Enterprise Data Storage

1. **As data evolves during its life cycle**, which of the following factors should ML pipelines address to operate properly?(check all that apply).

   - <mark>Account for anomaly detection.</mark>

     For example data errors must be handled in the same way as code bugs.

   - <mark>Provide resilient mechanisms for disruptions.</mark>

     For example ML pipelines should incorporate resilient mechanisms to deal with inconsistent data.

   - Use feature engineering.

   - Monitor model and data provenance.

   - <mark>Account for scalable solutions.</mark>

     Production traffic will vary from day to day and thus your pipeline must scale accordingly.

2. Many modeling problems use identical or similar features, and there is substantial value in enabling teams to share features between their own projects and for teams in different organizations to share features with each other. Which of the following storage solutions is deliberately designed to address these user cases?

   - Relational database

   - Data lake

   - Data warehouse

   - <mark>Feature Store</mark>

3. Which are the main advantages of using a cloud-based data warehouse ?(check all that apply).

   - User needs to handle all maintenance

   - <mark>Provides easy on-demand scalable solution</mark>

     Cloud solutions are really flexible for scaling up.

   - <mark>They are cost efficient</mark>

     Otherwise all the software and hardware costs will be handled by your organization.

   - User owns and controls data governance.

4. About data lakes it's only true that:

- Can handle both structured and unstructured data.
- Handles only processed data
- Aggregates data from a single source only.
- Handles only structured data.

Data lakes are really flexible in the type of data they can handle.