 03:30 UTC	Marketing Ops
dbt Staging Models	dbt Cloud	Daily — 04:30 UTC	Data Engineering
dbt Mart Models (Revenue & Customers)	dbt Cloud	Daily — 05:15 UTC	Data Engineering
Power BI Dataset Refresh	Power BI Service	Daily — 06:30 UTC	BI Team

Monitoring Pipeline Health

- Fivetran sends email failure alerts to data-alerts@apexretail.com automatically.
- dbt Cloud sends Slack notifications to #data-ops on any job failure.

- Power BI Service sends dataset refresh failure emails to the BI Team distribution list.
- Snowflake query history is available under Activity > Query History in the Snowflake console for ad hoc diagnostics.

1.2 Incident Response

The table below documents the most common failure scenarios identified during the build and UAT phases of this engagement, along with their root causes and recommended resolution steps.

Symptom	Likely Root Cause	Recommended Resolution
Fivetran connector shows 401 Unauthorized error	API credentials or OAuth token expired in the source system	Re-authenticate the connector: Fivetran > Connectors > [Source] > Settings > Reconnect. Re-enter API credentials. Trigger a manual sync to confirm.
dbt job fails: "relation does not exist"	A source table was renamed or removed in Snowflake's RAW layer following a Fivetran schema change	Check the Fivetran schema tab for recent changes. Update the relevant source name in dbt's sources.yml. Run dbt build --select +affected_model to propagate the fix.
Power BI dashboard shows stale or yesterday's data	dbt job or Fivetran sync did not complete before the Power BI refresh window	Check dbt Cloud job status. Trigger a manual dbt run if needed. Then trigger a manual Power BI dataset refresh from the Power BI Service.
Snowflake virtual warehouse is suspended	Auto-suspend triggered after 60 seconds of inactivity (by design)	The warehouse resumes automatically on the next query. To resume manually: Snowflake Console > Admin > Warehouses > Resume. No data is lost.
dbt test fails: "not_null violation" on a key column	Source system sent a record with a null primary key or unexpected null in a required field	Review the failing test output in dbt Cloud logs. Identify the offending record in the RAW schema. Notify the source system owner. If isolated, document and re-run.
Fivetran sync shows "Schema Drift" warning	The source system (e.g., Shopify) added or removed a column	Review the drift in Fivetran > Connectors > Schema tab. Accept or block the new column. If accepted, re-run affected dbt models to propagate the schema change.

1.3 Deployment Procedures

All dbt model changes follow a Git-based CI/CD workflow. Code is version-controlled in GitHub and deployed to production automatically via dbt Cloud upon merge to the main branch.

Promoting a dbt Model Change to Production

1. **Create a new branch in GitHub** from main: git checkout -b feature/your-change-description
2. **Edit the relevant .sql or .yml file** in the /models directory. All models live under /models/staging, /models/intermediate, or /models/marts.
3. **Test locally in dbt Cloud's IDE** using: dbt build --select your_model+ to validate logic and run data tests.

4. **Open a Pull Request to main** and assign a peer reviewer from the data team. Include a description of the change and its downstream impact.
5. **Upon PR merge,** dbt Cloud's CI/CD job triggers automatically. If all tests pass, the change is live in production within minutes.
6. **Verify in Power BI** that the updated model is reflected correctly in the relevant report or dataset.

> **IMPORTANT:** Never modify dbt models directly in the dbt Cloud IDE in the production environment. All changes must route through the GitHub Pull Request process to ensure peer review, automated testing, and a full version history.

02 TECHNICAL DOCUMENTATION

This section documents the structure, field definitions, data lineage, and architecture of the Apex Retail Group data platform as delivered by One Big Table Consulting.

2.1 Data Dictionary

The following tables represent the core semantic layer models available in Snowflake and exposed to Power BI. All production models reside in the APEX_PROD.MARTS schema.

mart_orders — Order-level revenue fact table

Column Name	Data Type	Description	Example Value
order_id	VARCHAR	Unique identifier for each order (Primary Key, sourced from Shopify)	ORD-20240415-8821
customer_id	VARCHAR	Foreign key linking to dim_customers. Consistent across all order records for the same customer.	CUST-10042
order_date	DATE	Calendar date the order was placed in UTC. Used as the primary date dimension join key.	2024-04-15
order_status	VARCHAR	Current fulfillment status of the order. Possible values: pending, fulfilled, returned, cancelled, partially_refunded.	fulfilled
gross_revenue	FLOAT	Total order value in USD before any discounts, refunds, or adjustments are applied.	149.95
discount_amount	FLOAT	Total discount value applied at checkout (includes promo codes and automatic discounts).	15.00
net_revenue	FLOAT	Final revenue after subtracting discount_amount and any post-purchase refunds. Primary metric for finance reporting.	134.95
channel	VARCHAR	Marketing acquisition channel attributed to the order. Values: organic_search, paid_search, email, social, referral, direct.	email
is_first_order	BOOLEAN	TRUE if this is the customer's first-ever completed order. Derived in dbt using a row_number window function over customer_id ordered by order_date.	TRUE
country_code	VARCHAR	ISO 3166-1 alpha-2 country code of the shipping destination.	US

mart_customers — Customer-grain dimension table

Column Name	Data Type	Description	Example Value
customer_id	VARCHAR	Unique customer identifier (Primary Key). Sourced from Shopify customer_id, consistent across all order records.	CUST-10042
email_hash	VARCHAR	SHA-256 hash of the customer's email address. PII is masked at the dbt staging layer; the raw email is not stored in any mart table.	a3f2c1...
first_order_date	DATE	Date of the customer's first completed order. Used to calculate customer cohorts and tenure.	2023-11-01
lifetime_orders	INTEGER	Total count of completed orders placed by this customer to date (excludes cancelled orders).	7
lifetime_revenue	FLOAT	Sum of net_revenue across all completed orders for this customer. Primary LTV metric.	834.60
customer_segment	VARCHAR	RFM-based segment assigned in dbt using recency, frequency, and monetary scoring. Values: VIP, Active, At-Risk, Lapsed, New.	VIP
acquisition_channel	VARCHAR	The channel attribute of the customer's first completed order. Used for cohort-level channel analysis.	paid_search

2.2 Data Lineage Maps

The table below describes the end-to-end lineage for the core revenue and customer intelligence pipelines. A full interactive lineage graph is available in dbt Cloud under Explore > Lineage.

Layer	Schema (Snowflake)	Key Models / Objects	Transformation Logic
Source (Raw)	APEX_RAW	shopify.orders, shopify.customers, shopify.line_items, salesforce.accounts, google_ads.campaigns, meta_ads.insights	No transformation. Fivetran loads data as-is from source APIs. Schema is managed by Fivetran.
Staging	APEX_STAGING	stg_shopify__orders, stg_shopify__customers, stg_salesforce__accounts, stg_google_ads__campaigns	Light cleaning only: rename columns to snake_case, cast data types, apply source filters, hash PII fields (email_hash). No business logic.
Intermediate	APEX_INTERMEDIATE	int_orders_enriched, int_customers_attributed, int_marketing_spend_daily	Cross-source joins and enrichment. Orders joined to customers; marketing spend joined to order attribution. RFM scoring calculated

Layer	Schema (Snowflake)	Key Models / Objects	Transformation Logic
			here.
Marts (Semantic Layer)	APEX_MARTS	mart_orders, mart_customers, mart_marketing_performance, mart_product_revenue	Final business-ready models. Flattened grain, pre-aggregated where needed for Power BI performance. These are the only tables exposed to the BI layer.
BI Layer	Power BI Service	Datasets: Revenue Overview, Customer Intelligence, Marketing Performance	Power BI connects via DirectQuery to APEX_MARTS only. No transformation logic is applied in Power BI; all logic lives in dbt.

2.3 Architecture Overview

The Apex Retail Group data platform is a fully cloud-native, modern data architecture. All components are SaaS-managed and require no on-premise infrastructure.

Component	Tool / Service	Purpose	Managed By
Data Ingestion	Fivetran	Automated ELT from 4 source systems (Shopify, Salesforce, Google Ads, Meta Ads). Schema management and sync scheduling handled by Fivetran.	Fivetran (SaaS)
Cloud Data Warehouse	Snowflake	Central cloud data store. All raw, staging, intermediate, and mart data resides here. Virtual warehouse auto-suspends after 60s of inactivity to control cost.	Apex Retail (Admin)
Data Transformation	dbt Cloud	SQL-based transformation, testing, documentation, and lineage. All business logic is defined here in version-controlled SQL models.	dbt Labs (SaaS)
Version Control	GitHub	Source of truth for all dbt model code. PR-based workflow enforces peer review before any production deployment.	Apex Retail (Admin)
Business Intelligence	Power BI	Dashboards, reports, and ad hoc Explore capability for business users. Connects directly to Snowflake marts via DirectQuery.	Microsoft (SaaS)
Alerting & Monitoring	Fivetran + dbt Cloud + Slack	Automated failure notifications routed to Slack #data-ops channel and email distribution list. No manual monitoring required for standard operations.	Pre-configured at delivery

03 GOVERNANCE & SECURITY

This section outlines how access is controlled across the data platform and how personally identifiable information (PII) is handled in compliance with applicable data privacy regulations including GDPR and CCPA.

3.1 Role-Based Access Control (RBAC)

Access to all Snowflake data is governed through Snowflake's native role-based access control (RBAC) system. No user has direct access to raw source data. Access is limited to the marts schema by default, and all new users must be provisioned by a SYSADMIN.

Role	Access Level	Who Holds This Role	Can Provision Others?
SYSADMIN	Full platform administration. Can create and manage all roles, warehouses, databases, and users.	Head of Data / IT Lead (1 person maximum recommended)	Yes
DATA_ENGINEER	Read and write access to all schemas: RAW, STAGING, INTERMEDIATE, MARTS. Used for dbt service accounts and internal engineers.	Internal data engineers + dbt Cloud service account	No
ANALYST_READ	Read-only access to APEX_MARTS schema only. Cannot see raw or staging data.	Business analysts, Power BI users, finance team	No
MARKETING_READ	Read-only access to mart_marketing_performance and mart_orders only. Scoped for the marketing team.	Marketing team members	No
FIVETRAN_SERVICE	Write access to APEX_RAW schema only. Cannot read from any other schema.	Fivetran service account (system user)	No
DBT_SERVICE	Write access to APEX_STAGING, APEX_INTERMEDIATE, and APEX_MARTS. Read access to APEX_RAW.	dbt Cloud service account (system user)	No

Provisioning a New User (Step-by-Step)

1. **Log in to Snowflake** as SYSADMIN.
2. **Create the user account:** CREATE USER firstname_lastname PASSWORD='TempPass123!' MUST_CHANGE_PASSWORD = TRUE;
3. **Assign the appropriate role:** GRANT ROLE ANALYST_READ TO USER firstname_lastname;
4. **Send the user their temporary credentials** and instruct them to reset their password on first login.
5. **Log the new user in the Access Control Register** (located in SharePoint > Data Governance > Access Log).

NOTE: Service account passwords for Fivetran and dbt Cloud are stored in the Credentials Register (SharePoint > Data Governance > Credentials). Only SYSADMIN should have access to this document. Rotate passwords annually or immediately if a team member with access departs.

3.2 PII & Compliance Standards

The platform is configured to protect customer PII at the transformation layer using dbt. No PII is exposed in the APEX_MARTS schema. Raw PII in APEX_RAW is accessible only to the DATA_ENGINEER and SYSADMIN roles.

PII Field	Source System	Treatment in dbt Staging	Availability in Marts
Customer Email	Shopify	Hashed at stg_shopify__customers using SHA-256. Stored as email_hash. Original field is dropped.	email_hash only — no plaintext email in any mart table
Customer Full Name	Shopify / Salesforce	Dropped at staging layer. Field is excluded from all stg_ models.	Not available in any mart table
Phone Number	Salesforce	Dropped at staging layer.	Not available in any mart table
IP Address	Shopify	Dropped at staging layer.	Not available in any mart table
Billing / Shipping Address	Shopify	Full address dropped. City and country_code retained for geographic analysis only.	city and country_code available in mart_orders (ANALYST_READ)
Date of Birth	Salesforce	Dropped at staging layer.	Not available in any mart table

GDPR / CCPA — Data Deletion Process: When a customer submits a verified data deletion request: (1) Delete the customer record from Shopify and Salesforce. (2) Trigger a Fivetran full re-sync on both connectors. (3) Run dbt build to propagate the deletion through all staging, intermediate, and mart models. (4) Confirm the customer_id no longer appears in mart_customers. This process must be completed within 72 hours of the verified request.

04 COST MANAGEMENT

This section provides guidance on monitoring and controlling cloud infrastructure spend across the Apex Retail Group data platform. Understanding your cost drivers early prevents unexpected billing growth as data volumes scale.

4.1 Cloud Spend Monitoring

The platform's cost footprint spans four SaaS vendors. Each has a distinct billing model and a specific location in their console where spend can be monitored.

Vendor	Billing Model	Where to Monitor	Primary Cost Driver
Snowflake	Credits (compute) + TB/month (storage). Credits consumed when a virtual warehouse is actively running a query.	Snowflake Console > Admin > Cost Management > Consumption	TRANSFORMING_WH warehouse during dbt build jobs. Auto-suspend (60s) is configured to minimize idle cost.
Fivetran	Monthly Active Rows (MAR). Each row synced from a source system in a calendar month counts toward your MAR limit.	Fivetran Console > Account > Usage > Monthly Active Rows	Shopify orders and Salesforce accounts are the highest-volume connectors. Historical re-syncs spike MAR significantly — avoid unless necessary.
dbt Cloud	Per-seat licensing (Team plan). Additional charges for CI/CD run minutes above plan threshold.	dbt Cloud > Account Settings > Billing	Developer seat count. CI/CD runs triggered by every GitHub PR — frequency is predictable and low-cost.
Power BI	Per-user Pro licenses + Premium capacity (if applicable).	Microsoft 365 Admin Center > Billing > Licenses	User seat count. Each Power BI Pro license is required per business user who accesses published dashboards.

How to Identify Expensive Queries in Snowflake

1. **Navigate to:** Snowflake Console > Activity > Query History.
2. **Filter by warehouse:** select TRANSFORMING_WH to focus on dbt-driven compute.
3. **Sort by Execution Time** (descending) to surface the longest-running queries.
4. **Any query running longer than 60 seconds** is a candidate for optimization. Click into the query to view the full SQL and the user or service account that triggered it.
5. **If a scheduled dbt model is the culprit,** consider converting it to an incremental materialization to reduce rows processed on each run. Contact One Big Table during the 30-day support window for guidance.

Estimated Monthly Platform Cost Benchmarks (as of Delivery)

Vendor	Estimated Monthly Cost (USD)	Plan / Tier at Delivery	Notes
Snowflake	$220 – $290	On-demand Standard Edition	Estimated at current data volumes (~80GB storage, ~300 dbt model runs/month). Costs will scale with data volume growth.
Fivetran	$400	Standard Plan	4 connectors, estimated ~10M MAR/month. Historical re-syncs are not included in this estimate.
dbt Cloud	$100	Team Plan — 1 developer seat	Includes 1 developer seat and scheduled job runs. Add $50/seat/month for additional developers.
Power BI	$120	Pro — 12 user licenses	$10/user/month x 12 licensed business users at time of delivery.
TOTAL (est.)	$840 – $910/month		This is the baseline at current data volumes. Review quarterly as data and user counts grow.

COST ALERT RECOMMENDATION: Set up a Snowflake Resource Monitor to alert your team when monthly credit consumption exceeds 80% of the agreed budget. Navigate to: Admin > Resource Monitors > Create Monitor. Set the threshold in credits, add your SYSADMIN email as the notification recipient, and select "Notify" (not "Suspend") as the action to avoid interrupting scheduled jobs.

Questions after handover? Our team is available for a 30-day post-launch support window. Contact us at **hello@onebigtable.us** or visit **onebigtable.us**.