

# Almedia's Data Engineer

## Hiring Case

Almedia is looking for a senior/principal data engineer to start building its data team.

This case aims toward providing a similar context as to what the company is living, while allowing the candidate to demonstrate his knowledge and skills.

You will be presented with a database, simulating real data, as well as context over the company's business model and operations.

It is your challenge to understand our system's raw data, design and implement a pipeline to create readily available data models and applications to serve business needs.

You are free to present your results in any format, but keep in mind that those will display not only your competencies and areas of expertise, but also your seniority.

Once completed, you may submit your solution to the HR member with whom you are in contact with. In normal circumstances, we should respond within a week.

Have fun, and good luck!

# The Company

Almedia's main product is [Freecash](#) (please do check it out).

Freecash allows users to earn real money by playing games, participating in surveys and promotions.

On signup, the user can explore an offerwall, pick and choose an offer of an app they might like, start playing & earning.

Each offer has customized tasks, which when achieved, reward the user with a certain amount of coins, which in turn, can be used to cash out \$\$ (1000c → 1\$).

Companies advertise with us, because we are able to provide quality customers at a competitive rate compared to traditional user acquisition channels.

In short, our business model is:

1. We spend money to acquire new users.
2. Some of these users engage in our partner's offers. We get paid for those users.
3. Users may achieve certain goals, in which case they get rewarded by us.

While some goals are easily achievable (such as installing an app), others might require the user to dedicate a considerable amount of time, or to make in-app purchases. Both of which are metrics our partners try to optimize for when considering which acquisition channel to invest in.

# Departments & Processes

The company is on a hyper growth stage, with most of its 100 employees having joined in the previous months.

Departments, teams, and processes are constantly being revised as we try to improve and optimize our operation.

While there are no specific constraints as to what data can work on, the core teams and processes that drive profitability are:

1. User Acquisition
2. Offer Management
3. Anti Fraud

Those teams have the need for some data dependencies (KPIs, analysis, monitoring).

Tech (development) and Data (all aspects) are separated verticals.

# The Data

In this case exercise we provide a small subset of our production data to the candidate.

There are three key entities in the shared dbase;

1. offers

Represents a promotion of an advertiser's app.

An offer is composed of distinct offer tasks, which an user can complete to receive rewards.

2. offer\_tasks

These are the different tasks available within an offer.

3. user\_offers

The completed offer tasks by an user.

Sensitive data has been anonymized; but it still retains pertinent information to the tasks at hand.

# Objective

We aim to **generate value** through data solutions.

We would like you to:

1. Understand our business model, and the raw data provided.
2. Define an architecture for our core data models.
3. Make a pipeline to build those core models.

When building your solution, it's important to keep in mind that:

- A. The data produced must serve a purpose.
- B. Our product is ever changing, and so is its data.
- C. We are expected to build, implement and maintain these solutions.

Bonus task(s):

## **1. Make data products out of the core models produced.**

Data produced in our data warehouse must serve a purpose.  
That is more often than not of an operational/analytical nature.

Can you generate a business solution on top of that data?

## **2. Redesign the data architecture of our source data (system).**

Right now, our system has a data architecture which did not consider that we might use the data for analytical and/or ML capabilities.

How can we improve its data architecture, while preserving the system functionality, and even improving its scalability and performance?

Consider for example that a user may start an offer, and subsequently an offer task, and we might change the nature of that offer and/or task. We might also want to personalize task displays to each user.

What should be the user experience? How would we design our data architecture to provide that, while maintaining reliable records of the user experience?

While we do not impose time constraints, a good solution usually takes the better part of a weekend. Spend as little or as much time as you believe is necessary.

Those tasks can be resolved in a minimalistic manner, or properly built taking into consideration pre-requisites and implications. The complexity and elegance of your solution will determine not only if you advance or not on the process, but our understanding of your level of seniority.

We currently use BigQuery. Feel free to use any stack for your solution.  
Feel free to refer to our job description to better understand our needs for this role.

Have a short summary or presentation of your solution.

If your solution satisfies our criterias, you should be invited to a technical/business presentation.

Have fun & good luck.