10Alytics

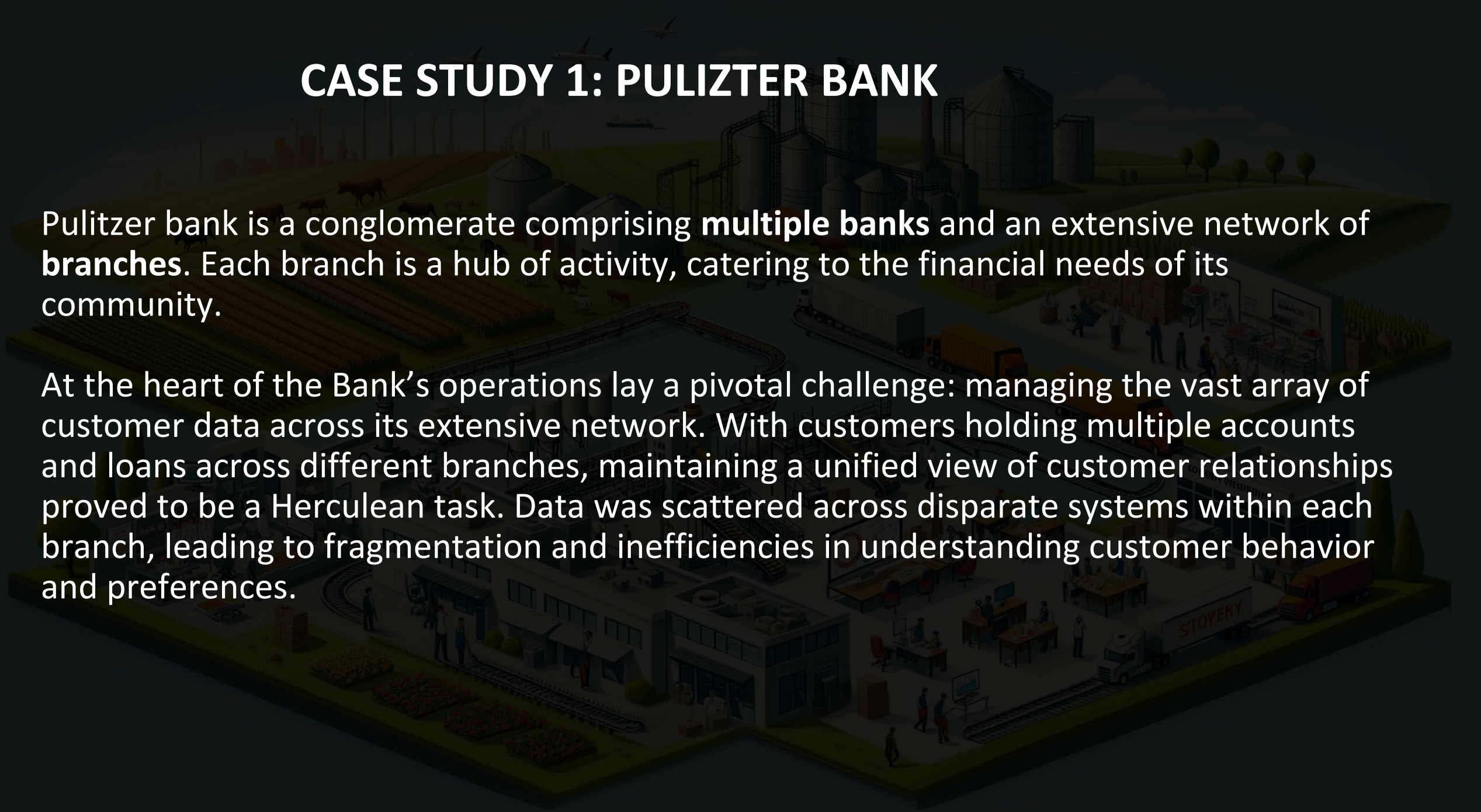
CASE STUDY: DATA MODELING

CHU NGWOKE

CASE STUDY 1: PULIZTER BANK

Pulitzer bank is a conglomerate comprising **multiple banks** and an extensive network of **branches**. Each branch is a hub of activity, catering to the financial needs of its community.

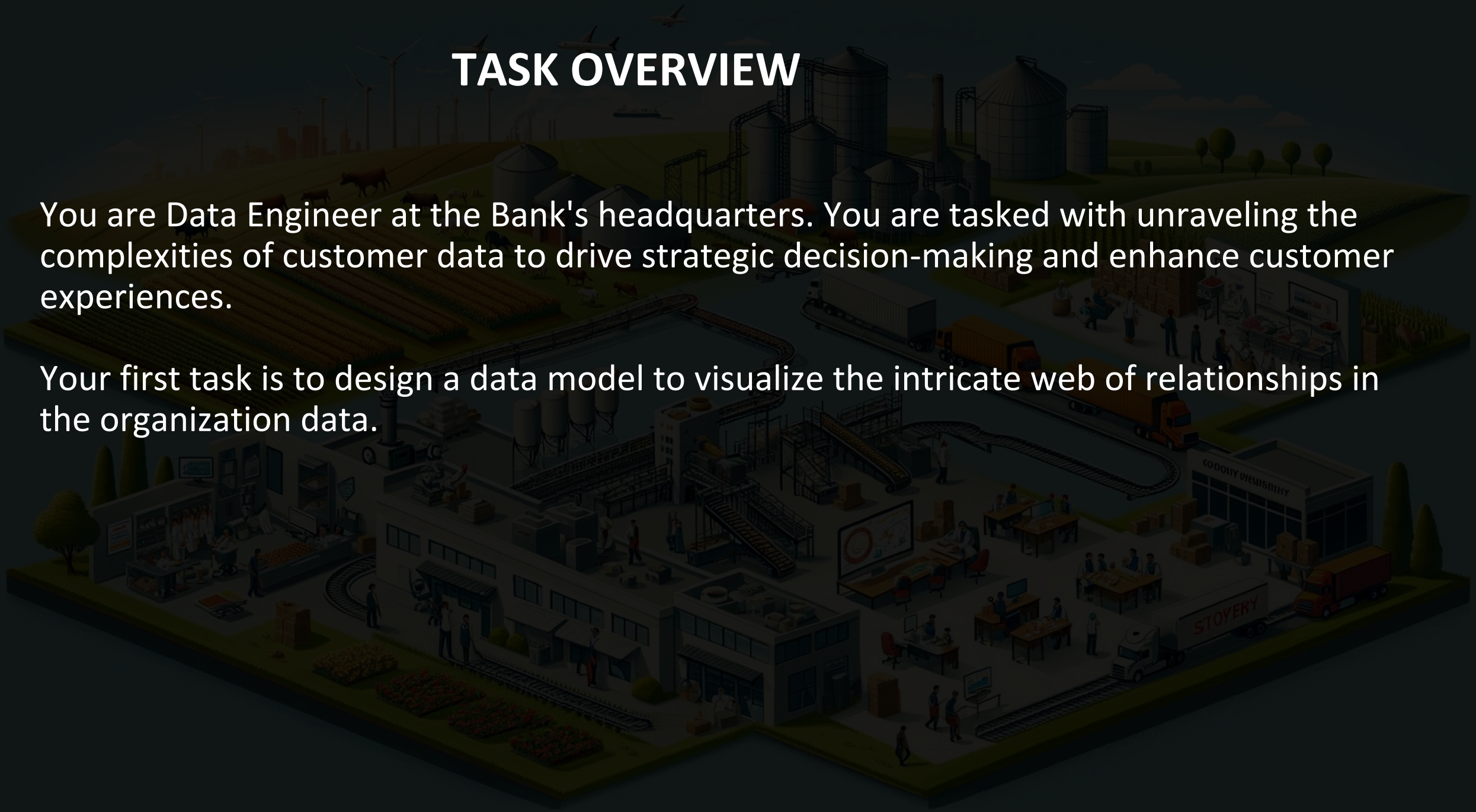
At the heart of the Bank's operations lay a pivotal challenge: managing the vast array of customer data across its extensive network. With customers holding multiple accounts and loans across different branches, maintaining a unified view of customer relationships proved to be a Herculean task. Data was scattered across disparate systems within each branch, leading to fragmentation and inefficiencies in understanding customer behavior and preferences.



TASK OVERVIEW

You are Data Engineer at the Bank's headquarters. You are tasked with unraveling the complexities of customer data to drive strategic decision-making and enhance customer experiences.

Your first task is to design a data model to visualize the intricate web of relationships in the organization data.



TASKS

1. Identify all Entities for Pulitzer Bank
2. Find and show all relationships between Entities
3. Identify the key attributes for every Entity
4. Identify the Keys for each attribute (Select a Primary Key)
5. Identify other relevant attributes for each Entity
6. Draw a complete E-R Diagram with all Entities, attributes, and relationships showing Primary and Foreign Keys
7. Review your Data Model with your Stakeholders and Business users

CASE STUDY 2: NATIONAL HOSPTIAL

A National Hospital consists of a number of specialized wards (such as Radiology ward, Oncology ward, etc.). Information about ward includes unique name, total number of current patients. Each ward hosts a number of patients, who are admitted by a consultant (Doctor) employed by the Hospital. On admission, the date and time are kept. The personal details of every patient includes name, medical recode number (MRN), set of phone number and address (city, street, code).

A separate register is held to store the information of tests undertaken. Each test has a unique episode No., category and the final result of the test. Number of tests maybe conducted for each patient.

Doctors are specialists in a specific ward and may be leading consultant for a number of patients. Each patient is assigned to one leading consultant, but maybe examined by other doctors, if required.

TASKS

You have been hired as a data engineer for this National Hospital. Your first task is to develop an ER diagram of a data model, to be used to build a digital database for this hospital. Your specific tasks are:

1. Identify the main entities
2. Identify the main relationships between the entities
3. Identify the key attributes of all the entities
4. Identify the candidate and primary key attribute of each entity
5. Determine the multiplicity constraint for each relationship.
6. State any assumption necessary to support your design.
7. Review your Data Model with your Stakeholders.

CASE STUDY 3: DWH MODELING

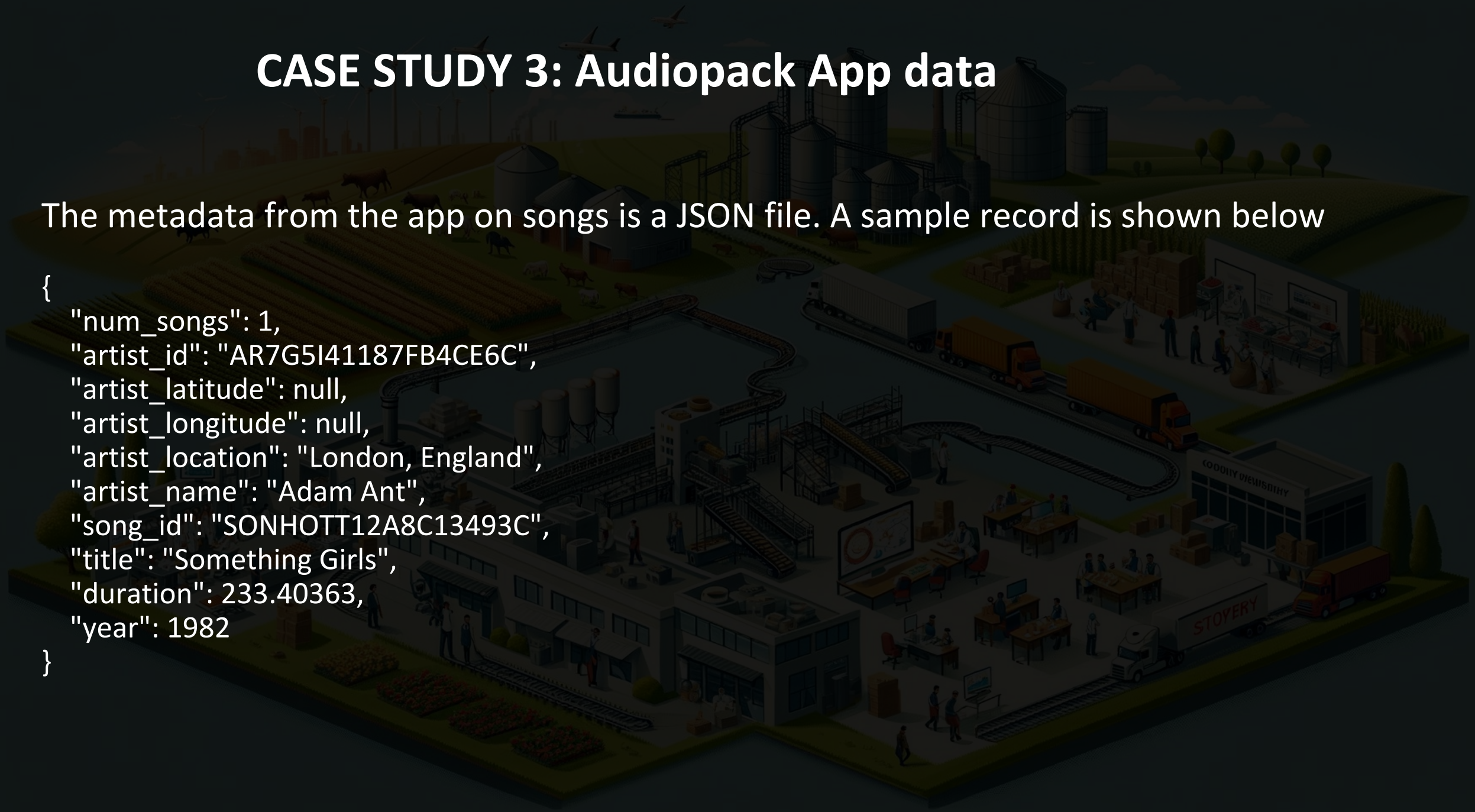
Audiopack is startup music streaming service. They want to analyze the data they've been collecting on songs and user activity on their new music streaming application. The analytics team is particularly interested in understanding what songs users are listening to. Currently, they don't have an easy way to query their data, which resides in a directory of JSON logs on user activity on the application, as well as a directory with JSON meta-data on the songs in their application.

You have been hired as a data engineer to create a Postgres DWH with tables designed to optimize queries on song play analysis. Your first task is to design a dimensional model.

CASE STUDY 3: Audiopack App data

The metadata from the app on songs is a JSON file. A sample record is shown below


```
{
    "num_songs": 1,
    "artist_id": "AR7G5I41187FB4CE6C",
    "artist_latitude": null,
    "artist_longitude": null,
    "artist_location": "London, England",
    "artist_name": "Adam Ant",
    "song_id": "SONHOTT12A8C13493C",
    "title": "Something Girls",
    "duration": 233.40363,
    "year": 1982
}
```

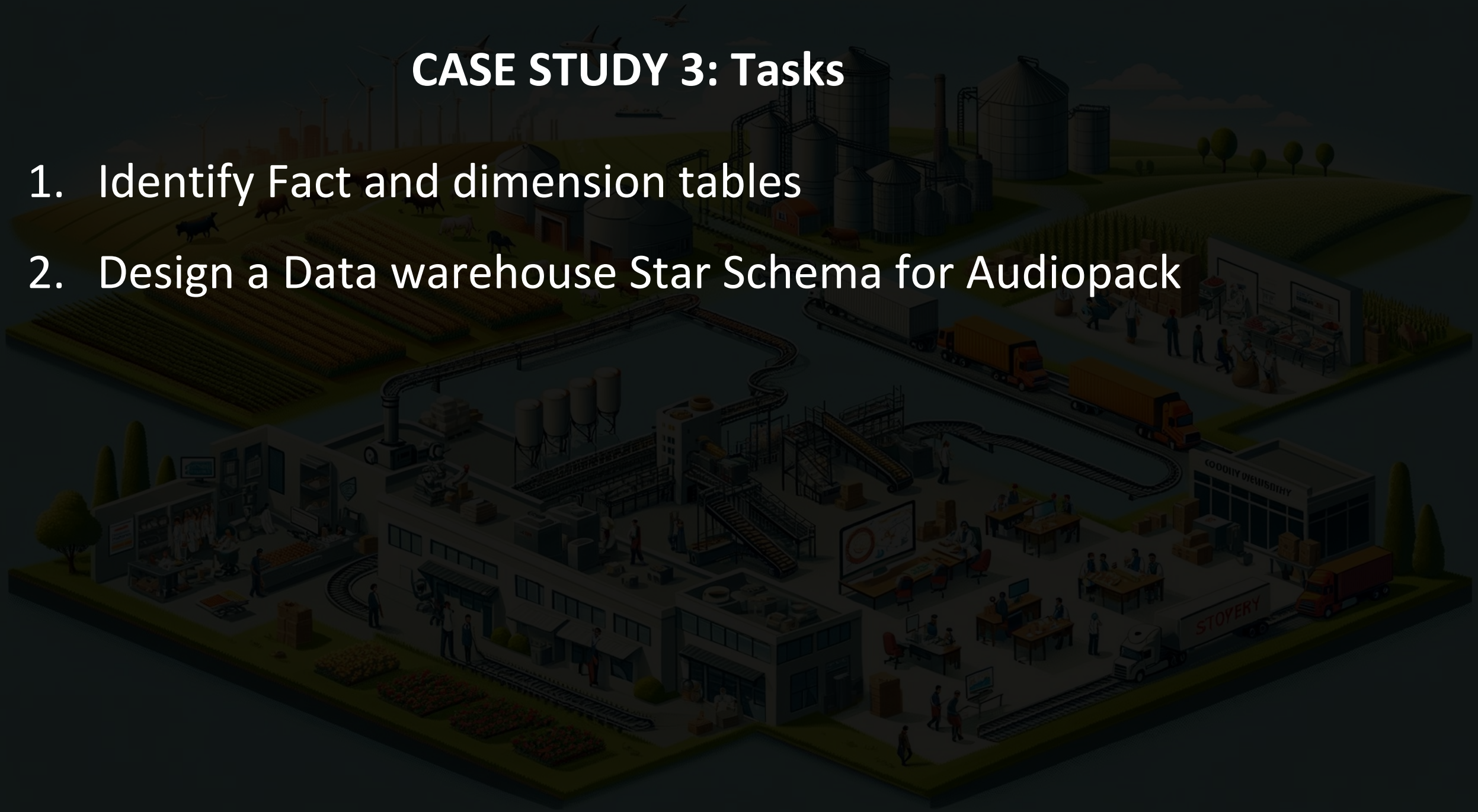
CASE STUDY 3: Audiopack App data

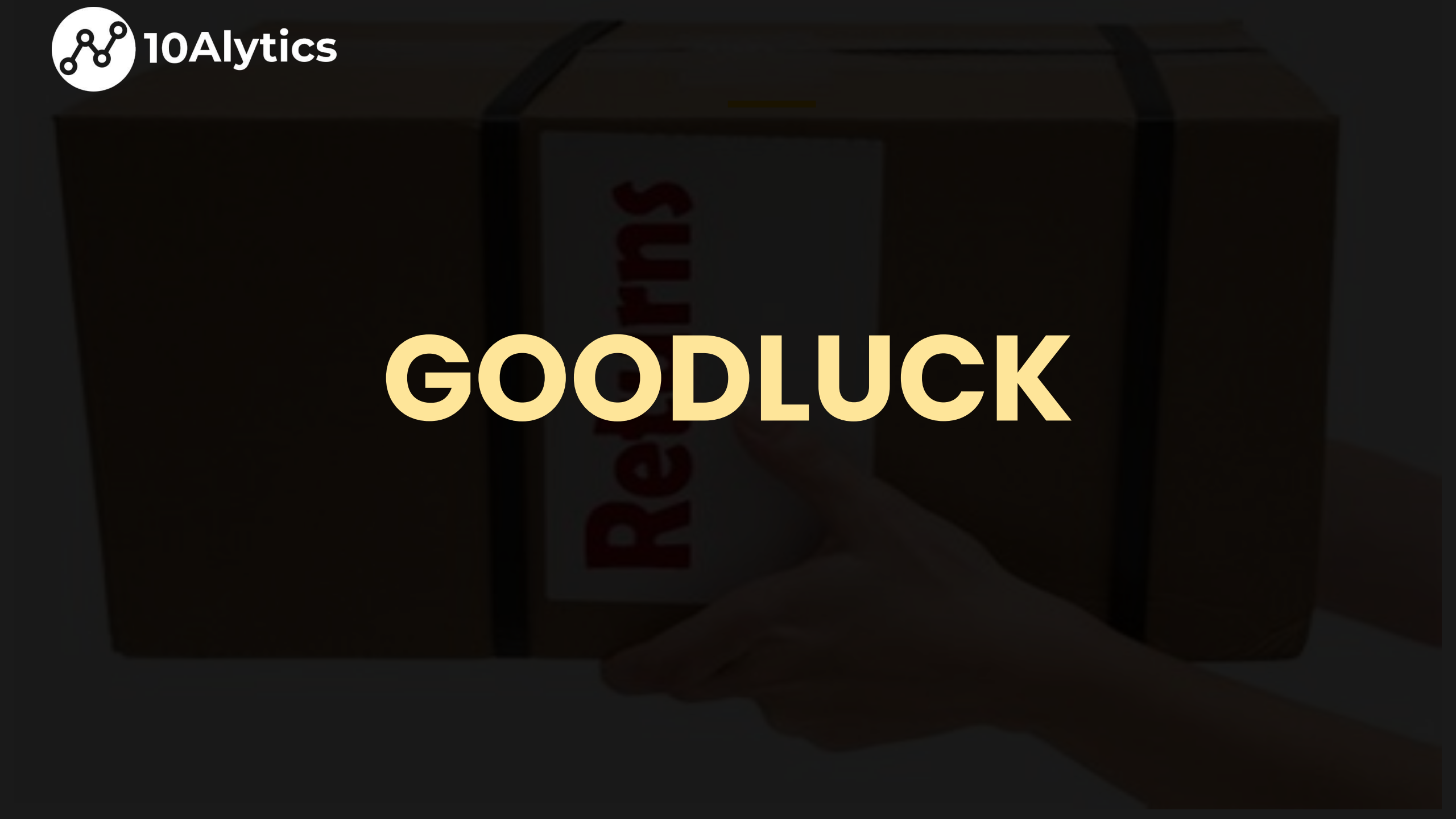
The logs of user activity on the app is also JSON file. A sample record is shown below


```
{
   "artist": null,
   "auth": "Logged In",
   "firstName": "Walter",
   "gender": "M",
   "itemInSession": 0,
   "lastName": "Frye",
   "length": null,
   "level": "free",
   "location": "San Francisco-Oakland-Hayward, CA",
   "method": "GET",
   "page": "Home",
   "registration": 1540919166796.0,
   "sessionId": 38,
   "song": null,
   "status": 200,
   "ts": 1541105830796,
   "userAgent": "\"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4) AppleWebKit/537.36",
   "userId": "39"
}
```

CASE STUDY 3: Tasks

1. Identify Fact and dimension tables
2. Design a Data warehouse Star Schema for Audiopack




10Alytics
GOODLUCK