

Introduction générale :

Des volumes considérables de données sont créés tous les jours par La communauté informatique. Lorsqu'on parle des données massives nous parlons sur des quantités qu'on ne peut imaginer. Elles représentent un volume à peine croyable d'environ 2,5 quintillions de données par jour.

Ce sont des informations qui provienne de partout, et sont disponibles dans les réseaux sociaux, les appareils mobiles, les messageries électroniques, les blogs, les transactions bancaires, les capteurs utilisés pour collecter les informations, les images numériques, les vidéos publiées en ligne, et les signaux GPS de téléphones mobiles...etc.

A partir des croissances de quantité énorme de l'information et la cause de l'inflation du volume des données,aujourd'hui les entreprises se trouvent avec un gros problème qui est celles de savoir comment collecter, stocker, analyser et exploiter ces grands volumes de données. Pour cette raison la technologie « Big Data » est apparu pour analyser les hautes et disparates volumétries de données.

Les moteurs de base de données basés sur le standard SQL ont de bonnes performances lors du traitement de petites quantités de données relationnelles mais ces outils sont très limités face à l'expansion des données en volume et en complexité.

Le besoin de rapidité et la bonne qualité dans le traitement des données a conduit à l'élaboration de solutions Big data telles que Hadoop qui est un outil efficace pour résoudre les problèmes de données massives. HadoopMapReduce est considéré comme la technique de traitement la plus efficace, comparée aux bases de données SQL.

L'objectif de notre travail est d'intégrer et d'exploiter cette technologie pour l'aide à l'apprentissage et à la maintenance informatique à travers une plateforme web. Il est bien connu que dans la vie courante l'informatique intervient maintenant presque dans tous les domaines possibles. Tout le monde fait recours à l'informatique, on dirait que ce dernier s'impose dans la vie quotidienne. La science de L'informatique est un ensemble vaste sur ses différentes spécialités. De ce fait, nous allons proposer une application à base de Big data permettant à l'apprenant informatique de trouver des réponses à ses questions liées à l'apprentissage ou à la maintenance des systèmes informatiques. Le rôle du Big data dans ce contexte est de supporter un volume gigantesque de texte, de tutoriels et de vidéo pour l'assistance des apprenants.

Pour atteindre cet objectif nous avons divisé ce mémoire en trois chapitres.

Dans le premier chapitre nous allons présenter des notions sur le BigData, nous allons également résumer l'historique du stockage de l'information ainsi que les Framework de traitement de donnée massive.

Dans le deuxième chapitre, nous allons exposer notre phase de conception tout en identifiant nos besoins à travers le langage de modélisations UML (Unified Modeling Language).

L'implémentation et la mise en œuvre de ce travail seront présentées dans le dernier chapitre avec l'ensemble d'outils techniques utilisés dans cette phase.

Finalement, nous terminons ce manuscrit par une conclusion et quelques perspectives.