(caravelo (

Introduction

Hi, I'm JoseLuis, co-founder and CPTO of Caravelo.

But first, congratulations, if you got this far, it means you're awesome, an A-player! As a former Engineer & Designer I know that most of the time, it is better to show rather than tell, so I have prepared an exercise for you to test which are the types of challenges that you might face in your day-to-day as well as demonstrate to us that you are a kick-ass **Analytics Engineer / Data Modeler** for our Platform team!

Remember, we don't want to make you invest a lot of hours or valuable free time in an assignment, so feel free to (i) share the time invested in the exercise and/or (ii) claim back your "assignment compensation". We usually do it in the form of a free meal (delivered via Glovo) or a couple of cinema tickets.

And of course, **there is no right or wrong answer to the challenges**, so just share your assumptions (if any).

Our product overview

Caravelo's platform helps airlines move towards the subscription economy and embrace a new way of doing airline retailing.

A subscription program can take on different forms: from clubs where subscribers pay a recurring fee to access benefits like discounts or other perks, to one of our more popular and recommended use-cases, flight subscriptions, or even the most provocative All-You-Can-Fly. I invite you to take a closer look at the possible models in this post or this other one.

Assignment overview

This assignment will be reviewed by both technical and business stakeholders. Please ensure your responses are understandable to stakeholders who possess a strong understanding of data and business processes, but may lack extensive technical knowledge.

🤖 Disclaimer: Feel free to utilize Large Language Models (LLMs) such as ChatGPT or Claude Code for assistance with this assignment. If you choose to do so, please provide the prompts you used and a brief explanation of your collaboration with the LLM.

Part 1: MRR Analytics Design and Implementation

Caravelo operates a B2B2C subscription platform serving multiple transportation providers including airlines and railways. Our enterprise customers (transportation providers) offer various subscription plans to their end-users, creating a complex multi-tenant revenue ecosystem that requires sophisticated analytics capabilities.

The platform supports 3 primary billing frequencies to accommodate diverse customer preferences and market requirements:

- **Monthly Plans:** Standard 30-day recurring billing cycles
- **Quarterly Plans:** 90-day billing cycles (every 3 months)
- **Annual Plans:** 365-day billing cycles (once yearly)

Each customer has complete autonomy in selecting which subscription frequencies to offer their end-users, resulting in a heterogeneous mix of billing patterns across the platform.

As a consequence, the platform operates across multiple dimensions of complexity:

- **Geographic Distribution:** Enterprise customers and their end-users span multiple time zones globally, creating challenges in temporal data alignment and revenue recognition timing.
- **Multi-Currency Environment:** End-users transact in their local currencies (USD, EUR, GBP, MXN, SAR, etc.), while enterprise customers require consolidated reporting in their preferred base currency for business analysis.
- **Dual Reporting Requirements:** The system must simultaneously support customer-specific reporting in their chosen currency and internal Caravelo's analytics standardized in EUR for consolidated performance analysis.

Design a comprehensive data modeling and processing solution to accurately calculate Monthly Recurring Revenue (MRR) across all subscription types, currencies, and geographic locations.

Please note that you might need to define MRR as a standardized metric tracking predictable, recurring revenue normalized to a monthly basis, regardless of actual billing frequency, enabling consistent performance analysis and forecasting across the entire platform.

Additional considerations and support to address the requirements.

- **Temporal Normalization**: Payment timestamps may not align with the intended billing month due to timezone differences and processing delays.
  An end-user in Tokyo (UTC+9) initiates a payment on August 1st at 02:00 local time. The system processes this payment as July 31st 17:00 UTC, potentially misallocating the revenue to July instead of August MRR.
- **Billing Frequency Normalization:** Quarterly and annual subscriptions generate revenue in large, infrequent chunks that must be smoothed into monthly metrics for consistent reporting. Some examples

    - Annual plan ($1,000 yearly) → $83.33 monthly MRR allocation
    - Quarterly plan ($300 every 3 months) → $100 monthly MRR allocation
    - Monthly plan ($50/month) → $50 monthly MRR (direct mapping)
- **Multi-Currency Conversion Framework:** Enterprise customers require revenue dashboards in their preferred reporting currency, necessitating near-real-time currency conversion for their operational decision-making. All revenue data must be normalized to EUR for internal performance analysis, strategic planning, and consolidated financial reporting.

The following examples illustrate the raw event data your solution must process:

Successful Subscription Renewal Event

JSON


```
{
  "eventDateTime": "31/Jul/2025:23:14:25 +0000",
  "airpassStatus": "CAN_BUY",
  "fundingId": "d78ceaff-b5db-46f9-b164-a56df199eaea",
  "addons": "",
  "productCollectionKey": "subscriptions",
  "ancillaries": "",
  "addCabinBag": false,
  "paidToDate": "2025-08-02",
  "quotaFromDate": "2025-07-01",
  "quotaUnits": 1,
  "externalPaymentIds": "d99f9bba-ddc4-4604-83de-7a695e18efd6",
  "currency": "MXN",
  "event": "SubscriptionRenewedEvent",
  "routeType": "DOM",
  "walletId": "f9e5dd3a-63ed-4f98-a131-adc3fee5e21f",
  "amount": "639.00",
  "baseProduct": "NETWORK-ALL-DOM-RT-2019_11",
  "passId": "bce94ab1-4665-4512-85d8-954d0cdce3b5",
  "quotaToDate": "2025-08-01",
  "creationDate": "2025-07-01",
  "invoiceDate": "2025-07-01",
  "externalCustomerIds": "3770067340",
  "subProduct": "default",
  "tripType": "RT",
  "hub": "NETWORK",
  "success": "true",
  "clientCode": "Y4",
  "installment": "639.00",
  "walletStatus": "ACTIVE",
  "paidFromDate": "2025-07-01",
  "invoiceId": "3774ea62-dc45-4ee4-a302-03ac08fd452e",
```

```
  "addCheckedBag": false,
  "subscriptionId": "d78ceaff-b5db-46f9-b164-a56df199eaea",
  "startDate": "2025-07-01",
  "eventId": "NAM_VFmfjoEEgng="
}
```


Failed Subscription Renewal Event:

JSON

```
{
  "eventDateTime": "01/Jul/2025:00:14:26 +0000",
  "airpassStatus": "CAN_VIEW",
  "fundingId": "95a8505e-8ff4-4b2a-ab85-1acee7e25fab",
  "addons": "F3-DOM-RT1-PRIORITYCABINBAG-MONTHLY-PLAN",
  "errorCode": "UNKNOWN",
  "productCollectionKey": "subscriptions",
  "ancillaries": "ANCILLARY_CONB",
  "addCabinBag": false,
  "paidToDate": "2025-08-02",
  "quotaFromDate": "2025-07-01",
  "quotaUnits": 1,
  "externalPaymentIds": "21bb3632-8a38-4154-814f-3c0ce2e8f951",
  "currency": "SAR",
  "event": "SubscriptionRenewalFailedEvent",
  "routeType": "RT1",
  "phase": "EVERGREEN",
  "walletId": "3255e7ad-f1bb-440b-b4a6-e76549d1f371",
  "amount": "327.00",
  "baseProduct": "F3-DOM-RT1-MONTHLY-PLAN",
  "passId": "2b2b1110-a8cb-4863-9b8a-27893f1297bd",
  "quotaToDate": "2025-08-01",
  "invoiceDate": "2025-07-01",
  "tripType": "RT",
  "hub": "F3",
  "clientCode": "F3",
  "walletStatus": "ACTIVE",
  "paidFromDate": "2025-07-01",
  "invoiceId": "3569f13d-fb15-4982-8d87-8364717a2ce6",
  "addCheckedBag": false,
  "subscriptionId": "95a8505e-8ff4-4b2a-ab85-1acee7e25fab",
  "eventId": "NAM_YHPSDoEEcYg="
}
```


The proposed solution should cover the following aspects

- Comprehensive Data Model Design (temporal, currency, lifecycle and geographic)

- MRR calculation framework (revenue recognition, frequency normalization, currency conversion and failed payment handling)
- Data Pipeline Architecture (processing and transformation)
- Reporting Data Structure (customer vs internal)

Part 2: Please choose one of the following two scenarios

Scenario A) Airline booking modeling

A critical requirement for our enterprise customers is the ability to integrate with major Passenger Service Systems (PSS) such as Amadeus Altea, Sabre, and other industry-standard platforms. These systems generate booking data with varying field structures and naming conventions, creating significant data integration challenges.

The core problem: Different platforms and airline customers use inconsistent field names and data structures for semantically equivalent information, making it difficult to create unified data processing pipelines and tables.

Design a comprehensive data model that can effectively map and normalize common booking elements across different platforms and airline customers.

Please feel free to make any assumptions, share the process followed and describe the solution chosen.

Find below, some sample booking data from multiple airline customers and some screenshots of booking receipts showing real-world data variations:

- Data_sample_1.pdf Data_sample_1.png

- Data_sample_2.pdf
- Data_sample_3.pdf
- Data_sample_4.pdf
- Data_sample_5.pdf

Important Note: All passenger personally identifiable information (PII) has been removed from the provided samples for privacy compliance. Please consider how passenger data would factor into your data modeling approach.

Present your solution using any of the following formats (or combination):

- Entity Relationship Diagrams (ERD) showing table relationships
- Data mapping tables with source-to-target field mappings
- Conceptual data models illustrating entity relationships
- Technical documentation explaining the normalization approach
- Visual diagrams showing data flow and transformation logic

Scenario B) Catalog version management and analytics integration

Caravelo operates a subscription program platform that enables airlines to offer customers recurring payment plans for flight packages. Customers pay a fixed monthly or annual fee for a predetermined number of flights and associated benefits within defined usage periods.

The platform utilizes a “Catalog service” to define plan types, pricing, and business rules. Only one catalog can be active at a time; updates require replacing it with a new version. On average, the Catalog is updated 9 times a year.

This Catalog is stored in JSON format in a NoSQL database. Here is a simplified example of the file: Catalog_sample.json

```JSON
{
  "type": "SUBS_VERSIONED_CATALOG",
  "effectiveDate": "2021-07-01T00:00:00.000Z",
  "items": [
    {
      "name": "FA-RT5-ALL-MONTHLY-LAUNCH-PLAN",
      "prettyName": "Fly10More Club Launch offer",
      "visible": true,
      "product": "NETWORK-ALL-DOM-RT-5",
      "phases": [
        {
          "type": "EVERGREEN",
          "billingPeriod": "MONTHLY",
          "duration": null,
          "fixedPrice": null,
          "recurringPrice": {
            "prices": [
              "ZAR 4000.00"
            ]
          }
        }
      ],
      "priceList": "DEFAULT",
      "quota": {
        "limit": 10,
        "installments": 1,
        "installmentsFrequency": "MONTHS"
      }
    },
    {
      "name": "FA-RT5-ALL-MONTHLY-NEXT-PLAN",
      "prettyName": "Fly10More Club",
      "visible": false,
      "product": "NETWORK-ALL-DOM-RT-5",
      "phases": [
        {
          "type": "EVERGREEN",
          "billingPeriod": "MONTHLY",
          "duration": null,
          "fixedPrice": null,
          "recurringPrice": {
```

```
            "prices": [
              "ZAR 4800.00"
            ]
          }
        }
      ],
      "priceList": "DEFAULT",
      "quota": {
        "limit": 10,
        "installments": 1,
        "installmentsFrequency": "MONTHS"
      }
    }
  ],
  "products": [
    {
      "name": "NETWORK-ALL-DOM-RT-5",
      "prettyName": "Fly10More",
      "category": "Flight subscriptions",
      "type": "BASE",
      "network": "DOMESTIC",
      "allowedAddons": [],
      "includedAddons": [],
      "rules": [],
      "tripType": "RT",
      "personalization": []
    }
  ],
  "currencies": [
    "ZAR"
  ]
}
```


Our Data Analytics Dashboards and Reports depend on event streaming to populate our Data Warehouse (DWH). However, catalog updates made directly to the NoSQL database do not generate events, creating a critical gap in our analytics data.

Design a comprehensive solution that enables catalog version tracking and analytics integration while maintaining our existing operational workflow.

Your solution must be built as an AWS-native architecture that leverages appropriate managed services to ensure scalability, reliability, and seamless integration with Caravelo's existing cloud infrastructure.

The system must efficiently handle complex nested JSON catalog structures with dynamic schema variations, supporting both schema evolution over time and backward compatibility to ensure existing integrations remain functional as the catalog format evolves.

From a performance perspective, the architecture must support near real-time detection of catalog changes.

Your solution should explicitly depict the high-level architecture, tooling and chosen strategy.

Deliverables

1. **Source Code Repository**: Provide a GitHub/GitLab repository with the implementation.
2. **README Documentation**: Instructions to set up, configure, and run the project.
3. **Brief Write-Up**: Explanation of assumptions, optimisation choices, and system design decisions.

Submission Guidelines

- Share your repository access with tech-assignment@caravelo.com and jlv@caravelo.com
- You can present your notes in the repo or in the form of a document.
- Include any assumptions or considerations you made while preparing your response.

- Focus on clarity, feasibility, and practical implementation strategies.

!! When delivering Google Docs/Sheets/Slides, please make sure "General access" is enabled for "Anyone with the link".

Assessment Criteria

We will evaluate your submission based on the following technical criteria:

- ✓ Representation of business processes through data.
- ✓ Soundness of data model and pipeline architecture.
- ✓ Quality, readability, and structure of code.
- ✓ Clarity and completeness of documentation.
- ✓ Clear communication of assumptions and design decisions.

This exercise is designed to give us insight into your skills and capabilities.

We look forward to your submission!