1_What is GitOps ?

GitOps is a set of practices to manage infrastructure and application configurations using Git, an open source version control system. GitOps works by using Git as a single source of truth for declarative infrastructure and applications.

GitOps uses Git pull requests to automatically manage infrastructure provisioning and deployment. The Git repository contains the entire state of the system so that the trail of changes to the system state are visible and auditable.

GitOps is built around the developer experience and helps teams manage infrastructure using the same tools and processes they use for software development. Other than Git, GitOps gives you the ability to choose the tools you need.

2_What is Big Data Tools ?

- Hadoop
- Atlas.ti
- HPCC
- Storm
- Qubole
- Cassandra
- Statwing
- CouchDB

3_ What is the 7 _VS of Big Data ?

1_Volume : Volume is how much data we have – what used to be measured in Gigabytes is now measured in Zettabytes (ZB) or even Yottabytes (YB). The IoT (Internet of Things) is creating exponential growth in data. The volume of data is projected to change significantly in the coming years.

2_Velocity : Velocity is the speed in which data is process and becomes accessible. I remember the days of nightly batches, now if it's not real-time it's usually not fast enough.

3_Variety **:** Variety describes one of the biggest challenges of big data. It can be unstructured and it can include so many different types of data from XML to video to SMS. Organizing the data in a meaningful way is no simple task, especially when the data itself changes rapidly.

4_Variability: is different from variety. A coffee shop may offer 6 different blends of coffee, but if you get the same blend every day and it tastes different every day, that is variability. The same is true of data, if the meaning is constantly changing it can have a huge impact on your data homogenization.

5_Veracity: Veracity is all about making sure the data is accurate, which requires processes to keep the bad data from accumulating in your systems. The simplest example is contacts that enter your marketing automation system with false names and inaccurate contact information. How many times have you seen Mickey Mouse in your database? It's the classic "garbage in, garbage out" challenge.

6_Visualization: Visualization is critical in today's world. Using charts and graphs to visualize large amounts of complex data is much more effective in conveying meaning than spreadsheets and reports chock-full of numbers and formulas.

7_Value: Value is the end game. After addressing volume, velocity, variety, variability, veracity, and visualization – which takes a lot of time, effort and resources – you want to be sure your organization is getting value from the data.

4_ What Is Overfitting?

Overfitting is a concept in data science, which occurs when a statistical model fits exactly against its training data

5_How to visualize Data that has more than 3D ?

1_ Parallel Coordinate Plot

2_Andrew's Plot

3_Scatter-Plot Matrix

6_How to convert Data from non_linear to linear ?

Log transformations is one of the ways in which this issue is dealt with ,However other transformations like sqrt,exponentiation also work at times.
However no data will be exactly linear but rather show a linear trend.
So you have to try out with different methods and see which one gives a better linearity.Also log transformations are a little different to interpret and hence that has to be taken care of if you are using log transform. And that help me to be able to customize the equation , but that will increase the traning time

7_ Why we use Integration in AI?

Integrals are concerned with the accumulation of quantities. In mathematics, integrals give numbers to function such that they can explain different quantities such as area, volume, displacement, and multiple other concepts by the continuous summation of the infinitesimal data. The process of finding integrals is known as integration. Along with differentiation, integration is a fundamental, essential operation of calculus, and serves as a tool to solve problems in mathematics.

8_difference between relational and non relational Database?

 relational databases : store data in rows and columns like a spreadsheet (Structured database)

 non-relational databases: store data don't, using a storage model (one of four) that is best suited for the type of data it's storing. (Unstructured Database)

9_what is genetic algorithm?

Optimization techniques are the techniques used to discover the best solution out of all the possible solutions available under the constraints present. The genetic algorithm is one such optimization algorithm built based on the natural evolutionary process of our nature. The idea of Natural Selection and Genetic Inheritance is used here. Unlike other algorithms, it uses guided random search, i.e., finding the optimal solution by starting with a random initial cost function and then searching only in the space with the least cost (in the guided direction). Suitable when you are working with huge and complex datasets.

Phases of Genetic Algorithm

1. Initialization of Population

Every gene represents a parameter (variables) in the solution. This collection of parameters that forms the solution is the chromosome.

2. Fitness Function

We have to select the best ones to reproduce offspring out of the available chromosomes, so each chromosome is given a fitness value. The fitness score helps to select the individuals who will be used for reproduction.

3. Selection

This phase's main goal is to find the region where getting the best solution is more.

Inspiration for this is from the survival of the fittest.

It should be a balance between exploration and exploitation of search space.

GA tries to move the genotype to higher fitness in the search space.

Too strong fitness selection bias can lead to sub-optimal solutions.

Too little fitness bias selection results in an unfocused search.

Thus, Fitness proportionate selection is used, also known as roulette wheel selection, as a genetic operator used in genetic algorithms to select potentially useful recombination solutions.

4. Reproduction

Generation of offsprings happen in 2 ways: Crossover Mutation

a) Crossover

Crossover is the most vital stage in the genetic algorithm. During crossover, a random point is selected while mating a pair of parents to generate offsprings.

There are 3 major types of crossover.

 Single Point Crossover: A point on both parents' chromosomes is picked randomly and designated a 'crossover point'. Bits to the right of that point are exchanged between the two parent chromosomes.

Two-Point Crossover: Two crossover points are picked randomly from the parent chromosomes. The bits in between the two points are swapped between the parent organisms.

Uniform Crossover: In a uniform crossover, typically, each bit is chosen from either parent with equal probability.The new offspring are added to the population.

b) Mutation

In a few new offspring formed, some of their genes can be subjected to a low random probability mutation. This indicates that some of the bits in the bit chromosome can be flipped. Mutation happens to take care of diversity among the population and stop premature convergence.

5. Convergence (when to stop)

Few rules which are followed which tell when to stop is as follows:

When there is no improvement in the solution quality after completing a certain number of generations set beforehand.

When a hard and fast range of generations and time is reached.

Till an acceptable solution is obtained.