

الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المدرسة العليا للإعلام الآلي - 08 ماي 1945 - بسبدي بلعباس
Ecole Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès



Mémoire de Fin d'étude

Pour l'obtention du diplôme d'ingénieur d'état

Filière : **Informatique**

Spécialité : **Systemes d'information et Technologies Web (SIW)**

Thème

Mise en œuvre d'une plateforme Hadoop pour l'analyse prédictive et temps réel des logs pour le réseau WAN de l'ANEM

Présenté par :

- Mr KHELLADI Abderrahmane
- Mr SELMI Anwar

Soutenu le : **06/07/2020**

Devant le jury composé de :

- | | |
|---------------------------------|-----------|
| - M. BELFEDHAL Alaa Eddine | Président |
| - M. MALKI Abdelhamid | Encadreur |
| - M. BENDAOUD Fayssal | Examineur |
| - M. KAZI TANI Mohammed Yassine | Examineur |
| - M. BOUGHEDDA Ahmed | Invité |

Année Universitaire : 2019 / 2020

Résumé

L'ANEM est l'organisme responsable du service public de l'emploi en Algérie. Sa mission est d'organiser et d'assurer la connaissance de la situation et de l'évolution du marché national de l'emploi et de la main d'oeuvre, en garantissant à tout demandeur et à toute entreprise un service de recrutement efficace et personnalisé.

Dans le cadre de modernisation des services de l'ANEM et suite à son passage vers le digital très réussi récemment, les données et l'information sont devenus au centre stratégique de la prise de décisions de l'organisme. Tous les directions de l'ANEM, partenaires externes et public à la recherche d'emploi sollicitent ses services numériques quotidiennement.

Cette situation a rendu la production de données et de l'information très riche (Variété), a flux très important (Volumétrie) avec une nécessité de disponibilité très élevée et très rapide (Vélocité) pour solliciter ces informations à des besoins de consultation et de recommandation, d'analyse et prise de décision.

Face à ce nouveau challenge des 3V, le concept de Big Data est né, cette approche pourra faire ces preuves au profit de leur institution toute en garantissant un retour sur investissement (ROI) très positif vis-à-vis les services offerts contre les ressources dédiées à cette solution.

Afin de confirmer les preuves de cette approche, l'ANEM s'est intéressé à l'alignement d'un écosystème de type Big Data au profit des différentes parties prenantes (IT et métier) qui seront impliqués à ce projet, en exploitant également cet écosystème pour la mise en oeuvre d'une solution IDS vue la volumétrie importante des logs qui sera capter et analyser en temps réel afin de prévenir contre les intrusions qui peuvent survenir au sein de leur réseau WAN.

Mots clés : Big Data, Streaming, Analyse des logs, Détection d'intrusions, IDS, Machine Learning.

Abstract

ANEM is the organization responsible for the public employment service in Algeria. Its mission is to organize and ensure the knowledge of the situation and the evolution of the national market of employment and manpower, by guaranteeing to all applicants and all companies an efficient and personalized recruitment service.

As part of the modernization of ANEM services and following its recent successful switch to digital, data and information have become at the strategic center of the organization's decision-making. All ANEM departments, external partners and the public looking for work use its digital services on a daily basis.

This situation has made the production of data and information very rich (Variety), has a very large flow (Volumetry) with a need for very high availability (Velocity) to request this information for consultation and recommendation needs, analysis and decision making.

Faced with this new challenge of 3V, the concept of Big Data was born, this approach will be able to prove its evidence to the profile of their institution while guaranteeing a very positive return on investment (ROI) when comparing the services offered against resources dedicated to this solution.

In order to confirm the evidence for this approach, ANEM focused on aligning a Big Data ecosystem to the profile of the various stakeholders (IT and business) who will be involved in this project, while exploiting this ecosystem for the implementation of an IDS solution for the large volume of logs that will be captured and analyzed in real time in order to prevent against intrusions that may occur within their WAN network.

Keywords : Big Data, Streaming, Log analysis, Intrusion detection, IDS, Machine Learning.