Q1.

We will be building a Cloud Computing Resources Prediction using ML/DL techniques. Here we will be creating a fully automated solution for B2B, which will give businesses an idea of how much resources(virtual machines) are being wasted. This will help businesses lower their OPEX/ CAPEX costs.
With the help of large dataset clusters from cloud service providers such as GCP, Azure, and AWS we will search for recurring patterns in utilisation and allocation of resources by these CSPs. As a further development to the ease of use we also intend to build a web interface to input customer data directly and make decisions to optimise resources.

Q2.
Since our freshman year, Srihan and I have been avid cloud enthusiasts. In the past, we have worked on multiple cloud platforms such as Amazon Web Services, Google Cloud, and Azure gaining experience with real-world cloud projects. We've done a vast amount of projects on both architecting and deploying a cloud infrastructure. However, we have had very limited experience in research in cloud computing and hence the inclination.
Another reason why we chose this project was that this was a problem that we had experienced ourselves. Often while deploying architecture over unpredictable traffic we would end up realizing that we had either used a lower performance instance and traffic wasn't being managed or load was being overfitted using an instance of higher performance, resulting in more costs.

Q3.

Before the modern age of high-performance computing, most of what we know today was achieved using physical servers, racks, rails, and data centers, which resulted in a large amount of capital and operational expenditure. The introduction of cloud computing virtualized almost all sorts of hardware that could be used at the industrial level. Companies shifted to a hybrid model of Annual Maintenance Contracts (AMCs), where cloud providers owned much of the physical assets such as data centers and servers. As a result this almost immediately brought down fixed capital costs by replacing them with variable operational costs. It was an optimization of the grandest level.

However, in recent times the goal has completely changed. Clients and cloud providers themselves are trying to achieve a limited amount of resource wastage. Existing solutions are very limited in scope, they provide resource optimization over metrics such as CPU and memory utilization. We on the other hand have a completely new outlook. Hence, we decided to build an resource optimiser for cloud providers and users.

Q4.

In the previous semesters my partner and I have chosen our previous projects based on our skill and the capability to complete the project before the deadlines.
This time we wanted to explore a field which was not taught in class and believed the fact a new project can help us learn new stuff as we go through different stages. So we chose cloud computing and resolving one of the core issues of its service.
what makes our project unique is that the issue is about efficiency of working services at the datacenters. This key issue has come into the limelight with the surged usage of cloud services which led to more usage and more pressure on the cloud's on premises center. Now how we want to tackle this issue is with the help of datasets from on premises data centers of various cloud vendors. With the help of various machine learning processes that can analyze the datasets and give the output of a threshold value for the usage of on premises center and now this value can help in increasing the efficiency of the services. If time permits we are thinking of automating the whole process so that it is not restricted to one particular cloud service vendor.

Q5.
As mentioned above the issue we are trying to resolve is based on the efficiency of cloud services at their on premises data centers which is a vital role that has been around for the past decade. Well the optimization of the resources at the cloud center can help companies to decrease their carbon footprint on the world. This can help vendors to save their costs and be eco friendly at the same time. The added advantage to this solution is that not only are cloud purveyors are benefitting from this issue but also their clients indirectly. Most of the firms follow economies of scale to measure their profit margin and sell their product at a feasible cost. Reducing this scale results in a cheaper bill at the end of the month for the current clients. By solving this issue there would be a possibility down the lane in the long run where cloud services may be used as a common utility by every company which has to deal with the tiresome data management just like computers.

Q6.
With this project we are trying to bring a new system in the industry with multiple innovations that different from pre-existing solutions in the market. Some of the innovations we are trying to introduce are:

1. Integrating Cloud Service Provider Platforms with our Machine Learning Model/Platform: Unlike other machine learning where most of the testing is solely done on pre-existing dataset, we will be having 2 test phases. Firstly alpha testing phase will comprise of the model to be tested on data we collected from various source, Beta testing on the other hand will comprise of use creating simulated cloud environments and then integrating the model to make decisions from real-time cloud watch metrics accessed from the cloud.

2. A solution for Cloud Providers: Where most people focus on optimisation of cloud resources for clients and customers. Our primary focus is to create a solution for cloud providers themselves. Since most of the hardware is present at the data centre level where optimisation is of almost importance, our solution will be a great aid for service providers to save money and resources.

Q7.
Adarsh Raghav (E19CSE063): My major contribution in this project will be towards the backend. The backend in our project consist of pipelining data and the entire process, making the machine learning model, evaluating performance and adjusting. In the pipeline phase I will basically be gathering data either through datasets/sources or with the help of web crawlers and scrappers, this data is to be cleaned, scaled and normalised. Moreover, I would also want to make some metrics, analysis and visualisation to observe few trends and changes in resource management that already exists.

Shrihan Chinta (E19CSE026): Shrihan's main focus would be towards the front end and cloud computing. Since Shrihan is well versed with various cloud service providers such as amazon web service and azure, he will build few infrastructures for us to test the machine learning model on. He will be guiding through the API calls and all the network features we will be using for this project. Moreover, he will make sure that integration between backend and front end is done smoothly and accurately.

Q8.
For the past few years many researchers have come up with different solutions to solve this is pondering issue. For example let us take into consideration Bayesian approach to virtual resource prediction. In this method the research author proposes this idea to predict resource usage with the help of network simulators and training the model with previous data center usage and predict the resource usage with accuracy. Now this idea which is quite closer to our project proposal but this has it's own cons as though it has all the advantages over traditional methods but it is expensive to implement and it performs poorly when it comes to high dimensional data it fails to provide an accurate output. As Cloud is majorly about data transfers. Another existing solution for

this issue which has been proposed by researchers is deep learning techniques with the help of some models like BiLSTM which capture dependency characteristics. This does solve the issue but just like the previous solution it cannot handle high dimensional data . There is one last solution which is quite similar to our approach intelligence ensemble algorithm with which it helps in feature selection as better features can help in taking in consideration more factors and an efficient output . Though it figured out the solution it took a lot of time. Now keeping all these drawbacks in mind we have shaped our idea to be more sustainable and adaptable in a generic way. We are trying to propose is a basic deep learning which takes into account all the features and what makes our solution different from the existing ones is that we are proposing the fact of training the real-time data along with the past data and predict an accurate threshold value and the reason behind using a deep learning model is that it is sustainable.

Q9.
The most questioning part of our project is how our solution will fit and solve the issue of the problem on the big scale. The first step towards solving our issue is to collect data from various cloud vendors which will take a couple of months for asking permissions and surf the internet for datasets of past server usage. Adding to taking the datasets we would have to clean the dataset and train the data as per the requirements which would require a deeper dive into machine learning as the basic knowledge that we have is not enough to push the project to the next stage and with the help of online sources we would be able to learn and implement the subject. Among all these constraints we would be able to build a prototype in four months as we have been able to find the resources and have been keeping track of our learning process. We believe with this strict regime we would be able to complete the project.

Q10.
February - Preliminary Research & Data Collection and Data Cleaning/ Pre-Processing and Documentation March - Implementation of Training and Evaluation, First Testing on basic machine learning model interface using the data collected April - Creating a web interface and API system too perform testing on real cloud platforms

Q11.
1. Collecting valid and authentic data of various cloud vendors:
https://github.com/google/cluster-data
https://github.com/alibaba/clusterdata
http://gwa.ewi.tudelft.nl/datasets/gwa-t-12-bitbrains
http://gwa.ewi.tudelft.nl/datasets/gwa-t-13-materna

2. Preprocessing of Data:
Data pre-processing in this project mainly consists of two parts: data cleaning and data transformation.
Data Pre-processing will be done to identify outliers. This can be achieved by creating box plots for each attribute. Moreover, we will be removing Nan values and replacing them where feasible, with imputation.
3. Feature Selection:
Using a heat map we will try to distinguish which features are important for the prediction of computing resources. Feature selection can improve prediction accuracy and reduce time consumption. Therefore, it becomes crucial to select the relevant set of features.

4. Training Machine Learning Models:
Since most of our data is already labeled we will be using supervised learning techniques to predict the apt number of resources to be used. Since most of our task is based on classifying customer resource usage we will first try to segment data using K-NN and SVMs. Further to improve on the accuracy we will tune the parameters a bit. We will also play around with other models like XGboost which are high-performance models for supervised learning.

5. Evaluation Model Selection:
Once the models are trained in step 4. we will test them and evaluate results using various metrics such as RMSE, MSE, Correlation, MAE, etc. This will give us a rough idea about which model performs the best.

6. Testing on actual cloud provisioning services:
We will try to run our models over various if not one cloud service such as AWS, GCP, Azure. Using API calls and functions we will be able to achieve this

7. A Web Interface (Future Work):
Using flask our model will be up and running on a web interface where customers can feed in data using API calls or data sheets to check if resources are being wasted.

Q12.
This project is a sign of what we have learned in the previous semesters. As the core concept of our project is Cloud computing. In our last semester the faculty has thrown some light on this subject and have showed us how this impossible idea has transformed into a common hub for all the technologies. Based on what we have studied we have seen the drawbacks of this subject that is what has become our project idea for this semester. To solve this problem we are using Machine Learning techniques. Now Machine Learning concepts have been taught to us in our previous semester. To sum up all the concepts this project is a combination of all the knowledge that we inculcated for the past three years to solve a realtime issue that has been part of the cloud industry for the past decade.

Q13.
Our project is based on solving one of the crucial problems existing in the industry. The success of our project will be on the basis on how far are we near to building our prototype. As solving such prominent issue can not be done in one shot but our humble attempt towards solving this. The major part of our solution is to use a Machine Learning model and based on that model it would predict a threshold value which is used by cloud service vendors this value can act as margin value for vendors to become more efficiently. Now the accuracy of our model is will be the success parameters for our project for this semester.

Q14.
First and foremost we will need to collect data from multiple online sources. Below are four links to collect and scrap data that will be necessary for this project:
https://github.com/google/cluster-data
https://github.com/alibaba/clusterdata
http://gwa.ewi.tudelft.nl/datasets/gwa-t-12-bitbrains
http://gwa.ewi.tudelft.nl/datasets/gwa-t-13-materna

Some other resources we will be requiring:
-> AWS/GCP/Azure and other CSPs: Since our Project entails developing a prediction system to detect wastage of resources for Cloud Service Providers, we will have to use different services provided by multiple providers such as amazon, google, Microsoft, and many others.
-> ReactJs: Eventually we will also try to build a web interface that will allow users to upload their data and predict world utilisation.
-> Swagger (Future) and ExpressJs: This is needed for a future feature where we intend to provide our services using APIs that can be built using swagger and expressJs.

Q15.
1. Integrated deep learning method for workload and resource prediction in cloud systems (Jing Bi, Shuang Li, Haitao Yuan, Meng Chu Zhou): In a more recent trend of 2018, it was concluded that more than 3.6 billion people are dependent on cloud computing in one form or another [Jing Bi et al. (2020)]. Owing to the size of this market, major players such as Amazon Web Services, Google Cloud and Microsoft Azure have stepped into this space. As the number of cloud service providers increase, over time more resources need to be set up for usage with a balance between performance and utilization.

2. Meisner, D., Gold, B., Whenish, T. 'Powernap: Eliminating server idle power' ACM SIGPLAN Notices 2013, 44(3): 205-216 : However, in modern times the goal has completely changed. Clients and cloud providers themselves are trying to achieve a limited amount of resource wastage with maximum performance. Most use cases of the cloud require dynamic traffic and usage patterns, where allocation of resources optimally and efficiently becomes very grueling.

Additionally increasing the number of resources can often result in a network mesh with high complexity and a large number of nodes. Keeping this in mind customers on the other hand still wish for servers and resources that are up and running almost every hour of the day. This leads to large amounts of carbon dioxide emmision and hence adding to environmental pollution [D. Meisner et al.(2013)].

3. Singh, Harvinder & Bhasin, Anshu & Kaveri, Parag & Chavan, Vinay. (2020). Cloud Resource Management: Comparative Analysis and Research Issues. International Journal of Scientific & Technology Research. 9. 18: Another underlying issue is that non-IT professionals often have a hard time understanding resource management and optimization. Existing solutions are very limited in scope, they provide resource optimization over metrics such as CPU and memory utilization. However, there are more points to consider while predicting resource utilization such as availability of resources, how dynamic a particular cloud environment is, heterogeneity of the environment, and the geographical distance between the user and the cloud resource [Harvinder Singh et al. (2020)].

Q16.
The main point of contention with this problem is that it is a completely new industry. Although Cloud Computing has become relatively well known and well heard off in todays time. However, the industry is still growing rapidly and there are new changes every day. This raises the bar to complexity a little higher as making a solution of this sorts will need constant changes and meaningful features. We will constantly have to work on different versions of our single project as changes are recurring in cloud computing.
Another issue that hinders the success of this project is that in todays time there are more than a dozen cloud providers and every Cloud Service Provider has a very different architecture which means that there are different needs. Hence to tackle this issue we will try to work around with one cloud service provider at the first stage and then try to manoeuvre from that point.

Q17.
1. Ms. Nupur Tomar (tomar.nupur01@gmail.com)(Friend): Nupur Tomar is data scientist with an experience of using amazon web services. While talking to her about our idea, she entailed how genuine our problem was and the solution was deeply in need. She herself had encountered many situations where she was using resources in an over throughput manner, and wasn't aware of this until her task got over. She also added that while we were mainly focusing on large cloud providers, we should also try and work around a method for smaller Cloud Service Providers.

2. Mr. Baljit Singh (baljitsinghdlf@gmail.com)(Family): In a one to one interaction with Mr. Baljit Singh he told us that our idea was a great step forward in the Cloud Computing Sector. In fact, he was not aware of the problem before we described it to him and released how large of a market such a project can carry or cater too. He added that we should build a web interface for this project so that it reaches out to more and more people and is easy to use.

3. Dr. Indrajeet Gupta (indrajeet.gupta@bennett.edu.in)(Teacher): On an impromptu conversation with Dr. Indrajeet Gupta we had discussed our interest in cloud computing and how we wanted to work on project in that field. Which is when Dr. Indrajeet talked us through the real world problem of cloud resource management. He guided us as to how we should approach the problem and think on the right track.

Q18.
In the past Shrihan and I have worked on a total of 3 projects together. Below are the descriptions of each project we have done:
1. Predicting Bird Migration using Machine Learning and Python (1st Year): Developed a software using machine learning and python to predict bird migration. Our main objective would be to predict the location to which a certain species of birds are most likely to migrate and the time of the year at which they will migrate. We will observe and analyse any change that have occurred in the past few years. Conclusively, the predictions from this program will be displayed on a map.
Team: Arnav Malhotra, Jarvis Prem Raj, Shrihan Chinta, Adarsh Raghav

2. ProBono (2nd Year): An app which assists the court in prioritizing which cases to take up first in order of severity. This app allows lawyers to safeguard evidence and case files by digitizing them

and giving easy access to the lawyers and courtroom officials. In this way justice won't get delayed and there is a reduced risk of tampering with evidence. This application comprises of two portals, one is for the judge and the other is for the lawyers. So, this app basically arranges all the cases to be taken up on a day in the order of severity of the offence and displays the same on the judge portal. This app also gives the ability to the lawyers to store their evidence safe and secure in their own portal.
Team: Arnav Malhotra, Jarvis Prem Raj, Shrihan Chinta, Adarsh Raghav

3. One League (3rd Year): We plan on developing a platform to make college applications more easy and efficient for students. This will act as a one stop solution to all students applying for either an undergraduate or postgraduate education. We plan on connecting universities throughout India on one platform. With one username and password we want to give power in the hands of future students. We are planning to eliminate a constant need to fill the same details several times for different colleges. WebApp will also act as a guide for students to gather insights and knowledge about the universities they wish to apply to and the requirements that need to be fulfilled.
Team: Adarsh Raghav, Shrihan Chinta, Vishal Gembali

Q19.
Our project will follow the following guidelines:
1. As our web service deals with users giving out some personal information or company related information about resource usage, so we are going to keep our user data encrypted and secure.
2.We may collect few information from user such as data regarding the resource performance for recommending the b test VM/resources for them. Gathering of basic company related data such as name, function, clients may also be gathered to make a better assessment of their use cases and functioning.
3.Our web service will use cookies or software like cookies which would like to gather and secure user data as a whole but the user can disable the cookies option individually but this may limit the user to access some features of our web services such as sticky session where data may not be restored from previous access or use.
4.The main reason we collect user info is because we would like to be more interactive experience with user. Moreover, we want to provide our clients with the best overall experience which is only possible when we have the necessary information and data.

Q20.
To understand how our project has the potential to be a startup, it is important to understand the objectives of the project or in other words what we intend to achieve out of this project.
Giving businesses an idea of how much resources(virtual machines) are being wasted: Our first objective is to create a solution that will optimise resource management of cloud service providers as well as customers who use these resources.
Here we will be creating a fully automated solution for B2B, B2C: Since our main target audience are cloud service providers/ large business as well as consumers. Hence, there is potential for this project to be a startup and has commercial viability.
Help businesses lower their OPEX/CAPEX costs: Almost every business facing industry requires some sort of optimization, where they want to lower their cost. Which are project will help them achieve hence it is viable for startup.
The selling point of our project is that it is made by developers for customers. Optimization is the need of the hour and saving millions by properly utilising cloud resources is what businesses are looking for. The betting bid is that we are trying to reduce operational/variable costs of cloud computing users. Moreover, there are very limited players in this market and our product is of a particular niche and very unique. All together all of these points make our project a viable business/startup in the coming future.

Q21.
Cloud computing is considered to be an evergreen field. With respect to the Gartner Hype Cycle, cloud computing will be relevant for years to come from today. Hence, it only makes sense that optimisation will always be necessary for clients and providers to reduce their costs over time. The problem we our addressing is the need of the current hour, however, it is a something that will always be needed be it 5, 10 or even 15 years.

Moreover, we are using machine learning and deep learning techniques that are part of most advanced technologies in todays world. This means that the technology we are creating will still provide a high throughput even after years to come.