INFINITYWORKS
ACADEMY

Generation

Final Project

Background

Good news, the work you undertook for your client was a huge success! The café has seen unprecedented growth and has expanded to hundreds of outlets across the country.

Due to the demand the company is receiving, they need to figure out how they can best target new and returning customers, and also understand which products are selling well.

They are experiencing issues with collating and analysing the data they are producing at each branch, as their technical setup is limited.

You have been asked to step in and provide consultation on what they need to do in order to grow their technical offerings, so that they can continue to accelerate their growth.

Current Setup

Every day, the following occurs for each branch:

- A CSV file containing data about every transaction they made for that day is generated.
- At 8pm, the data is uploaded to a piece of software installed in the back office computers.
- Daily, weekly or monthly reports for sales figures and other related business metrics are created.

The Problem

- The software currently being used only generates reports for single branches.
- It is time consuming to collate data on all branches.
- Gathering meaningful data for the company on the whole is difficult, due to the limitations of the software.

The Bigger Problem

The company currently has no way of identifying trends, meaning they are potentially losing out on major revenue streams.

They are in desperate need of help putting together a platform that will allow them to easily understand all of the data they are producing.

Due to the highly professional work you completed for them in the past, they are keen to work alongside you in creating a solution to solve the problem they're facing.

The Solution

After a thorough analysis and discovery of what the client is looking for, a plan has been fleshed out. You have been tasked with building a fully scalable ETL (Extract, Transform, Load) pipeline to handle large volumes of transaction data for the business. This pipeline will collect all the transaction data generated by each individual café and place it in a single location. By being able to easily query the company's data as a whole, the client will drastically increase their ability to identify company-wide trends and insights.

You and your team will be in charge of figuring out how to obtain, process, store and analyse this data.

The Vision

Below is a vision of what both parties would like to produce as the end result.

- Each night a CSV for each branch will be uploaded to the cloud
- The system we have developed will read each file and perform ETL steps
- Data will be stored in a data warehouse
- Data analytics software will be used to create Business Intelligence analytics for the client
- Application monitoring software will be used to produce operational metrics, such as system errors, up-time and more

The Agile Approach

As we have recently learned, agile is a fantastic way to approach a project of this size.

So, in true agile nature, we will be building upon the project slowly, and adding more complexity to it as time progresses.

We have communicated our ways of working with the client, so they already know how we will be operating as a team day-to-day.

Ways of Working

- The project will consist of five sprints, where each sprint is a week in length (weeks 8-12).
- Each week you'll attempt to complete a series of tasks from a backlog provided to you by the product owner.
- Sprint planning will take place every Monday morning, to fully understand what is required for that week.
- You and your team will have a retrospective at the end of each week, to get together and discuss things that went well, what could be improved, and any actions you can take.
- You will be expected to organise yourselves and assign tasks to work on individually or in pairs/groups.

Data Engineering Skills

You will need to apply a range of software and data engineering skills in order to successfully build the project. You will be trained up on cutting edge technology, which you can then use to apply to the project. Here are some examples:

- ETL
- Data Warehousing
- Data Analytics / Visualisations / Business Intelligence
- Monitoring
- Development Operations (DevOps)

Technologies

You will be using a wide range of technologies to support the development of your ETL application.

Some of them you will be familiar with, others will require you to research how best to use them.

- Python for developing your application
- GitHub for source control
- Trello for project management
- AWS Services such as S3, Lambda, Redshift, Quicksight and Cloudwatch
- Grafana for application monitoring

Role of Instructor

Your instructor will simulate the role of both product owner and technical lead.

Product Owner: Representing the product's stakeholders and the voice of the customer, and is responsible for delivering good business results.

Technical Lead: A software engineer responsible for leading a team and alignment of the technical direction.

That's all for now, folks

More details will be released at the start of sprint 1. For now, sit back and relax (but not too much!).

When we start, have fun, and most importantly, enjoy yourself.