Data Engineer Coding Exercise: Affiliate Data Processing

Scenario

You're a Data Engineer at an affiliate marketing platform. Your team is building a resilient and scalable data pipeline to:

1. Ingest raw affiliate transaction data from multiple sources.
2. Normalize and clean the data.
3. Apply affiliate-specific commission logic.
4. Store data into a central repository.
5. Sync commissions with a third-party API
6. Maintain audit and idempotency guarantees.

Please limit your time on this exercise to 3 hours.

Functional Requirements

Build a data processing solution that does the following:

1. Ingest Transactions:

- Read from **multiple CSV files**, each representing a different affiliate network
    - See example files at the of this document: network_A.csv network_B.csv

2. Normalize Data:

- Standardize fields: `transaction_id`, `affiliate_id`, `network`, `amount`, `timestamp`.

- Parse and clean inconsistent formats (e.g., timestamps in various formats, amounts as strings).

3. Apply Commission Rules:

Use external config (`commission_config.json`) to calculate commission logic:

```
{
```

```
  "networkA": {"baseRate": 0.1},
  "networkB": {"baseRate": 0.12, "bonus": {"minAmount": 100,
"additionalRate": 0.02}},
  "default": {"baseRate": 0.05}
}
```

Each transaction amount should be multiplied by the `baseRate` defined for the network to determine the commission amount.

If a bonus rate is defined, the transaction amount should be multiplied by `baseRate + additionalRate`, only if the transaction amount is greater than or equal to the `minAmount.`

4. Store Data in a Persistent DB:

- Persist the data from the CSV files into a relational database. The transaction data should go into a table with the following definition (you can add extra columns to support your implementation if necessary):

```
CREATE TABLE affiliate_transactions (
    transaction_id INTEGER PRIMARY KEY,
    timestamp TIMESTAMP NOT NULL,
    affiliate_id INTEGER NOT NULL,
    network VARCHAR(255) NOT NULL,
    amount DECIMAL(10,2) NOT NULL,
    commission DECIMAL(10,2) NOT NULL
);
```

- Ensure **idempotency** — reprocessing the same file or transaction should not create duplicates.

- Maintain a `commission_log` table with metadata (processed_at, file_name, success/failure, error_message). Maintain a log in a new table ingestion_log that shows the result of each run:

```
CREATE TABLE ingestion_log (
    log_id INTEGER PRIMARY KEY,
    processed_at TIMESTAMP NOT NULL,
    file_name VARCHAR(255) NOT NULL,
```

```
    network VARCHAR(255) NOT NULL,
    num_rows_imported INTEGER NOT NULL,
    num_rows_errored INTEGER NOT NULL
);
```

5. Send to External API (Simulated):

For each processed transaction tagged with affiliate `aff_01`, simulate sending a POST request to an external API:

```
POST /sync-commission

{
  "transaction_id": "tx123",
  "amount": 100.00,
  "commission": 10.00
}
```

- Assume the API can:
    - Timeout
    - Fail with `500` or `429`
    - Return a `200` on success
- Retry failed API requests with exponential backoff (max 3 retries).

Technical Requirements

1. Platforms

The solution will be deployed into **AWS**. You can choose any technologies that would run on that platform. For this exercise you can either deploy to AWS or simulate locally. *If you do not have access to AWS services, you can implement using an alternative stack, but please be prepared to talk about what technologies you would use if deploying to AWS.*

For persistent storage you may choose any relational database.

There is a strong preference for using platform features that will minimise the operational work required.

2. Deployment

For this exercise there is no need to automate deployment with CI/CD or Infrastructure as Code.

3. Security

For this exercise you do not need to worry about security requirements such as limiting which users can upload files or read or write data; the focus is on the data integration.

4. Integration testing

Write integration tests for these scenarios:

1. Successful ingestion of a valid CSV file from network A
2. Partially successful ingestion of a CSV file from network A with 2 valid and 2 malformed rows
3. Complete failure of a malformed CSV file.

Deliverables

Supply a Zip file or public Github repository link containing:

- Documentation
  - Assumptions
  - Architecture diagram
  - Technologies chosen
  - Deployment instructions
  - Running instructions
- Source code
  - Code and tests, arranged in a logical hierarchy

Evaluation Criteria

- Clear documentation
- Sound technology choices
- Meets business and technical requirements
- Software design & modularity
- Fault-tolerance and idempotency
- Use of best practices (e.g., retries, logging, config-driven logic)
- Code quality and readability
- Approach to testing

Sample input files

network_A.csv

```
id,aff_id,value,created_at
tx001,aff_01,125.50,2025-05-01 09:15:23
tx002,aff_02,87.25,2025-05-01 09:32:41
tx003,aff_03,203.75,2025-05-01 09:45:12
tx004,aff_01,156.00,2025-05-01 10:12:33
tx005,aff_04,94.80,2025-05-01 10:28:56
tx006,aff_05,78.90,2025-05-01 10:41:19
tx007,aff_02,189.25,2025-05-01 11:03:44
tx008,aff_03,112.60,2025-05-01 11:19:07
tx009,aff_01,298.45,2025-05-01 11:34:52
tx010,aff_04,145.30,2025-05-01 11:47:18
tx011,aff_05,67.85,2025-05-01 12:05:36
tx012,aff_02,234.70,2025-05-01 12:21:49
tx013,aff_03,98.15,2025-05-01 12:38:12
tx014,aff_01,176.25,2025-05-01 12:54:37
tx015,aff_04,89.60,2025-05-01 13:11:25
tx016,aff_05,201.90,2025-05-01 13:27:48
tx017,aff_02,134.35,2025-05-01 13:44:13
tx018,aff_03,267.80,2025-05-01 14:00:56
tx019,aff_01,92.40,2025-05-01 14:17:19
tx020,aff_04,158.75,2025-05-01 14:33:42
tx021,aff_05,83.20,2025-05-01 14:50:08
tx022,aff_02,219.65,2025-05-01 15:06:31
tx023,aff_03,145.90,2025-05-01 15:22:54
tx024,aff_01,78.35,2025-05-01 15:39:17
tx025,aff_04,192.80,2025-05-01 15:55:43
tx026,aff_05,126.45,2025-05-01 16:12:06
tx027,aff_02,87.70,2025-05-01 16:28:29
tx028,aff_03,254.15,2025-05-01 16:44:52
tx029,aff_01,103.60,2025-05-01 17:01:18
tx030,aff_04,178.25,2025-05-01 17:17:41
tx031,aff_05,94.90,2025-05-01 17:34:04
tx032,aff_02,161.35,2025-05-01 17:50:27
tx033,aff_03,89.80,2025-05-01 18:06:53
```

```
tx034,aff_01,237.45,2025-05-01 18:23:16
tx035,aff_04,112.70,2025-05-01 18:39:39
tx036,aff_05,186.25,2025-05-01 18:56:02
tx037,aff_02,98.60,2025-05-01 19:12:28
tx038,aff_03,159.95,2025-05-01 19:28:51
tx039,aff_01,74.30,2025-05-01 19:45:14
tx040,aff_04,223.85,2025-05-01 20:01:37
tx041,aff_05,137.40,2025-05-01 20:18:03
tx042,aff_02,85.75,2025-05-01 20:34:26
tx043,aff_03,201.20,2025-05-01 20:50:49
tx044,aff_01,118.65,2025-05-01 21:07:12
tx045,aff_04,76.90,2025-05-01 21:23:38
tx046,aff_05,195.35,2025-05-01 21:40:01
tx047,aff_02,142.80,2025-05-01 21:56:24
tx048,aff_03,91.25,2025-05-01 22:12:47
tx049,aff_01,268.70,2025-05-01 22:29:13
tx050,aff_04,105.15,2025-05-01 22:45:36
tx051,aff_05,179.60,2025-05-01 23:01:59
tx052,aff_02,93.45,2025-05-01 23:18:22
tx053,aff_03,167.90,2025-05-01 23:34:48
tx054,aff_01,81.35,2025-05-01 23:51:11
tx055,aff_04,239.80,2025-05-02 00:07:34
tx056,aff_05,124.25,2025-05-02 00:23:57
tx057,aff_02,88.70,2025-05-02 00:40:23
tx058,aff_03,212.15,2025-05-02 00:56:46
tx059,aff_01,156.60,2025-05-02 01:13:09
tx060,aff_04,79.95,2025-05-02 01:29:32
tx061,aff_05,184.40,2025-05-02 01:45:58
tx062,aff_02,108.85,2025-05-02 02:02:21
tx063,aff_03,97.30,2025-05-02 02:18:44
tx064,aff_01,253.75,2025-05-02 02:35:07
tx065,aff_04,131.20,2025-05-02 02:51:33
tx066,aff_05,86.65,2025-05-02 03:07:56
tx067,aff_02,175.10,2025-05-02 03:24:19
tx068,aff_03,119.55,2025-05-02 03:40:42
tx069,aff_01,92.90,2025-05-02 03:57:08
tx070,aff_04,206.35,2025-05-02 04:13:31
tx071,aff_05,148.80,2025-05-02 04:29:54
```

```
tx072,aff_02,84.25,2025-05-02 04:46:17
tx073,aff_03,229.70,2025-05-02 05:02:43
tx074,aff_01,113.15,2025-05-02 05:19:06
tx075,aff_04,77.60,2025-05-02 05:35:29
tx076,aff_05,191.05,2025-05-02 05:51:52
tx077,aff_02,135.50,2025-05-02 06:08:18
tx078,aff_03,89.95,2025-05-02 06:24:41
tx079,aff_01,264.40,2025-05-02 06:41:04
tx080,aff_04,102.85,2025-05-02 06:57:27
tx081,aff_05,176.30,2025-05-02 07:13:53
tx082,aff_02,95.75,2025-05-02 07:30:16
tx083,aff_03,159.20,2025-05-02 07:46:39
tx084,aff_01,83.65,2025-05-02 08:03:02
tx085,aff_04,227.10,2025-05-02 08:19:28
tx086,aff_05,116.55,2025-05-02 08:35:51
tx087,aff_02,91.00,2025-05-02 08:52:14
tx088,aff_03,204.45,2025-05-02 09:08:37
tx089,aff_01,138.90,2025-05-02 09:25:03
tx090,aff_04,75.35,2025-05-02 09:41:26
tx091,aff_05,189.80,2025-05-02 09:57:49
tx092,aff_02,123.25,2025-05-02 10:14:12
tx093,aff_03,86.70,2025-05-02 10:30:38
tx094,aff_01,251.15,2025-05-02 10:47:01
tx095,aff_04,129.60,2025-05-02 11:03:24
tx096,aff_05,94.05,2025-05-02 11:19:47
tx097,aff_02,167.50,2025-05-02 11:36:13
tx098,aff_03,111.95,2025-05-02 11:52:36
tx099,aff_01,88.40,2025-05-02 12:08:59
tx100,aff_04,213.85,2025-05-02 12:25:22
```

network_B.csv

```
transaction,affiliate,amount,timestamp
tx201,aff_01,$142.75,05/01/2025 08:15 AM
tx202,aff_03,$198.30,05/01/2025 08:32 AM
tx203,aff_05,$89.45,05/01/2025 08:47 AM
tx204,aff_02,$267.80,05/01/2025 09:03 AM
tx205,aff_04,$134.20,05/01/2025 09:19 AM
tx206,aff_01,$76.95,05/01/2025 09:34 AM
```

```
tx207,aff_03,$223.60,05/01/2025 09:50 AM
tx208,aff_05,$156.85,05/01/2025 10:06 AM
tx209,aff_02,$91.40,05/01/2025 10:21 AM
tx210,aff_04,$285.15,05/01/2025 10:37 AM
tx211,aff_01,$118.75,05/01/2025 10:53 AM
tx212,aff_03,$87.20,05/01/2025 11:08 AM
tx213,aff_05,$204.95,05/01/2025 11:24 AM
tx214,aff_02,$145.60,05/01/2025 11:40 AM
tx215,aff_04,$78.35,05/01/2025 11:55 AM
tx216,aff_01,$192.50,05/01/2025 12:11 PM
tx217,aff_03,$126.85,05/01/2025 12:27 PM
tx218,aff_05,$83.70,05/01/2025 12:42 PM
tx219,aff_02,$259.25,05/01/2025 12:58 PM
tx220,aff_04,$167.90,05/01/2025 01:14 PM
tx221,aff_01,$94.45,05/01/2025 01:29 PM
tx222,aff_03,$231.80,05/01/2025 01:45 PM
tx223,aff_05,$138.15,05/01/2025 02:01 PM
tx224,aff_02,$85.60,05/01/2025 02:16 PM
tx225,aff_04,$276.35,05/01/2025 02:32 PM
tx226,aff_01,$112.90,05/01/2025 02:48 PM
tx227,aff_03,$89.25,05/01/2025 03:03 PM
tx228,aff_05,$195.70,05/01/2025 03:19 PM
tx229,aff_02,$124.85,05/01/2025 03:35 PM
tx230,aff_04,$82.40,05/01/2025 03:50 PM
tx231,aff_01,$248.95,05/01/2025 04:06 PM
tx232,aff_03,$156.30,05/01/2025 04:22 PM
tx233,aff_05,$91.75,05/01/2025 04:37 PM
tx234,aff_02,$213.40,05/01/2025 04:53 PM
tx235,aff_04,$135.85,05/01/2025 05:09 PM
tx236,aff_01,$88.20,05/01/2025 05:24 PM
tx237,aff_03,$264.65,05/01/2025 05:40 PM
tx238,aff_05,$147.10,05/01/2025 05:56 PM
tx239,aff_02,$79.55,05/01/2025 06:11 PM
tx240,aff_04,$202.90,05/01/2025 06:27 PM
tx241,aff_01,$119.35,05/01/2025 06:43 PM
tx242,aff_03,$86.80,05/01/2025 06:58 PM
tx243,aff_05,$239.25,05/01/2025 07:14 PM
tx244,aff_02,$158.70,05/01/2025 07:30 PM
```

tx245,aff_04,$93.15,05/01/2025 07:45 PM
tx246,aff_01,$271.60,05/01/2025 08:01 PM
tx247,aff_03,$126.05,05/01/2025 08:17 PM
tx248,aff_05,$84.50,05/01/2025 08:32 PM
tx249,aff_02,$197.95,05/01/2025 08:48 PM
tx250,aff_04,$141.40,05/01/2025 09:04 PM
tx251,aff_01,$78.85,05/01/2025 09:19 PM
tx252,aff_03,$216.30,05/01/2025 09:35 PM
tx253,aff_05,$153.75,05/01/2025 09:51 PM
tx254,aff_02,$97.20,05/01/2025 10:06 PM
tx255,aff_04,$284.65,05/01/2025 10:22 PM
tx256,aff_01,$131.10,05/01/2025 10:38 PM
tx257,aff_03,$88.55,05/01/2025 10:53 PM
tx258,aff_05,$245.90,05/01/2025 11:09 PM
tx259,aff_02,$166.35,05/01/2025 11:25 PM
tx260,aff_04,$81.80,05/01/2025 11:40 PM
tx261,aff_01,$208.25,05/02/2025 12:56 AM
tx262,aff_03,$142.70,05/02/2025 01:12 AM
tx263,aff_05,$95.15,05/02/2025 01:27 AM
tx264,aff_02,$262.60,05/02/2025 01:43 AM
tx265,aff_04,$137.05,05/02/2025 01:59 AM
tx266,aff_01,$89.50,05/02/2025 02:14 AM
tx267,aff_03,$224.95,05/02/2025 02:30 AM
tx268,aff_05,$159.40,05/02/2025 02:46 AM
tx269,aff_02,$73.85,05/02/2025 03:01 AM
tx270,aff_04,$191.30,05/02/2025 03:17 AM
tx271,aff_01,$116.75,05/02/2025 03:33 AM
tx272,aff_03,$84.20,05/02/2025 03:48 AM
tx273,aff_05,$251.65,05/02/2025 04:04 AM
tx274,aff_02,$148.10,05/02/2025 04:20 AM
tx275,aff_04,$92.55,05/02/2025 04:35 AM
tx276,aff_01,$279.90,05/02/2025 04:51 AM
tx277,aff_03,$123.35,05/02/2025 05:07 AM
tx278,aff_05,$86.80,05/02/2025 05:22 AM
tx279,aff_02,$214.25,05/02/2025 05:38 AM
tx280,aff_04,$155.70,05/02/2025 05:54 AM
tx281,aff_01,$98.15,05/02/2025 06:09 AM
tx282,aff_03,$237.60,05/02/2025 06:25 AM

tx283,aff_05,$132.05,05/02/2025 06:41 AM
tx284,aff_02,$87.50,05/02/2025 06:56 AM
tx285,aff_04,$204.95,05/02/2025 07:12 AM
tx286,aff_01,$169.40,05/02/2025 07:28 AM
tx287,aff_03,$75.85,05/02/2025 07:43 AM
tx288,aff_05,$243.30,05/02/2025 07:59 AM
tx289,aff_02,$128.75,05/02/2025 08:15 AM
tx290,aff_04,$91.20,05/02/2025 08:30 AM
tx291,aff_01,$268.65,05/02/2025 08:46 AM
tx292,aff_03,$145.10,05/02/2025 09:02 AM
tx293,aff_05,$83.55,05/02/2025 09:17 AM
tx294,aff_02,$220.90,05/02/2025 09:33 AM
tx295,aff_04,$157.35,05/02/2025 09:49 AM
tx296,aff_01,$94.80,05/02/2025 10:04 AM
tx297,aff_03,$231.25,05/02/2025 10:20 AM
tx298,aff_05,$138.70,05/02/2025 10:36 AM
tx299,aff_02,$76.15,05/02/2025 10:51 AM
tx300,aff_04,$195.60,05/02/2025 11:07 AM