DATA 100: Vitamin 12 Solutions

April 28, 2019

1 Data Warehouse

A data warehouse ____ and organizes historical data from ____ source(s).

- ☑ collects, multiple

- ☐ distributes, a single

- ☐ gathers, large

- ☐ collects, water

2 ETL

ETL stands for:

- ☐ extra, transform, load

- ☑ extract, transform, load

- ☐ extract, transition, load

- ☐ extra, transform, look

3 Data Lake

Which of the following issues are associated with data lakes?

- ☑ Noise dominates signal in data

- ☐ Too much data governance and planning

- ☑ Plenty of uncleaned data

- ☑ Limited backwards compatibility with existing tools

4   Distributed File Systems

Distributed file systems should be fault tolerant because:

- ☐ they support massive files spread across multiple machines.

- ☐ it allows parts of the files to be performed in parallel.

- ☑ they are often built with cheap commodity machines that have high failure rates.

5   Parallelism

Which of the following programming paradigms is most amenable to parallelism?

- ☐ object oriented

- ☑ functional

- ☐ mathematical

- ☐ procedural