The following revision questions are designed to help you study for the upcoming online open-book exam.

1. Introduction

Define what is meant by the term distributed system.
A distributed system typically satisfies certain criteria. What are these criteria?
Give three reasons for the growing importance of distributed systems.
Mutual exclusion is a challenge when designing a distributed system. Describe this challenge.
Replica management is a challenge when designing a distributed system. Describe this challenge.
What is a failure?
Name and describe three types of failures which can occur in a distributed system.
Synchronisation of processes is a challenge when designing a distributed system. Describe this challenge. Illustrate this challenge with an example.
What does it mean to say that a distributed system is size scalable?
What does it mean to say that a distributed system is geographical scalable?
What is middleware and what is the purpose of middleware?
What is the aim distribution transparency?
Name and describe three types of distribution transparency.
Describe the relationship between distributed and parallel systems?
Define what is meant by the term cluster computing.
What is a sensor network?
Define what is meant by the term cloud computing.
Cloud computing employs a measured service model. Define this model.
There are three main types of cloud computing. What are these?

2. MapReduce

Why is MapReduce important in the context of big data?
Why can MapReduce be considered a form of parallel computing?
Describe some properties of MapReduce.
Describe the MapReduce programming model.
Provide pseudocode for a MapReduce algorithm which computes the number of occurrences of each word in a large collection of documents.

3. Architectures

What is a software architecture?
What is a system architecture?
Describe the layered software architecture.
Describe the object-based software architecture.
Describe the Service-Oriented Architecture (SOA) which is a type of software architecture.
Describe the client-server system architecture.
What is meant by the terms thin and fat clients?
What does the network topology of a distributed system describe?

What is an overlay network?
What is the difference between a structured and unstructured topology?
Define what is meant by vertical and horizontal distribution in a distributed system.
Peer-to-peer systems support horizontal distribution. What does this statement mean?
Describe the peer-to-peer (P2P) system architecture.
Describe the difference between a structured peer-to-peer architecture and an unstructured peer-to-peer architecture?
In a peer-to-peer system, keys corresponding to data are often determined using a hash function. Explain how a hash function works.
In a structured peer-to-peer system the chord algorithm can be used to efficiently determine the location of data. Describe with the aid of an example how this algorithm works.
Describe with the aid of an example an algorithm for locating data in an unstructured peer-to-peer system.
Describe with the aid of an example how a flooding algorithm can be used to locate data in an unstructured peer-to-peer system.
Describe with the aid of an example how a random walk algorithm can be used to locate data in an unstructured peer-to-peer system.
Describe how a hierarchically organized peer-to-peer architecture can be used to efficiently determine the location of data.
Give an example of a hybrid system architecture?
How would one describe the system architecture used by BitTorrent. Justify your answer.

4. Processes

What is a process?
What is a thread?
What is the difference between a process and a thread?
Describe one advantage of using processes as opposed to threads.
Describe two advantages of using threads as opposed to processes.
Explain how a multithreaded server functions.
Explain what is meant by the term virtualization.
Why is cloud computing an important application of virtualization?
Virtualization plays a key role in Infrastructure as a Service (IaaS). Justify this statement.
What is code migration?
Give an example of how code migration allows distributed systems to be dynamically configured.
Describe two challenges of code migration.

5. Communication

Low-level communication facilities of computer networks do not offer distribution transparency. Justify this statement.
Explain how middleware can help to achieve distribution transparency.
Describe the difference between connection-oriented and connectionless communication.
Describe the difference between persistent and transient communication.
Describe the difference between asynchronous and synchronous communication.

What is the purpose of a Remote Procedure Call (RPC)?
Describe how a Remote Procedure Call (RPC) works?
Using a Remote Procedure Call (RPC) presents some challenges. Describe two such challenges.
What is the purpose of a Remote Method Invocation (RMI)?
Describe two reasons why one might use a Message Oriented Middleware (MOM) for communication rather than a Remote Procedure Call (RPC) or Remote Method Invocation (RMI).
What is the purpose of multicast communication in a distributed system?
With the aid of an example, describe how application-level tree-based multicasting works.

6. Web Services

List the three major roles in a web service environment.
A web service may be used to implement a Service-Oriented Architecture (SOA). What is an SOA?
Describe how a SOAP web service works.
What is the purpose of the Web Service Description Language (WSDL) in the context of a SOAP web service?

7. Coordination

Give a real world example of a problem that can occur in a distributed system which does not exhibit process synchronization.
What is clock drift in the context of a distributed system?
What is Universal Coordinated Time (UTC)?
How might one use Universal Coordinated Time (UTC) to achieve process synchronization in a distributed system?
If a single machine in a distributed system has a UTC receiver, how can the problem of achieving process synchronization be solved?
If no machine in a distributed system has a UTC receiver, how can the problem of achieving process synchronization be solved?
What is clock drift rate in the context of a distributed system?
How can process synchronization in a distributed system be achieved using the network time protocol?
How can process synchronization in a distributed system be achieved using the Berkeley algorithm?
Explain how a logical clock may be used to achieve process synchronization in a distributed system.
The Lamport logical clock is based on the concept of a happens-before relation. Describe this relation.
Lamport proposed an algorithm for assigning timestamps to events in a distributed system such that they satisfy the happens-before relation. Describe this algorithm.
Why may it be necessary to ensure the mutual exclusive access to resources by processes in a distributed system? Use a real world example to illustrate your answer.

Describe how mutual exclusive access to resources in a distributed system can be achieved using a centralized coordinator.
Describe how mutual exclusive access to resources in a distributed system can be achieved using a token-ring algorithm.
What is the purpose of an election algorithm in a distributed system?
With the aid of an example, describe how the bully election algorithm works.

8. Security

Security in distributed systems may be considered as consisting of two parts. Describe each of these parts.
What is communication security concerned with?
What is authorisation security concerned with?
Describe what is meant by a security incident and a security threat. In doing so describe the relationship between both concepts.
Describe three types of security threats.
What is a security policy? How may a security policy be enforced?
Describe three important security mechanisms in the context of distributed systems.
What do the terms plaintext and ciphertext refer to in the context of cryptography? How are these terms related?
A cryptosystem consists of three algorithms. What are these algorithms?
Describe how a symmetric (secret-key) cryptosystem works.
Describe how an asymmetric (public-key) cryptosystem works.
What is the main difference between a symmetric (secret-key) cryptosystem and an asymmetric (public-key) cryptosystem?
How may one send a confidential message using an asymmetric (public-key) cryptosystem?
How may one perform message authentication using an asymmetric (public-key) cryptosystem?
What is a digital signature?
Provide a real world application of digital signatures.
What is the purpose of access control?
What is the purpose of a reference monitor in the context of access control?
Explain how access control may be implemented using an access control matrix.
Describe the purpose of a firewall in a distributed system.
Explain why mobile code introduces a number of security threats.