What is your experience with Hbase, Cassandra, Redshift, Lambda?

**Apache HBase** is an open source NoSQL database that provides real-time read/write access to those large datasets. A non-relational (NoSQL) database that runs on top of HDFS

HBase scales linearly to handle huge data sets with billions of rows and millions of columns, and it easily combines data sources that use a wide variety of different structures and schemas. HBase is natively integrated with Hadoop and works seamlessly alongside other data access engines through YARN.

[**Apache Cassandra**](http://cassandra.apache.org/), a top level Apache project [born at Facebook](http://en.wikipedia.org/wiki/Apache_Cassandra#History) and built on [Amazon’s Dynamo and Google’s BigTable](http://www.datastax.com/documentation/articles/cassandra/cassandrathenandnow.html), is a distributed database for managing large amounts of structured data across many commodity servers, while providing highly available service and no single point of failure. Apache Cassandra™ offers capabilities that relational databases and other NoSQL databases simply cannot match such as: continuous availability, linear scale performance, operational simplicity and easy data distribution across multiple data centers and cloud availability zones.

**Amazon Redshift** is a fully managed [petabyte](https://searchstorage.techtarget.com/definition/petabyte)-scale [data warehouse](https://searchdatamanagement.techtarget.com/definition/data-warehouse) service. Redshift is designed for analytic workloads and connects to standard SQL-based clients and business intelligence tools.

**AWS Lambda** lets you run code without provisioning or managing servers. You pay only for the compute time you consume - there is no charge when your code is not running.

With Lambda, you can run code for virtually any type of application or backend service - all with zero administration. Just upload your code and Lambda takes care of everything required to run and scale your code with high availability. You can set up your code to automatically trigger from other AWS services or call it directly from any web or mobile app.

Whats the difference between Rds and s3?

S3 provides standalone storage. We don't need to run an EC2 instance. we can also store lots and lots of data here.

RDS is a service that handles a MySQL database.  
RDS is a computer. EC2 is also a computer. EBS is a hard drive that can be accessed by EC2. S3 is a hard drive with no computer, so we can't do any processing.

What is a vpc?

Yes, I worked with Amazon vpc, Amazon Virtual Private Cloud enables customers to launch AWS resources into a virtual network that they have defined. This virtual network closely resembles a traditional network that you'd operate in customers’ data center, with the benefits of using the scalable infrastructure of AWS.

How do you do ML 

Im familiar with on premise machine learning deployment and do the machine learning on cloud.

have you used ML lib? Spark in general?

MLlib is Spark's machine learning library, focusing on learning algorithms and utilities, including classification, regression, clustering, collaborative filtering, dimensionality reduction, as well as underlying optimization primitives.

MLlib It is built on Apache Spark, which is a fast and general engine for large scale processing. Supposedly, running times or up to 100x faster than Hadoop MapReduce, or 10x faster on disk. Supports writing applications in Java, Scala, or Python.

how do you define a problem in deep learning and how do you solve it.

location? Willing to relocate?

I am now working in san Francisco california, ya I can relocate if necessary