Your grade: 100%

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

Next item

1. What are the key elements of any requirements gathering conversation as discussed in the videos? Select all that apply. 1 / 1 point

- [x] Learn what existing data systems or solutions are already in place.

  Correct
  That's right, it's important to first understand any existing systems or solutions before starting your work to modify or replace them. This was described as one of the key elements of requirements gathering in the "Translate Stakeholder Needs into Specific Requirements" video

- [x] Learn what problems or pain points there are with existing systems.

  Correct
  Yes, this is important because your stakeholder's pain points with any existing systems will be important to consider when designing a new system. This was described as one of the key elements of requirements gathering in the "Translate Stakeholder Needs into Specific Requirements" video

- [x] Learn what action stakeholders plan to take based on the data you serve them.

  Correct
  Indeed, this is critical because stakeholders may describe their needs using phrases like "real-time" but only when you understand the action they plan to take can you start to translate their needs into actual system requirements. This was described as one of the key elements of requirements gathering in the "Translate Stakeholder Needs into Specific Requirements" video

- [x] Identify any additional stakeholders you will need to talk to.

  Correct
  That's right. Oftentimes stakeholder conversations will lead you to discover there are others you need to talk to as well. This was described as one of the key elements of requirements gathering in the " Translate Stakeholder Needs into Specific Requirements" video

2. Which of the following best describes data engineering as defined in this course? 1 / 1 point

- (•) Data engineering is the development, implementation, and maintenance of systems and processes that take in raw data and produce high-quality, consistent information that supports downstream use cases, such as analysis and machine learning.
- ( ) Data engineering is about setting up storage systems to store raw data from different sources before serving it to its end-users.
- ( ) Data engineering is focused on deriving insights from data that can add value for the business..
- ( ) Data engineering is about using tools and technologies to build data pipelines that move data through the stages of the data engineering lifecycle.

Correct
Nice job! The complete definition as stated in this course and in the book also included a second sentence mentioning the undercurrents of the data engineering lifecycle: "Data engineering is the intersection of security, data management, DataOps, data architecture, orchestration, and software engineering."

3. What is the difference between a region and an availability zone on the AWS cloud? 1 / 1 point

- ( ) An availability zone spreads across multiple regions and each region contains one or more data centers.
- ( ) A data center consists of multiple regions and each region consists of multiple availability zones.
- ( ) A data center consists of multiple availability zones and an availability zone is spread across multiple regions.
- (•) A region consists of multiple availability zones and an availability zone contains one or more data centers.

Correct
Each AWS region contains at least 3 isolated and physically separated availability zones. Each availability zone contains a group of one or more discrete data centers with redundant power, networking and physical security.

4. Which of the following statements are correct about virtual servers? Select all that apply. 1 / 1 point

- [x] Multiple virtual servers can share the same underlying physical resources, which helps achieve efficient and cost-effective use of resources.

  Correct
  Nice work! The benefit of this virtualization is that you can create more than one virtual machine that shares the same underlying physical resources.

- [x] An Amazon Elastic Compute Cloud or EC2 instance is a virtual server in the AWS Cloud

  Correct
  Nice work!

- [ ] A virtual server is an actual physical server that is hosted on the cloud and to which you can connect to virtually through the internet.
- [x] A virtual server is a software representation or emulation of an actual physical server.

  Correct
  Nice work!

5. What's the difference between business requirements, stakeholder requirements, and system requirements as described in this week's videos? 1 / 1 point

- ○ There are functional and nonfunctional business, stakeholder, and system requirements.
- ○ Business and stakeholder requirements are the same thing, you gather both of these in the first stage of any data engineering project. System requirements are what you define after collecting high level requirements.
- ◉ Business requirements define the high level goals of the business. Stakeholder requirements define the needs of the individuals within the organization to meet those business goals. System requirements define what a system needs to be able to do in order to meet business and stakeholder requirements.
- ○ Business requirements are defined by the company leadership team, stakeholder requirements are defined by individual contributors within the organization, and system requirements are defined by the data engineer.

Correct
Great work! These are the definitions defined in the video about system requirements.

6. What are the key steps you should take in the first stage of any data engineering project as presented in the "framework for thinking like a data engineer"? Select all that apply. 1 / 1 point

- ☑ Identify the business goals and stakeholders you will serve.

Correct
That's right. Before starting any data engineering project it's important to understand the goals of the business and know who the stakeholders are that you will serve. This will inform the requirements for the system you need to build.

- ☐ Translate stakeholder needs into functional requirements for your system.
- ☑ Explore existing systems and stakeholder needs.

Correct
Good job! In the first stage of any data engineering project, you should have conversations with individual stakeholders to learn what systems are currently in place and what needs stakeholders have beyond what the current systems are providing. This will inform the requirements for the system you need to build.

- ☐ Perform a cost-benefit analysis to choose between comparable tools and technologies.
- ☑ Ask what action stakeholders plan to take based on the data you serve them.

Correct
That's right. Before designing and building any systems, you should understand what action stakeholders plan to take, as this will inform the requirements for your system.

7. Which of the following statements are true about VPC (Virtual Private Cloud)? Select all that apply. 1 / 1 point

- ☐ A VPC can only contain private subnets.
- ☑ A VPC exists inside a region and can span multiple availability zones inside the region.

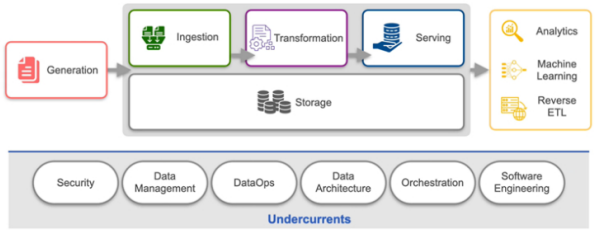
Correct
Nice work! A VPC exists inside a region, which can contain more than one VPC, and a VPC spans multiple availability zones inside the region. VPC is a way to isolate your resources (for example EC2) from the outside world.

- ☑ A VPC is an isolated private network in which you can launch your AWS resources.

Correct
Nice work!

- ☐ A VPC exists within one subnet and can belong to a single availability zone.

8. This week you were introduced to the diagram of the data engineering lifecycle, depicted below. 1 / 1 point



Which of the following statements are true about the diagram of the data engineering lifecycle? Select all that apply.

- ☑ The undercurrents are not stages of the lifecycle but rather represent aspects of data engineering that cut across the entire lifecycle.

Correct
That's right. The undercurrents represent aspects of data engineering that will be relevant no matter which stage you're working in.

- ☑ Storage is part of all stages of the data engineering lifecycle.

Correct
That's right, data storage is involved in every stage of the data engineering lifecycle, even in the generation stage, although that's not shown in the diagram since data generation is not your focus as a data engineer.

- ☑ The middle gray box (around ingestion, transformation, serving and storage) represents the stages that

will be the focus of your work as a data engineer.

Correct
Nice work! These are the stages that are the focus of your work as a data engineer.

- [ ] Analytics, machine learning, and reverse ETL represent the final stage of the data engineering lifecycle.

9. Consider the following scenario: 1 / 1 point

You work as a data engineer at a financial institution. You were tasked with building a pipeline to serve data for the data science team of the "clients services" department. You learned that the data scientists are interested in predicting customer churn, or in other words, being able to predict whether a given customer is likely to stop doing business with the bank. You met with the data scientists who told you that they'd like you to provide them with historical information about active and inactive customers, such as their past transactions, credit information, and account information. The data scientists will then analyze this data and build a churn prediction model. You learned that the information requested by the data scientists exists in several databases maintained by software engineers from different bank departments.

Identify the upstream and downstream stakeholders.

- ○ The downstream stakeholders are both the data scientists and the software engineers that maintain the source databases. The upstream stakeholders are the bank clients.
- ◉ The upstream stakeholders are the software engineers that maintain the source databases and the downstream stakeholders are the data scientists of the "client services" department.
- ○ The downstream stakeholders are the software engineers that maintain the source databases and the upstream stakeholders are the data scientists of the "client services" department.

Correct
Good job! Upstream stakeholders represent the stakeholders that create and/or maintain the source systems you will ingest data from. Downstream stakeholders are the end-users to whom you're serving the data.

10. Which of the following statements best describes the AWS "shared responsibility model"? 1 / 1 point

- ◉ AWS is responsible for the security of the cloud and you are responsible for security in the cloud.
- ○ AWS is responsible for security behind the scenes in cloud systems and you are responsible for the security of any public facing components of your systems.
- ○ AWS is responsible for security in the cloud and you are responsible for the security of the cloud.
- ○ Depending on the resources and services you choose to build your systems, AWS will be responsible for the security of some of them and you will be responsible for the security of others.

Correct
That's right, and this means that the security of the infrastructure and services that make up the cloud is AWS's responsibility and the security of the systems you build on the cloud is your responsibility.