

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



جامعة بجاية
Tasdawit n Bgayet
Université de Béjaïa

Université Abderrahmane Mira de Bejaia
Faculté des sciences exactes
Département d'Informatique



Mémoire de fin d'études

En Vue De L'obtention du Diplôme

Master 2 professionnel en Informatique

Option : Administration et sécurité des réseaux informatiques

Thème

Mise en place d'une réplication de bases de données

Cas d'étude : Entreprise Portuaire de Bejaia

Réalisé par :

Mlle SALI Rabha

Mlle HADROUG Rosa

Président	Mme AIT ABDELOUHAB Karima	Maître de conf. B	U. A/Mira Béjaia.
Rapporteur	Mme EL BOUHISSI Houda	Maître de conf. A	U. A/Mira Béjaia.
Examineur	Mme OUYAHIA Samira	Maître de conf. B	U. A/Mira Béjaia.

Année universitaire : 2019/2020

Remerciements

Le grand remerciement pour notre Dieu le tout puissant et miséricordieux, Qui nous a donné le courage et la patience pour mener à bien ce travail.

On tient à remercier toutes les personnes qui ont contribué de manière directe ou indirecte à l'aboutissement de ce travail.

*En premier lieu, Nous exprimons notre gratitude à Madame **ELBOUHISSI Houda** notre encadreur, on apprécie son enthousiasme, sa gentillesse, ses conseils et ses orientations remarquables.*

On tient à remercier également tous les membres du jury qui ont accepté de juger ce travail.

*Merci à notre encadreur de stage monsieur **MAKHELOUFI Hichem**, et à tous les enseignants pour leurs efforts tout au long de nos années d'étude.*

Enfin, un grand Merci à tous nos proches qui nous ont apporté leurs soutiens durant l'élaboration de ce travail.

Liste des abréviations

BDD	Bases De Données.
CPU	Central Processing Unit
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
DSI	Direction des Systèmes d'Information.
EPB	Entreprise Portuaire de Bejaia.
GED	Gestion Electronique Documentaire
HTTP	Hypertext Transfer Protocol
LAN	Local Area Network
MySQL	My Structured Query Language.
OS	Operating System
PHP	Hypertext Preprocessor
SGBD	Système de Gestion de Bases de Données.
VM WAR	Virtual machine WAR.
WIMAX	Worlwide Interoperability for Microwave Access

Liste des figures

Figure 1 : L'organigramme de l'EPB.	3
Figure 2 : Missions du système d'information de l'EPB	6
Figure 3 : Organigramme de la direction des systèmes d'information.....	7
Figure 4 : Réseau fibre optique de l'EPB	8
Figure 5 : Architecture actuelle du réseau local de l'entreprise	9
Figure 6 : Réplication unidirectionnelle	22
Figure 7 : Réplication bidirectionnelle	23
Figure 8 : Réplication asymétrique synchrone	24
Figure 9 : Réplication symétrique synchrone.....	25
Figure 10 : Réplication asymétrique asynchrone.....	25
Figure 11 : réplication symétrique asynchrone.....	26
Figure 12 : Architecture de la solution	28
Figure 13 : Protocole de réplication MYSQL	32
Figure 14: Interface de Virtualbox.....	34
Figure 15: Interface de réseau local	34

Table des matières

Liste des abréviations

Liste des figures

Introduction générale.....	1
Chapitre 1 : Etude de l'existant	2
Introduction.....	2
1.1 Présentation de l'organisme d'accueil.....	2
1.2 Missions et activités de l'EPB.....	2
1.2.1 Les Missions	2
1.2.2 Les activités.....	2
1.3 Organisation de l'EPB.....	3
1.3.1 Direction Générale	3
1.3.2 Directions Générale Adjointe opérationnelles	4
1.3.3 Direction Générale Adjointe Fonctionnelles.....	5
1.4 L'infrastructure informatique	7
1.5 Le réseau local de l'EPB.....	8
1.5.1 Architecture du réseau local de l'entreprise	8
1.5.2 Etude de l'architecture	9
1.6 Le parc informatique de l'EPB.....	12
1.7 Problématique	12
1.7.1 Solution proposée	12
Conclusion	13
Chapitre 2 : Généralité sur les bases de données	14
Introduction.....	14
2.1 Base de données	14
2.1.1 Définition.....	14
2.1.2 Modèle de base de données.....	14
2.2 Le Système de Gestion de Base de Données	16
2.2.1 Définition.....	16
2.2.2 FONCTIONS DES SGBD	16
2.2.3 Architecture d'un SGBD	17

2.2.4	Objectifs des SGBD	18
2.2.5	Les Avantages d'un SGBD	19
2.2.6	Caractéristiques de la conception de BDD en SGBD.....	19
2.2.7	Méthode d'accès aux données	19
2.2.8	Les SGBD plus connus et utilisé	20
2.3	Réplication	20
2.3.1	Réplication de Base de Données.....	20
2.3.2	Principe de la réplication	21
2.3.3	Modes de Réplication	22
2.3.4	Méthode de propagation.....	23
2.3.5	Les différents types de Réplication.....	24
2.3.6	Objectifs de la Réplication de Base de Données.....	26
Conclusion		27
Chapitre 3 : Réalisation		28
Introduction.....		28
3.1	L'architecture de la solution proposée	28
3.2	Présentation des outils utilisés	28
3.2.1	La virtualisation.....	28
3.2.2	Son fonctionnement	29
3.2.3	La virtualisation des serveurs.....	29
3.2.4	L'hyperviseur.....	30
3.2.5	Les machines virtuelles	30
3.3	Les principaux logiciels de virtualisation	31
3.3.1	Logiciel choisi	31
3.4	Système de gestion de bases de données	32
3.5	Linux.....	33
3.6	Mise en place de l'architecture	33
3.7	Présentation de processus de mise en place de la réplication (maitre-maitre)....	35
3.7.1	Installation de MYSQL server	35
3.8	Les étapes de la création de la réplication	36
3.9	Les étapes de vérification du fonctionnement de la Réplication	38
Conclusion		39

Conclusion générale	40
Bibliographie.....	41

Introduction générale

L'arrivée des nouvelles technologies de l'information au sein des entreprises a révolutionné les manières de travailler.

De nos jours, il est primordial dans une entreprise de s'assurer qu'aucune donnée ne sera perdue. Il est donc important de faire des sauvegardes régulières et de mettre en place tous les mécanismes qui participeront à minimiser le risque de perte.

La réplication est une puissante fonctionnalité des SGBD qui permet de distribuer les données et d'exécuter les procédures stockées sur plusieurs serveurs de l'entreprise. La technologie de réplication a considérablement évolué et permet maintenant de copier, déplacer les données à différents endroits et de synchroniser automatiquement les données.

C'est dans cette optique que nous nous sommes vus attribués durant notre stage à la Direction des systèmes d'informations de l'entreprise portuaire de Bejaia, le sujet intitulé « la mise en œuvre d'une réplication de base de données ».

Notre projet aura pour objectif de réaliser une réplication de base de données Maître-Maitre qui visera essentiellement à éliminer le problème de perte de données et améliorer la tolérance aux pannes.

Pour notre travail nous avons opté pour MySQL comme système de gestion des bases de données, linux comme système d'exploitation et deux machines virtuelles.

Le premier chapitre consiste à donner une vue générale sur la structure de l'Entreprise Portuaire De Bejaia en illustrant ses multiples problèmes et la solution envisagée.

Le deuxième chapitre présente les généralités sur les bases de données et les réplications.

Le troisième chapitre fera l'objet de la réalisation de la réplication, nous présentons les outils utilisés ainsi que le processus de mise en place.

En conclusion, nous finirons par dresser quelques perspectives qui peuvent être envisagées afin d'améliorer notre solution.

1 Chapitre 1 : Etude de l'existant

Introduction

L'étude de l'existant est la phase initiale de notre travail, elle consiste à mettre à plat d'une façon aussi claire que possible l'analyse qualitative et quantitative du fonctionnement actuelle de l'entreprise.

Dans le présent chapitre, nous allons présenter un aperçu général de l'entreprise portuaire de Bejaia, et cela en mettant l'accent sur la direction des systèmes d'information (DSI) afin de repérer ses insuffisances et extraire les défaillances de cette dernière pour pouvoir poser d'éventuelles solutions.

1.1 Présentation de l'organisme d'accueil

Le port de Bejaia est un port algérien situé dans la région de Kabylie dans le nord du pays. Il est notamment consacré au commerce international et aux hydrocarbures. Il joue un rôle très important dans les transactions internationales vu sa situation géographique.

Son positionnement au cœur de la méditerranée occidentale et au centre de la côte algérienne lui confère d'importants atouts sur le plan économique et une place privilégiée sur les routes maritimes où il a su y capter un important trafic.

1.2 Missions et activités de l'EPB

1.2.1 Les Missions

La gestion, l'exploitation et le développement du domaine portuaire sont les charges essentielles de la gestion de l'EPB, c'est dans le but de promouvoir les échanges extérieurs du pays. Elle se doit assumer la police et la sécurité au sein du pays. Elle est chargée des travaux d'entretien, d'aménagement, de renouvellement et de création d'infrastructures. L'EPB assure également des prestations à caractère commercial, à savoir ; le remorquage, la manutention et l'aconage.

1.2.2 Les activités

Les principales activités de l'entreprise sont :

- L'exploitation de l'outillage et des installations portuaires.
- L'exécution des travaux d'entretien, d'aménagement et de renouvellement de la super structure portuaire.
- L'exercice du monopole des opérations d'aconage et de manutention portuaire.
- L'exercice du monopole des opérations de remorquage, de pilotage et d'amarrage.
- La police et la sécurité portuaire dans la limite géographique du domaine public portuaire.

1.3 Organisation de l'EPB

La structure de l'EPB est dotée :

- D'une direction générale ;
- De 04 directions fonctionnelles ;
- De 07 directions opérationnelles ;

Son organigramme peut être schématisé ainsi :

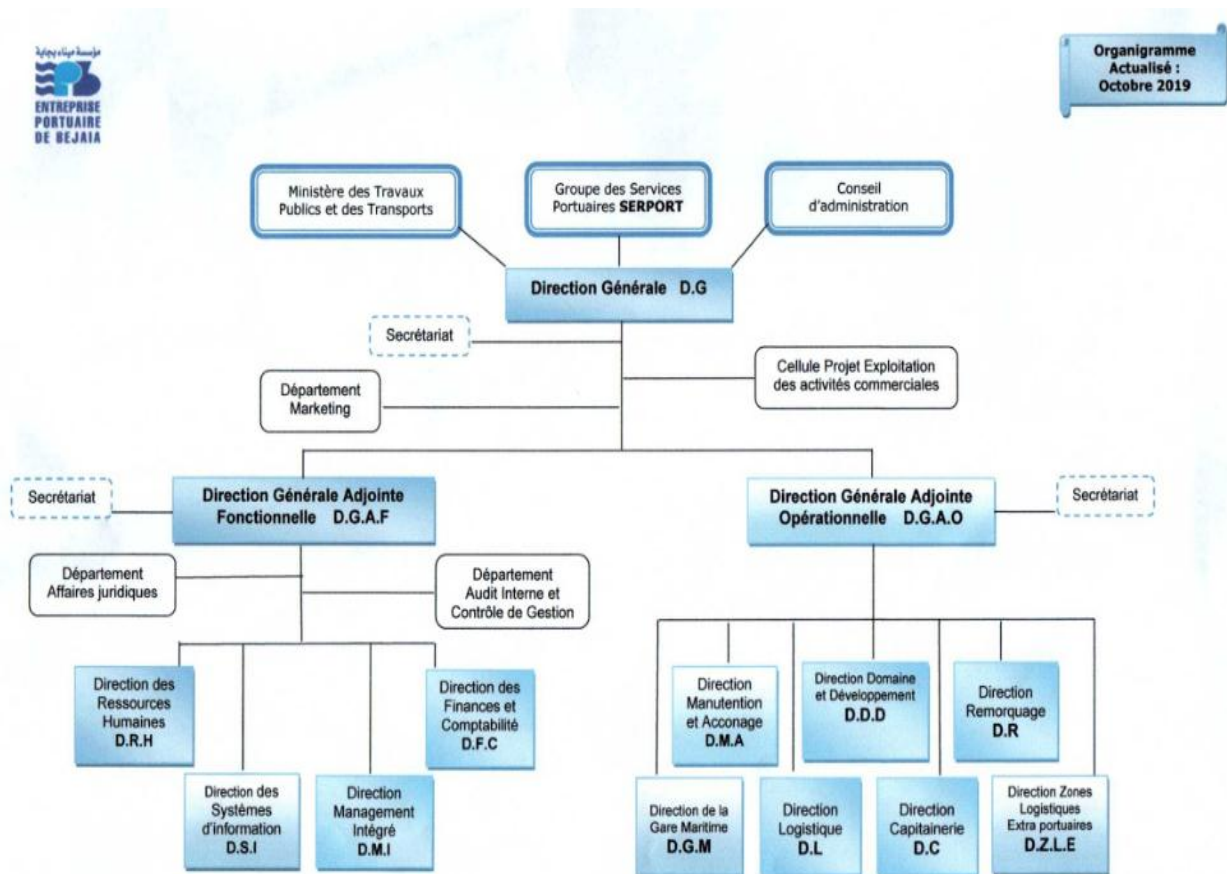


Figure 1 : L'organigramme de l'EPB.

1.3.1 Direction Générale

Elle est chargée de concevoir, coordonner et contrôler les actions liées à la gestion et au développement de l'entreprise. L'EPB est organisée selon des directions fonctionnelles et opérationnelles.

1.3.2 Directions Générale Adjointe opérationnelles

Il s'agit des structures qui prennent en charge les activités sur le terrain et qui ont une relation directe avec les clients.

1.3.2.1 Direction MANUTENTION et ACCONAGE

Elle est chargée de prévoir, organiser, coordonner et contrôler l'ensemble des actions de manutention et d'aconage liées à l'exploitation du port.

1.3.2.2 Direction Domaine et Développement

Qui a pour tâches :

- Amodiation et location de terre-pleins, hangar, bureaux, immeubles, installation et terrains à usage industriel ou commercial.
- Enlèvement des déchets des navires et assainissement des postes à quai.
- Pesage des marchandises (pont bascule).
- Avitaillement des navires en eau potable.

1.3.2.3 Direction Capitainerie

La Capitainerie du Port est chargée de la sécurité portuaire, ainsi que de la bonne régulation des mouvements des navires et la garantie de sauvegarde des ouvrages portuaires. Elle assure également les activités suivantes :

LE PILOTAGE, LE LAMANAGE, L'ACCOSTAGE et LE *REMORQUAGE*

1.3.2.4 Direction Remorquage

Elle est chargée d'assister le pilote du navire lors de son entrée et de sa sortie du quai.

1.3.2.5 Direction Logistique

Elle consiste à gérer tout ce qui concerne le transport et le stockage des produits de l'entreprise : véhicules nécessaires au transport, fournisseurs de l'entreprise, entrepôts, manutention, en optimisant leur exploitation pour minimiser les coûts et les délais.

1.3.2.6 Direction Zone Logistique Extra-Portuaire

Elle a pour objet de gérer les flux physiques, informationnels et financiers d'une organisation, dans le but de mettre à disposition les ressources correspondantes aux besoins, et ce, aux conditions économiques et pour une qualité de service déterminées, dans des conditions de sécurité et de sûreté satisfaisantes.

1.3.2.7 Direction de la Gare maritime

Dans l'objectif de développer le trafic passager, une nouvelle infrastructure spécialement dédiée pour l'accueil et le transit des passagers dans les meilleures conditions, a été conçue.

Ce nouveau terminal à passagers répond aux normes internationales d'accueil et de transit des voyageurs et des véhicules et s'inscrit dans l'optique de modernisation des infrastructures pour offrir les meilleures prestations aux usagers.

1.3.3 Direction Générale Adjointe Fonctionnelles

Il s'agit des structures de soutien aux structures opérationnelles.

1.3.3.1 Direction Du Management Intégré

Elle est chargée de :

- La mise en œuvre, le maintien et l'amélioration continue du Système de Management Intégré (plans projets et indicateurs de mesures).
- L'animation et la coordination de toutes les activités des structures dans le domaine QHSE (Qualité, Hygiène, Sécurité, Environnement).
- La Contribution dans des actions de sensibilisation et de formation à la prévention des risques de pollution, à la protection de l'environnement, la santé des travailleurs et à l'intervention d'urgence.

1.3.3.2 Direction Finance et Comptabilité

Elle est chargée de :

- La tenue de la comptabilité, La gestion de la trésorerie (dépense, recettes et placements), La tenue des inventaires et Le contrôle de gestion (comptabilité analytique et contrôle budgétaire).

1.3.3.3 Direction Ressources Humaines

Elle est chargée de prévoir, d'organiser et d'exécuter toutes les actions liées à la gestion des ressources humaines en veillant à l'application rigoureuse des lois et règlements sociaux.

1.3.3.4 Direction des systèmes d'information

1.3.3.4.1 Présentation de la DSI

Le système d'information (SI) est un ensemble organisé de ressources qui permet de collecter, stocker, traiter et distribuer de l'information. La DSI est une direction de l'EPB

rattachée directement à la direction générale, elle a pour mission l'automatisation des métiers de l'entreprise portuaire de Bejaïa, et cela en mettant en place les logiciels et l'infrastructure nécessaire pour la gestion du système d'information.

1.3.3.4.2 Missions

Durant le stage effectué au niveau de la direction des systèmes d'information nous avons pu identifier les missions du système d'informations comme la montre la figure ci-dessous :

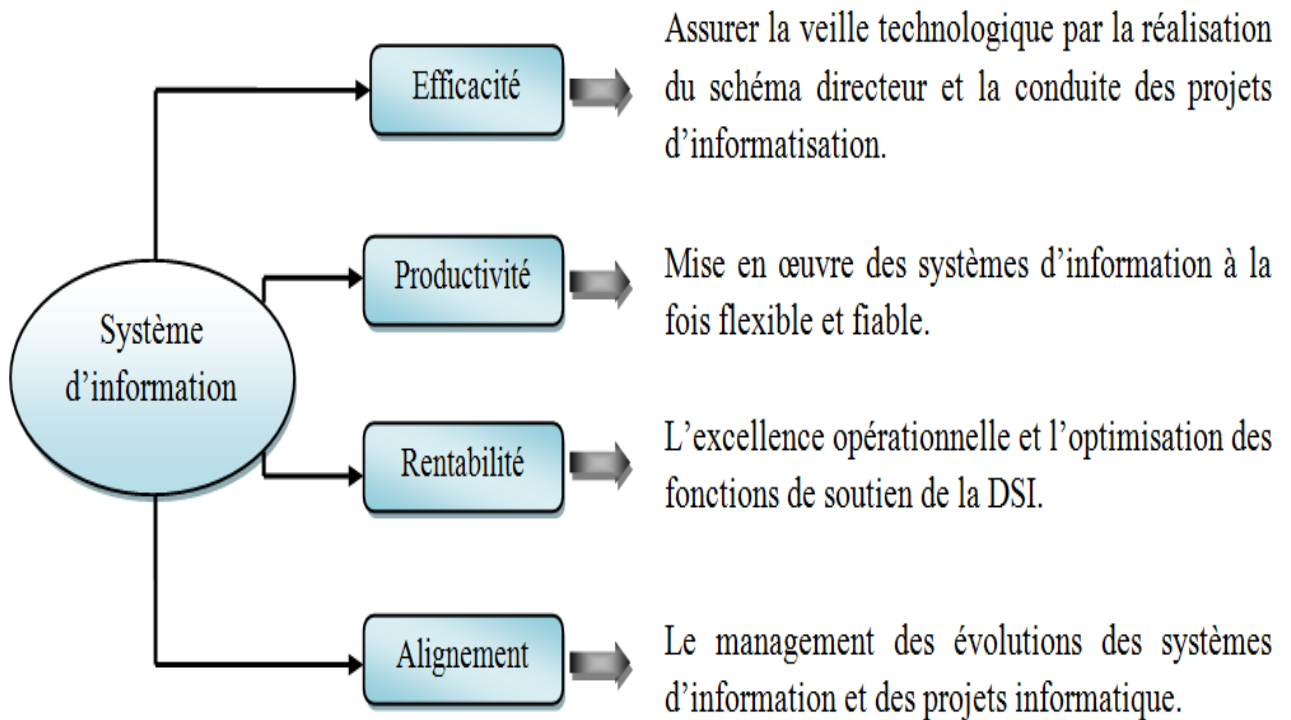


Figure 2 : Missions du système d'information de l'EPB

1.3.3.4.3 Organisation humaine de la direction des systèmes d'information

La direction se compose de trois départements comme le montre l'organigramme suivant :

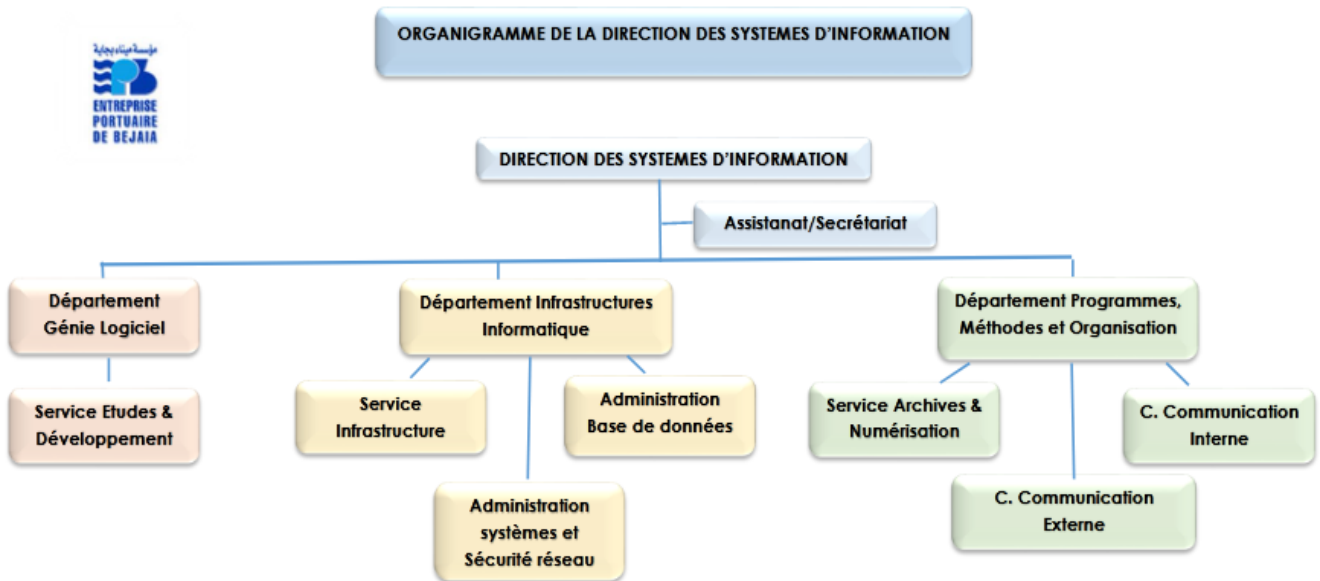


Figure 3 : Organigramme de la direction des systèmes d'information

1.4 L'infrastructure informatique

Le réseau du port de Bejaia s'étend du port pétrolier (n°16) aux ports 13 et 18 (port à bois). La salle machine du réseau local de l'EPB contient principalement une armoire de brassage et une autre armoire optique de grande taille, ces deux armoires servent à relier les différents sites de l'entreprise avec le département informatique par fibres optiques de type 4, et 12 brins, comme l'illustre la figure (Fig. 4).

Chaque site a une armoire de brassage contenant un/des convertisseur(s) media, un/plusieurs Switch où sont reliés les différents équipements par des câbles informatiques.

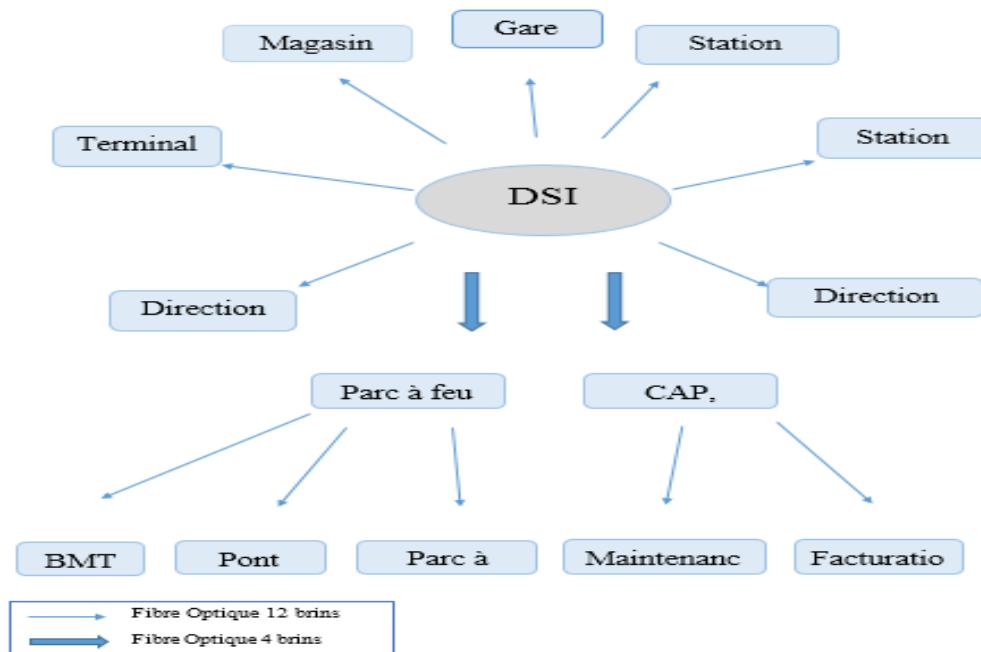


Figure 4 : Réseau fibre optique de l'EPB

1.5 Le réseau local de l'EPB

Le réseau local de l'EPB permet aux différents postes de travail d'échanger des informations, de se connecter vers l'extérieur et d'utiliser des applications hébergées en interne nécessaire à l'exécution des tâches quotidiennes des employés. Le réseau du port de Bejaïa s'étend du port pétrolier (N16) aux ports 13 et 18 (port à bois).

1.5.1 Architecture du réseau local de l'entreprise

L'architecture du réseau LAN de l'entreprise est représentée dans la figure ci-dessous :

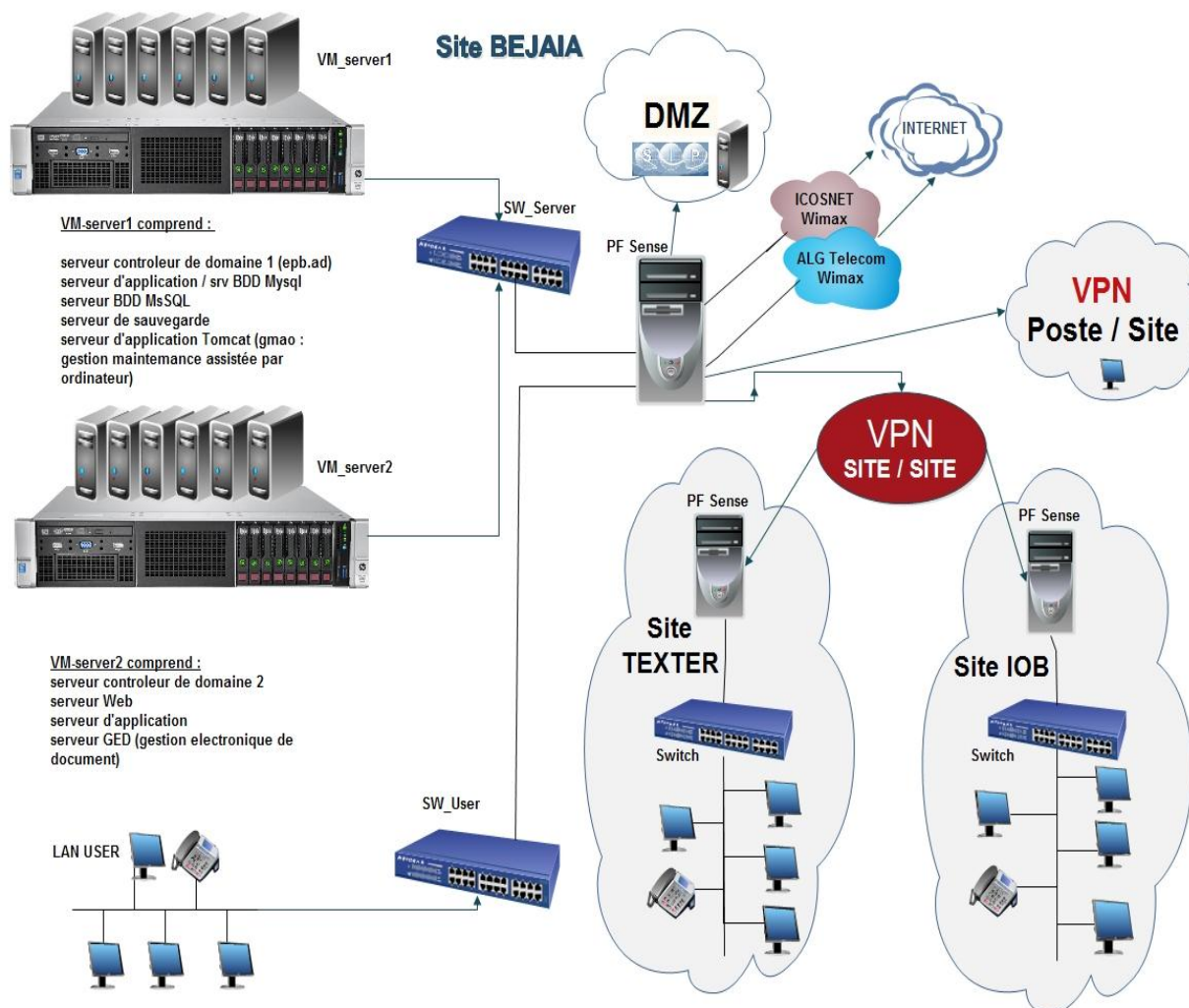


Figure 5 : Architecture actuelle du réseau local de l'entreprise

1.5.2 Etude de l'architecture

1.5.2.1 Connexion Internet

L'entreprise portuaire de Bejaia est dotée de deux connexions internet, Icosnet et Algérie télécom.

En utilisant l'accès haut débit sans fil WIMAX (Worldwide Interopérabilité for Microwave Access) qui est une solution alternative pour le déploiement des réseaux haut débit, l'entreprise se connecte à Internet sans-fil à partir d'un poste fixe qui communique par ondes hertziennes via une antenne-relais appelée station de base.

1.5.2.2 Sécurité

La sécurité est assurée par un pare-feu pour appliquer les stratégies d'accès et les règles de routages déterminant la manière dont les clients accèdent à Internet.

- **Pfsense (pare- feu) :** La sécurité est assurée par deux serveurs virtuel pare-feu qui agissent comme un filtre afin de définir les règles d'accès à un réseau comme Internet à cause des risques que peut représenter une connexion normale dans certains cas.
- **DMZ :** une zone démilitarisée qui consiste à créer une « zone » de telle sorte que l'échange entre le réseau interne et le réseau externe transite par cette dernière, dont l'objectif principale et que tous les échanges passent par cette zone qui offre les différents services de sécurité tels le filtrage réseaux entre les différents réseaux ainsi interconnecter mais aussi des serveurs relais dans la DMZ pour gérer tous le trafic interne et externe.
- **VPN (Virtuel Private Network) :** Pour répondre aux besoins d'interconnexion de l'entreprise portuaire de Bejaia aux différents sites distants (BORDJ BOU ARRERIDJ et IGHIL OUBEROUAK TALA HAMZA).

1.5.2.3 Salle machine

Les activités du port se rattachent à cette salle, qui rassemble tous les serveurs nécessaires pour répondre aux attentes des utilisateurs afin d'aboutir à un réseau qui accompagne la croissance de l'entreprise. En plus des switches elle comporte une armoire de brassage qui contient des différents serveurs tels que :

- **Serveur de base de données (SQL server and MARIA DB) :** un serveur de base de données répond à des demandes de manipulation de données stockées dans une ou plusieurs bases de données. Il s'agit de demande de recherche, de tri, d'ajout, de modification ou de suppression de données. Ces données sont utilisées par des serveurs web et des utilisateurs.
- **Serveur de contrôleur de domaine DC1 (Active Directory) :** sous Windows Server 2012 R2 l'objectif principal d'Active Directory est de fournir des services centralisés d'identification et d'authentification à un réseau d'ordinateur utilisant le système Windows. Il répertorie les éléments de ce réseau administré tel que les comptes des utilisateurs, les serveurs, les postes de travail, les dossiers partagés, les imprimantes etc.
- **Serveur de contrôleur de domaine redondant DC2 (Active Directory) :** il permet de conserver des répliques de données de l'annuaire sur un autre contrôleur de domaine, cela garantie la disponibilité et la continuité.
- **Serveur de sauvegarde de données :** il a pour rôle de sauvegarder en continue les données générées par l'entreprise. Si un employé efface par erreur un document

ou qu'il y a un dysfonctionnement d'un ordinateur, il est possible de rétablir le premier fichier.

- **Serveur de messagerie électronique (Exchange)** : est un serveur qui est connecté à internet, il permet à ses utilisateurs d'envoyer et de recevoir des courriers électroniques. Pour se connecter au serveur de messagerie, l'utilisateur a recours à un logiciel client, capable de gérer l'adressage (ou envoi) du courriel mais aussi sa réception tel que Outlook de Ms office, ou via navigateur web (Web mail).
- **Serveur application/fichier** : C'est un serveur sur lequel sont installées les applications utilisées par les usagers. Ces applications sont chargées sur le serveur d'application pour y accéder à distance. Un serveur d'application peut être un serveur qui centralise toutes les applications utilisées par les postes clients.
- **Serveur antivirus (Eset)** : est un outil de gestion centralisé de protection contre les programmes malveillants dans l'entreprise et d'administration de l'antivirus chez les utilisateurs.
- **Serveur DNS : (Domain Name System)** : est le service de résolution de nom d'hôte, il permet d'associer un nom d'une machine à un adresse IP.
- **Serveur DHCP** : pour un adressage dynamique, afin d'avoir une configuration fiable des paramètres de connexion et d'assurer la réduction de gestion de configuration.
- **Serveur GED (Gestion Electronique de Documents)** : Le terme GED désigne les logiciels permettant la gestion de contenus documentaires, qui est un point important pour l'entreprise EPB permettant l'organisation ainsi que la gestion des informations et des documents électroniques au sein de cette dernière. En effet il fonctionne sur un serveur en mode Intranet, Extranet, il est simple à installer et à manipuler. Pour accéder à la GED on utilise un navigateur web (Firefox, Chrome, Internet Explorer).
- **Serveur web** : est spécifiquement un serveur multi-service utilisé pour publier des sites web sur Internet ou un intranet. L'expression « serveur Web » désigne également le logiciel utilisé sur le serveur pour exécuter les requêtes HTTP, le protocole de communication employé sur le World Wide Web.
- **Serveur d'impression** : est un serveur qui permet de partager une ou plusieurs imprimantes entre plusieurs utilisateurs (ou ordinateurs) situés sur un même réseau informatique ainsi la centralisation de l'administration.
- **La Technologie de stockage avec les RAID : (Redundant Array of Independent Disk)** pour une meilleure gestion de stockage et de disponibilité des données

l'entreprise EPB s'est dotée de la technologie RAID et plus précisément le RAID de niveau 5.

1.6 Le parc informatique de l'EPB

L'EPB dispose de 250 PC HP et ACER répartis à travers les différentes directions de l'entreprise et interconnecté à un réseau informatique interconnecté par fibre optique et de câbles à paires torsadés.

Les systèmes d'exploitation utilisés sur les postes de travail sont Windows et Linux sous différentes distributions.

La majorité des PC est reliée à des imprimantes de plusieurs types (matricielle, laser et à jet d'encre couleur).

- Chaque ordinateur est branché à un onduleur APC de 400 à 1000 VA.
- Tous les PC sont dotés d'un anti-virus ESET END point.
- Tous les PC sont connectés à l'internet.

1.7 Problématique

Comme dans chaque entreprise, la direction des systèmes d'information de l'entreprise portuaire de Bejaia est contrainte à augmenter et assurer la haute disponibilité des données.

Au bout de la période de stage effectué au niveau de cette dernière nous avons remarqué que le seul moyen utilisé pour assurer la disponibilité des données en cas de pannes est la sauvegarde des données qui se fait manuellement chaque mois.

Cette méthode a engendré des problèmes majeurs lors de la manipulation de leurs bases de données comme le non accès à la dernière mise à jour, la perte de données et la non tolérance aux pannes.

1.7.1 Solution proposée

L'une des solutions proposées pour gérer ces problèmes est la réplication des données. Ce mécanisme est considéré comme un outil très puissant dans la gestion des données pour donner plus de performance aux systèmes et une haute disponibilité des données.

Notre solution consiste à Mettre en place une réplication de base de données en temps réel entre deux serveurs en mode bidirectionnelles (symétrique), Plus précisément un autre serveur peut prendre la relève lorsque le serveur principal s'écroule sans interrompre le travail.

Conclusion

Ce chapitre nous a permis de présenter le fonctionnement de l'entreprise portuaire de Bejaia d'une manière générale afin de proposer une solution optimale qui convient aux besoins du service.

Cette étape sera le point de départ de l'étude détaillée que nous présenterons dans le prochain chapitre.

2 Chapitre 2 : Généralité sur les bases de données

Introduction

Les bases de données et la réplication des données sont reconnues aujourd'hui comme étant des moyens efficaces et essentiels dans le monde de l'informatique, plus particulièrement en gestion de données.

Depuis toujours les chercheurs ont essayé de donner plus de performance à leurs systèmes, les problèmes de gestion des données sont devenus de plus en plus compliqués, surtout en ce qui, la disponibilité des données partagées.

L'une des solutions proposées pour gérer ce problème est la réplication des données, elle offre aux utilisateurs de meilleures performances et une plus grande disponibilité des données.

Dans cette partie de notre recherche, nous allons passer en revue les concepts liés aux bases de données, les systèmes de bases de données (SGBD), ainsi les réplifications.

2.1 Base de données

2.1.1 Définition

Une base de données est un ensemble de données modélisant les objets d'une partie du monde réel et servant de support à une application informatique et pouvant être interrogées et mises à jour par une communauté d'utilisateurs. [1]

2.1.2 Modèle de base de données

Un modèle de base de données illustre la structure logique d'une base de données, y compris les relations et les contraintes qui déterminent comment les données peuvent être stockées et accessibles. [1]

2.1.2.1 Modèle hiérarchique

Le modèle hiérarchique organise les données dans une structure arborescente, où chaque enregistrement dispose d'un seul parent (racine). Les enregistrements frères et sœurs sont triés dans un ordre particulier. Cet ordre est suivi pour le stockage physique de la base de données. Ce modèle convient à la description de nombreuses relations du monde réel. [1]

2.1.2.2 Modèle réseau

Le modèle réseau est une extension du modèle hiérarchique qui autorise des relations plusieurs-à-plusieurs entre des enregistrements liés, ce qui implique plusieurs enregistrements parents. Basé sur la théorie mathématique des ensembles, ce modèle s'articule autour d'ensembles d'enregistrements connexes.

Chaque ensemble se compose d'un enregistrement propriétaire (ou enregistrement parent) et d'un ou plusieurs enregistrements membres (ou enfants).

Un enregistrement peut être un membre ou un enfant dans plusieurs ensembles, ce qui permet à ce modèle de traduire des relations complexes. [1]

2.1.2.3 Modèle relationnel

Le modèle le plus courant, appelé modèle relationnel, trie les données dans des tables, que l'on appelle aussi des relations, dont chacune se compose de colonnes et de lignes. Chaque colonne contient un attribut de l'entité en question, comme le prix, le code postal ou la date de naissance. L'ensemble des attributs d'une relation est appelé domaine. La clé primaire est constituée par un attribut spécifique ou une combinaison d'attributs. On peut y faire référence dans d'autres tables : elle est alors appelée clé étrangère.

Chaque ligne, également appelée tuple, comprend des données sur une instance spécifique de l'entité en question, comme un employé en particulier. Le modèle tient également compte des types de relations entre ces tables, notamment les relations un-à-un, un-à-plusieurs et plusieurs-à-plusieurs. [1]

2.1.2.4 Modèle orienté objet

Ce modèle définit une base de données comme une collection d'objets, ou d'éléments logiciels réutilisables, associés à des caractéristiques et des méthodes. Il existe plusieurs types de bases de données orientées objet :

Une base de données multimédia réunit des médias, tels que des images, qui ne peuvent pas être stockés dans une base de données relationnelle.

Une base de données hypertextuelle permet à n'importe quel objet d'être relié à un autre. Elle est utile pour organiser de nombreuses données disparates, mais peu adaptée à l'analyse numérique. [1]

2.1.2.5 Modèle relationnel-objet

Ce modèle de base de données hybride associe la simplicité du modèle relationnel à certaines des fonctionnalités avancées du modèle de la base de données orientée objet. En substance, il permet aux concepteurs d'intégrer des objets dans la structure bien connue des tables. [1]

2.1.2.6 Modèle entité-association

Ce modèle reproduit les relations entre les entités du monde réel de façon très similaire au modèle réseau, mais il n'est pas lié aussi directement à la structure physique de la base de données. Il est souvent utilisé pour la création d'une base de données d'un point de vue conceptuel. [1]

2.2 Le Système de Gestion de Base de Données

2.2.1 Définition

Afin de pouvoir contrôler les données ainsi que les utilisateurs, le besoin d'un système de gestion s'est vite fait ressentir.

Un système de gestion de base de données (SGBD) est un ensemble de services (applications logicielles) qui permet d'inscrire, de retrouver, de modifier, de trier, de transformer ou d'imprimer les informations de la base de données. Il permet d'effectuer des comptes rendus des informations enregistrées et comporte des mécanismes pour assurer la cohérence des informations, éviter des pertes d'informations due à des pannes, assurer la confidentialité et permettre son utilisation par d'autres logiciels.

Ils sont utilisés pour de nombreuses applications informatiques, notamment les guichets automatiques bancaires, les logiciels de réservation, les bibliothèques numériques, les logiciels d'inventaire, les progiciels de gestion intégrés ou la plupart des blogs et sites web. Il existe de nombreux systèmes de gestion de base de données.

Les SGBD sont souvent exploités par d'autres logiciels. Ils peuvent être sous forme de composant logiciel, de serveur, de logiciel applicatif ou d'environnement de programmation.

[2]

2.2.2 Fonctions des SGBD

Le SGBD supporte de plus en plus fréquemment des fonctions avancées pour la gestion de procédures et d'événements. [3]

Parmi ces fonctions on trouve :

- Description des données ;
- Recherche de données ;
- Mise à jour des données ;
- Transformation des données ;
- Contrôle de l'intégrité des données ;
- Gestion de transactions et sécurité ;

2.2.3 Architecture d'un SGBD

La conception d'une base de donnée nécessite de considérer et de spécifier trois niveaux d'abstractions.

2.2.3.1 Le niveau conceptuel

Le niveau central est le niveau conceptuel. Il correspond à la structure canonique des données qui existent dans l'entreprise, c'est-à-dire leur structure sémantique inhérente sans souci d'implantation en machine, représentant la vue intégrée de tous les utilisateurs. La définition du schéma conceptuel d'une entreprise ou d'une application n'est pas un travail évident. Ceci nécessite un accord sur les concepts de base que modélisent les données. Par exemple, le schéma conceptuel permettra de définir :

1. Les types de données élémentaires qui définissent les propriétés élémentaires des objets de l'entreprise (cru d'un vin, millésime, qualité, etc.) ;
2. Les types de données composés qui permettent de regrouper les attributs afin de décrire les objets du monde réel ou les relations entre objets (vin, personne, buveur, etc.) ;
3. Les types de données composés qui permettent de regrouper les attributs afin de décrire les associations du monde réel (abus de vin par un buveur, production d'un vin par un producteur, etc.) ;
4. Éventuellement, des règles que devront suivre les données au cours de leur vie dans l'entreprise (l'âge d'une personne est compris entre 0 et 150, tout vin doit avoir un producteur, etc.). [4]

2.2.3.2 Le niveau interne

Le niveau interne correspond à la structure de stockage supportant les données. La définition du schéma interne nécessite au préalable le choix d'un SGBD. Elle permet donc de décrire les données telles qu'elles sont stockées dans la machine, par exemple :

- Les fichiers qui les contiennent (nom, organisation, localisation...) ;
- Les articles de ces fichiers (longueur, champs composants, modes de placement en fichiers...) ;
- Les chemins d'accès à ces articles (index, chaînages, fichiers inversés...) ;

[4]

2.2.3.3 Le niveau externe

Au niveau externe, chaque groupe de travail utilisant des données possède une description des données perçues, appelée schéma externe. Cette description est effectuée selon la manière dont le groupe voit la base dans ses programmes d'application. Alors qu'aux niveaux conceptuel et interne les schémas décrivent toute une base de données, au niveau externe ils décrivent simplement la partie des données présentant un intérêt pour un utilisateur ou un groupe d'utilisateurs. En conséquence, un schéma externe est souvent qualifié de vue externe. Le modèle externe utilisé est dépendant du langage de manipulation de la base de données utilisé. [4]

2.2.4 Objectifs des SGBD

Des objectifs principaux ont été fixés aux SGBD dès l'origine de ceux-ci et ce, afin de résoudre les problèmes causés par la démarche classique. [5]

Ces objectifs sont ci-après :

- **Indépendance physique** : La façon dont les données sont définies doit être indépendante des structures de stockage utilisées.
- **Indépendance logique** : Un même ensemble de données peut être vu différemment par des utilisateurs différents.
- **Accès aux données** : Se fait par intermédiaire d'un langage de manipulation de données (LMD).
- **Administration centralisée de données (intégration)** : toutes les données doivent être centralisées dans un réservoir unique commun à toutes les applications.
- **Non redondance** : Chaque donnée doit être présente qu'une seule fois dans la base.
- **Cohérence de données** : Les données sont soumises à un certain nombre de contraintes d'intégrité qui définissent un état cohérent de la base.
- **Partage de données** : Il s'agit de permettre à plusieurs utilisateurs d'accéder aux données au même moment de manière transparente.
- **Sécurité des données** : Les données doivent pouvoir être protégées contre les accès non autorisés.
- **Résistance aux pannes**.

2.2.5 Les Avantages d'un SGBD

- Le partage des données s'améliore au sein de l'organisation : Obtenir un meilleur accès aux données ainsi qu'une meilleure gestion des données.
- Amélioration de la sécurité des données : un meilleur cadre est fourni pour l'application des politiques de confidentialité et de sécurité des données.
- Les systèmes de gestion de base de données minimisent l'incohérence des données : L'utilisation d'un système de gestion approprié et d'outils de gestion de la qualité des données permet de réduire au minimum le problème de l'incohérence des données.
- Augmentation de la productivité de l'utilisateur : en déployant les meilleurs outils de qualité de données et systèmes de gestion de bases de données, la productivité de l'utilisateur est accrue.
- Intégration efficace des données : lorsque la gestion des données est améliorée, elle favorise une image intégrée des opérations d'une organisation.
- Gagner du temps : une base de données centralisée, facilement accessible à tous, peut fournir des réponses rapides à des questions redoutables. [6]

2.2.6 Caractéristiques de la conception de BDD en SGBD

- Les entités deviennent des tables dans le modèle relationnel ;
- Les identifiants deviennent des clés primaires et secondaires ;
- Les propriétés deviennent des attributs de la table ;
- L'ensemble des propriétés devient une ligne ;
- Les types de relations qu'on peut rencontrer sont au nombre de quatre :
 1. un à un (1,1) ;
 2. un à plusieurs (1,n) ;
 3. plusieurs à plusieurs (n,n) ;
 4. plusieurs à un (n,1).

2.2.7 Méthode d'accès aux données

La base de données étant créée, il sera nécessaire que les utilisateurs trouvent bien un moyen d'y accéder par diverses opérations telles que :

- L'ajout de données ;
- La mise à jour des données ;
- La suppression des données et tant d'autres.

Cela se fait grâce à une méthode d'accès aux données.

2.2.8 Les SGBD plus connus et utilisé

Oracle

- Le plus « portable », que ce soit en matière de machines ou d'architectures ;
- Le plus « adaptable », puisqu'il permet de créer des applications ultra-personnalisées pour éliminer les fonctionnalités superflues et gagner en performance ;
- Il est accessible à tous les langages et à toutes les technologies. [7]

MySQL

- Il est open source et gratuit ;
- Il gère des masses de données énormes ;
- Pas de contraintes d'intégrité référentielle (*foreign keys*) ;
- Le plus populaire dans les sites web et l'intégration PHP. [7]

Microsoft SQL server

- Il est pensé pour la mise en place d'applications de gestion, d'applications décisionnelles et de DataWarehouses complexes ;
- Édité et commercialisé par Microsoft depuis 1994 ;
- MS SQL Server inclut par défaut des outils de gestion, d'administration et de développement de bases de données. [7]

PostgreSQL

- PostgreSQL est un système de gestion de base de données relationnelle ;
- Il est orienté objet ;
- Il est open source ;
- Il est puissant puisqu'il est capable de prendre en charge en toute sécurité les charges de travail de données les plus complexes ; [7]

Access

- Access est un système passif de gestion de base de données (SGBD) relationnelle ;
- Il permet de créer des formulaires de saisie et des fonctionnalités propres à l'application ; [7]

2.3 Réplication

2.3.1 Réplication de Base de Données

En informatique, la réplication est un processus de partage d'informations pour assurer la cohérence de données entre plusieurs sources de données redondantes, pour améliorer la fiabilité, la tolérance aux pannes ou l'accessibilité.

On parle de réplication de données si les mêmes données sont dupliquées sur plusieurs périphériques. Une application de base de données repose sur un modèle client-serveur. Dans ce modèle, le client se connecte au SGBD pour passer des ordres. Ces ordres sont de deux natures : lecture (on parle alors de requêtes) ou mise à jour (on parle alors de transactions). Pour les transactions, il y a une modification des données sur le serveur, mais cela reste des ordres de courte durée.

A l'inverse, dans le cas d'une lecture, il n'y a pas de modification des données mais les traitements peuvent être longs et porter sur une grande masse de données. On comprend donc aisément que, dans le cadre d'un site web par exemple, un nombre important de requêtes peut surcharger partiellement (ou complètement) le serveur. Il existe plusieurs solutions pour remédier à ce genre de problèmes et, ça tombe bien, la réplication en est une.

La réplication n'est pas à confondre avec une sauvegarde : la réplication est une sauvegarde suivie d'une synchronisation, Les données sauvegardées ne changent pas dans le temps, reflétant un état fixe des données, tandis que les données répliquées évoluent sans cesse à mesure que les données sources changent.

La réplication est aussi considérée comme la capacité à maintenir à jour une base de données distribuée sur plusieurs machines reliées en réseau, en recopiant à intervalles réguliers des morceaux ou l'intégralité de la base d'une machine à l'autre. Plusieurs méthodes de réplication existent selon la configuration matérielle présente. [8]

2.3.2 Principe de la réplication

Le but principal de la réplication est de faciliter l'accès aux données en augmentant la disponibilité. Soit parce que les données sont copiées sur différents sites permettant de répartir les requêtes, soit parce qu'un site peut prendre la relève lorsque le serveur principal s'écroule.

Une autre application tout aussi importante est la synchronisation des systèmes embarqués non connectés en permanence. Ce qui peut se résumer à l'aide des trois types de scénario suivants :

- Deux serveurs distants sur lesquels les données doivent être consistantes ;
- Deux serveurs, un comme serveur principal, l'autre comme serveur de backup à chaud ;
- Plusieurs serveurs en cluster utilisés pour de l'équilibrage de charge et de la tolérance à la panne ;

Le principe de la réplication, qui met en jeu au minimum deux SGBD, est assez simple et se déroule en trois temps :

- La base « maître » reçoit un ordre de mise à jour (INSERT, UPDATE ou DELETE).

- Les modifications faites sur les données sont détectées et stockées (dans une table, un fichier, une queue) en vue de leur propagation.
- Un processus de réplication prend en charge la propagation des modifications à faire sur une seconde base dite esclave. Il peut bien entendu y avoir plus d'une base esclave.

Bien entendu, il est tout à fait possible de faire de la réplication dans les deux sens (de l'esclave vers le maître et inversement). On parlera dans ce cas-là de réplication bidirectionnelle ou symétrique.

Dans le cas contraire la réplication est unidirectionnelle (seulement du maître vers l'esclave) et on parle de réplication en lecture seule ou asymétrique. [8]

2.3.3 Modes de Réplication

2.3.3.1 Réplication unidirectionnelle (asymétrique)

Consiste à effectuer uniquement une mise à jour du replica maître vers le replica esclave ou le replica esclave vers le replica maître, comme l'illustre la figure (Figure 6). [9]

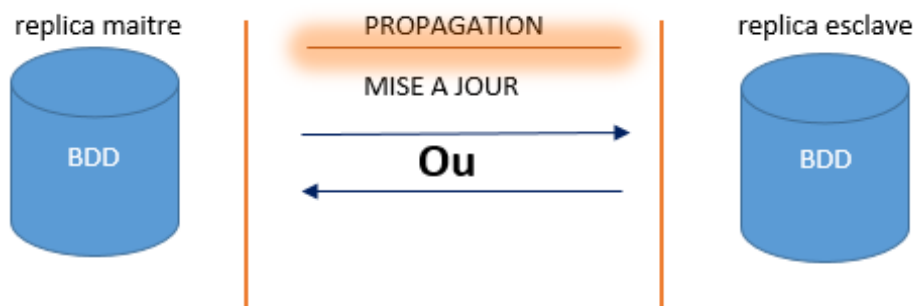


Figure 6 : Réplication unidirectionnelle

2.3.3.2 Réplication bidirectionnelle (symétrique)

Consiste à effectuer une mise à jour du replica maître vers les replicas esclaves et des replicas esclaves vers le replica maître, comme l'illustre la figure (Figure 7). [9]

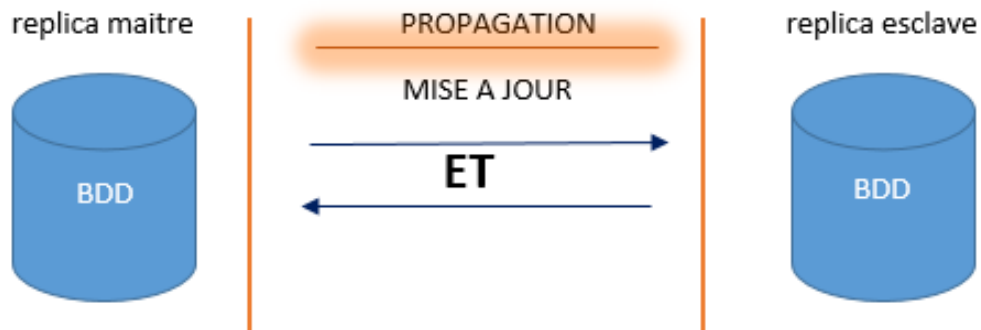


Figure 7 : Réplication bidirectionnelle

2.3.4 Méthode de propagation

Que ce soit la réplication unidirectionnelle ou la réplication bidirectionnelle, la propagation ne peut se faire que selon un mode de mise à jour :

2.3.4.1 Mise à jour asynchrone (temps différé)

C'est le mode de distribution dans lequel certaines sous opérations locales effectuées suite à une mise à jour globale sont accomplies dans des transactions indépendantes en temps différé. Le temps de mise à jour des copies peut être plus au moins différé : les transactions de report peuvent être lancées dès que possible ou à des instants fixes, par exemple le soir ou en fin de semaine.

Les avantages sont la possibilité de mettre à jour en temps choisi des données, tout en autorisant l'accès aux versions anciennes avant la mise à niveau.

Les inconvénients sont bien sûr que l'accès à la dernière version n'est pas garanti. Ce qui limite les possibilités des mises à jour. [10]

2.3.4.2 Mise à jour synchrone (temps réel)

C'est le mode de distribution dans lequel toutes les sous opérations locales effectuées suite à une mise à jour globale sont accomplies pour le compte de la même transaction.

Dans le contexte des copies, ce mode de distribution est très utile lorsque les mises à jour effectuées sur un site doivent être prises en compte immédiatement sur les autres sites.

L'avantage essentiel de la mise à jour synchrone est de garder toutes les données au dernier niveau de mise à jour. Le système peut alors garantir la fourniture de la dernière version des données quel que soit la copie accédée.

Les inconvénients sont cependant multiples, ce qui conduit beaucoup d'applications à éviter la gestion de copies synchrones. Ce sont d'une part, la nécessité de gérer des transactions multiples coûteuses en ressources et d'autre part, la complexité des algorithmes de gestion de concurrence et de panne d'un site. C'est pour cela que l'on préfère souvent le mode de mise à jour asynchrone (encore appelé mise à jour différée). [10]

2.3.5 Les différents types de Réplication

Il est possible d'avoir quatre types de réplication :

- Réplication asymétrique avec propagation synchrone ;
- Réplication symétrique avec propagation synchrone ;
- Réplication asymétrique avec propagation asynchrone ;
- Réplication symétrique avec propagation asynchrone ;

Dans ce qui suit, nous allons voir ces types avec plus de détail ;

2.3.5.1 Réplication asymétrique avec propagation synchrone

La copie est instantanément mise à jour à chaque modification de la base de données maître ou esclave. Si la copie échoue c'est toute la transaction qui est annulée, comme l'illustre la figure (Figure 8). [9]

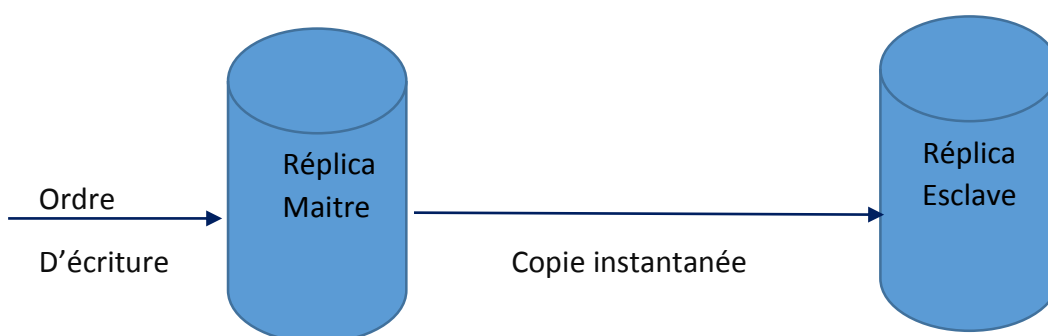


Figure 8 : Réplication asymétrique synchrone

2.3.5.2 Réplication symétrique avec propagation synchrone

Les deux bases peuvent être modifiées .si l’une est mise à jour la propagation sur l’autre base se fait directement. Il est noté que la réplication fait partie de la transaction ce qui ne ralentit que très peu le système, comme l’illustre la figure (Figure 9). [9]

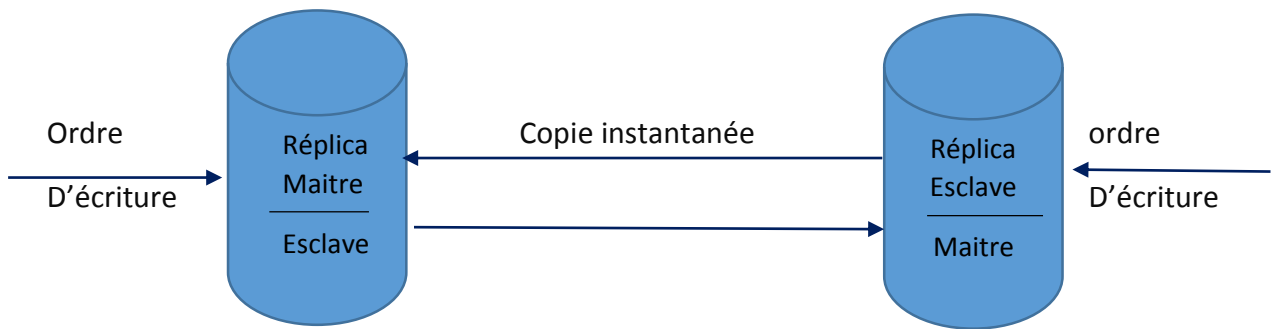


Figure 9 : Réplication symétrique synchrone

2.3.5.3 Réplication asymétrique avec propagation asynchrone

La mise à jour de la base répliquée est différée ➡ utilisation d’un programmeur de tâches pour déclencher la réplication avec une horloge, comme l’illustre la figure (Figure 10). [9]

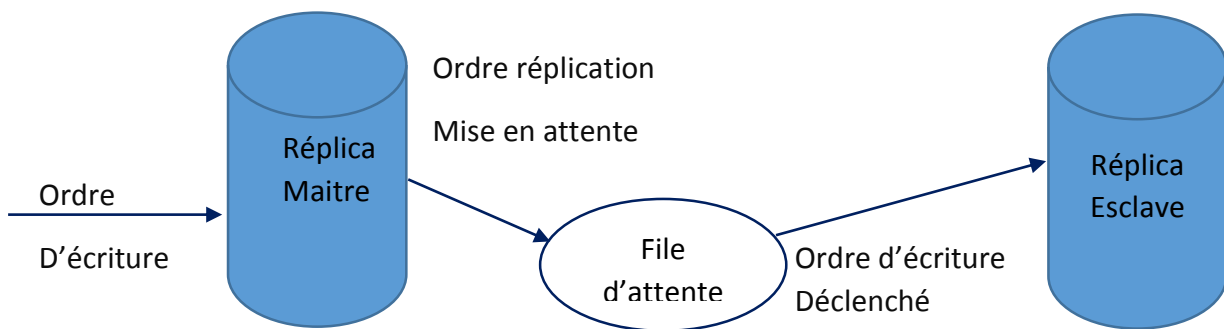


Figure 10 : Réplication asymétrique asynchrone

2.3.5.4 Réplication symétrique avec propagation asynchrone

La mise à jour des bases répliquées est différée → on peut mettre en place deux programmeurs ou utiliser le même, comme l'illustre la figure (Figure 11). [9]

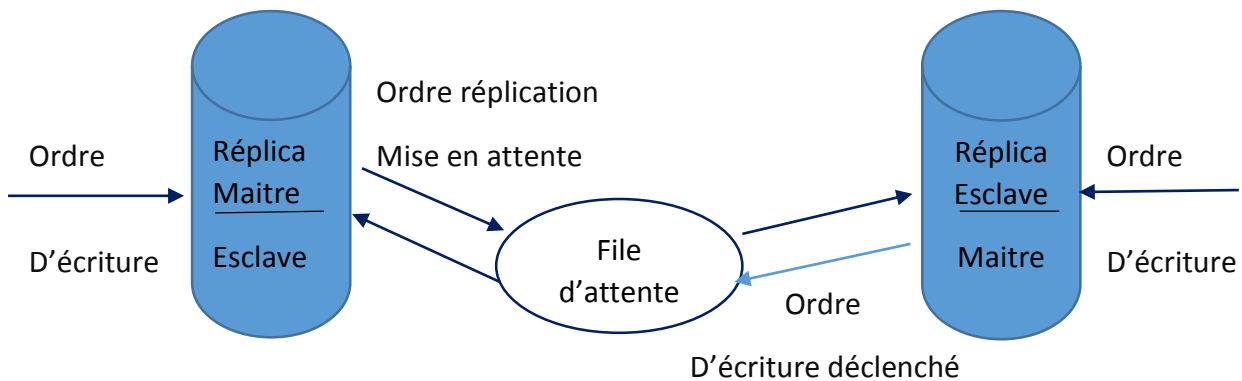


Figure 11 : réplication symétrique asynchrone

2.3.6 Objectifs de la Réplication de Base de Données

- La Haute Disponibilité : la base étant présente dans différentes sites l'accessibilité est préservée en cas de pannes ou catastrophes naturelles.
- La performance : les performances sont augmentées car la charge de travail est répartie entre les sites.
- Trafic réseau : du moment que le travail est reparté (localement) la charge réseau ne peut que diminuer.
- Traitement hors ligne : il est possible de travailler sur une machine non connectée au réseau.
- Equilibrage des charges : la réplication permet de mieux répartir les données donc la charge de travail du serveur hôteur de la base de données principale sera allégée. [11]

Conclusion

A travers les différents points développés dans le présent chapitre, nous avons pu, détailler et expliquer l'intérêt des bases de données et le système de gestion de base de données.

Nous avons pu également constater l'intérêt particulier porté à la réplication des bases de données et aux différents problèmes auxquels ce type de solution a pu remédier.

Dans le chapitre suivant, Nous allons présenter le support matériel et logiciel qui nous a servi d'appui pour la réalisation de notre réplication.

3 Chapitre 3 : Réalisation

Introduction

De nos jours, la réplication des données est reconnue comme étant un moyen efficace pour augmenter la disponibilité et la fiabilité des bases de données au sein d'une entreprise.

Cette phase consiste à transformer nos connaissances et toutes les informations récoltées durant l'étude théorique établie précédemment en des composants logiciels formant notre système.

Dans ce chapitre, nous allons commencer par proposer une architecture de solution puis à dégager et élaborer les composants de notre système qui nous serviront d'appui pour la réalisation de notre réplication.

3.1 L'architecture de la solution proposée

Après l'étude de l'architecture actuelle qui a engendré un grand nombre de problèmes au sein de l'entreprise, nous avons opté pour la solution suivante :

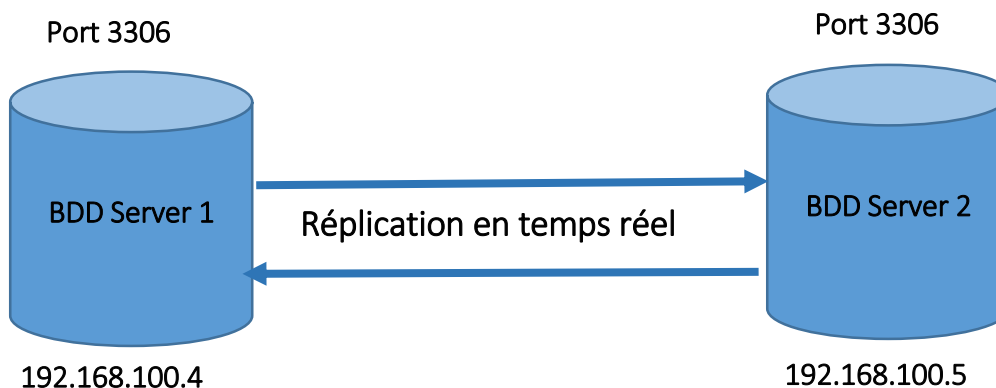


Figure 12 : Architecture de la solution

3.2 Présentation des outils utilisés

3.2.1 La virtualisation

La virtualisation consiste à créer une représentation virtuelle, basée logicielle, d'un objet ou d'une ressource telle qu'un système d'exploitation, un serveur, un système de stockage ou un réseau. Ces ressources simulées ou émulées sont en tous points identiques à leur version physique.

Les machines virtuelles sont exécutées sur une couche d'abstraction matérielle. Ainsi, les entreprises sont en mesure d'exécuter plusieurs systèmes d'exploitation et applications simultanément sur un serveur unique. Les ressources et capacités du serveur sont réparties entre les différentes instances. Ceci permet de réaliser d'importantes économies et d'utiliser les serveurs plus efficacement puisque leurs capacités sont pleinement exploitées. De plus, la virtualisation permet d'exécuter des programmes de façon isolée pour effectuer des tests en toute sécurité sans risquer de compromettre les autres machines virtuelles exécutées sur le même serveur hôte. Pour les particuliers et les utilisateurs de PC, il est aussi possible d'utiliser la virtualisation pour exécuter des applications nécessitant différents systèmes d'exploitation sur un seul et même ordinateur. [12]

3.2.2 Son fonctionnement

La technologie de virtualisation repose sur l'abstraction d'une application, d'un système d'exploitation ou du stockage de données du véritable logiciel ou du hardware sous-jacent. L'un des principaux cas d'usage est la virtualisation de serveur, reposant sur une couche logicielle appelée « hyperviseur » pour émuler le hardware.

L'hyperviseur permet de créer et d'exécuter des machines virtuelles. Le logiciel se charge de diviser les ressources entre les différents environnements virtuels en fonction de leurs besoins.

Par le passé, on distinguait les hyperviseurs natifs, directement intégrés au hardware du logiciel, et les hyperviseurs « hébergés » semblables à des applications logicielles. Cependant, sur les systèmes modernes, cette distinction a perdu de l'importance. Pour cause, il existe désormais des systèmes comme le KVM (kernel-based virtual machine), qui est intégré au kernel Linux et peut exécuter les machines virtuelles directement bien qu'il soit possible de continuer à utiliser le système lui-même comme un ordinateur normal.

Les utilisateurs peuvent ensuite interagir et lancer des applications ou des calculs au sein de l'environnement virtuel, que l'on appelle généralement machine virtuelle. La machine virtuelle est l'équivalent émulé d'un ordinateur, et est exécutée par-dessus un autre système. Elle peut exploiter la puissance de calcul du CPU et la mémoire de la machine hôte, un ou plusieurs disques virtuels pour le stockage, une interface réseau virtuelle ou réelle, ainsi que les autres composants tels que les cartes graphiques ou même les clés USB.

La machine virtuelle fonctionne comme un fichier de données unique. Elle peut donc être transférée d'un ordinateur à l'autre, et fonctionner de la même manière sur les deux machines. Si la machine virtuelle est stockée sur un disque virtuel, on parle souvent d'une image disque. [12]

3.2.3 La virtualisation des serveurs

La virtualisation de serveurs est une architecture logicielle qui permet à plusieurs systèmes d'exploitation de serveur de fonctionner comme hôte sur un serveur physique spécifique. Lorsque le logiciel de serveur n'est plus pris en considération sur la machine physique, le serveur devient une "machine virtuelle" détachée du dispositif physique, bien

qu'il se représente toujours fonctionner exclusivement avec ses ressources informatiques et sa mémoire. En fait, il fonctionne sur une imitation virtuelle du serveur physique.

La virtualisation des serveurs permet une utilisation plus efficace des ressources informatiques que précédemment. Avant la virtualisation des serveurs, il était courant d'avoir du matériel surutilisé ou sous-utilisé dans le même datacenter. Avec la virtualisation, il est possible de déplacer des charges de travail entre les machines virtuelles selon la charge. Le même serveur physique peut aussi fonctionner sur plusieurs systèmes d'exploitation de serveur et plusieurs configurations, ce qui augmente son efficacité. La virtualisation des serveurs est la base du cloud computing et de l'informatique hybride.[13]

3.2.4 L'hyperviseur

La présentation de la virtualisation se focalise très souvent sur le concept de machine virtuelle (VM). Toutefois, lorsqu'on s'attache à mieux appréhender la virtualisation,

L'hyperviseur est une brique essentielle qui fait fonctionner l'ensemble. Un hyperviseur est une plate-forme de virtualisation qui permet à plusieurs systèmes d'exploitation et donc plusieurs VMs de travailler sur une même machine physique en même temps. Il s'agit donc d'un moniteur de machine virtuelle (VM). Il existe deux types d'hyperviseur :

- Un hyperviseur de type 1 est un système qui s'installe directement sur la couche matérielle du serveur. Ces systèmes sont allégés de manière à se « concentrer » sur la gestion des OSs invités c'est-à-dire ceux utilisés par les VMs qu'ils contiennent. Ceci permet de libérer le plus de ressources possibles pour les VMs. Toutefois, il est possible d'exécuter uniquement un hyperviseur à la fois sur un serveur.

- Un hyperviseur de type 2 est un logiciel qui s'installe et s'exécute sur un OS déjà en place. De ce fait, plus de ressources sont utilisées étant donné qu'on fait tourner l'hyperviseur et l'OS qui le supporte, il y a donc moins de ressources disponibles pour les machines virtuelles. L'intérêt qu'on peut trouver c'est le fait de pouvoir exécuter plusieurs hyperviseurs simultanément vu qu'ils ne sont pas liés à la couche matérielle. [14]

3.2.5 Les machines virtuelles

Une machine virtuelle est un fichier informatique, généralement appelé image, qui se comporte comme un ordinateur réel. En d'autres termes, il s'agit d'un ordinateur créé à l'intérieur d'un ordinateur. Elle s'exécute dans une fenêtre, comme tout autre programme, en offrant à l'utilisateur final une expérience identique à celle qu'il aurait sur le système d'exploitation hôte. La machine virtuelle est placée dans un « bac à sable » qui l'isole du reste du système, de sorte que les logiciels installés sur la machine virtuelle ne peuvent ni s'échapper, ni modifier l'ordinateur hôte. Cela produit un environnement idéal pour tester d'autres systèmes d'exploitation, dont des versions bêta, l'accès à des données infectées par des virus, la création de sauvegardes de système d'exploitation et l'exécution de logiciels ou d'applications sur des systèmes d'exploitation auxquels ils ne sont pas destinés à l'origine.

Il est possible d'exécuter plusieurs machines virtuelles simultanément sur un même ordinateur physique. Pour les serveurs, les divers systèmes d'exploitation fonctionnent côte à côte, avec un composant logiciel appelé hyperviseur pour les gérer, alors que les ordinateurs de bureau classiques n'utilisent qu'un seul système d'exploitation pour exécuter d'autres systèmes d'exploitation dans des fenêtres de programme qui leur sont propres. Chaque machine virtuelle fournit son propre matériel virtuel, à savoir les processeurs, la mémoire, les disques durs, les interfaces réseau et les autres périphériques nécessaires. Le matériel virtuel est ensuite mappé au matériel réel sur la machine physique, ce qui permet de réaliser des économies en réduisant les besoins en matériel, ainsi que les coûts de maintenance, d'alimentation et de refroidissement associés. [15]

3.3 Les principaux logiciels de virtualisation

1. **Vmware** : VMware Workstation est un logiciel de machine virtuelle utilisé pour les Ordinateurs pour exécuter plusieurs systèmes d'exploitation sur un seul ordinateur hôte physique. Chaque machine virtuelle peut exécuter simultanément une seule instance de tout système d'exploitation (Microsoft, Linux, etc.). VMware Workstation prend fortement en charge la compatibilité matérielle et fonctionne comme un pont entre l'hôte et la machine virtuelle pour toutes sortes de ressources matérielles, y compris les disques durs, les périphériques USB et les CD-ROM. Tous les pilotes de périphériques sont installés via la machine hôte. [16]
2. **Virtualbox** : Virtualbox est un logiciel libre de virtualisation développé par Oracle, il permet une virtualisation en hyperviseur de type 2, qui permet de virtualiser plusieurs machines virtuelles sur la même machine physique. Il permet de virtualiser plusieurs systèmes d'exploitation (Mac OS X, Linux, Windows, etc...) et peut être installé sur tous les OS. [17]

3.3.1 Logiciel choisi

Pour la réalisation de notre travail nous avons opté pour le Virtualbox version 6.1.12 comme logiciel de virtualisation

- **Ses avantages**

VirtualBox a de nombreux avantages [17]

- C'est un logiciel libre et gratuit ;
- Il permet de virtualiser n'importe quelle OS ;
- Il permet de virtualiser plusieurs systèmes d'exploitation sans faire de multiboot et avoir une seule configuration machine ;
- Il permet aussi de configurer la quantité de mémoire, l'espace disque dur et l'accès ou pas au USB, réseau, carte son, etc...

- VirtualBox intègre un serveur RDP (Remote Desktop Control), ce qui permet de démarrer une machine virtuelle sur un ordinateur, et d'utiliser cette même machine virtuelle à partir d'un autre ordinateur. Ceci s'avère extrêmement pratique pour l'aide à distance multiplateforme ;

3.4 Système de gestion de bases de données

Dans le cadre de notre travail, nous avons travaillé avec une base de données MySQL Parce que c'est un choix adopté par l'entreprise.

MySQL est un serveur de bases de données relationnelles Open Source. Un serveur de bases de données stocke les données dans des tables séparées plutôt que de tout rassembler dans une seule table, Cela améliore la rapidité et la souplesse de l'ensemble. Les tables sont reliées par des relations définies, qui rendent possible la combinaison de données entre plusieurs tables durant une requête. Le SQL dans "MySQL" signifie "Structured Query Language" : le langage standard pour les traitements de bases de données.

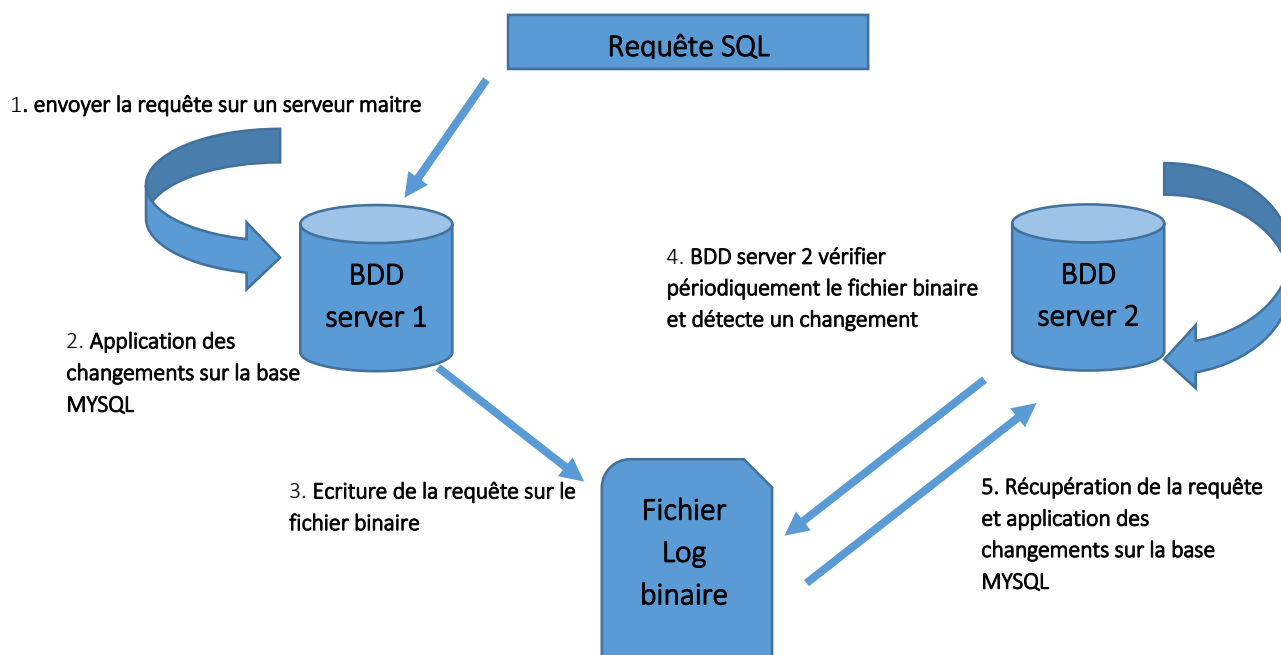


Figure 13 : Protocole de réplication MYSQL

3.5 Linux

Linux est un système d'exploitation (l'équivalent de MAC/OS ou Windows). Avec Linux, vous n'êtes pas obligé d'avoir MS Windows sur votre machine, mais les deux peuvent très bien cohabiter. Sous Mac, vous ne pouvez pas lancer un programme fait pour Windows (sauf par l'intermédiaire d'un émulateur), sous Linux, il en est de même. Mais Linux est compatible UNIX et il existe beaucoup de programmes pour Linux. [18]

- **Ses avantages**

- Il est gratuit ;
- Il fonctionne sur une dizaine de plateformes dont les plus classiques PC et Power-PC ;
- Il permet de faire un serveur de mail, news, ftp, disques pour windows et mac, web, nfs, firewall ;
- Il plante tellement rarement ;
- Il est rapide ;
- Il est multitâche préemptif (réellement), 32 bits (64 sur DEC/ALPHA), multi-utilisateurs et supporte X-Windows. Par exemple, cela permet, si vous avez plusieurs machines, de les connecter en réseau pour avoir une machine en tant que serveur et les autres en tant que simples terminaux X. Pour une puissance équivalente, on fait de grosses économies (de temps, d'argent).

3.6 Mise en place de l'architecture

Nous avons créé deux machines virtuelles linux nommés Serveur_1 et Serveur_2 comme l'illustre la figure 14.

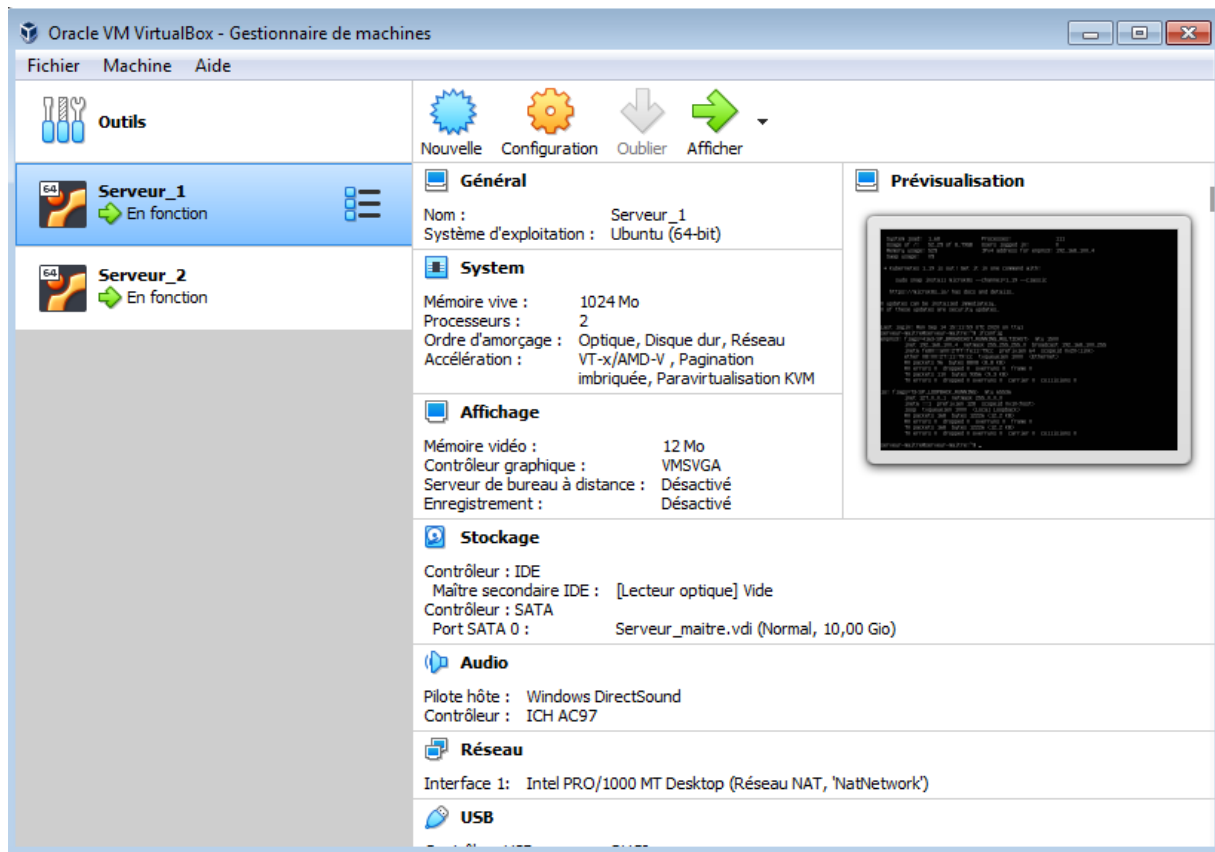


Figure 14: Interface de Virtualbox

Nous avons créé un réseau local qui a pour adresse 192.168.100.0/24 pour interconnecter nos deux machines

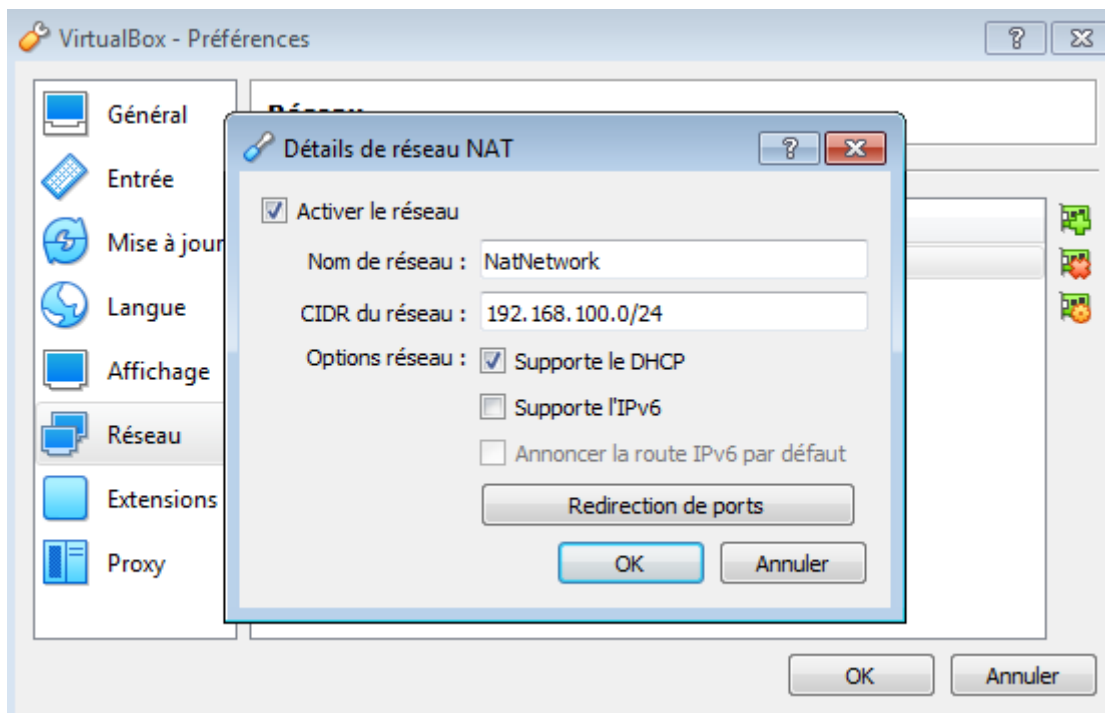


Figure 15: Interface de réseau local

Nos deux machines virtuelle Serveur_1 et Serveur_2 ont respectivement les adresses IPV4 suivantes :

Virtuelle machine 1 :

```
serveur-maitre@serveur-maitre:~$ ifconfig
enp0s3: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.100.4 netmask 255.255.255.0 broadcast 192.168.100.255
    inet6 fe80::a00:27ff:fe11:79cc prefixlen 64 scopeid 0x20<link>
    ether 08:00:27:11:79:cc txqueuelen 1000 (Ethernet)
    RX packets 4890 bytes 369454 (369.4 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 4916 bytes 363898 (363.8 KB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1000 (Local Loopback)
    RX packets 8572 bytes 608978 (608.9 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 8572 bytes 608978 (608.9 KB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
```

Virtuelle machine 2 :

```
serveur-maitre2@serveur-maitre2:~$ ifconfig
enp0s3: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 192.168.100.5 netmask 255.255.255.0 broadcast 192.168.100.255
    inet6 fe80::a00:27ff:fe53:9360 prefixlen 64 scopeid 0x20<link>
    ether 08:00:27:53:93:60 txqueuelen 1000 (Ethernet)
    RX packets 4857 bytes 365382 (365.3 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 4866 bytes 360463 (360.4 KB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536
    inet 127.0.0.1 netmask 255.0.0.0
    inet6 ::1 prefixlen 128 scopeid 0x10<host>
    loop txqueuelen 1000 (Local Loopback)
    RX packets 8574 bytes 609294 (609.2 KB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 8574 bytes 609294 (609.2 KB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
```

3.7 Présentation de processus de mise en place de la réplication (maitre-maitre)

3.7.1 Installation de MYSQL server

Mise à jour des serveurs à l'aide du gestionnaire de paquets

```
root@serveur-maitre:/home/serveur-maitre# aptitude update -y && aptitude upgrade -y
```

Installation ensuite du serveur MYSQL en tant que serveur à l'aide de la commande ci-dessous :

```
root@serveur-maitre:/home/serveur-maitre# aptitude install mysql-server -y
```

Cette opération devra être répéter sur chacun des deux serveurs pour que les serveurs MySQL soient fonctionnels.

3.8 Les étapes de la création de la réplication

Processus de la mise en place de la réplication entre le serveur_1 (192.168.100.4) et le serveur_2 (192.168.100.5)

Etape 1 : Création d'un utilisateur replica sur chaque base :

```
mysql> CREATE USER replica IDENTIFIED BY 'motdepasse';  
Query OK, 0 rows affected (0.12 sec)
```

Etape 2 : Création de la réplication en attribuant le privilège de la réplication à l'utilisateur replica :

```
mysql> GRANT REPLICATION SLAVE ON *.* TO replica;  
Query OK, 0 rows affected (0.03 sec)
```

Etape 3 : configuration de la réplication sur le serveur 2 :

```
mysql> SHOW MASTER STATUS;  
+-----+-----+-----+-----+-----+  
| File          | Position | Binlog_Do_DB | Binlog_Ignore_DB | Executed_Gtid_Set |  
+-----+-----+-----+-----+-----+  
| mysql-bin.000239 | 156 | bdd_test     | mysql              |                    |  
+-----+-----+-----+-----+-----+  
1 row in set (0.01 sec)
```

Récupérer le fichier log binaire et sa position sur le serveur 1

```
mysql> stop slave;  
Query OK, 0 rows affected, 1 warning (0.00 sec)  
  
mysql> CHANGE MASTER TO MASTER_HOST='192.168.100.4', MASTER_USER='replica', MASTER_PASSWORD='motdepasse', MASTER_LOG_FILE='mysql-bin.000239', MASTER_LOG_POS=156;  
Query OK, 0 rows affected, 2 warnings (0.12 sec)  
  
mysql> START SLAVE;  
Query OK, 0 rows affected (0.05 sec)
```

A ce stade le Serveur_1 est maitre et le Serveur_2 est slave

Etape 4 : configuration du slave sur le serveur 1 :

```
mysql> show master status;
+-----+-----+-----+-----+-----+
| File           | Position | Binlog_Do_DB | Binlog_Ignore_DB | Executed_Gtid_Set |
+-----+-----+-----+-----+-----+
| mysql-bin.000069 |      156 | bdd_test     | mysql              |                    |
+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)
```

Récupérer le fichier log binaire et sa position sur le serveur 2

```
mysql> stop slave;
Query OK, 0 rows affected, 1 warning (0.00 sec)

mysql> CHANGE MASTER TO MASTER_HOST='192.168.100.4', MASTER_USER='replica', MASTER_PASSWORD='motdepasse', MASTER_LOG_FILE='mysql-bin.000239', MASTER_LOG_POS=156;
Query OK, 0 rows affected, 2 warnings (0.12 sec)

mysql> START SLAVE;
Query OK, 0 rows affected (0.05 sec)
```

Activation de la réplication au niveau des fichiers de configuration MySQL

```
/etc/mysql/my.cnf
```

Sur le serveur 1 :

```
server-id          = 1
log_bin            = /var/log/mysql/mysql-bin.log
binlog_expire_logs_seconds = 2592000
max_binlog_size    = 100M
binlog_do_db       = bdd_test
binlog_ignore_db   = mysql
```

Sur le serveur 2 :

```
server-id          = 2
log_bin            = /var/log/mysql/mysql-bin.log
binlog_expire_logs_seconds = 2592000
max_binlog_size    = 100M
binlog_do_db       = bdd_test
binlog_ignore_db   = mysql
```

Les paramètres à ajouter/modifier sont les suivants :

- server-id : c'est l'identifiant de la base. il doit être différent entre les deux bases
- log_bin : le chemin vers le fichier log de mysql.
- expire_logs_days: durée pour laquelle mysql conserve les logs
- binlog_do_db: spécifier la base de données qui concernée par la réplication

- `binlog_ignore_db`: pour exclure la réplication de la base de données mysql (qui tiré à l'environnement le fonctionnement de mysql)

3.9 Les étapes de vérification du fonctionnement de la Réplication

Afin de vérifier que la réplication est fonctionnelle sur chacun des deux serveurs. On exécute d'abord la commande suivante afin de lister les bases et tables existantes :

```
mysql> show databases;
+-----+
| Database |
+-----+
| bdd_test |
| information_schema |
| mysql |
| performance_schema |
| sys |
+-----+
5 rows in set (0.01 sec)
```

```
mysql> use bdd_test;
Database changed
```

Sur le serveur 1 nous allons par exemple crée une table dont la base de donnée `bdd_test` qui sera utiliser pour montrer la réplication du premier serveur vers le second

```
mysql> create table tb_test( id INT(11) AUTO_INCREMENT PRIMARY KEY, username VARCHAR(128) NOT NULL, age INT(11));
Query OK, 0 rows affected, 2 warnings (0.19 sec)
```

Sur le second serveur on affiche les tables de la base de donnée `bdd_test`, on trouve que la table est créée sur le serveur 1 et aussi créé sur le serveur 2

```
mysql> show tables;
+-----+
| Tables_in_bdd_test |
+-----+
| tb_test |
+-----+
1 row in set (0.01 sec)
```

Afin de prouver que la réplication s'effectue aussi dans le sens inverse on va insérer deux lignes sur la table `tb_test` à partir du serveur 2

```
mysql> insert into tb_test (username,age) values ( 'Rosa',23);  
Query OK, 1 row affected (0.09 sec)
```

```
mysql> insert into tb_test (username,age) values ( 'Rabha',25);  
Query OK, 1 row affected (0.06 sec)
```

Ainsi faudra obtenir le résultat suivant sur chacun des deux serveurs, ce qui montre que la réplication fonctionne correctement

```
mysql> select * from tb_test;  
+----+-----+-----+  
| id | username | age |  
+----+-----+-----+  
|  1 | Rosa     |  23 |  
|  2 | Rabha    |  25 |  
+----+-----+-----+  
2 rows in set (0.00 sec)
```

Conclusion

Dans ce dernier chapitre, nous avons présenté les aspects pratiques liés à la mise en place de la réplication des bases de données maître-maître à travers MySQL à savoir l'environnement du travail, les outils de virtualisations utilisés, ainsi que quelques lignes de commandes.

Conclusion générale

De nos jours, les bases de données informatiques sont utilisées dans un grand nombre d'entreprises pour stocker, organiser et analyser les données.

On rappelle que le but de notre projet est de réaliser une réplication des bases de données sous linux à travers MYSQL permettant de résoudre à la fois des problèmes de sécurité et de disponibilité des données.

Dans notre cas, la réplication de serveur à serveur permet une meilleure intégration ou rapprochement des données entre plusieurs serveurs de base de données et l'avantage de cette dernière par rapport à la méthode actuellement utilisée à l'Entreprise Portuaire de Bejaia est que les utilisateurs qui travaillent sur les bases qui participent à la réplication peuvent ainsi consulter des données de meilleure qualité.

La réalisation de ce projet nous a été très bénéfique car ce fut l'occasion d'enrichir nos connaissances dans le domaine des bases de données et d'acquérir une expérience professionnelle complète au sein d'une grande entreprise qui a son importance avec une envergure non négligeable dans son domaine au niveau national telle que l'Entreprise Portuaire De Bejaia.

Notre réalisation permettra à l'entreprise d'avoir :

- Un équilibrage de charge
- Un meilleur temps de réponse
- Une probabilité de panne plus faible

Cependant des perspectives d'améliorations restent envisageable telle que :

- Mettre en place un programme qui gère automatiquement le basculement de bases de données en cas de crash.
- Mettre en place d'une tierce alternative en cas de crash peu probable des deux serveurs simultanément.

Bibliographie

- [1] : Types de modèles de bases de données. https://www.lucidchart.com/pages/fr/quest-ce-quun-modele-de-base-de-donnees#section_2, page consultée le 17/03/2020.
- [2] : Les SGBD. <https://sgbdslam5.wordpress.com/les-sgbd/>, page consultée le 17/03/2020.
- [3] : Georges Gardartin, 2003. Base de Données. Eyrolles, 832p.
- [4]: Tschritzis D.C., Klug A. 1978. ANSI/X3/SPARC DBMS Framework Report on Database Management Systems », *Information Systems*, vol. 3: p. 173-191
- [5] : Laurent Audi Bert, 2009. Base de Données de la modélisation au SQL. Ellipses
- [6] : Les avantages d'un système de gestion de base de données.
<https://www.oracle.com/fr/database/pourquoi-utiliser-sgbd.html>, page consulter le 01/04/2020.
- [7] : Elodie Djegadisvarane : Quels sont les SGBD les plus plébiscités par les entreprises. (21/03/2019), <https://e-mag.xxe.fr/sgbd-plus-populaires/>, page consultée 01/04/2020.
- [8] : Djedje NKONGOLO : Etude d'une réplique symétrique asynchrone dans une base de données répartie Application. (2011), https://www.memoireonline.com/10/12/6180/m_Etude-dune-replication-symetrique-asynchrone-dans-une-base-de-donnees-repartie-Application-24.html, page consultée le 14/04/2020.
- [9] : <https://doc.pcsoft.fr/fr-fr/?3044266>, page consultée le 20/04/2020.
- [10]: Jennifer L. Welch: Proceedings of the 15th International Conference on Distributed Computing. (2001), <https://dl.acm.org/doi/proceedings/10.5555/645958?fbclid=IwAR3o4gz1KOcgL6ZxalEzbeMyNoLHp-GJUlSVNllb1dVMAOGOy3XSi9F6TM>, page consultée le 20/04/2020.
- [11] : replication. [PDF], <http://www-inf.int-evry.fr/cours/BD/MSCurriculum/FRENCH/PDF/replication.pdf>, page consultée le 23/04/2020.
- [12] : <https://www.lebigdata.fr>, page consultée le 25/04/2020.
- [13] : <https://www.hpe.com>, page consultée le 27/04/2020.
- [14] : <http://www.efort.com>, page consultée le 01/05/2020.
- [15] : <https://www.azure.microsoft.com>, page consultée le 02/05/2020.

[16] : Mise en place d'un environnement de virtualisation mémoire 2013.

[17] : Damien Bouyeure : VirtualBox. [PDF], <http://damien-bouyeure.fr/doc/VirtualBox.pdf>, page consultée le 03/05/2020.

[18] : <https://www.maretmanu.org>, page consultée le 10/05/2020.

Résumé

En informatique, la réplication est un processus de partage d'informations pour assurer la cohérence de données entre plusieurs sources de données redondantes, pour améliorer la fiabilité, la tolérance aux pannes, ou la disponibilité, nous nous sommes fixé l'objectif de résoudre les problèmes liés à la sauvegarde des données qui se fait manuellement chaque mois à l'Entreprise Portuaire de Bejaia reflétant un état fixe des données, La réplication n'est pas à confondre avec une sauvegarde car les données répliquées évoluent sans cesse à mesure que les données sources changent.

L'utilisation des machines virtuelles, et linux comme système d'exploitation nous a permis d'atteindre l'objectif que nous nous étions fixé qui est la réalisation de la réplication.

Mots clés : Machine virtuelle, MySQL, Réplication, SGBD.

Abstract

Replication in computing involves sharing information so as to ensure consistency between redundant resources, to improve reliability, fault-tolerance, or accessibility, we have set ourselves the goal of solving problems related to the data backup which is done manually each month at the Port Company of Bejaia reflecting a fixed state of the data, Replication is not to be confused with a backup because the replicated data constantly evolves as the data sources change.

The use of virtual machines and Linux as the operating system, allowed us to achieve the objective we had set ourselves, which is the realization of replication.

Keywords: MySQL, Replication, SGBD, virtual machine.