Practice Interview Questions

# Data Science at Scale

* What is MapReduce? What kind of problems is it used for?
* Do you know what the advantages are of Spark over MapReduce?
* Can you explain the difference between batch processing and stream processing?
* Do you know what is a Spark Driver?
* What is a Spark Executor?
* Can you explain the difference between a Spark RDD and a Spark DataFrame?
* What programming languages does Spark support with APIs?
* What Machine Learning libraries do you know for Spark?
* What is sparkContext?
* What is the difference between a Transformer and an Estimator?
* What is the Pipeline API used for?
* What is a Fully Connected Layer?
* What is a pooling layer?
* What are Recurrent Neural Networks?
* What sort of improvement does an LSTM bring over RNNs?
* Can you explain the TensorFlow programming model?
* What types of activation functions do you know?
* Why do Deep Neural Networks train faster on GPUs?