

## Tek Düğüm ve Çok Düğüm Hadoop Ortamı İle Sunucu Erişimi Günlükleri Analiz Edilerek Sunucu Bakımı İçin Doğru Zamanlama Tespiti

Ferhat OFLEZER<sup>1\*</sup>, Turgay Tugay Bilgin<sup>2</sup>

<sup>1</sup>Bilgisayar Programcılığı Bölümü, Bozok Üniversitesi, Türkiye

<sup>2</sup>Bilgisayar Mühendisliği, Bursa Teknik Üniversitesi, Türkiye

\*Corresponding author: ferhat.oflezer@bozok.edu.tr

\*Speaker: ferhat.oflezer@bozok.edu.tr

Presentation/Paper Type: Oral / Abstract

**Özet-** Sürekli kayıt altına alınan günlük (log) dosyaları web sunucuların ürettiği en önemli bilgilerden biridir. Hadoop, donanım kümeleri kullanılarak büyük miktarda veri üzerinde çalışan açık kaynaklı uygulama platformudur. Web sunucusu günlükleri anlamlı şekilde yapılandırılmış geniş hacimli düz metin dosyalarıdır. Sunucular, kaynak paylaşımı için özelleştirilmiş işletim sistemleri kullanırlar. Zaman içerisinde sunucuların yük durumu izlenerek güvenlik güncelleştirmeleri gibi bazı önlemlerin önceden planlanması son derece önemlidir. Bu yüzden sunucuya erişimin en az olduğu zamanın tespit edilmesi gerekmektedir. Hadoop ortamında log dosyalarını saklamak için Hadoop dosya sistemi kullanılmakta ve sonuçlar map ve reduce fonksiyonları kullanılarak değerlendirilmektedir. Bu çalışmada, web günlüklerini analiz etmek için Hadoop mapreduce programlama modeli tek düğüm ortamında ve dağıtılmış ortamda iki ayrı senaryo ile kullanılmış ve sunucuya gün içerisinde hangi saatlerde ne kadar erişim yapıldığı tespit edilerek bakım için en uygun zaman bulunmaya çalışılmıştır.

*Anahtar Kelimeler-* Hadoop; Haritalama-İndirgeme; Dosya sistemi; Büyük veri; Veri işleme

## Analysis Of Server Logs To Determine The Right Timing For Server Maintenance Using A Single Node And Multi-Node Hadoop Environment

**Abstract-** The log files that are continuously recorded are one of the most important information generated by web servers. Hadoop is an open source application platform that works on large amounts of data using hardware clusters. Web server logs are large-volume plain text files that are meaningfully structured. The servers use customized operating systems for resource sharing. It is extremely important that, some precautions like security updates are formerly planned by monitoring the load status of the servers over time. Therefore, it is necessary to determine when the server workload is minimal. In the Hadoop environment, the Hadoop file system is used to store the log files and the results are appreciated using the map and reduce functions. In this study, the Hadoop MapReduce programming model was used with two mode for analyzing weblogs. These modes are single node environment mode and distributed environment mode. By using our approach, we have determined how many connections made to the server during the day and we have tried to find the best time for maintenance.

*Keywords-* Hadoop; Mapreduce; File System; Bigdata; Data Processing