

Your grade: 100%

Your latest: 100% • Your highest: 100% • To pass you need at least 80%. We keep your highest score.

Next item →

1. Which of the following statements is true about the transformation stage of the data engineering lifecycle?

1 / 1 point

- ☒ Data transformation is the "turn data into something useful" stage of the data engineering lifecycle.
- ☐ The transformation stage of the data engineering lifecycle is made up of two parts: queries and transformation.
- ☐ Data transformation only occurs after data has been ingested and before it is stored.

✓ Correct

That's right. The transformation phase of the data engineering lifecycle is where you as a data engineer really start to add value to data by transforming it into a useful form that serves downstream use cases.

2. Which of the following statements are true about data generation and source systems? Select all that apply.

1 / 1 point

- ☒ As a data engineer, you will typically not create or maintain the source systems from which you ingest data.

✓ Correct

In general, source systems are created and maintained by upstream stakeholders so they're typically out of your control.

- ☐ As a data engineer, you are typically responsible for setting up a source system and generating data before the ingestion stage.

- ☒ Databases are the most common source systems you'll interact with.

✓ Correct

The most common source systems are databases, which could be a relational database or other type of NoSQL systems.

- ☒ As a data engineer, you'll be most successful if you work directly with source system owners to understand how those systems work.

✓ Correct

That's right. You'll be most successful if you work with the source system owner to understand how those systems generate data, how the data might change over time, and how those changes will impact the downstream systems you build.

3. Which of the following are examples of storage abstractions?

1 / 1 point

- ☐ Object Storage
- ☐ Relational Database
- ☒ Data Lake

✓ Correct

A data lake is an example of a storage abstraction that combines other storage systems.

- ☒ Data Warehouse

✓ Correct

A data warehouse is an example of a storage abstraction that combines other storage systems.

4. Imagine you are a data engineer at a company that provides online courses through a mobile app. Which of the following examples represents an embedded analytics use case?

1 / 1 point

- ☒ A user-facing dashboard that shows learners how many courses they have completed and the total time they have spent learning each week.
- ☐ An internal dashboard showing course enrollments and ratings.
- ☐ A real-time dashboard that tracks critical performance metrics in the mobile app – like the load time for course content pages and videos – and sends alerts to engineers when errors occur or app features are not performing as expected.
- ☐ A recommendation engine that provides learners with suggestions on the next course they should take.

✓ Correct

This is an example of embedded analytics, which is typically an externally focused, customer-facing type of analytics.

5. When it comes to the security of your data systems, what does the principle of least privilege imply?

1 / 1 point

- ☐ You should always ingest sensitive data, but only allow access to sensitive data to those individuals or applications that require it.
- ☒ You give users or applications access to only the essential data and resources to perform their job or intended function, and only for the duration that is required.
- ☐ Only the data engineers should operate as "admin", "superuser" or from the root shell.
- ☐ Only the most senior members of the team should be given admin access to systems, while more junior members should be given restricted access.

✓ Correct

That's right!

6. Which of the following statements are true about magnetic disk drives, solid-state drives, and Random Access Memory (RAM), which are some of the raw ingredients of storage systems? Select all that apply. 1 / 1 point

☒ Magnetic disk drives are significantly cheaper than solid-state drives.

☒ **Correct**
At the time of the creation of these courses, disk storage is 2-3 times cheaper than solid-state storage.

☐ Magnetic disk drives are not used anymore in modern storage systems.

☒ RAM is volatile, meaning that if your system loses power, data stored in RAM is typically lost very quickly.

☒ **Correct**
That's right.

☐ Solid-state drives typically have faster read and write speeds than RAM.

7. True or False: data engineers need to know how to code. 1 / 1 point

☒ True

☐ False

☒ **Correct**
That's right, within the software engineering undercurrent but also across all stages of the data engineering lifecycle you will need to be able to write clean performant code in your job as a data engineer.

8. For which of the following use cases would you consider streaming ingestion over batch ingestion? Select all that apply. 1 / 1 point

☐ In a pipeline to serve data to data analysts who are interested in analyzing the company's weekly sales. Each Monday, the data analysts would like to look at the sales from the previous week.

☒ To serve data to a data analyst who wants to create dashboards showing online user activity on the company's website within a second or less after the activity data has been recorded.

☒ **Correct**
Since the data has to be shown on the dashboard shortly after it is produced, this is a good use case for streaming ingestion.

☒ To serve data for an operational analytics use case, where instantaneous alerts must be issued if a website goes down or a live product feature stops working.

☒ **Correct**
Operational analytics is typically about monitoring real-time data for immediate action, so this is a good use case for streaming ingestion.

☐ To serve data to a data scientist who is interested in training a machine learning model. The data scientist would like to use historical data to train the model.

☒ To serve data for a fraud detection system in bank transactions, where fraudulent transactions or attempted transactions must be detected and mitigated immediately while they are underway.

☒ **Correct**
Since the transaction fraud has to be detected and dealt with as soon as possible after it occurs, this is a good use case for streaming ingestion.

9. What are the three pillars of DataOps? ? 1 / 1 point

☐ Automation, Version Control, and Data Quality

☒ Automation, Observability & Monitoring, and Incident Response

☐ Automation, Observability, and Monitoring

☐ Orchestration, Version Control, and Incident Response

☒ **Correct**
That's right!

10. In the lab this week, you used AWS Glue, which is an ETL service that helps you prepare and integrate data from multiple sources. Which stage(s) of the data engineering lifecycle did you implement using AWS Glue? Select all that apply. 1 / 1 point

☐ Data generation in source systems

☐ Storage

☒ Ingestion

☒ **Correct**
In the lab, you used AWS Glue to ingest and transform data before storing the data in object storage.

☒ Transformation

☒ **Correct**
In the lab, you used AWS Glue to ingest and transform data before storing the data in object storage.

11. In the lab, one of the AWS services you used was Amazon Athena. Which of the following statements best describes this service? 1 / 1 point

☒ Amazon Athena is a query service that allows you to directly query data from Amazon S3 using SQL queries.

☐ Amazon Athena is a visualization tool that allows you to create Business Intelligence dashboards.

- ☐ Amazon Athena is a relational database service for normalized data.
- ☐ Amazon Athena is a type of object storage system.

☒ **Correct**

That's right. Amazon Athena allows you to query data from S3 without needing to store the data in a traditional database system.