



RAPPORT DU PROJET DE FIN D'ÉTUDES

Pour l'obtention du diplôme de
Post Graduation Spécialisée en Big Data et Calcul Intensif

**Plateforme Big Data pour l'analyse de sécurité
du trafic réseau**

Réalisé par : FATHI Boualem & ELATSAMENIA Hamid

Encadreur : Dr YAHIAOUI Said & Dr ZEGHACHE Linda

Soutenu le : 13/12/2021. **Devant le jury** composé de :

Président : Mr.BELAZZOUGUI Djamal

Examineurs : Mr.BOUCENNA Fateh

Année Universitaire : 2019 /2020

Remerciements

Avant tout nous remercions Dieu le tout puissant de nous avoir donné la santé, le courage, la force et la volonté d'entamer et d'accomplir ce modeste travail, et de nous avoir éclairé le chemin de la réussite.

En deuxième lieu nos remerciements vont en particulier à nos encadreurs : Dr YAHIAOUI Saïd & Dr ZEGHACHE Linda qu'ils trouvent ici le témoignage de notre profonde gratitude pour avoir proposé ce thème, et pour la confiance qu'ils nous ont accordée et les conseils fructueux qu'ils n'ont cessé de nous prodiguer tout au long de la réalisation de ce travail.

Nous tenons à exprimer nos profondes gratitudes et nos remerciements les plus vifs au président et aux membres de jury pour l'honneur qu'ils nous font en acceptant de juger notre travail.

Enfin, que tous ceux qui nous ont aidé, de près ou de loin, tout au long la réalisation de ce travail, les cadres et les enseignants du CERIST.

Dédicaces

Mes très chers parents : pour leur amour, soutien et énormes sacrifices ;

À ma chère femme Yamna ;

À mes chers enfants : Djohayna, Amani, Khawla et Djawad ;

À mes chers frères et sœurs ;

À tous mes collègues de la formation PGS de l'année 2020/2021 ;

À tous mes Amis.

FATHI Boualem



Mes très chers parents : pour leur amour, soutien et énormes sacrifices ;

À ma chère femme Karima ;

À mes chers enfants : hadil, , ritedj et ahmed ;

À mes chers frères et sœurs ;

À tous mes collègues de la formation PGS de l'année 2020/2021 ;

À tous mes Amis.

ELASMENIA Hamid



ملخص

أصبح تحليل بيانات الشبكة من أهم مجالات معالجة البيانات الضخمة، لأن حجم وتعقيد البيانات المتعلقة بتحليل بيانات الشبكة أصبح كبيرًا جدًا بحيث لا يمكن معالجتها بواسطة الأدوات والوسائل التقليدية. بالإضافة إلى ذلك، هناك مجموعة كبيرة من أدوات معالجة البيانات الضخمة المستخدمة في تحليل بيانات الشبكة والتي يتم انتقاؤها حسب الواقع.

في هذا العمل، نقترح معالجة إشكالية إنشاء أرضية خاصة بتحليل بيانات الشبكة تستند على أدوات معالجة البيانات الضخمة، بطريقة مستنيرة تعتمد على دراسة الهدف المسطر والمعلومات المتوفرة حاليا. حيث قمنا بإجراء دراسة مقارنة لمختلف الأرضيات المتوفرة حاليا والمتعلقة بتحليل بيانات الشبكة وبشكل أكثر تحديداً، أدوات المعالجة الخاصة بالبيانات الضخمة، سمحت لنا هذه الدراسة المقارنة بتسليط الضوء على معايير الاختيار لتصميم حل يتم تكييفه حسب متطلبات التطبيق المقصود.

في النهاية، ولشرح الوضع في الخدمة لنظام خاص بتحليل بيانات الشبكة قائم على أدوات معالجة البيانات الضخمة من الناحية العملية، قمنا بتصميم والوضع في الخدمة لتطبيق يتمثل في مراقبة تدفق بيانات الشبكة حيث يتم أولاً التقاط وحفظ البيانات بصفة آنية، بعدها يتم تحليلها وعرضها مع إظهار حالة استغلال الشبكة.

Abstract

Network traffic analysis is becoming a Big Data problem along with the growing size and complexity of security data that cannot be manipulated by traditional security tools. In addition, there is panoply of Big Data tools proposed for use in Network traffic analysis and choosing the right ones depends on reality.

In this work, we propose to treat the problem of setting up a Big data-based Network traffic analysis platform, in as knowledgeable as possible way, after an objective and well-documented study of the solutions proposed in the current state of the art. Thus, we made a comparative study of existing Big Data platforms dedicated to Network traffic analysis and more specifically, tools dedicated to the Big Data processing. This comparative study allowed us to highlight selection criteria for the design of a solution that is adapted to the requirements of the intended application.

At the end, and to explain in practical terms the implementation of a Network traffic analysis platform based on Big Data, we implemented a network traffic monitoring application consisting in capturing and saving, in real time, the network flow, and then making some analysis and visualization on the collected data to give an image on the network status.

Résumé

L'analyse du trafic réseau est en train de devenir un problème de Big Data en raison de la taille et de la complexité croissante des données de sécurité qui ne peuvent être manipulées par les outils de sécurité traditionnels. De plus, il existe une panoplie d'outils Big Data proposés pour la cybersécurité et dont le choix dépend de la réalité.

Dans ce travail, nous proposons de traiter la problématique de l'utilisation des outils Big data dans l'analyse du trafic réseau basée sur les données, de la manière la plus informée possible. Après une étude objective et bien documentée des solutions proposées dans l'état actuel des connaissances, nous avons réalisé une étude comparative des plates-formes Big Data existantes dédiées à l'analyse du trafic réseau et plus particulièrement des outils dédiés au traitement du Big Data. Cette étude comparative nous a permis de mettre en évidence des critères de sélection pour la conception d'une solution adaptée aux besoins de l'application envisagée.

Enfin, pour expliquer concrètement la mise en place d'une plateforme d'analyse du trafic réseau basée sur le Big Data, nous avons implémenté une application de surveillance du trafic réseau consistant à capturer et à sauvegarder, en temps réel, le flux du réseau, puis à effectuer des analyses et une visualisation sur les données collectées mettant en évidence l'état d'utilisation du réseau.

Table des matières

Remerciements	ii
Dédicaces	iii
ملخص	v
Abstract	vi
Résumé	vii
Table des matières	viii
Liste des Tableaux	xi
Liste des Figures	xii
Introduction Générale	1
Chapitre I : La sécurité informatique	5
1.1 Introduction	6
1.2 Les origines de l'insécurité	7
1.3 Les différents types d'attaque informatique	8
1.4 La sécurité informatique	8
1.4.1 Définitions de la sécurité informatique	8
1.4.2 Objectifs de la sécurité	9
1.4.3 Gestion des risques	10
1.4.4 Politique de sécurité	10
1.4.5 Techniques de sécurité	12
1.4.6 Processus de sécurité :	16
1.5. Insuffisance des solutions classiques de Cybersécurité	18
1.6. Applications du Big Data dans le domaine de la cybersécurité	18
1.7 Conclusion	20
Chapitre II : Big Data	21
II.1 Introduction	22
II.2 Définition	23
II.3. Caractéristiques du Big Data	23
II.3.1. Volume	23
II.3.2. La Vitesse (Vitesse)	24
I.3.3. La Variété	24
I.3.4. La Vérité (fiabilité)	24
II.3.5. La Valeur	25
II.4. Sources du Big Data	25
II.4.1. Internet of Things (IoT)	25

I.4.2. réseaux sociaux	26
II.4.3. Evolution des équipements mobiles	27
II.4.4. Les systèmes gouvernementaux.....	28
II.4.5. Les villes intelligentes	28
II.5. Architecture du « pipe » Big Data	29
II.6. Domaines d'application du Big Data	30
II.7. Conclusion	31
Chapitre III : Les Outils Big Data	32
III.1. Introduction.....	32
III.2. Hadoop	33
III.2.1. L'Histoire de Hadoop	33
III.2.2. Concepts	34
III.2.3. Architecture	35
III.2.4. Caractéristiques de Hadoop :	44
III.2.5. Avantages de Hadoop.....	45
III.2.6. Les défis liés à l'utilisation de Hadoop	46
III.2.7. Hadoop dans les entreprises	46
III.2.8. Les distributions commerciales	48
III.3. Les outils d'ingestions en temps réel	48
III.3.1 apache Kafka	48
III.3.2 apache Flume.....	53
III.3.3 Comparaison entre Apache Kafka et Flume.....	55
III.4. Les outils de Traitements en temps réel	57
III.4.1. Spark.....	57
III.4.2. Storm	65
III.4.3. Discussion	69
III.5. Critères de choix des outils de traitement	72
III.5.1. Comment le Big Data peut-il aider en l'analyse de sécurité du trafic réseau ?.....	72
III.5.2. Critères d'analyse du Trafic Réseau	73
III.6. Les outils Big Data utiliser pour la surveillance du trafic réseau.....	75
III.6.1. OpenSOC	75
III.6.2. Apache Metron	78
III.6.3. Discussion	81
III.7. les outils BIG Data possibles à utiliser pour Notre implémentation	82
III.8. Conclusion	83
Chapitre IV : Une solution d'analyse de sécurité du trafic réseau basée sur les outils Big Data.....	84
IV.1. Introduction	85

IV.2. Scénario généraliste d'une application de l'analyse de sécurité du trafic réseau	85
IV.3. Architecture d'une plateforme d'Analyse du Trafic Réseau basée sur le Big Data	88
VI.4. Préparation de l'environnement :	88
IV.5. Implémentation de notre solution d'Analyse du Trafic Réseau à base d'outils Big Data	90
IV.5.1. Scénario d'exécution :	92
IV.6 Conclusion	104
Conclusion Générale	105
Bibliographie	108
Les Annexes	112

Liste des Tableaux

Tableau 1: Récapitulatif de La comparaison entre Kafka et Flume56
Tableau 2: Récapitulatif de La comparaison entre Spark et Storm.....70
Tableau 3: Les besoin en traitement des données du trafic réseau.....72

Liste des Figures

Figure 1: le volume et la variété du Big Data [27]	22
Figure 2 : Les caractéristiques du Big Data	24
Figure 3 : Les Sources du Big Data.....	25
Figure 4: Internet of Things (IoT) 2021[20]	26
Figure 5: Une minute Internet 2020/2021	26
Figure 6: Evolution des équipements mobiles [21].....	27
Figure 7: Schéma d'une ville intelligente [21]	28
Figure 8: Outils Big Data par catégories	32
Figure 9: Architecture de l'HDFS.	35
Figure 10: Rôle du NameNode Secondaire	37
Figure 11: Stockage et réplication des blocs dans l'HDFS.	37
Figure 12: Gestion des répliquions des blocs dans l'HDFS	38
Figure 13: Gestion ressources dans Hadoop 1.0	39
Figure 14: YARN dans Hadoop 2.0	40
Figure 15: Composants de YARN.....	42
Figure 16: Enchaînement des étapes et interaction des composants YARN.....	43
Figure 17: Principe de MapReduce	43
Figure 18: Principe de traitements parallèle dans MapReduce.	44
Figure 19: fonctionnalités de Kafka	49
Figure 20: L' architecture de KAFKA	51
Figure 21: groupe de consommateurs.....	52
Figure 22: les composants du topic	52
Figure 23: Distribution des Partitions	53
Figure 24: L'ingestion des données avec flume.....	53
Figure 25: L'architecture de flume	54
Figure 26: les composants de flume	55
Figure 27: Écosystème Spark	58
Figure 28: Opérations sur les RDDs.....	60
Figure 29: Persistance des RDDs	60
Figure 30: Exemple d'un DAG	61
Figure 31: Composantes d'une application Spark distribuée	62

Figure 32: Conversion du programme utilisateur en tâches	63
Figure 33: Planification des taches sur les Executors	63
Figure 34: Topologie de Storm.	66
Figure 35: Architecture de Storm	67
Figure 36: Composants de OpenSOC	77
Figure 37: fonctionnalités de Metron	79
Figure 38: Composants de Metron	80
Figure 39: schéma général du flux de données dans l'entreprise.	86
Figure 40: Schéma de notre solution d'Analyse du Trafic Réseau basée sur les outils big data.	89
Figure 41: l'hyperviseur VMware ESXI 5.	91
Figure 42: VMware vSphere Client.	92
Figure 43: les différentes statistiques et job dans spark streaming lors de d'exécution de programme.....	98
Figure 44: Interfaces tableau de bord Kibana visualisation des trafics /*ssh-tcp-udp et autres*/.	100
Figure 45: L'opération de l'indexation au niveau de elasticsearch	101
Figure 46: consommation du trafic réseau en utilise consumer kafka	101
Figure 47: Limiter le nombre de connexion SSH à 120 connexions.....	102
Figure 48: contacter le cluster Ibn Bdis au niveau du Cerist.....	102
Figure 49: traitement du trafic qui contient le Protocol SSH	102
Figure 50: augmentation du nombre de connexions SSH de 113 à 136.....	103
Figure 51: les différentes informations du trafic réseau	103