Tell me a bit about yourself (Make sure to show off your skills and

make it clear that you will be a great data engineer.

Don't forget to include a hobby at the end!)

I recently graduated from the University of Birmingham in Maths which is a degree I really enjoyed. I chose to study maths because I really liked the problem-solving aspect but also the sense of reward from completing a difficult question. At University I gained a lot of skills in different areas, specifically in logical thinking and being able to view problem with different perspectives. I also got a lot of experience in group work which I found to be really valuable as now I have the communication and team-player skills and I can apply them in future. In my spare time I enjoy running and drawing, because I find they’re both good ways to relax.

[Boring?]

Why do you want to be a Data Engineer?

When I left University, I wasn’t 100% what I wanted to do with my degree, I just knew which areas of maths weren’t my stronger points but also which parts were my stronger areas and which I enjoyed and wanted to carry on with in the future. One of the modules I enjoyed most at uni involved coding (in C++) so when I graduated, I looked into different languages and found Python to be a widely used language so I started to learn python and found that I really liked it, even just that route of playing around with code to find different ways of using it and so yeah I figured that in future I wanted a career involving coding. So when I researched careers involving both aspects of coding and also group work [because that was another part of some uni modules I found I liked as well] and I came across Data Engineering and ended up here with Sparta.

[Too generic?]  
  
What is the role of a data engineer consultant? (Two sides to this,

the data engineering side and the consultancy side- good people person, flexible, good with change, [consultant is a very broad term])

Data engineers work together to essentially develop, test and maintain databases in order to improve the reliability, efficiency and quality of the data. Data scientists then interact with these databases and interpret and analyse the data to help businesses make important decisions. In terms of a consultant, you’d be flexible because you are working with different groups so then you also need good teamwork skills and have good people skills. Also good communication skills because you’d be working between two companies.

[More?]

What attributes make a great data engineer consultant?

I would say communication is a big one, in addition to maths and computing skills like extraction and transformation of data, but I say communication because you need to have the ability to not only report things to the other data engineers you are working with but the data scientists and other people working on or using the data.

Problem solving is also a good attribute to have as well as critical thinking in order to be able to come across any hurdles in terms of the coding side of the job.

[Generic?]

Why should we hire you?

From my experience I gained at university and skills I’ve acquired I can apply them to this role. For example, I can easily adapt to working in groups with new people and work well in group environments. I am a very determined individual, when I have a goal, I see it through and actually set myself a lot of personal goals. This helps me to be well organised and also good with time-management. I’m excited to start my work as a Data Engineer following training as I’m excited to put what I’ve learnt into practice.

[Pahahaha not good]

Can you give an example of when you have worked in a team?

One of my university modules required us to work in randomized groups and in these groups, we were assigned a project. Ours was to analyse given data on engine failures which would help to identify where the issue was and how to fix it. We had to present this in the form of a report and also a short presentation. The way in which we decided to carry this out, was by assigning each person a different role, mine being trying to find any consistencies with engine failure in the data and once I found the relevant data I informed the group on what I had. The group worked really well together and the only issue we had was that one member fell ill meaning they were behind on their work but the team quickly resolved this just by working together to produce her part of the report and in the end we achieved a first on this project so it was really successful.

[Only answer worth telling]

How do you handle stressful situations?

Try to stay calm and maybe take a minute or two to step back and just clear my head a little. If it was a situation where I was stressed about having lots to do or something I don’t know how to do I would list it out in smaller, simpler tasks of what needs to be done and take it from there.

I’ve also got an app on my phone with breathing exercises which I found really helpful at uni.

[Meh]

What makes you unique? Can use that I’m a good people person, willingness to learn, wanting/ needing to understand things and doing further research

I would say that I am a fast learner and

[Brill.]

Where do you see yourself in 5 years?

[Fuck knows]

**Agile**

1. What are the SCRUM events?

They are the daily scrum, sprint planning, the sprint review.   
[Need to add more detail]

1. What are the SCRUM artefacts?

The product backlog, the sprint backlog and the increments.

1. What is your understanding of Agile delivery? What are the benefits and drawbacks?

Agile is an approach to delivering a project in a group by a repetitive cycle of steeps. Benefits I think are that it makes it easier to list and identify ways in which to improve next time and also it promotes continuous improvement and allows ideas and concepts to be tested and either rejected or accepted early on. A drawback would be that it focuses on small incremental changes

**Agile** focuses on small incremental changes and the challenge is that the bigger picture can become lost and create uncertainty amongst stakeholders.

1. How does Agile compare with other methodologies like waterfall?
2. Are there any drawbacks to Agile?
3. In what situations would another methodology work better over Agile? Can you give an example?
4. What are user stories?
5. How can we prioritise our product backlog?
6. What information radiators can we use during retrospectives?
7. What is the purpose of a review?
8. What is the definition of ready?
9. What is a definition of done?
10. Why is it important to have good acceptance criteria for user stories?
11. What is the INVEST criteria for good user stories?
12. Why is it a good idea to use the INVEST criteria?
13. What do you need to consider when gathering requirements?

# Data Concepts

1. What is ETL?

ETL is a type of pipeline and it stands for extract, transform, load. Where extract is just extracting the data from the original source, transform is cleaning the data and putting it into the right format (which could include combining it, removing blank entries or duplicates etc.) and load is then just loading the data into the database.

1. What is ELT?

ELT is a different type of pipeline which stands for extract, load, transform. So it is similar to ETL except you extract then load it as it is into the cloud data warehouse or a data lake and you only transform it when you need to read it.

1. When might you want to use ELT over ETL?

ELT tends to be used for AI/ machine-learning scenarios so when for massive projects when there’s lots of data and you don’t know when or if you’ll need to read it so you store in for later use.

1. What is a json file? What does it look like?

Json means JavaScript Object Notation which is based on two structures: 1. A collection of key-value pairs; 2. An ordered list of values. It is primarily used for transmitting data between a web app and a server. They’re lightweight, text-based, human-readable and can be edited using a text editor.

1. What is OLAP?

OLAP is a type of database which stands for Online Analytic Processing. They’re designed to extract business intelligence from OLTP by working with the OLTP. You would design an OLAP for data analysts because it has an extra layer of abstraction and aggregation (i.e. takes away some of the techy bits) and has a semantic data model so it describes the meaning of data elements and has data integrated from multiple sources and aggregated together across multiple dimensions.

1. What is OLTP?

OLTP is a type of database which stands for Online Transaction Processing. It’s the management of transactional data like payments received, orders taken, services delivered etc. It requires a high degree of normalisation meaning there’s lots of tables but is easier to interpret.

1. What are ACID principles?

They are Atomicity, Consistency, Isolation and Durability.

Atomicity means each transaction must be treated as a single unit, i.e. if part of the transaction fails then the it all fails.   
Consistency means each transaction must results in a valid database state i.e. primary-foreign key relationships are intact.  
Isolation means that concurrent read/write executions should produce the same results as sequential executions.  
Durability means the transactions should remain committed in the event of system failure.  
[This is gonna sound v rehearsed so change it]

1. What are the two schools of thought when it comes to Data +Warehouse Design?

The Inmon Approach and the Kimball Approach.  
[What else do I say here?]

1. What's the difference between a Database and a Data Warehouse?

One main difference is that Data Warehouses are designed for data analysis whereas Databases are designed for transactional processing. Also databases are structured as efficiently as possible, with no duplicate information in multiple tables. Data Warehouses typically denormalise their data, prioritizing read operations over write operations.   
[Is this right?]

1. Explain the Inmon Architecture for Data Warehouse Design

This uses the top-down approach so a normalised data warehouse is created first then data is summarised and distributed out to Data Marts. The data is non-volatile, so never overwritten or deleted.

1. Explain the Kimball Architecture for Data Warehouse Design

This uses the bottom-up approach so you separate the Data Marts first by aligning with single departments and then you combine the data from Data Marts. This approach is quicker and less expensive to set-up but harder to maintain.

1. What is Dimensional Modeling?

It’s a data structuring technique used to optimise data so you can retrieve it faster. Its based on two types of table, fact and dimensional, where fact tables represent events and dimension tables represent people/items or objects.   
[Do I need to say more?]

1. What are the differences between databases, data warehouses, and data lakes?

Lake is messy.

# Database Design

1. What is an ERD?

An entity relationship diagram shows the relationships between tables in a database. Each entity is represented by a table in the database and each has a list of attributes (which would be the columns in the table). The primary and foreign keys are represented and relationships are shown using lines. Uses crows feet notation to describe relationships in more detail.

1. What is a primary key?

Primary keys help to uniquely identify records in a table. They can’t be empty or changeable but can consist of multiple columns in which case it’s a composite key.

1. What is a foreign key?

They reference primary keys of other tables and don’t have to be unique and also a table can have any number of foreign keys.

1. What is the difference between a foreign and primary key?

One difference is that you can only have one primary key per table but you can have multiple foreign keys. Another difference is that primary keys must be unique but foreign keys don’t necessarily have to be.

1. Explain Normalisation.

Normalisation is a way to organise data in a database by creating tables and establishing relationships between the tables.

1. What are the advantages and disadvantages of normalisation?

An advantage is it helps to reduce redundancies which could waste disk space and also makes it easier say part of the data needs to be changed. Disadvantages are that it takes more time and also it’s harder to read because you can’t get all your information from one table you may have to refer to multiple different ones.

1. What does it mean for a database to be in 3rd normal form?

A database is in 3rd normal form if it is in 2NF and there is no transitive functional dependency.   
[Do I need to explain transitive functional dependency? When a non-key column is dependent on another non-key column which is dependent on the primary key.]

1. What is a junction table?

Contains all the primary keys?

# SQL

1. What command would you use if you want to add a table?

CREATE TABLE name\_of\_table (var1, var2, …)

1. What command would you use if you want to delete a row of data

DELETE FROM name\_of\_table  
 WHERE var1 < whatever you want

1. What command would you use if you want to insert data into a table

INSERT INTO name\_of\_table (names of columns you are inserting into)   
 VALUES (data in same order as column names)

1. What are DML, DDL, DCL and TCL?

They are types of commands, so DML (Data Manipulation Language) allows you to modify data using the commands INSERT, UPDATE & DELETE, DDL (Data Definition Language) helps to define the structure so examples of specific commands are CREATE, DROP & ALTER. DCL is (Data Control Language) give permission to other users by using GRANT and REVOKE. Finally, TCL (Transaction Control Language) deals with the transactions within the database using commands like ROLLBACK (to undo transactions that haven’t already been saved) and SAVEPOINT (to set a savepoint within a transaction).

1. What does NULL mean?

Null is used to show that a value doesn’t exist in the database for that variable.

1. Can a primary key be NULL?

No because primary key are used to reference so if there’s no value then you wouldn’t be able to find the referenced data values.

1. How many joins are there in SQL and difference between them?

There are 4: left, right, inner and outer. Left will join

1. Tell me about the different types of SQL statements?

[Is this not the same as Q4?]

1. What is your experience with SQL?

# Python

1. What is a function? Why are they useful?
2. What are some good practices when writing functions?
3. What are the benefits of using classes? Are there disadvantages?
4. What is instantiation?
5. What is initialisation?
6. Is it a good idea for Data Engineers to use classes?
7. What is OOP?
8. What are the 4 pillars of OOP?
9. How can OOP be useful for Data Engineers?
10. What is a lambda function?
11. What is unit testing?
12. What are the advantages of unit testing?
13. What is Test Driven Development (TDD)?
14. Is unit testing useful for Data Engineers? Why?
15. In what way can we manage errors in python?