



République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene

Faculté d'Electronique et d'Informatique

Département Informatique

Projet de fin d'étude pour l'obtention du diplôme Master

Option

Architectures parallèles et calcul intensif

**Stratégies de parallélisation d'un Algorithme Génétique
sur le framework Apache Hadoop avec le modèle
map/reduce :**

Application au problème de Flow Shop de Permutation.

Sujet proposé par :

Me : MEHDI Malika

Mr : BENDJOUDI AHCÈNE

Présenté par :

BELBEZZA Hamza

Soutenu le : 22/06/2016

Devant le Jury composé de :

Mr BABA ALI Riadh

Mr ISLI Amar

Président

Membre

Monôme N° 048/2016

Dédicaces

À ma famille, mes amis et collègues ;)

REMERCEMENTS

Tout d'abord, louanges à DIEU le tout puissant pour m'avoir aidé à accomplir ce modeste travail.

Je remercie très fort mes deux respectueux encadreurs : Mr Ahcène Bendjoudi et Mme Malika Mehdi, pour m'avoir encadré pendant toute la période du projet de fin d'études, et qui m'ont beaucoup apporté par leur savoir.

Je remercie tous mes enseignants de Master 2 APCI, pour leur savoir dispensé, efforts et encouragements surtout à mon endroit.

Je ne saurais trop remercier, sans exception aucune, tous les étudiants et étudiantes du Master 2 APCI, promotion 2015-2016 USTHB.

Je n'oublierai jamais de signifier ma reconnaissance à mes amis et collègues de l'IFEP de Birkhadem.

Enfin, je remercie toutes les personnes qui m'ont aidées de près ou de loin pour aller au bout de ce projet de fin d'études.

Sommaire

Introduction générale :	1
--------------------------------	----------

Chapitre 01 : Optimisation combinatoire et algorithmes génétiques

1- Introduction :	4
2- Optimisation combinatoire :	4
3- Problèmes combinatoires :	4
3-1 Formulation d'un problème d'optimisation combinatoire :	5
3-2 Complexité combinatoire :	5
4- Méthodes de résolution :	6
4-1 Méthodes exactes :	6
4-2 Méthodes heuristiques (incomplètes) :	7
4-3 Méthodes métaheuristiques :	7
5- Algorithmes évolutionnaires :	7
5-1 Fonctionnement :	8
5-1-1 Sélection :	8
5-1-2 Croisement :	9
5-1-3 Mutation :	11
5-2 Structure typique d'un algorithme génétique :	12
5-3 Paramètres de contrôle d'un algorithme génétique :	12
6- Stratégies de parallélisation des Algorithmes Génétiques :	12
6-1 Le modèle d'évaluation parallèle de la population :	13
6-1-1 Le mode synchrone :	14
6-1-2 Le mode asynchrone :	14
6-2 Modèle en îles :	14
6-2-1 Modèle en îles avec migration :	15
6-2-2 Modèle en îles sans migration :	17
6-3 Modèle d'évaluation parallèle d'une solution :	18
6-4 Modèle de Grille:	18
6-5 Modèle hybride :	19
Conclusion :	19

Chapitre 02 : Le modèle de programmation Map Reduce sous le Framework Apache Hadoop

1- Introduction :	20
--------------------------	-----------

2- MapReduce :	20
2-1 Apparition de mapreduce :	21
2-2 Principe de fonctionnement mapreduce :	21
2-3 Le Workflow MapReduce:	22
2-4 Problématiques liées à l'implémentation de Mapreduce :	23
2-4-1 Mapreduce et parallélisation :	23
2-4-2 Scalability (évolutivité) :	23
2-4-3 Tolérance aux pannes :	23
2-4-4 Equilibrage de charge :	24
2-4-5 Optimisation sur les transferts disques et réseau	24
3- Présentation du framework Hadoop	24
3-1 Big data	24
3-1-1 Sources de données :	24
3-1-2 Rapidité de la croissance des données :	25
3-1-3 Données non structurées :	25
3-2 La plateforme (framework) Hadoop	26
3-2-1 Architecture maître esclave de Hadoop:	27
3-2-2 Ecosystème de Hadoop Apache :	28
3-2-3 HDFS : Hadoop Distributed File System (Système de fichiers distribués de Hadoop) :.....	29
3-2-4 Apache Hadoop YARN : (Yet Another Resource Negotiator).....	32
3-2-5 Hadoop Mapreduce :	33
3- Versions Hadoop:	36
3-1 Apports d'Apache Hadoop 2.x :	36
Conclusion :	37

Chapitre 03 : Les architectures parallèles

1- Introduction :	38
2. Classification des architectures parallèles des ordinateurs :	38
2.1 Classification de Flynn :	39
2-1-1- SISD: Single instruction stream, single data stream :	39
2-1-2- SIMD: Single instruction stream, multiple data streams :	39
2-1-3- MISD : Multiple instruction streams, single data stream,	40
2-1-4- MIMD: Multiple instruction streams, multiple data streams,.....	41
2-2- Classification par types de mémoires :	41
2-2-1 Systèmes à mémoires partagées :	42
2-2-2- Systèmes à mémoire distribuée : (Multi-ordinateur) :	43

3- Les grappes d'ordinateurs (clusters) :	43
3-1 Cluster IBNBADIS :	44
Conclusion :	45

Chapitre 04 : Conception

1- Introduction :	47
2- Problème de Flow Shop de permutation (PFSP) :	47
3- Résolution du problème de Flow Shop de Permutation par les Algorithmes Génétiques :	48
3-1 Codage des individus (chromosomes) :	48
3-2 Génération de la population :	49
3-3 Calcul du Makespan (fonction objectif) :	49
3-4 Sélection :	49
3-5 Croisement :	49
3-6 Mutation :	50
3-7 Remplacement :	50
4- Stratégies de parallélisation d'un Algorithme Génétique sur Hadoop pour la résolution du PFSP	50
4-1 Modèle Maître/Esclaves en MapReduce :	51
4-2- Modèle insulaire en MapReduce:	54
Conclusion :	57

Chapitre 05 : Implémentation et expérimentations

1- Introduction :	58
2- Environnements et outils de développement :	58
2-1- Logiciels :	58
2-1-1- Système d'exploitation :	58
2-1-2- Kit de développement Java :	58
2-1-3- Hadoop Apache :	59
2-1-4- Langage de programmation :	59
2-1-5- Eclipse :	59
2-2- Matériel :	59
3- Implémentation :	60
3-1 Structures de données et variables :	60
3-1-1Classes :	60
3-2 Pseudo-Algorithmes :	61
3-2-1 Différence entre le modèle Maître/esclaves et le modèle Insulaire :	62

Sommaire

4- Expérimentations et résultats :	62
4-1 Choix des Benchmarks :	62
4-2- Paramètres des deux algorithmes :	62
4-3 Scénarios des expérimentations :	63
5- Discussion des résultats des trois premiers scénarios:	66
Conclusion :	67
Conclusion générale :	68