

Systemes distribués

Lionel Souop

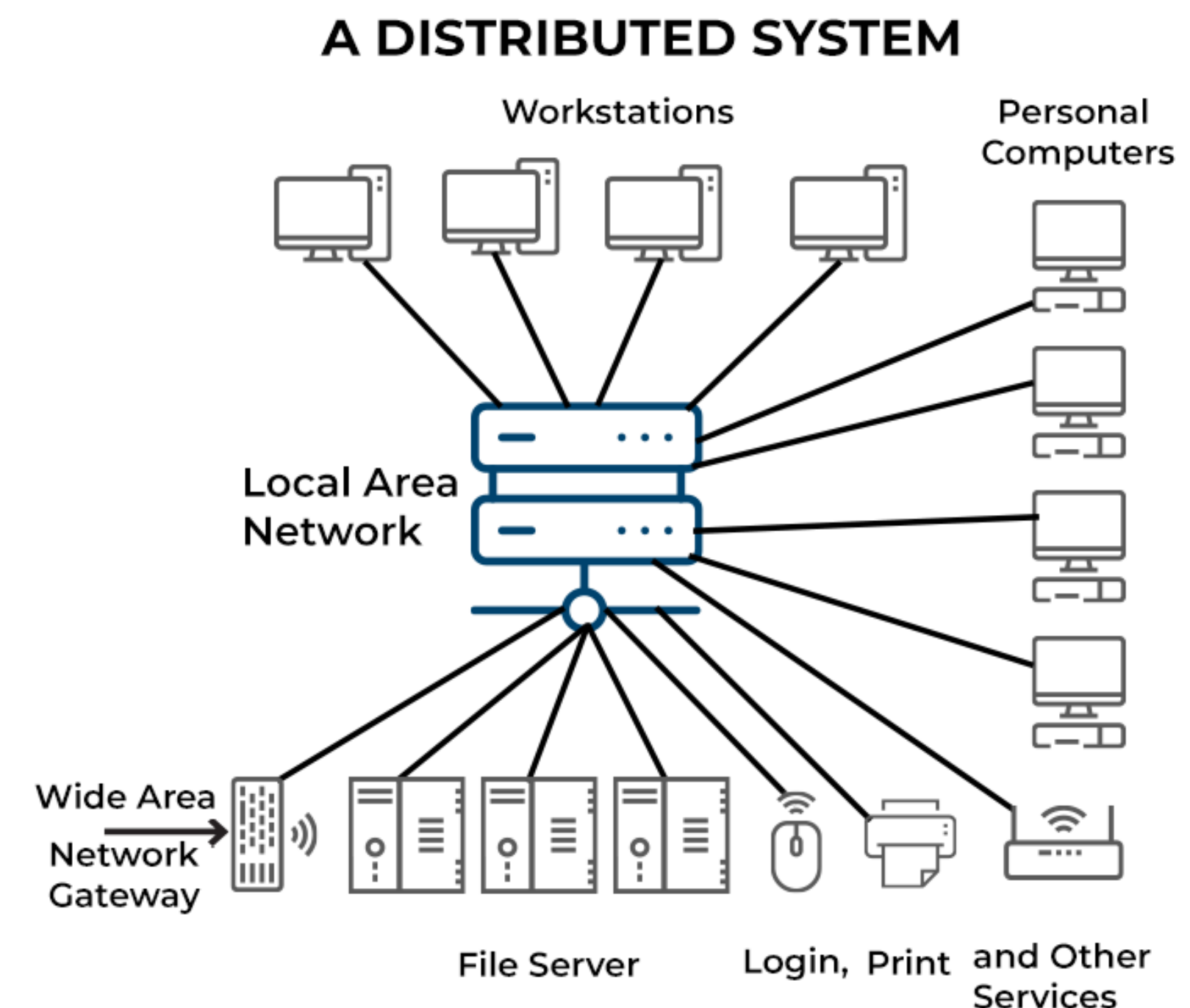
Systemes distribués

Introduction

Un système distribué est un ensemble de composants logiciels ou matériels interconnectés par un réseau, communiquant et coordonnant leurs actions pour atteindre un objectif commun



Local computation



Systemes distribués

Avantages

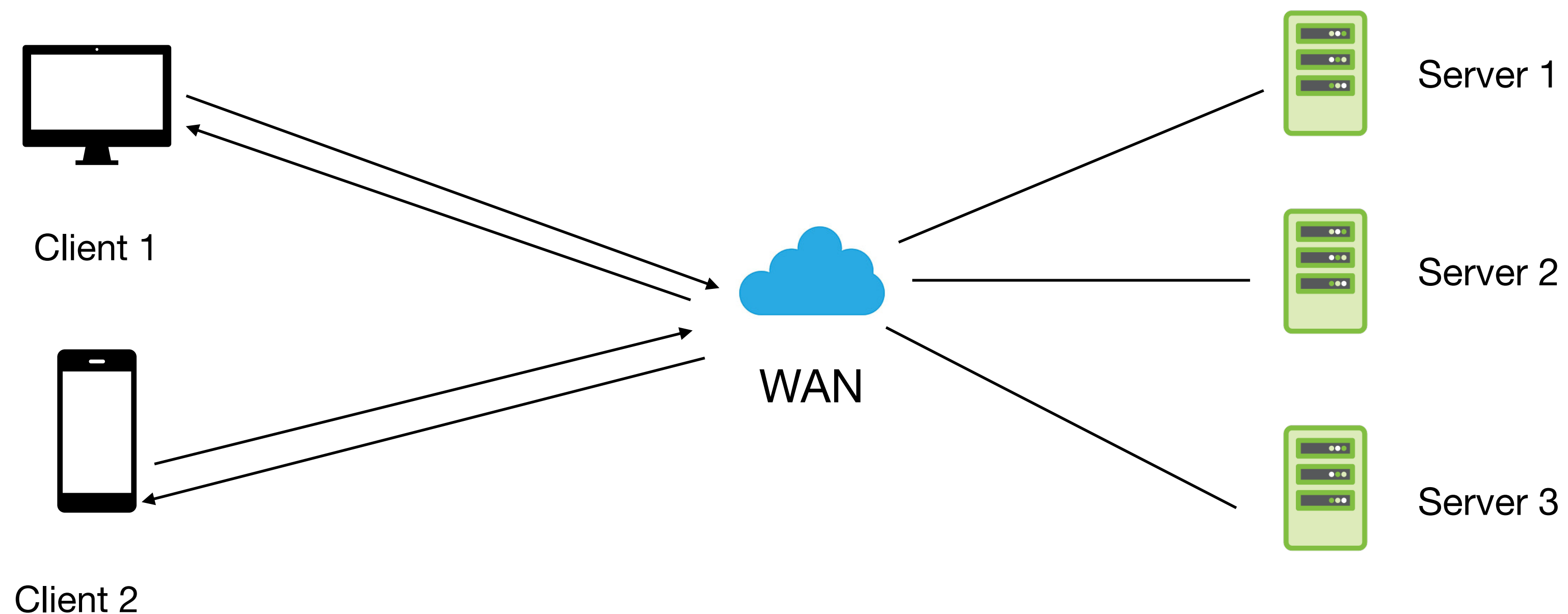
- Scalabilité
 - Capacité à augmenter les ressources et les performances de manière proportionnelle à la demande.
- Fiabilité et Tolérance aux pannes
 - Répartition de la charge et duplication des données pour assurer la disponibilité.
- Flexibilité
 - Capacité à intégrer de nouveaux composants et services facilement.

Architectures distribuées

Architectures distribuées

Client-serveur

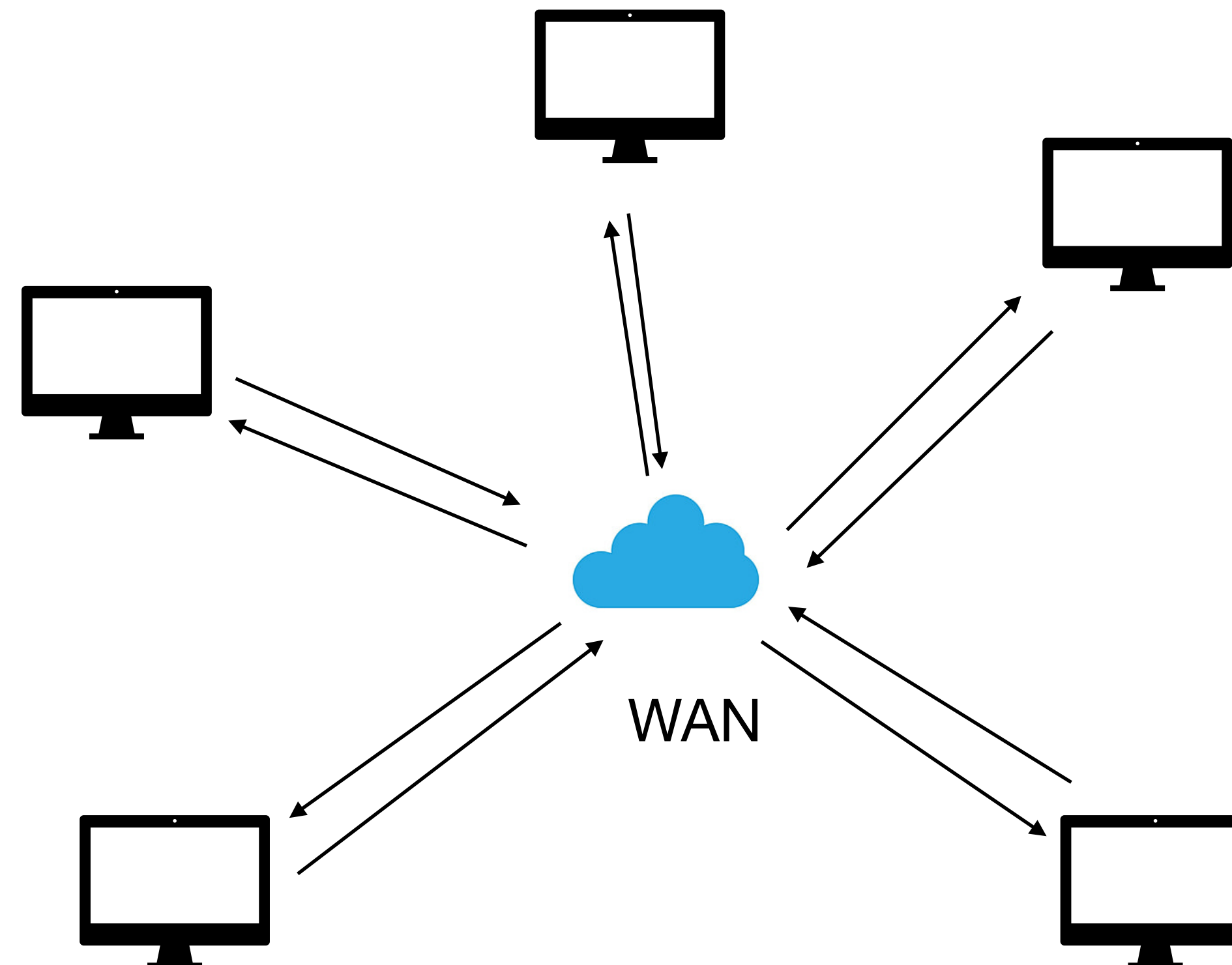
- C'est la forme la plus basique parmi les systèmes distribués
- Un ou plusieurs serveurs fournissent des services à un ou plusieurs clients



Architectures distribuées

P2P - Peer-to-peer

- Chaque nœud est à la fois client et serveur, partageant les ressources et les responsabilités de manière équitable.



Modèle de communication

Modèle de communication

RPC - Remote Procedure Call

RPC (Remote Procedure Call)

- Permet aux programmes d'exécuter des procédures sur des machines distantes.

Messaging

- Communication asynchrone via des messages

RESTful Services

- Communication basée sur HTTP et JSON, utilisée dans les architectures web modernes.

Synchronisation et coordination

Introduction

Locks et Mutex

- Utilisés pour gérer l'accès concurrent aux ressources partagées.

Élections de leader

- Algorithmes pour sélectionner un coordinateur parmi les nœuds (ex. Paxos, Raft).

Consensus

- Accord entre nœuds pour assurer la cohérence des données (ex. Zookeeper).
-

Consistence et Cohérence

Modèles de Consistance

- Forte consistance, éventuelle consistance.

CAP Theorem

- Concilier Consistency, Availability, Partition Tolerance (on ne peut en avoir que deux sur trois).

Bases de données distribuées

- Modèles et garanties de consistance (ex. BASE(basically available, soft state, and eventually consistent) vs ACID).

Tolérance aux pannes

Redondance

- Duplication des données et des services pour prévenir les pannes.

Repli et Reprise

- Techniques pour la reprise après panne (ex. Checkpointing, journaling).

Détection de pannes

- Utilisation de timeouts et d'algorithmes de heartbeat pour détecter les nœuds défaillants.

Sécurité

Authentication et Autorisation

- Contrôle d'accès aux ressources et services.

Cryptographie

- Chiffrement des données en transit et au repos.

Audit et Surveillance

- Journalisation des événements et détection des anomalies.

Exemples de systèmes distribués

Hadoop Distributed File System (HDFS)

- Système de fichiers distribué pour le stockage et le traitement des big data.

Apache Kafka

- Plateforme de streaming distribuée pour le traitement des flux de données.

Kubernetes

- Orchestrateur de conteneurs pour le déploiement et la gestion des applications distribuées.

Révision



1. Quelle sont les caractéristiques d'un système distribué?

- a) Il fonctionne sur une seule machine
- b) Les machines sont interconnectées par un réseau
- c) Il ne permet pas la communication entre les composants
- d) Il est résistant aux pannes

2. Quel est l'avantage principal de la scalabilité dans un système distribué?

- a) Réduction des coûts
- b) Diminution des performances
- c) Augmentation proportionnelle des ressources et des performances
- d) Complexité accrue du système

3. Dans une architecture client-serveur, que représente le serveur?

- a) Un client supplémentaire
- b) Un nœud qui demande des services
- c) Un nœud qui fournit des services
- d) Un nœud qui ne communique pas avec les clients

4. Quel modèle de communication utilise des appels de procédures distantes?

- a) Messaging
- b) RESTful Services
- c) RPC
- d) Peer-to-peer



5. Quelle est la fonction principale d'un lock ou d'un mutex dans un système distribué?

- a) Augmenter la bande passante
- b) Gérer l'accès concurrent aux ressources partagées
- c) Réduire le temps d'exécution
- d) Synchroniser les horloges des nœuds

6. Qu'est-ce que l'algorithme de Paxos dans un système distribué?

- a) Un protocole de transfert de fichiers
- b) Un algorithme de chiffrement
- c) Un algorithme d'élection de leader
- d) Un modèle de base de données

7. Le théorème CAP indique qu'un système distribué ne peut pas simultanément garantir:

- a) Cohérence, Disponibilité, Tolérance aux pannes
- b) Performances, Sécurité, Disponibilité
- c) Scalabilité, Fiabilité, Flexibilité
- d) Redondance, Sécurité, Cohérence

9. Quelle technique est utilisée pour reprendre l'exécution après une panne dans un système distribué?

- a) Compression des données
- b) Repli et Reprise
- c) Cryptographie
- d) Partitionnement des données



12. Que représente le terme "éventuelle consistance" dans un système distribué?

- a) Les données sont toujours consistantes
- b) Les données peuvent ne jamais être consistantes
- c) Les données finiront par être consistantes après un certain temps
- d) La consistance n'est pas une préoccupation

14. Dans une architecture P2P, chaque nœud:

- a) Est uniquement un serveur
- b) Est uniquement un client
- c) Est à la fois client et serveur
- d) N'interagit pas avec les autres nœuds

15. Quelle méthode permet de détecter une panne de nœud dans un système distribué?

- a) Chiffrement
- b) Élections de leader
- c) Algorithme de heartbeat
- d) Compression des données

16. Que signifie le terme "checkpointing" dans le contexte de la tolérance aux pannes?

- a) Suppression des anciennes données
- b) Sauvegarde de l'état actuel d'un système
- c) Migration des données vers un autre serveur
- d) Vérification de la cohérence des données



18. Dans un système distribué, que permet la duplication des données?

- a) Améliorer la sécurité
- b) Prévenir les pannes
- c) Réduire la charge réseau
- d) Augmenter la complexité des opérations