PSP Synthetic Use Case

1. Objective

This document describes how to simulate a **Payment Service Provider (PSP)** end-to-end data ecosystem using synthetic data. The goal is to help teams understand the payment lifecycle, domain boundaries, and how data flows through the **Medallion Architecture** (Bronze → Silver → Gold). The use case focuses on a fictional PSP similar to **KurvPay** that combines in-person and online payments.

---

2. What is a PSP?

A **Payment Service Provider (PSP)** enables merchants to accept digital payments securely through multiple channels (POS terminals, online checkouts, and mobile wallets). PSPs handle the entire lifecycle:

- **Merchant onboarding and verification (KYB)**
- **Payment processing (authorization and capture)**
- **Settlement and payouts**
- **Refunds and disputes**
- **Reporting, risk, and compliance**

---

3. Payment Lifecycle (End-to-End)

Step	Description	Typical Data Entities
1. Merchant Onboarding	Merchant submits business details, bank info, and passes KYB verification.	`merchants`
2. Payment Initiation	Customer makes a purchase via card, mobile wallet, or terminal.	`customers`, `payment_instruments`, `orders`
3. Authorization	PSP routes the payment to issuer; receives approve/decline response.	`transactions`

4. Capture & Settlement	Funds are captured and prepared for settlement in daily batches.	`transactions`, `payouts`
5. Payout	PSP pays merchant net funds (after fees and reserves).	`payouts`
6. Refunds & Disputes	Adjustments occur for returns or chargebacks.	`transactions`, `disputes`
7. Reporting & Analytics	Data consolidated for dashboards and merchant statements.	Silver & Gold layers

---

4. Schema

For this synthetic demo, we use **six core tables** that capture the entire PSP lifecycle. These can easily be extended, but remain minimal for clarity.

Merchant Domain

`merchants` – Represents businesses that accept payments via the PSP.

JSON

```json
{
  "merchant_id": "m_12345",
  "legal_name": "Acme Coffee LLC",
  "mcc": "5812",
  "country": "US",
  "kyb_status": "approved",
  "pricing_tier": "standard",
  "created_at": "2024-07-01T10:00:00Z"
}
```

Customer Domain

`customers` – End users making purchases.

JSON
```json
{
  "customer_id": "c_98765",
  "email_hash": "hash_abcd1234",
  "phone_hash": "hash_efgh5678",
  "created_at": "2024-06-25T15:30:00Z"
}
```

`payment_instruments` – Tokenized cards/wallets associated with customers.

JSON
```json
{
  "instrument_id": "pi_5678",
  "customer_id": "c_98765",
  "brand": "visa",
  "bin": "411111",
  "last4": "1234",
  "expiry_year": 2026,
  "wallet_type": "applepay",
  "status": "active",
  "first_seen_at": "2024-06-25T15:30:00Z"
}
```

Commerce Domain

`orders` – Represents each customer purchase.

JSON
```json
{
  "order_id": "ord_1111",
  "merchant_id": "m_12345",
  "customer_id": "c_98765",
  "currency": "USD",
  "total_amount_cents": 1520,
  "tax_cents": 120,
  "tip_cents": 200,
  "channel": "ecommerce",
  "created_at": "2024-07-01T11:15:00Z"
}
```

Payments Domain

`transactions` – Captures the state transitions of payments (authorization → capture → settlement → refund).

```json
{
  "txn_id": "txn_5555",
  "order_id": "ord_1111",
  "instrument_id": "pi_5678",
  "amount_cents": 1520,
  "currency": "USD",
  "state": "captured",
  "response_code": "00",
  "three_ds": "frictionless",
  "authorized_at": "2024-07-01T11:15:01Z",
  "captured_at": "2024-07-01T11:15:10Z",
  "settled_at": "2024-07-02T09:00:00Z",
  "fees_total_cents": 42
}
```

Settlement Domain

`payouts` – Represents daily settlements sent to merchants.

```json
{
  "payout_id": "pay_8888",
  "merchant_id": "m_12345",
  "batch_day": "2024-07-02",
  "currency": "USD",
  "gross_cents": 150000,
  "fees_cents": 4200,
  "reserve_cents": 300,
  "net_cents": 145500,
  "status": "paid",
  "paid_at": "2024-07-03T09:00:00Z"
}
```

Disputes Domain

`disputes` – Represents chargebacks or contested payments.

```json
{
  "dispute_id": "cb_9999",
  "txn_id": "txn_5555",
  "reason_code": "FRAUD",
  "amount_cents": 1520,
  "stage": "chargeback",
  "opened_at": "2024-07-05T12:00:00Z",
  "closed_at": null,
  "liability": "issuer"
}
```

---

5. Medallion Architecture View

Bronze Layer (Raw Sources)

Each domain ingests raw data separately from multiple sources.

- `merchant_bronze/merchants_raw_*.json`
- `customer_bronze/customers_raw_*.json`
- `commerce_bronze/orders_raw_*.json`
- `payments_bronze/transactions_raw_*.json`
- `settlement_bronze/payouts_raw_*.json`
- `disputes_bronze/disputes_raw_*.json`

Silver Layer (Conformed, Domain-Oriented)

The Silver layer consolidates and massages data from the raw Bronze sources into **domain-aligned datasets** rather than separating into dimensions and facts. Each Silver dataset represents a business domain with a unified schema, consistent naming, standardized timestamps, and normalized currencies.

- `merchant_silver` – consolidated merchant and pricing information
- `customer_silver` – unified customer and payment instruments
- `commerce_silver` – cleaned and normalized order data
- `payments_silver` – conformed transaction lifecycle with consistent states
- `settlement_silver` – merged settlement and payout records
- `disputes_silver` – cleaned chargeback and dispute information

This layer ensures consistency and readiness for analytical modeling while preserving business context by domain.

Gold Layer (Analytics & Dashboards) (Analytics & Dashboards)

Purpose-built datasets for:

- Merchant financial statements
- Daily revenue and fee analysis
- Authorization and capture success rates
- Take rate trends and settlement performance
- Fraud and chargeback monitoring

---

6. Summary

This PSP synthetic dataset lets you:

- Demonstrate a **complete payment lifecycle** using only 6 tables.
- Simulate **multi-source ingestion → conformed Silver layer consolidation**.
- Drive **realistic analytics and dashboards** from synthetic yet meaningful data.
- Use tools like **ShadowTraffic** to generate and orchestrate data streams for each domain.

In short, this is the perfect dataset for showing *how a modern PSP system works from data to insight*.