# Pinterest Data Warehouse

Entrepôt de données et big data

Auteurs : Massila LAKAF, Imane LAZIZI, Kanzy YOUSSEF

# Table des matières

# 1 Business Needs Analysis

## 1.1 Comprehensive Case Analysis

### 1.1.1 Overview and Market Position

Pinterest is a global visual discovery platform where users explore and save images and videos called *pins* grouped into thematic collections known as boards. Positioned between social media, search, and e-commerce, Pinterest focuses on inspiration rather than social interaction, helping users discover ideas and products aligned with their interests. With over 480 million monthly active users across 195+ countries, Pinterest turns early-stage user intent into measurable actions such as merchant-site visits and purchases.

### 1.1.2 Objectives and Economic Model

Pinterest aims to increase user engagement while maximizing advertising profitability. Delivering relevant recommendations and accurately measuring commercial impact requires extensive behavioral data to optimize content ranking, ad targeting, and merchant analytics.

### 1.1.3 Revenue Streams

**Advertising** is Pinterest's main source of income. The platform earns revenue when advertisers pay to display promoted content within users' feeds, search results, and recommendation surfaces.

Pinterest also engages in **commerce partnerships**, which support and enhance advertising performance.

### 1.1.4 Internal and External Data Sources

**Internal data** includes on-platform interactions (impressions, clicks, saves, searches, sessions) that fuel recommendation algorithms and trend detection.

**External data** is collected from merchant sites via the *Pinterest Tag* and the *Conversions API*, tracking off-platform events (product views, add-to-cart, checkouts, purchases). Standardized event codes enable attribution of conversions to campaigns and measurement of ad effectiveness.

### 1.1.5 Key Data for the Data Warehouse

The warehouse must unify engagement and conversion data around entities such as users, pins, campaigns, impressions, clicks, sessions, and conversions.

## 1.2 Actions / Operations to Track

**Impressions and Pin Views :** Display of pins on a user's feed or search results.
**Clicks on Pins and Merchant Site Visits.**
**Saves and Pin Creations.**
**Conversions and External Purchases :** Off-platform transactional events.
**Creator Engagement.**

## 1.3 Analytical Queries and Business Needs

*(All analytical queries for each action are provided in Appendix 1)*5.4

## 1.4 Ranking by Importance / Profitability Potential

1. Conversions and External Purchases 2. Clicks on Pins and Merchant Site Visits 3. Impressions and Pin Views 4. Saves and Pin Creations 5. Creator Engagement

## 1.5 Most Important Actions to Analyze

The two most important actions are **Conversions and External Purchases** and **Clicks on Pins and Merchant Site Visits**. The first is prioritized because it directly reflects revenue and advertising performance, while the second captures the transition from user engagement to commercial intent.

## 1.6 Dependency of the Data Warehouse Design

The warehouse design is therefore centered on these two key actions, as they drive Pinterest's monetization model and support the most critical decision-making needs.

# 2 Detailed Design of the First Model

## 2.1 Star Schema for the Most Important Action

The most important operation identified earlier is **Conversions and External Purchases**. The corresponding fact table contains the following attributes :

**Fact table (fact_conversions) :** conversion_id, date_id, time_id, user_id, pin_id, campaign_id, merchant_id, event_type_id, conversion_value, conversion_count, ad_spend.

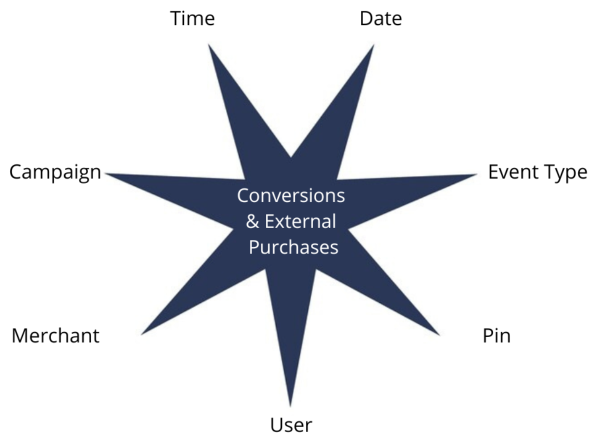


FIGURE 1 – Star Schema for the Conversions Data Mart

## 2.2 Dimension Tables

The dimensions are summarised below (full detailed tables appear in Appendix 2). 1.5

**dim_user :** user_id, signup_date, cohort_month, country, language, age_bucket, gender, signup_channel, device_preference, follower_count, account_type.

**dim_campaign :** campaign_id, advertiser_id, campaign_name, objective, start_date, end_date, budget, bid_strategy, targeting_summary, status, placement_type, total_spend.

**dim_pin :** pin_id, creator_user_id, created_at, pin_type, category_id, tags_list, is_promoted, media_format, external_url_domain, content_language, nb_saves, pin_title.

**dim_merchant :** merchant_id, domain, merchant_name, industry, country, store_currency, integration_method, lifetime_spend_est, avg_order_value_est, merchant_tier, contact_region.

**dim_date :** date_id, full_date, day, month, year, day_of_week, week_of_year, is_weekend, is_holiday, fiscal_period, quarter.

**dim_time :** time_id, hour, minute, second, time_bucket, period_of_day, minute_of_day.

**dim_event_type :** event_type_id, event_type_name, event_category, is_monetisable, conversion_credit, default_value, description.

## 2.3 Measures and Additivity

| Measure | Type | Additivity | Explanation |
|---|---|---|---|
| `conversion_value` | DECIMAL | Additive | Sum across all dimensions (time, user, pin, campaign, merchant). |
| `conversion_count` | INTEGER | Additive | Sum across all dimensions to compute total conversions. |

## 2.4 Ability to Answer the Main Business Need

Yes. The model supports a complete evaluation of advertising performance by linking on-platform behavior with off-platform results and enabling the analysis of conversion trends, monetary impact, and performance across all relevant dimensions.

## 2.5 Example Instance and Volume Estimation

**Sample rows :**

**fact_conversions (example)**

| conversion_id | date_id | time_id | user_id | pin_id | campaign_id | merchant_id | event_type_id | conversion_count | conversion_value |
|---|---|---|---|---|---|---|---|---|---|
| 1001 | 20241115 | 1430 | U123 | P456 | C789 | M555 | E186 | 1 | 89.99 |

**Estimated annual volume :** 115M conversions, with dimensions ranging from 1,440 rows (dim_time) to 10B rows (dim_pin). This volume justifies a data warehouse over Excel due to storage limits, query performance requirements, and concurrent user access needs.

## 2.6 Measure Validation

— `conversion_value` : Monetary value of each conversion event. Fully compatible with all dimensions. Used for : total revenue, AOV [1], revenue by campaign.

— `conversion_count` : Number of conversions. Fully compatible with all dimensions. Used for : conversion volume, trend analysis.

Both measures are additive across all dimensions.

**Metrics requiring external data :** ROAS [2] requires `dim_campaign.total_spend` (campaign-level). Conversion rate requires data coming from a separate data mart.

# 3 Design of Two Less Detailed Models

## 3.1 Snapshot Data Mart

A snapshot model is appropriate for analysing **Clicks on Pins and Merchant Site Visits**, because this action evolves continuously and can be measured periodically. The grain of the snapshot is : *one row per pin, placement, device, and day.*

1. AOV : Average Order Value = Total Revenue ÷ Number of Orders
2. ROAS : Return on Ad Spend

**Fact table (fact_clicks_snapshot) :** snapshot_date_id, pin_id, placement_id, device_id, campaign_id, clicks_count, impressions_count, unique_users. Datamart : 4

## 3.2 Updated-Records Data Mart

An updated-records model is appropriate for **External Purchases**, because each transaction evolves through a sequence of states (in cart, checkout, purchased, refunded) but only the latest state is required for analysis.

The grain is : *one row per transaction kept at its most recent state.*

**Fact table (fact_external_purchase_current) :** transaction_id, pin_id, platform_id, status_id, payment_id, last_update_date_id, total_purchase_value, shipping_cost, quantity, is_purchase_done. Datamart : 4

## 3.3 Dimension Tables

At least five dimensions are included for each model. Full detailed tables are provided in the Appendix.(see full tables in appendix 3.1and 3.2)

**Snapshot model dimensions :**

**dim_date :** (same dimension as in conversion's model)

**dim_pin :** (same dimension as in conversion's model)

**dim_campaign :** (same dimension as in conversion's model)

**dim_placement :** placement_id, placement_type, position_rank, algorithm_type, is_promoted, campaign_id, placement_name, traffic_source, feed_type, surface.

**dim_device :** device_id, device_type, os, browser, country, language, app_version, screen_resolution, manufacturer, connectivity_type.

**Updated-records model dimensions :**

**dim_pin :** (same as snapshot model).

**dim_date :** (same as snapshot model).

**dim_platform :** platform_id, platform_name, platform_country, platform_type, partnership_level, reputation_score, return_policy, commission_rate, currency, main_category.

**dim_purchase_status :** status_id, status_name, is_final_status, is_success_status, requires_payment, requires_shipping, cancellation_possible, refund_eligible, is_active_state, status_priority.

**dim_payment :** payment_id, payment_method, payment_risk_level, currency, conversion_rate, fraud_flag, issuer_country, issuer_type, authentication_level, payment_gateway.

## 3.4 Measures and Additivity

**Snapshot fact (fact_clicks_snapshot) :**

| Measure | Type | Additivity | Explanation |
|---|---|---|---|
| clicks_count | INTEGER | Additive | Sum across all dimensions. |
| impressions_count | INTEGER | Additive | Sum across all dimensions. |
| unique_users | INTEGER | Non-additive | Cannot be summed across time without duplication. |

**Updated-records fact (fact_external_purchase_current) :**

| Measure | Type | Additivity | Explanation |
|---|---|---|---|
| total_purchase_value | DECIMAL | Additive | Represents cumulative purchase value. |
| shipping_cost | DECIMAL | Additive | Can be aggregated across all dimensions. |
| quantity | INTEGER | Additive | Total item quantities purchased. |
| is_purchase_done | BOOLEAN | Non-additive | Indicates final state ; cannot be summed. |

## 3.5 Ability to Answer Secondary Business Needs

Both models support the intended analytical goals. The snapshot model enables periodic monitoring of pin visibility and engagement across devices, placements, and campaigns, while the updated-records model captures the latest state of each external transaction, allowing the analysis of purchase outcomes, platform performance, and payment behaviour.

## 3.6 Example Instances

**Snapshot fact example :**

| snapshot_date_id | pin_id | placement_id | device_id | campaign_id |
|---|---|---|---|---|
| 20241110 | P123 | PL45 | D01 | C789 |

| clicks_count | impressions_count | unique_users |
|---|---|---|
| 152 | 3400 | 120 |

**Updated-records fact example :**

| transaction_id | pin_id | platform_id | status_id | payment_id |
|---|---|---|---|---|
| T55621 | P789 | PL12 | ST03 | PAY44 |

| last_update_date_id | total_purchase_value | shipping_cost | quantity |
|---|---|---|---|
| 20241111 | 59.90 | 4.99 | 1 |

## 3.7 Volume Estimation and Excel Justification

he snapshot model creates hundreds of millions of rows per year from daily updates on pins and placements, while the updated-records model only tracks active transactions, adding tens of millions of rows yearly. These amounts are way beyond what Excel can handle.

# 4 Design – Advanced Modeling Techniques

## 4.1 4.1.Bridge Table

Pins can belong to multiple categories simultaneously. This many-to-many relationship is modeled using a bridge table `bridge_pin_category`.

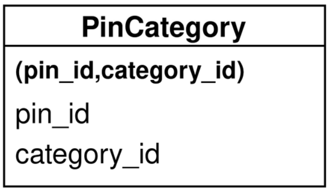


FIGURE 2 – Bridge Table for Pin and Category

To properly implement this N :M relationship, we introduce `dim_category` as a separate dimension. The `category_id` attribute is removed from `dim_pin` and becomes the primary key of the new category dimension.(see Appendix E.2 6.2).

**Analytical query example :**

```
SELECT c.category_name, SUM(f.conversion_value) as revenue
FROM fact_conversions f
JOIN bridge_pin_category b ON f.pin_id = b.pin_id
JOIN dim_category c ON b.category_id = c.category_id
WHERE f.date_id >= '2025-01-01'
GROUP BY c.category_name
ORDER BY revenue DESC;
```

This query analyzes revenue by category, accounting for pins that belong to multiple categories.

## 4.2 The most voluminous dimension

**dim_Pin** is the most voluminous dimension with an estimation of **5 to 10 billion** lines.

**attributes That are static :** pin_id, creator_user_id, created_at, pin_type, media_format, content_language.
**attributes That are dynamic :** pin_id, tags_list, is_promoted, external_url_domain, nb_saves, pin_title.

| Attribute | SUD [3] | Justification |
|---|---|---|
| `tags_list` | type 1 | A user can change their pin's tags list from '#fashion' to '#fashion #DIY', but that information is not analytically relevant, so only the newest values are kept. |
| `is_promoted` | type 3 | We only need to know whether the pin was previously promoted or not ; keeping one previous value is enough. |
| `external_url` | type 1 | Tracking URL changes is not useful for analytical purposes and would generate excessive data, so we overwrite the value. |
| `nb_saves` | type 1 | This is a measure that grows over time, only the current value of the number of saves is important for analysis. |
| `pin_title` | type 1 | Not analytically important, overwrite |

## 4.3 Hybrid Partitioning of `dim_pin`

`dim_pin` is the largest dimension. A hybrid partitioning strategy is applied by combining vertical (column-based) and horizontal (row-based) partitioning.

### Column Partitioning

Attributes are separated according to the way they may evolve (see appendix for the result 6.1) :

— **dim_pin_static** : stable attributes (creator, creation date, category, media format)

— **dim_pin_dynamic** : frequently changing attributes (is_promoted, nb_saves, tags_list)

This avoids scanning dynamic attributes when queries only require static metadata (historical or creative-level analysis).

### Row Partitioning

— **dim_pin_static** : RANGE partitioning by `created_at` (e.g., monthly), enabling partition pruning in time-based analytical queries.

— **dim_pin_dynamic** : HASH partitioning by `pin_id`, with a local index on `is_promoted` to support filtering on promotion status.

3. SUD : Strategies for Updating Dimensions

### Why This Reduces Data Reads

- Column partitioning ensures queries only read the subset of attributes they need (e.g., trend analysis reads only the static table).
- Range partitioning removes entire time partitions from the scan when filtering by creation date.
- Hash partitioning allows parallel execution and balanced partition sizes.
- Local indexing on `is_promoted` avoids full scans when filtering promoted pins.

### Example of Query Optimization

```sql
SELECT COUNT(*), AVG(nb_saves)
FROM dim_pin_dynamic d
JOIN dim_pin_static s ON d.pin_id = s.pin_id
WHERE s.created_at BETWEEN '2025-01-01' AND '2025-03-31'
  AND d.is_promoted = TRUE
  AND d.nb_saves > 1000;
```

**Without partitioning** : full scan of both 10B-row tables. **With hybrid partitioning** : only Q1-2025 partitions are read + indexed filtering on `is_promoted`. Approx. ~30M rows read instead of 10B.

# 5 The Implementation And Querying, Materialized Views, and Bitmap Indexes

## 5.1 SQL Implementation Of Fact Conversions

We implemented the star schema using Oracle SQL. Shared dimensions (`dim_date`, `dim_pin`, `dim_user`) are accessed via virtual views to enable reuse across multiple data marts.

Listing 1 – SQL Code Example

```sql
//Creation Of Virtual Views For Shared Dimensions
CREATE OR REPLACE VIEW DATEDIM
AS SELECT * FROM DIM_DATE ;

CREATE OR REPLACE VIEW PINDIM
AS SELECT * FROM DIM_PIN;

CREATE OR REPLACE VIEW USERDIM
AS SELECT * FROM DIM_USER;
```

### 5.2 Analytical Queries

The following queries address key business questions for revenue optimization :

**Q1 : Monthly revenue trend by campaign objective**

```sql
SELECT
    d.year, d.month, c.objective,
    SUM(f.conversion_value) as revenue,
    SUM(f.conversion_count) as conversions
FROM fact_conversion f
JOIN dim_date d ON f.date_id = d.date_id
JOIN dim_campaign c ON f.campaign_id = c.campaign_id
GROUP BY d.year, d.month, c.objective
ORDER BY d.year, d.month;
```

**Q2 : Top performing merchants by country**

```sql
SELECT
    m.country, m.merchant_name,
    SUM(f.conversion_value) as revenue,
    COUNT(DISTINCT f.user_id) as unique_customers
FROM fact_conversion f
JOIN dim_merchant m ON f.merchant_id = m.merchant_id
WHERE f.date_id >= '2025-01-01'
GROUP BY m.country, m.merchant_name
ORDER BY revenue DESC;
```

**Q3 : Weekend vs weekday conversion performance**

```sql
SELECT
    d.is_weekend,
    SUM(f.conversion_value) as revenue,
    SUM(f.conversion_count) as conversions,
    AVG(f.conversion_value) as avg_order_value
FROM fact_conversion f
JOIN dim_date d ON f.date_id = d.date_id
GROUP BY d.is_weekend;
```

**Q4 : User cohort analysis - revenue by signup month**

```sql
SELECT
    u.cohort_month, u.country,
    COUNT(DISTINCT f.user_id) as converting_users,
    SUM(f.conversion_value) as total_revenue
FROM fact_conversion f
JOIN dim_user u ON f.user_id = u.user_id
```

```sql
GROUP BY u.cohort_month, u.country
ORDER BY u.cohort_month;
```

**Q5 : Category performance with time-of-day analysis**

```sql
SELECT
    c.category_id, t.period_of_the_day,
    SUM(f.conversion_value) as revenue,
    COUNT(*) as conversion_events
FROM fact_conversion f
JOIN dim_category c ON f.category_id = c.category_id
JOIN dim_time t ON f.time_id = t.time_id
GROUP BY c.category_id, t.period_of_the_day;
```

## 5.3 Materialized Views Design

we propose the following materialized views :

**MV1 : Daily revenue by campaign and merchant**

```sql
CREATE MATERIALIZED VIEW mv_daily_campaign_merchant AS
SELECT
    f.date_id, f.campaign_id, f.merchant_id,
    SUM(f.conversion_value) as total_revenue,
    SUM(f.conversion_count) as total_conversions
FROM fact_conversion f
GROUP BY f.date_id, f.campaign_id, f.merchant_id;
```

*Supports queries : Q1, Q2*

**MV2 : User cohort performance**

```sql
CREATE MATERIALIZED VIEW mv_user_cohort AS
SELECT
    u.cohort_month, u.country, u.device_preference,
    COUNT(DISTINCT f.user_id) as user_count,
    SUM(f.conversion_value) as total_revenue
FROM fact_conversion f
JOIN dim_user u ON f.user_id = u.user_id
GROUP BY u.cohort_month, u.country, u.device_preference;
```

*Supports queries : Q4*

**MV3 : Pin category by time dimension**

```sql
CREATE MATERIALIZED VIEW mv_category_time AS
SELECT
    c.category_id, t.period_of_the_day, d.is_weekend,
```

```sql
    SUM(f.conversion_value) as total_revenue,
    COUNT(*) as conversion_count
FROM fact_conversion f
JOIN dim_category c ON f.category_id = c.category_id
JOIN dim_time t ON f.time_id = t.time_id
JOIN dim_date d ON f.date_id = d.date_id
GROUP BY c.category_id, t.period_of_the_day, d.is_weekend;
```

*Supports queries : Q3, Q5*

**Lattice coverage :**

- 3 materialized views cover 6 queries
- Each MV supports at least 2 queries (avoids one-to-one mapping)
- No single MV answers all queries (avoids over-aggregation)

## 5.4 Bitmap Join Indexes

We create bitmap join indexes to optimize queries filtering on dimensional attributes with low cardinality.

**Index 1 : Campaign objective on fact table**

```sql
CREATE BITMAP INDEX bji_campaign_objective
ON fact_conversion(event_type_id)
```

**Optimizes :** Q1 (filters/groups by campaign objective)

**Justification :** Event type has low cardinality ( 5-10 values : PAGE_VISIT, ADD_TO_CART, PURCHASE, etc.), making bitmap indexes highly efficient for filtering and grouping operations on event types in analytical queries.

**Index 2 : Weekend flag on fact table**

```sql
CREATE BITMAP INDEX bji_is_weekend
ON fact_conversion(is_weekend)
```

**Optimizes :** Q3 (filters by is_weekend boolean)

**Justification :** Boolean attribute (2 values) is ideal for bitmap indexing, enabling fast filtering between weekday and weekend conversions.

# Bibliographie

[1] Pinterest Help Center. *Install the Pinterest Tag*. Available sur : https://help.pinterest.com/en/business/article/install-the-pinterest-tag

[2] Pinterest Help Center. *Track Conversions with Pinterest Tag*. Available sur : https://help.pinterest.com/en/business/article/track-conversions-with-pinterest-tag

[3] Pinterest Help Center. *Add Event Codes*. Available sur : https://help.pinterest.com/en/business/article/add-event-codes

[4] Pinterest Business. *Pinterest for Business*. Available on : https://business.pinterest.com/en-gb/

[5] Anthropic. *Claude AI*, utilisé pour relecture ponctuelle (2025). Available on : https://claude.ai

[6] Kimball, R. & Ross, M. *The Data Warehouse Toolkit*. 3rd Edition, Wiley, 2013.

[7] Inmon, W. H. *Building the Data Warehouse*. 4th Edition, Wiley, 2005.

[8] Oracle Documentation. Available on : https://docs.oracle.com/en/database/oracle

[9] Pinterest statistics available on https://truelist.co/blog/pinterest-statistics/?utm_source=chatgpt.com

# Appendix

## 1 Analytical Queries for All Actions

### 1.1 Impressions and Pin Views

**Top categories by impression volume :** Identifies high-visibility themes and supports content prioritization.

**Volume of impressions per user segment :** Reveals which audience groups have the highest exposure, supporting audience targeting strategies.

**Correlation between impressions and pin clicks :** Measures visual effectiveness and supports optimization of ad creatives.

### 1.2 Clicks on Pins and Merchant Site Visits

**CTR by campaign :** Measures content effectiveness and informs budget allocation.

**User path from clicks to merchant sites :** Identifies friction points in the conversion journey and guides UX or campaign adjustments.

**Comparison between organic and sponsored click rates :** Evaluates the relative performance of paid vs. organic strategies to optimize advertising investment.

### 1.3 Saves and Pin Creations

**Most saved pins by topic :** Detects trends and supports recommendation tuning.

**Ranking of top users by number of repins generated :** Helps identify influential users to reinforce community or partnership programs.

**Temporal evolution of created pins :** Detects emerging interests and seasonal trends for strategic content planning.

### 1.4 Conversions and External Purchases

**Conversion rate by audience segment :** Evaluates profitability and refines targeting strategies.

**Average basket value (ABV) and customer lifetime value (CLV) per segment :** Measures long-term profitability of user segments.

**Attribution of conversions by interaction channel :** Identifies the most impactful touchpoints to optimize multi-channel marketing strategies.

### 1.5 Creator Engagement

**Creator interaction rate ranking :** Identifies high-performing creators for partnership decisions.

**Performance of brand–creator partnerships :** Measures collaboration effectiveness and supports future partnership planning.

**Identification of creators with fast audience growth :** Detects rising influencers to expand Pinterest's creator ecosystem.

## 2 Detailed Dimension Tables

### 2.1 B. Dimensions for the Conversions Data Mart

#### B.1 dim_user

| **Attribute** | **Description** |
|---|---|
| user_id | Unique user identifier |
| signup_date | Date the user created the account |
| cohort_month | Monthly cohort grouping |
| country | User country |
| language | Preferred interface language |
| age_bucket | Age segment |
| gender | Declared gender |
| signup_channel | Acquisition channel |
| device_preference | Most frequent device used |
| follower_count | Total number of followers |
| account_type | Personal or professional account |

### B.2 dim_campaign

| Attribute | Description |
|---|---|
| campaign_id | Unique campaign identifier |
| advertiser_id | Advertiser owning the campaign |
| campaign_name | Campaign label |
| objective | Traffic, conversions, awareness, etc. |
| start_date | Start date |
| end_date | End date |
| budget | Total campaign budget |
| bid_strategy | CPC, CPM, dynamic bidding |
| targeting_summary | Segment definitions |
| status | Active, paused, ended |
| placement_type | Where the ads appear |

### B.3 dim_pin

| Attribute | Description |
|---|---|
| pin_id | Unique pin identifier |
| creator_user_id | Author of the pin |
| created_at | Creation timestamp |
| pin_type | Image, video, idea pin, etc. |
| category_id | Content category |
| tags_list | Keywords associated with the pin |
| is_promoted | Paid or organic pin |
| media_format | Resolution/format |
| external_url_domain | Outbound destination domain |
| content_language | Language of the content |
| nb_saves | Number of saves |
| pin_title | Title displayed to users |

### B.4 dim_merchant

| Attribute | Description |
|---|---|
| merchant_id | Unique merchant identifier |
| domain | Website domain |
| merchant_name | Store name |
| industry | Vertical category |
| country | Merchant country |
| store_currency | Primary currency |
| integration_method | Feed, API, manual upload |
| lifetime_spend_est | Estimated spend on Pinterest |
| avg_order_value_est | Estimated AOV |
| merchant_tier | Bronze, silver, gold |
| contact_region | Region of business representative |

### B.5 dim_date

| Attribute | Description |
|---|---|
| date_id | Surrogate key (YYYYMMDD) |
| full_date | Calendar date |
| day | Day of month |
| month | Month number |
| year | Year |
| day_of_week | 1–7 |
| week_of_year | ISO week |
| is_weekend | Boolean |
| is_holiday | Boolean |
| fiscal_period | Fiscal mapping |
| quarter | Quarter in the year |

### B.6 dim_time

| Attribute | Description |
|---|---|
| time_id | Surrogate key (HHMMSS) |
| hour | Hour of the day |
| minute | Minute |
| second | Second |
| time_bucket | 5m / 15m / 30m buckets |
| period_of_day | Morning / afternoon / evening / night |
| minute_of_day | 0–1439 |

#### B.7 dim_event_type

| Attribute | Description |
|---|---|
| event_type_id | Identifier |
| event_type_name | Click, add-to-cart, purchase, etc. |
| event_category | Engagement, conversion, browsing |
| is_monetisable | Boolean |
| conversion_credit | Attribution weight |
| default_value | Standardised value |
| description | Detailed meaning |

# 3 C. Dimensions for the Snapshot and Updated-Records Models

## 3.1 C.1 Snapshot Data Mart Dimensions

#### C.1.1 dim_date

(same as B.5)

#### C.1.2 dim_pin

(same as B.3)

#### C.1.5 dim_campaign

(same as B.2)

#### C.1.3 dim_placement

| Attribute | Description |
|---|---|
| placement_id | Identifier |
| placement_type | Home feed, search, related pins |
| position_rank | Rank in feed |
| algorithm_type | Retrieval algorithm used |
| is_promoted | Paid placement flag |
| campaign_id | Associated ad campaign |
| placement_name | Internal name |
| traffic_source | Source category |
| feed_type | Organic or ads feed |
| surface | UI surface |

### C.1.4 dim_device

| Attribute | Description |
|---|---|
| device_id | Identifier |
| device_type | Mobile, desktop, tablet |
| os | Operating system |
| browser | Browser name |
| country | Country of usage |
| language | Interface language |
| app_version | App version |
| screen_resolution | Device resolution |
| manufacturer | Device brand |
| connectivity_type | WiFi / 4G / 5G |

## 3.2 C.2 Updated-Records Data Mart Dimensions

### C.2.1 dim_platform

| Attribute | Description |
|---|---|
| platform_id | Identifier |
| platform_name | E-commerce platform |
| platform_country | Operating country |
| platform_type | Marketplace, brand store |
| partnership_level | Standard, premium |
| reputation_score | Quality score |
| return_policy | Policy summary |
| commission_rate | Rate applied |
| currency | Default currency |
| main_category | Main product domain |

### C.2.2 dim_purchase_status

| Attribute | Description |
|---|---|
| status_id | Identifier |
| status_name | In cart, checkout, purchased... |
| is_final_status | Boolean |
| is_success_status | Indicates successful purchase |
| requires_payment | True if payment step required |
| requires_shipping | True if shipping needed |
| cancellation_possible | Boolean |
| refund_eligible | Boolean |
| is_active_state | State still in progress |
| status_priority | Ordering of statuses |

### C.2.3 dim_payment

| Attribute | Description |
|---|---|
| payment_id | Identifier |
| payment_method | Visa, PayPal, Klarna... |
| payment_risk_level | Risk score |
| currency | Currency used |
| conversion_rate | FX rate |
| fraud_flag | Fraud suspicion |
| issuer_country | Bank country |
| issuer_type | Bank or fintech |
| authentication_level | 3DS, SCA, none |
| payment_gateway | Processor name |

### C.2.4 dim_date

(same as B.5)

### C.2.5 dim_pin

(same as A.1.2)

# 4 Appendix D : Snapshot Data Mart Schema

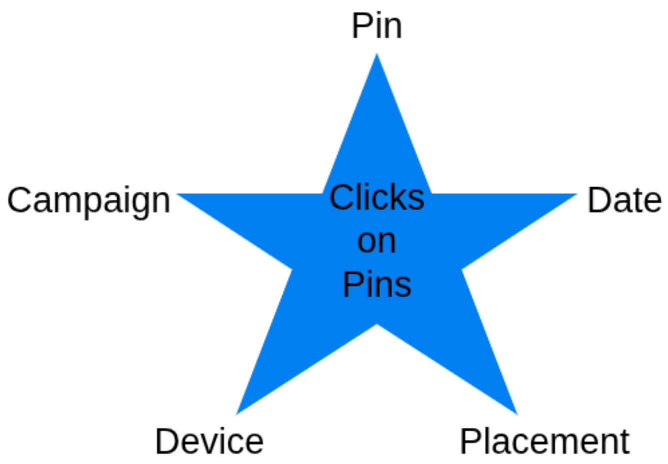


FIGURE 3 – Star Schema for the Snapshot Data Mart (Clicks on Pins)

# 5 Appendix E : Updated-Records Data Mart Schema

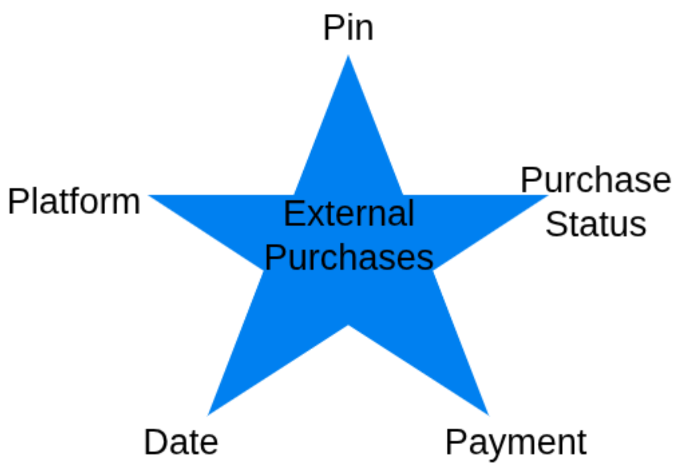


FIGURE 4 – Star Schema for the Updated-Records Data Mart (External Purchases)

3.1

# 6 Annex F : Design and Advanced Modeling

## 6.1 Annex F.1 : Hybrid Partitioning

**dim_pin_static**

| Attribute |
|---|
| pin_id |
| creator_user_id |
| created_at |
| pin_type |
| media_format |
| content_language |

**dim_pin_dynamic**

| Attribute |
|---|
| pin_id |
| tags_list |
| is_promoted |
| external_url_domain |
| nb_saves |
| pin_title |

## 6.2 F.2 dim_category

| Attribute | Description |
|---|---|
| `category_id` | Unique category identifier (PK) |
| `category_name` | Category name (e.g., "Food", "Fashion") |
| `parent_category_id` | Parent category for hierarchical organization |
| `category_level` | Depth level in category hierarchy |
| `description` | Detailed category description |
| `is_active` | Boolean indicating if category is currently active |