Cloud Computing - MapReduce Interview Questions And Answers Guide.

Global Guideline.
http://www.globalguideline.com/
Cloud Computing - MapReduce Job Interview Preparation Guide.

Question # 1
Do you know how is MapReduce related to cloud computing?

Answer:-
The mapreduce framework contains most of the key architecture principles of cloud computing such as:
Scale: The framework is able to expand itself in direct proportion to the number of machines available.
Reliable: The framework is able to compensate for a lost node and restart the task on a different node.
Affordable: A user can start small and over time can add more hardware.
Due to the above features the mapreduce framework has become the platform of choice for the development of cloud applications.

Read More Answers.

Question # 2
How does fault tolerance work in mapreduce?

Answer:-
In a mapreduce job the master pings each worker periodically. In case a worker does not respond to that system then the system is marked as failed. Even completed tasks are rescheduled because the output was stored in a in a local disk of a worker which failed. Hence mapreduce is able to handle large-scale failures easily by simply restarting a task. The master node always saves itself at checkpoints and in case of any failure it simply restarts from that checkpoint.

Read More Answers.

Question # 3
Can you please explain in MapReduce what is a scarce system resource?

Answer:-
A scarce resource is one which is available in limited quantities for the system. In mapreduce the network band-with is a scarce resource. It is conserved by making use of local disks and memory in cluster to store data during tasks. The function uses the location of the input files into account and aims to schedule a task on a system which has the input files.

Read More Answers.

Question # 4
What are the various input and output types supported by MapReduce?

Answer:-
Mapreduce framework provides a user with many different output and input types.
Ex. Each line is a key/value pair. The key is the offset of the line from the beginning of the file and the value are contents of the line. It is up-to the will of the user. Also a user can add functionality at his will to support new input and output types.

Read More Answers.

Question # 5
What is task granularity?

Answer:-
In mapreduce the map phase if subdivided into M pieces and the reduce phase into R pieces. Each worker is assigned a group of tasks this improves dynamic load balancing and also speeds up the recovery of a worker in case of failures.

Read More Answers.

Question # 6
With the help of two examples name the map and reduce function purpose?

Answer:-
Distributed grep: A line is emitted by the map function if it matches a pattern. The reduce function is an identity function that copies supplied intermediate data for output.
Term-vector per host: In this the map function emits a hostname, vector pair for every document (input). The reduce function adds all the term vectors pairs generated and discards any infrequent terms.

Question # 7
Do you know the general mapreduce algorithm?

Answer:-
The mapreduce algorithm has 4 main phases:
1. Map,
2. Combine,
3. Shuttle and sort
4. Phase output
Mappers simply execute on unsorted key/values pairs. They create the intermediate keys. Once these keys are ready the combiners pair the key/value pairs with the right key. The shuffle/sort is done by the framework their role being to group data and transfer it. Once completed, it will proceed for the output via the phase output process.

Question # 8
Write a short note on the disadvantages of Mapreduce?

Answer:-
Some of the shortcomings of mapreduce are:
One input two-phase data flow is rigid i.e. it does not allow for multiple step processing of records.
Being based on a procedural programming model this framework requires code for simple operations.
The map and reduce functions being opaque does not allow for optimization easily.

Question # 9
What do you understand by MapReduce?

Answer:-
MapReduce is a software framework that was created by Google. It’s prime focus was to aid in distributed computing, specifically large sets of data on a group of many computers. The frameworks took its inspiration from the map and reduce functions from functional programming.

Question # 10
Tell me how mapreduce works?

Answer:-
The processing can occur on data which are in a file system (unstructured) or in a database (structured). The mapreduce framework primarily works on two steps:
1. Map step
2. Reduce step
Map step: During this step the master node accepts an input (problem) and splits it into smaller problems. Now the node distributes the small sub problems to the worker node so that they can solve the problem.
Reduce step: Once the sub problem is solved by the worker node, the node returns a solution to the master node which accepts all the solutions of the worker node and re-compiles them into a solution. This solution is for the input that was provided to the master node.

Question # 11
Tell me what is an input reader in reference to mapreduce?

Answer:-
The input reader as the name suggests primarily has two functions:
1. Reading the Input
2. Splitting it into sub-parts
The input reader accepts a user entered problem and then it divides/splits the problem into parts which then each are assigned a map function. Also an input reader will always read data from a stable storage source only to avoid problems.

Question # 12
Explain the purpose of the Partition function in mapreduce framework?

Answer:-
In mapreduce framework each map function generates key values. The partition function accepts these key values and in return provides the index for a reduce. Generally the key is hashed and a modulo is done to the number of reducers.

Question # 13
What is combiners?

Answer:-
Combiners codes are used to increase the efficiency of a mapreduce process. They basically help by reducing the amount of data that needs to be shifted across to reducers. As a safe practice the mapreduce jobs should never depend upon combiners execution.

**Question # 14**
Can you explain what you understand by speculative execution?

**Answer:**
Mapreduce works on the basis of large number of computers connected via a network also known as node. In a large network there is always a possibility that a system may not perform as quickly as others. This results in a task being delayed. By speculative execution this can be avoided as multiple instances of the same map are run on different systems.

**Question # 15**
When do reducers play their role in a mapreduce task?

**Answer:**
The reducers in a mapreduce job do not begin before all the map jobs are completed. Once all the map jobs are completed the reducers begin copying the intermediate key-value pairs from the mappers. Overall reducers start working as soon as the mappers are ready with key-value pairs.
Cloud Computing Most Popular Interview Topics.

1: Cloud Computing - Amazon Frequently Asked Interview Questions and Answers Guide.
2: Basic Cloud Computing Frequently Asked Interview Questions and Answers Guide.
4: Cloud Computing Frequently Asked Interview Questions and Answers Guide.
5: Ubuntu Cloud Frequently Asked Interview Questions and Answers Guide.
Global Guideline is a platform to develop your own skills with thousands of job interview questions and web tutorials for fresher's and experienced candidates. These interview questions and web tutorials will help you strengthen your technical skills, prepare for the interviews and quickly revise the concepts. Global Guideline invite you to unlock your potentials with thousands of Interview Questions with Answers or begin a tutorial right away, such as HTML, XML, XSLT, Cascading Style Sheet (CSS), Search Engine Optimization (SEO), JavaScript, Structure Query Language (SQL), Database Articles, Web Hosting Guide and much more. Learn the most common technologies Interview Questions and Answers. We will help you to explore the resources of the World Wide Web and develop your own skills from the basics to the advanced. Here you will learn anything quite easily and you will really enjoy while learning. Global Guideline will help you to become a professional and Expert, well prepared for the future.

* This PDF was generated from http://www.GlobalGuideline.com at January 29th, 2017

* If any answer or question is incorrect or inappropriate or you have correct answer or you found any problem in this document then don't hesitate feel free and e-mail us we will fix it.

You can follow us on FaceBook for latest Jobs, Updates and other interviews material. www.facebook.com/InterviewQuestionsAnswers

Follow us on Twitter for latest Jobs and interview preparation guides http://twitter.com/InterviewGuide

Best Of Luck.

Global Guideline Team
http://www.globalguideline.com
Support@globalguideline.com