**Premise**

* Put together an ETL pipeline for some mock data from a fictionalised version of Sparta Global
* Candidates for the academy will have their contact details collected and will be invited to a particular Sparta Day
* During the Sparta Day each candidate’s performance is recorded (psychometric and presentation), each candidate is interviewed, and notes are made (strengths, weaknesses, programming languages), final decision (pass or fail) and course interest are noted
* If they pass the interview candidates are invited to join the Sparta academy where each week a score out of 8 is given for the 6 behaviours
* Candidates with low scores may be removed from the academy

**Tasks**

* Design an appropriate data store to house all the relevant data that can be drawn from the files on S3
* The database should provide a single person view – we should be able to track a person’s journey from the beginning to the end
* Write production level code to Extract, Transform and Load data into the database
* Should be possible to add new data to the S3 bucket and have it picked up and incorporated in the pipeline
* Produce a dashboard on the data to help with analysis of the key business questions

**Key Business Questions**

* What useful insights can be drawn?
* Which members of talent team are performing best?
* How many trainees are removed at each stage of the course?
* What factors influence removal?
* What factors at the interview stage predict high performance on the course?
* How are courses on a given date performing?

**Academy Files**

* One CSV file per course that contains the name of the trainer and behavioural competency scores for each trainee

**Talent Files**

* Three different file types
* One CSV file per month for 2019 with information on candidates invited to a Sparta assessment day
* One JSON file per candidate with notes on their answers to key interview questions during the Sparta Day and the final pass or fail decision
* One TXT file per Sparta Day with the date and location, a list of candidates and their psychometric and presentation scores