

**T.C.  
SÜLEYMAN DEMİREL ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**HADOOP KULLANARAK METEOROLOJİ VERİLERİNDEN BİR  
İKLİM DEĞİŞİMİ EĞİLİM ANALİZİ**

**Mustafa KAYA**

**Danışman  
Prof. Dr. Tuncay AYDOĞAN**

**YÜKSEK LİSANS TEZİ  
ELEKTRONİK-BİLGİSAYAR EĞİTİMİ ANABİLİM DALI  
ISPARTA - 2018**



© 2018 [Mustafa KAYA]

## TEZ ONAYI

**Mustafa KAYA** tarafından hazırlanan "**Hadoop Kullanarak Meteoroloji Verilerinden Bir İklim Değişimi Eğilim Analizi**" adlı tez çalışması aşağıdaki jüri üyeleri önünde Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü **Elektronik-Bilgisayar Eğitimi Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak başarı ile savunulmuştur.

**Danışman**

**Prof. Dr. Tuncay AYDOĞAN**  
Süleyman Demirel Üniversitesi



**Jüri Üyesi**

**Dr.Öğr.Üyesi Ahmet TOKGÖZLÜ**  
Süleyman Demirel Üniversitesi



**Jüri Üyesi**

**Dr.Öğr.Üyesi Ali KAVURUR**  
Mehmet Akif Ersoy Üniversitesi



**Enstitü Müdürü**

**Prof. Dr. Yasin TUNCER**

## **TAAHHÜTNAME**

Bu tezin akademik ve etik kurallara uygun olarak yazıldığını ve kullanılan tüm literatür bilgilerinin referans gösterilerek tezde yer aldığını beyan ederim.

**Mustafa KAYA**

## İÇİNDEKİLER

İÇİNDEKİLER.....	i
ÖZET .....	ii
ABSTRACT .....	iv
TEŞEKKÜR.....	vi
ŞEKİLLER DİZİNİ .....	vii
ÇİZELGELER DİZİNİ .....	viii
SİMGELER VE KISALTMALAR DİZİNİ .....	ix
1. GİRİŞ.....	1
2. Kaynak Özetleri .....	8
3. MATERYAL VE YÖNTEM .....	18
3.1. Veri Seti.....	18
3.2. Hadoop Büyük Veri İşleme Aracı .....	20
3.2.1. Mapreduce mimarisi.....	21
3.2.1.1. Map fonksiyonu .....	25
3.2.1.2. Reduce fonksiyonu .....	26
3.2.1.3. Mapreduce görevi tanımlama .....	27
3.2.2. HDFS (Hadoop Dağıtık Dosya Sistemi).....	28
3.2.2.1. HDFS namenode ve datanode yazılımları .....	29
3.2.2.2. HDFS girdi çıktı mekanizması (HDFS I/O).....	30
3.3 Trend Analiz Yöntemleri.....	32
3.3.1. Mann-Kendall Testi .....	32
3.3.2. Sperman'ın Rho Testi .....	34
4. BULGULAR.....	35
4.1. Hadoop Ortamının Hazırlanması .....	35
4.2. Veri Setinin Hadoop İle İşlenmesi .....	42
4.3. Verilerin Analizi.....	50
5. SONUÇ VE ÖNERİ .....	58
KAYNAKLAR .....	61
EKLER.....	71
ÖZGEÇMİŞ.....	73

## ÖZET

**Yüksek Lisans Tezi**

### **HADOOP KULLANARAK METEOROLOJİ VERİLERİNDEN BİR İKLİM DEĞİŞİMİ EĞİLİM ANALİZİ**

**Mustafa KAYA**

**Süleyman Demirel Üniversitesi  
Fen Bilimleri Enstitüsü  
Elektronik-Bilgisayar Eğitimi Anabilim Dalı**

**Danışman: Prof. Dr. Tuncay AYDOĞAN**

İklim değışikliklerini tespit edebilmek için, uzun yıllar boyunca yapılan gözlemler sonucunda elde edilen veriler günümüzde büyük veri sınıfında olmaktadır. Bu veri yığınlarının analizi geleneksel yöntemlerle mümkün olmadığından yeni modeller geliştirilmiştir. Map-reduce kütüphanesi de Google tarafından bilişim dünyasına kazandırılan dağıtık mimari üzerinde çok büyük verilerin kolay bir şekilde analiz edilebilmesini sağlayan bu programlama modellerinden biridir. Mapreduce programlama modeli, büyük miktardaki verileri işlemek için yüksek maliyetli süper bilgisayarlar ve veri tabanı sunucuları yerine sıradan bilgisayarlardan oluşan bir ağ kullanır. Ağ üzerindeki bilgisayarlar üzerinde map ve reduce fonksiyonları paralel olarak çalıştırılır. Mapreduce mimarisi ağ üzerinde veriyi taşımak yerine verinin bulunduğu makinelere ilgili map veya reduce fonksiyonunu gönderir. Böylece ağ üzerinde oluşacak bant genişliği problemini ortadan kaldırır. Madreduce modeli farklı alanlarda kullanılabilir. Bu çalışmada, Mapreduce modeli günlük meteorolojik verilerden anlamlı sonuçlar üretebilmek için kullanıldı.

Tüm dünyanın en belirgin sorunu olan iklim değışiklikleri ülkemiz için de büyük bir problemdir. Çünkü iklim değışikliklerinin etkisi ile kullanılabilir tatlı su kaynaklarının azalacağı, ortalama sıcaklıklarının artacağı, az ılıman alanlarda uzun kuraklıklara ve çöller oluşturmaya sebep olabileceği düşünülmektedir. Yapılan tez çalışmasında Karadeniz Bölgesine ait meteorolojik veriler analiz

edilerek söz konusu iklim deęişikliklerin olup olmadığı saptanılmaya çalışılmıştır.

Meteoroloji 10. Bölge Müdürlüğünden günlük yağış, sıcaklık, rüzgâr, nem, basınç, 20 cm toprak altı sıcaklığı, 50 cm toprak altı sıcaklığı, açık yüzey buharlaşma miktarı, güneşlenme süresi ve güneş şiddeti bilgileri alındı. Alınan bu verilerin yıllık ortalamaları tam dağıtık modda çalıştırılan Hadoop/Mapreduce yöntemi kullanılarak hesaplandı. Elde edilen verilerin trend analizleri için parametrik olmayan Mann-Kendall testi ve Spearman'ın Rho testi uygulandı. Çalışma sonunda günlük ortalama sıcaklık, günlük ortalama toprak altı sıcaklığı (50 cm için), günlük ortalama nem ve günlük ortalama açık yüzey buharlaşma verilerinin artan yönde bir trende sahip olduğu görülmüştür.

**Anahtar Kelimeler:** HADOOP, Spearman'ın Rho, Mann-Kendall.

**2018, 72 sayfa**

## **ABSTRACT**

**M.Sc. Thesis**

### **A CLIMATE CHANGE TREND ANALYSIS FROM METEOROLOGICAL DATA USING HADOOP**

**Mustafa KAYA**

**Süleyman Demirel University  
Graduate School of Natural and Applied Sciences  
Department of Electronic-Computer Education**

**Supervisor: Prof. Dr. Tuncay AYDOĞAN**

In order to detect climate changes, the data obtained as a result of observations made over many years are in the large data class nowadays. Since the analysis of these data stacks is not possible with conventional methods, new models have been developed. The Mapreduce library is one of those programming models that allows very large data to be easily analyzed on the distributed architecture that is gained to the world of informatics by Google. The Mapreduce programming model uses a network of ordinary computers instead of high-cost supercomputers and database servers to handle large amounts of data. The map and reduce functions are executed in parallel on the computers on the network. Mapreduce architecture sends relevant map or reduce function to the machines where the data is stored instead of moving the data over the network. This situation eliminates the problem of bandwidth on the network. Madreduce model can be used in different areas. In this study, Mapreduce model was used to produce meaningful results from daily meteorological data.

Climate change, which is the most obvious problem of the whole world, is a major problem in our country, too. It is thought that because of the effects of climate change, available freshwater resources can be reduced, mean temperatures will increase, long term droughts in less temperate areas and deserts. In the thesis study, meteorological data belonging to the Black Sea Region were analyzed and it was tried to determine whether there are climate changes.

Daily precipitation, temperature, wind, humidity, pressure, 20 cm soil temperature, 50 cm soil temperature, open surface evaporation amount, sunshine duration and solar intensity information were obtained from the Meteorology 10th Regional Directorate. This data is calculated using the Hadoop / Map-reduce method, which runs in full distributed mode. For the trend analysis of the obtained data, Mann-Kendall test and Spearman's Rho test, which are non-parametric, were applied. At the end of the study, daily mean temperature, daily average mean soil temperature (for 50 cm), daily mean moisture and daily mean open surface evaporation data were found to have an increased trend.

**Keywords:** HADOOP, Spearman's Rho, Mann-Kendall.

**2018, 72 pages**

## TEŐEKKÜR

Yüksek Lisans Eğitimim boyunca gerek bana karşı tutumuyla gerekse de tecrübeleriyle bana yol gösteren danışman hocam Prof. Dr. Tuncay AYDOĐAN'a teşekkürlerimi sunarım. Ayrıca karşılaştığım zorluklarda her zaman yanımda olan biricik aileme teşekkür ederim.

Mustafa KAYA  
Isparta, 2018



## ŞEKİLLER DİZİNİ

	<b>Sayfa</b>
Şekil 3.1. Tez çalışmasının kapsadığı iller .....	18
Şekil 3.2. Verilerin Hadoop ve trend analiz teknikleri ile işlenmesinin şematik gösterimi .....	20
Şekil 3.3. Map-Reduce veri akış diyagramı .....	22
Şekil 3.4. Örnek programın çalışma mantığı .....	24
Şekil 3.5. Map Fonksiyonu.....	26
Şekil 3.6. Reduce Fonksiyonu.....	27
Şekil 3.7. Map-Reduce görevi tanımlama program kodu .....	28
Şekil 3.8. HDFS'de istemci tarafından dosya okunma şeması .....	30
Şekil 3.9. HDFS'ye istemci tarafından dosya yazılma şeması .....	31
Şekil 4.1. Hadoop ortamı için Tam dağıtık mod yapısı.....	38
Şekil 4.2. Makinenin çalışacağı portu belirleme .....	41
Şekil 4.3. NameNode konfigürasyonu.....	41
Şekil 4.4. Map-reduce konfigürasyonu .....	42
Şekil 4.5. Meteoroloji Genel Müdürlüğü tarafından gönderilen sıcaklık verileri .....	43
Şekil 4.6. Mapreduce projesi oluşturma .....	44
Şekil 4.7. Sıcaklık verilerini işleyen map fonksiyonu için gerekli kodlar .....	45
Şekil 4.8. Sıcaklık verilerini işleyen reduce fonksiyonu için gerekli kodlar..	46
Şekil 4.9. Map ve reduce fonksiyonları için görev tanımlama.....	47
Şekil 4.10. Hadoop kümesi çalıştırıldıktan sonra sistemin genel özellikleri	48
Şekil 4.11. Hadoop kümesi çalıştırıldıktan sonra izleyicinin özet bilgileri ...	48
Şekil 4.12. Sıcaklık verilerinin yıllara göre değişim grafiği.....	54
Şekil 4.13. Mann-Kendall korelasyon testi sonucunda $u(t)$ - $u'(t)$ grafiği.....	54

## ÇİZELGELER DİZİNİ

	<b>Sayfa</b>
Çizelge 3.1. 10. Bölge Müdürlüğünden alınan veriler .....	18
Çizelge 4.1. Karadeniz Bölgesi yıllık ortalama sıcaklık sonuçları .....	49
Çizelge 4.2. Yıllara göre sıcaklık değerleri .....	50
Çizelge 4.3. Sıcaklık için Mann-Kendall Korelasyon Testine Göre u(t) Çizelgesi.....	52
Çizelge 4.4. Sıcaklık için Mann-Kendall Korelasyon Testine Göre u'(t) Çizelgesi.....	53
Çizelge4.5. Mann-Kendall ve Spearman'ın Rho test sonuçları.....	56



## SİMGELER VE KISALTMALAR DİZİNİ

GB	Giga byte
Ghz	Giga hertz
HDFS	Hadoop Distributed File Systemı
IPCC	Intergovernmental Panel on Climate Change
Jvm	Java Virtual Machine
kg	Kilogram
kw	Kilo watt
m <sup>2</sup>	Metre kare
mm	Mili metre
n	Yıl uzunluğu
Sgn	İşaret fonksiyonu
sn	Saniye
VNI	Visual Networking Index



## 1. GİRİŞ

Dünya atmosferi farklı gazlardan oluşmaktadır. Bu gazlar, güneşten gelip atmosferi geçerek yeryüzünü ısıtan ışınların bir kısmını tutarak yeryüzünün ısı kaybını engellemiş olurlar. Atmosferin ısıyı tutma yeteneği sayesinde suların sıcaklığı dengede kalır. Böylece nehirlerin ve okyanusların donması engellenmiş olur. Bu şekilde oluşan atmosferin ısıtma ve yalıtma etkisine sera etkisi denir (Akın, 2006). Sera gazlarının bazıları şunlardır: Su (H<sub>2</sub>O), karbondioksit (CO<sub>2</sub>), metan (CH<sub>4</sub>) ve ozon (O<sub>3</sub>). Bu gazların sera etkisine katkıları ise yaklaşık şu şekildedir: su %36-70, karbondioksit %9-26, metan %4-9, ozon %3-7. Sera gazları dünyayı daha yaşanabilir bir hale getirmiştir. Sera gazlarının etkisi dünyayı yaklaşık 33°C daha sıcak olmasını sağlamaktadır (Çılgın Yamanoglu, 2006). Ancak İnsan aktiviteleri ve sanayileşmenin bir sonucu olarak bu gazların miktarları zamanla artış göstermiştir.

Endüstri devriminden sonra sera gazı emisyonlarında meydana gelen artışlar atmosfer üzerinde değişikliklere sebep olmuştur. Sera gazlarının atmosferdeki oranlarının yükselmesi, yüzey sıcaklıklarında artışa neden olmaktadır (Türkeş vd., 2000 ). Birleşmiş Milletlerin küresel iklim değişikliği konusundaki uzman kuruluşu Hükümetler Arası İklim Değişikliği Paneli (Intergovernmental Panel on Climate Change-IPCC) tarafından yayınlanan 2013 Değerlendirme Raporu'na göre, ortalama yüzey sıcaklığında artış görülmüştür. Raporu göre 1901-2012 döneminde 0,89°C'lik bir artış meydana gelmiştir (Türkeş vd., 2013). Bu sıcaklık artışları okyanusları, buzulları ve küresel deniz yüzeyini de etkilemektedir. Küresel ölçekte, okyanusların üst 75 m'lik katmanı 1971-2010 döneminde her 10 yılda yaklaşık 0,11°C ısınmıştır. IPCC 1. Çalışma Grubu 5. Değerlendirme Raporu'na göre, Grönland ve Antarktika buz kalkanları geçtiğimiz yirmi yıllık dönem içinde kütle kaybına uğramaktadır, buzullar neredeyse küresel ölçekte küçülmeye devam etmektedir. Arktik deniz buzu ve Kuzey Yarımküre'de ilkbahar kar örtüsü alansal olarak azalmasına devam etmektedir. Yine aynı rapora göre, 19. yüzyıl ortasından beri gözlenmiş olan deniz düzeyi yükselmesi oranı (hızı), önceki iki bin yıllık dönemdeki ortalama yükselme oranından daha büyüktür. Küresel ortalama deniz düzeyi 1901-2010 döneminde yaklaşık 19 cm

yükselmiştir (Türkeş vd., 2013). İklimdeki bu değişimler günümüzde de devam etmektedir. Eskiden daha yavaş bir şekilde değişim gösteren iklim, günümüzde hissedilebilecek seviyede hızlı bir değişim içerisindedir. Belirli bir yerin iklim durumu hakkında yorum yapılabilmesi için en az 30 yıllık meteorolojik verilerine ihtiyaç vardır (Beşel ve Kayıkçı,2016).

Dünyada bu değişimler olurken ülkemizin de etkilenmesi kaçınılmazdır. Yurdumuzda iklim değişikliğinin boyutu incelendiğinde, 2050 yılına kadar, Türkiye üzerinde sıcaklık artışları beklenmektedir. Sadece sera gazlarındaki artışlar bile dikkate alınıp hesaplar yapıldığında sıcaklıkta, 1-3°C arasında bir değişim öngörülmektedir. Sera gazlarındaki ve sülfat parçacıklarındaki değişimler ile birlikte dikkate alındığında ise sıcaklıkta 1-2°C arasında bir artış olacağı tahmin edilmektedir. Yağışta ise durum sıcaklığın aksi yönündedir. Ülkemizde yağışın Akdeniz iklim sahasında azalma eğilimi şeklinde olacağı tahmin edilmektedir (Apak ve Ubay, 2007). Bu konuda Meteoroloji Genel Müdürlüğü tarafından daha kapsamlı bir çalışma yapılmış, 2100 yılına kadar olan olası iklim senaryoları üzerinde çalışılmıştır. Çalışma IPCC 5. Değerlendirme Raporu için hazırlanan RCP4.5 senaryosu temelinde üretilen HadGEM2-ES küresel dolaşım modeli çıktılarının, RegCM4.3.4 bölgesel iklim modeli kullanılarak ölçek küçültme yöntemi ile daha yüksek çözünürlükte iklim parametrelerinin yeniden üretilmesini kapsamaktadır (Demir vd., 2013). Bu kapsamda 1971-2000 yıllarına ait 30 yıllık referans verileri kullanılarak 2013-2099 yılları arası olası iklim değişikliği senaryosu oluşturulmuştur. Bu senaryo sonucuna göre 2099'a kadar olan sürecin ilk periyodunda sıcaklık artışı 1-1.5°C arasında olmasına rağmen özellikle 2070-2099 yılları arasında Kıyı Ege ve Güney Doğu Anadolu'da yaz sıcaklıklarında artış 4-5°C arası olması beklenmektedir. Yağışlarda ise, Marmara, Kıyı Ege ve Batı Akdeniz'de sonbahar ve kış yağışlarında ilk dönemde artışlar beklenmektedir. Bu senaryoya göre, 2070-2099 arası zaman dilinde ise Doğu Akdeniz, Güneydoğu Anadolu ve Doğu Anadolu'nun güneyinde kış ve ilkbahar yağışlarında azalmalar beklenmektedir (Demir vd., 2013).

Görüldüğü üzere iklim değışikliklerini tespit etmek için uzun süreli verilere ihtiyaç duyulmaktadır. Belli bir yerin iklim değışikliği durumunu inceleyebilmek için en az 30 yıl gibi uzun süreli ve birçok istasyondan alınan meteorolojik verilerin analizi gerekmektedir. Bu kapsamda elde edilen bu tür veriler günümüzde büyük veri (big data) olarak adlandırılmaktadır.

“Bilgi 21. yüzyılın petrolüdür” der Peter Sondergaard. Bilgi, verinin işlenmiş halidir. Karmaşık ve büyük boyutlu bir veriden çıkarılan anlamlı içerik bilgi olarak adlandırılır. Günümüzde teknolojinin hızlı gelişimine paralel olarak elde edilen veri miktarları da gün ve gün büyümektedir. Sadece Google günde 1.5 trilyon sayfayı analiz etmektedir (Oracle ,2016). CISCO tarafından yayınlanan, Görsel Ağ Endeksi'nin (Visual Networking Index-VNI) 2014- 2019 Küresel Mobil Veri Trafığı Öngörü Raporuna göre 2021 yılına kadar dünya çapında mobil veri trafiğinin yılda 1000 Eksabayta (EB) ulaşacaktır (CISCO, 2017). Büyük veri kavramı geçmiş dönemlerde de karşımıza farklı şekillerde çıkmıştır. Son yıllarda internet kullanımının yaygınlaşması, sosyal paylaşım platformlarının artması, akıllı cihazların yaygınlaşması, teknolojinin ulaşılabilirliğinin artması ve ucuzlaması veri miktarındaki artışını en önemli sebebidir. Eski yaklaşıma göre iş sahibinin ihtiyaçları belirlenir ve bu ihtiyaçlara göre geliştirmeler yapılırdı. Günümüzde ise bilgi teknolojilerinin sunduğu imkânlar ile iş sahibinin özgür bir şekilde veriyi yorumlayabilmesini sağlayacak altyapılar sunulmaktadır (Göksu, 2011).

Geleneksel veri tabanları oldukça işlevsel olmasına rağmen veri miktarı katlanarak artmaya devam ettiği için veriyi bulunduğu yerde işleyecek yazılımlara gereksinim duyulmuştur. Sürekli olarak artmaya devam eden veriler insan yapımı veriler ve makine yapımı veriler olmak üzere iki sınıfta gruplandırılabilir. E-postalar, dosyalar, fotoğraflar, videolar insan yapımı verilere örnektir. Makine yapımı verilere ise sensörler, web logları, ağ verileri, tıklama logları örnek verilebilir. İnsan yapımı veriler makine yapımı verilere oranla daha küçük boyutludurlar. Büyük verinin (Big Data) oluşumunda beş bileşenden söz edilmektedir. 5V şeklinde de adlandırılan bu bileşenler sırasıyla şunlardır: “variety”, “velocity”, “volume”, “verification” ve “value”.

- **Variety (Çeşitlilik):** Veriler(data) farklı ortamlardan ve değişik formatlardan elde edilmektedir. Elde edilen bu verilerin anlamlı bir bütün teşkil edebilmesi için dönüştürülmesi ve bütünleşik olması gerekmektedir.
- **Velocity (Hız):** Sosyal medya gelişimiyle bilgi muazzam bir hızda artmaktadır ve bu hız yüksek ivmeyle artmaya devam etmektedir. Bunun sonucu olarak bu verileri işleyebilecek farklı yöntemlere gereksinim duyulmuştur.
- **Volume (Veri Büyüklüğü):** Yukarıda belirtilen hız sonucunda verilerin büyüklüğü de her geçen gün daha da artmaktadır. Her gün dünyada üretilen veri miktarının yaklaşık 2.5 kentilyon bayt (2.3 trilyon gigabayt) olduğu tahmin edilmektedir
- **Verification (Doğrulama):** Böylesine yoğun bir bilgi akışının olduğu dünyada en önemli konulardan biri veri güvenliğidir. Veri akışının, “bilmesi gereken prensibi”ne göre saklı kalması gerekmektedir.
- **Value (Değer):** Her konuda olduğu gibi, büyük veride de en önemli konu değerdir. Büyük veri, yukarıdaki bileşenlerden filtrelendikten sonra, üretim ve işleme katmanları sonucunda, “şirketiniz ya da kurumunuz için artı bir değer yaratıyor mu?”, “Karar süreçlerinizde anlık olarak istediğiniz bilgiye/veriye ulaşabiliyor musunuz?”, “Doğru kararı almanızda size fayda sağlıyor mu?” gibi sorulara cevap verebilmelidir.

Artık hayatımızın her alanında yer edinen büyük veriler dinamik bir yapıya sahip olduğundan bunları analiz ederken bazı problemlerle karşılaşılmaktadır. Bu problemler şu şekilde sıralanabilir:

- **Verinin elde edilmesi ve depolanması:** Değişik aygıtlardan gelen verilerin sayısı günlük yaklaşık olarak 2,5 Kentilyon byte boyutundadır ve bu rakam her üç yılda yaklaşık olarak ikiye katlanarak devam

etmektedir (Hilbert ve López 2011). Bu kadar büyük miktarda verinin saklanması için depolama alanları oluşturmak ve bakımını sağlamak yüksek maliyetlidir. Bu duruma alternatif olarak bulut bilişim (Cloud Computing) alt yapısı bu sorunun üstesinden gelebilmek için yaygın olarak kullanılmakta olunan bir yöntemdir.

- **Ölçeklenebilirlik:** Elde edilen veri sayısının hızlı bir şekilde artmasıyla, herhangi bir işlem yapılırken taranan ve işlenen veri miktarı da buna paralel olarak artmaktadır. Hatta bir süre sonunda sistem, yapılan işlemlere cevap veremez hale gelebilmekte ve ölçeklenebilirlik problemi meydana gelmektedir. Büyük verilerin bu olumsuz durumu NoSQL (Not Only SQL) gibi ölçeklenebilir veri tabanlarının kullanılmalarının yaygınlaşmasına sebep olmuştur (Cattell,2011). Ayrıca Bulut Bilişim alt yapıları ve Dağıtık Sistemler de (Distributed Systems) büyük verilerdeki bu soruna bir çözüm önerisi getirmektedir.
- **Verinin bütünlüğü:** Farklı kaynaklardan alınan verilerin anlamlı hale getirilmesi için yönetilmesi, analiz edilmesi ve işlenmesi gerekmektedir. Büyük verinin karakteristikleri (büyüklük-verinin elde edilen kaynakların miktarının çok fazla olması, hız-verilerin çok hızlı bir şekilde değişmesi, çeşitlilik- verilerin elde edilen kaynakların çeşitliğinin fazlalığı, güvenilirlik-verilerin güvenilirliği) veri bütünlüğünü bir zorluk haline getirmektedir (Dong, ve Srivastava, 2013). Bütünlük, verilerin niteliklerine ve güvenilirliklerine göre yorumlanmaktadır (Khan vd., 2014).
- **Verinin niteliği:** Girdi verileri kalitesi ne kadar yüksek ise işlem sonunda elde edilen sonuçlar da istenilen kalitede olur. Daha önceki yıllarda veriler iyi bilinen ve sınırlı kaynaklardan geldiğinden işlemlerden daha doğru sonuçlar elde ediliyordu (Redman ve Blanton, 1997). Günümüzde durum daha farklıdır. Veriler güvenilirliği şüpheli birçok farklı kaynaktan elde edilmektedir. Bu durumun sonucu olarak

verinin kalitesi dolayısı ile sonuçların doğruluğu olumsuz etkilenmektedir.

- **Verinin görselleştirilmesi:** Veri boyutunun büyüklüğü ve yüksek boyutlu olmasından dolayı büyük veri uygulamalarında verinin görselleştirilmesini gerçekleştirmek zordur (Chen ve Zhang, 2014). Bu sorunun çözümü için farklı grafik ara yüzleri kullanılarak; daha verimli ve daha ölçeklenebilir ara yüzler geliştirilmesi gerektiği düşünülmektedir. Ayrıca büyük verinin sonucu olan dinamik verilerin sürekli olarak işlenmesi ve işlenen sonuçlarının tekrar görsel olarak gösterilmesi oldukça önemlidir.

Büyük verilerin sahip oldukları bu sıkıntılardan dolayı klasik veri tabanı yöntemleri yerine farklı yöntemler tercih edilmeye başlandı. Günümüzde büyük verileri işlemek için Riak, MongoDB, CouchDB, CassandraDB, HBase, Simpledb, Hadoop gibi yazılım araçları ve teknikleri geliştirilmiştir.

Bu tez çalışmasında Hadoop yazılım aracı meteorolojik verilerin analizi için kullanılmıştır. Karadeniz Bölgesine ait günlük yağış, sıcaklık, rüzgâr, nem, basınç, 20 cm toprak altı sıcaklığı, 50 cm toprak altı sıcaklığı, açık yüzey buharlaşma miktarı, global güneş radyasyonu, güneşlenme süresi ve güneş şiddeti bilgileri önce Hadoop/Mapreduce yöntemi kullanılarak işlenebilir hale getirilmiştir. Daha sonra işlenmiş bu veriler üzerinde Mann-Kendall ve Spearman Rho trend test teknikleri uygulanarak Karadeniz Bölgesi iklim değişikliklerinin analiz edilmesi amaçlanmıştır.

Verileri işlemede kullanılan Hadoop, yüksek maliyetli bilgisayarlar yerine sıradan bilgisayarlardan oluşan küme (cluster) üzerinde büyük verileri işleyen bir sistemdir. Bu sistem Hadoop Distributed File System (HDFS) olarak isimlendirilen bir dağıtık dosya sistemi ile Hadoop Mapreduce özelliklerini bir araya getiren açık kaynaklı bir kütüphanedir. Bu kütüphane Java programlama diliyle geliştirilmiştir. Dean ve Ghemawat çalışmasında HDFS'nin oldukça düşük hata toleransı, yüksek verimlilik ve güvenilirlikle sunucular üzerinde işlem

yapmak için tasarlandığını, Map-Reduce'un ise Google tarafından büyük veri grupları için önerilen bir programlama modeli olduğunu belirtmişlerdir (Aktaran Padhy, 2013). Map-reduce, büyük verilerin dağıtık bilgisayar kümelerinde işlem yapmalarına olanak sağlar. Daha sonra paralel şekilde işlenmiş dağıtık işlemlerin sonucu tekrar birleştirilir (Daniel Q. Duffy vd.,2011 ). Çalıştırılacak programın çalıştığı düğüm (node) ile işlenecek verinin farklı bilgisayarlarda olması dağıtık sistemlerde paralel işlemede verimli çalışmaya engel olan ve işlemlerin daha yavaş gerçekleşmesine sebep olan bir durumdur. HDFS'de bu sorunu çözmek için işleyicinin (task) çalışacağı düğüm ile işleyeceği veri yığınının bulunduğu düğümün yakın olması için çalışılmıştır. Veri yerelliği (data locality) olarak da isimlendirilen bu durum paralel işlemede çalışma hızını ciddi ölçüde artırmaktadır (White, 2009).

Tezin çalışması aşağıdaki konu başlıkları çerçevesinde ilerlemiştir.

- 2. bölümünde konu hakkında literatür özeti verilmiştir.
- 3. bölümde çalışmada kullanılan materyal ve yöntemler tanıtılmıştır.
- 4. bölümde meteorolojik verilerden iklim değişikliklerini ortaya koymak için yapılan Hadoop temeli veri işleme ve trend analizleri anlatılmıştır.
- 5. bölümde çalışma sonunda elde edilen bilgiler sunulmuştur.

## 2. Kaynak Özetleri

Bu bölümde tez konusuna yön veren YÖK'ün tez veritabanında bulunan tezler ile ulusal ve uluslar arası makaleler incelenmiştir.

Alkan (2010), yaptığı tez çalışmasında belli bir ağ içinde önceden belirlenen eşik değerde kendisini tekrar eden küçük ağları bulmaya çalışmıştır. Bu çalışma için açık kaynak kodlu bir ortam olan Hadoop tercih edilmiştir. Tez sonunda, yöntemin büyük ağlara ölçeklenebildiği ve veri miktarının artmasına paralel olarak yöntemin daha yüksek verimle çalıştığı belirlenmiştir.

Akkoyun (2011), yaptığı tez çalışmasında Hadoop ortamını sağlık alanında kullanmıştır. Protein-protein etkileşim ağları görevi henüz bilinmeyen proteinlerin hücredeki biyolojik fonksiyonlarının ne olabileceği konusunda yardımcı olmaktadır. Bu amaçla bilgisayar üzerinde "fonksiyonel akış" denilen bir yöntem kullanılmaktadır. Oldukça başarılı olan bu yöntem gelişmiş canlılar için üretilmekte olan karmaşık ve büyük etkileşim ağları üzerinde bellek yetersizliğinden dolayı uygulanamamaktadır. Bu sorunu çözmek için her biri 8 çekirdekten oluşan 18 makineye Hadoop kurulmuştur. Yapılan çalışma sonunda uygulamanın yüksek bir başarımla çalıştığı belirlenmiştir.

Demir (2012), yaptığı tez çalışmasında, görüntü dosyalarının Mapreduce ile işlenebilmesi amacıyla yeni bir Hadoop eklentisi geliştirmiştir. Çalışmada HDFS'nin daha performanslı çalışabilmesi için küçük imgeleri birleştirilmesi amaçlanmıştır. Böylece daha büyük boyutlu dosyalar elde edilip performans kaybı engellenmiştir. Çalışma sonunda yüz saptama uygulaması yapılarak önerilen tekniğin başarısı gösterilmiştir.

Er (2013), yaptığı tez çalışmasında, Hadoop ortamında genetik algoritmanın paralelleştirilmesini çalışmıştır. Algoritma olarak "Gezgin Satıcı Problemi" seçilmiştir. Oraya çıkan sonuçlar daha önceki uygulamalarda elde edilen sonuçlarla karşılaştırılmıştır. Çalışma sonunda, Mapreduce programlama modeli ile daha hızlı ve daha iyi sonuçlar elde edildiği gözlemlenmiştir.

Söylev (2013), yaptığı tez çalışmasında QNET yöntemini Java programlama dili kullanarak Hadoop ortamında kodlamıştır. QNET, hücre için de meydana gelen olayları açıklamaya yarayan biyolojik ağların sorgu ağıyla karşılaştırıldıktan sonra benzerliğin belli bir eşik değerinin üstünde veya altında olan ağları ayırtmaya yarayan bir yöntemdir. Bu çalışma kapsamında 10 makineye Hadoop kurulmuştur. Yapılan çalışma sonunda bir sorgu ağının tüm veritabanıyla hizalanmasının 14 dakika 42 saniye sürdüğü gözlemlenmiştir. Aynı çalışmanın C++ programlama diliyle gerçekleştirildiğinde ise 6 saat 35 dakika 50 saniye sürdüğü belirtilmiştir.

Çetin (2014), yaptığı tez çalışmasında Hadoop üzerinde kullanılan Hive, Mapreduce ve Pig teknolojileriyle dağıtık RDFS çıkarsama işlemi gerçekleştirip, performanslarını gözlemlemiştir. Ayrıca doküman indeksleme yöntemi de kullanılmıştır. Bunun amacı performansları yükseltmektir. Yapılan çalışmalar sonunda Mapreduce yönteminin performans açısından daha iyi olduğu görülmüştür. Ayrıca doküman indeksleme yöntemi Mapreduce'un çalışma performansını daha da arttırmıştır.

Koca (2014), yaptığı tez çalışması ile firmaların büyük verileri saklamada yaşadığı zorlukları ve çözüm önerilerini sunmuştur. Yapılan çalışmada Hadoop ortamında yapılan mimari düzenlemelerle, verilerin daha hızlı ve daha düşük maliyetle işlenebileceği tespit edilmiştir.

Hallaç (2015), yaptığı tez çalışmasında büyük veri ve bulut bilişim teknolojilerini, büyük veri üzerinde paralel algoritmaların çalıştırılmasını ve dağıtık makine öğrenmesi algoritmalarının büyük veriye uygulanmasını incelemiştir. Ayrıca büyük miktarda log verisinin analizini hem Hadoop hem de Spark yöntemleriyle gerçekleştirip performanslarını karşılaştırmıştır.

Selçuk (2015), yaptığı tez çalışmasında büyük veri ortamında şifreli bulut verisi üzerinde güvenli kelime-tabanlı arama işleminin zor olmasının sebeplerini belirtmiştir. Hızlı ve güvenilir bir arama işlemi için farklı çözümler denenmiştir. Tez çalışmasında HDFS gibi dağıtık dosya sistemi ve Mapreduce

modeli gibi paralel programlama yöntemi kullanılarak bu sorunun etkin bir şekilde çözülebileceği önerilmiştir. Çalışma sonucunda yapılan kapsamlı deneysel sonucunda bu önerinin doğru olduğu belirtilmiştir.

Şarkışla (2015), yaptığı tez çalışmasında minimum donanım maliyeti ile Hadoop sistemine en uygun değer yapılandırma parametrelerini tespit etmeye çalışmıştır. Bu sonucu olarak da uygun “Konfigürasyon Parametre Modeli” ’ni belirlenmesi amaçlanmıştır. Bu çalışma için birbirlerine 1GB/s ile haberleşen anahtarla bağlanmış 5 bilgisayarla bir Hadoop kümesi oluşturulmuştur.

Yılboğa (2015), yaptığı tez çalışmasında kod kitabı tekniği ile gerçek dünya uygulamaları için işlevsel bir içerik tabanlı görüntü sorgulama sistemi geliştirmiştir. Bu çalışmanın iskelet yapısı için Hadoop tercih edilmiştir. Yapılan çalışma sonucunda istenilen görüntüye çok daha hızlı bir şekilde ulaşılabilmesi sağlanmıştır.

Alewiwi (2015), yaptığı tez çalışmasında şifrelenmiş metin verilerini çözmek için yapılan karşılaştırma sayısını büyük ölçüde azaltan yeni bir teknik geliştirmiştir. Yapılan çalışma için Hadoop dağıtık dosya sistemleri (HFDS) ve Mapreduce programlama modelinden faydalanılmıştır. Çalışma sonunda yöntemin diğer tekniklere göre çok daha hızlı olduğu gözlemlenmiştir.

Bayhan (2015), yaptığı tez çalışmasında sosyal medya üzerinde veri toplayan ve toplanan bu veriler üzerinde zamansal analiz yaparak, bu analizleri görselleştiren DeuSocial web uygulaması geliştirmiştir. Büyük miktardaki verilerin analizi işlemleri için Hadoop Mapreduce mimarisini kullanılmıştır.

Yılmaz (2015), yaptığı tez çalışmasında, büyük verilerin analizini yüksek maliyetli ve büyük sistemli yapılar yerine düşük maliyetli ve çok daha az yer kaplayan ARM mimarili mikro bilgisayarlarla nasıl yapılabileceğini belirtmiştir. Çalışmada Hadoop Dağıtık Dosya Sistemi (HDFS) ve Mapreduce programlama modeli kullanılarak bu sistem gerçekleştirilmiştir.

Çetinkaya (2016), yaptığı tez çalışmasında HADOOP yöntemini incelemiştir. Çalışmada k-means algoritmasının Mapreduce metoduna uyarlanması ve etkin analizlerin yapılabilirliği incelenmiştir. Yapılan incelemelerde k-means algoritmasının uyarlanması başarılı bir şekilde gerçekleştirilmiştir. Ayrıca elde edilen ve sonuçların analizi neticesinde mevcut müşteri gruplarının oluşturulma dinamikleri yerine k-means ile müşteri gruplarının oluşturulmasının karlılığı artıracığı ve daha net sonuçlar üreteceği gözlemlenmiştir.

Uzun (2016), yaptığı tez çalışmasında mobil ağlarda, iletişim paketlerini uçtan uca analiz ederek saniye başına yüklenen ve indirilen veri miktarını hesaplayan bir sistem geliştirmiştir. İşlem sonunda elde edilen test sonuçları Hadoop ortamında analiz edilmiştir. Yapılan çalışma ile internet hız karakteristik haritası çıkarılmıştır. Ayrıca mobil ağda oluşabilecek problemlerin önceden tespit edilmesi sağlanmıştır.

Ali (2016), yaptığı tez çalışmasında Hadoop ve Spark sistemlerini karşılaştırıp performanslarını incelemiştir. Farklı parametre ayarlarıyla 4 farklı uygulama test edilmiştir. Çalışma sonunda WordCount ve Sort uygulamalarında Spark daha iyi sonuç verirken Naive-Bayes uygulaması için Hadoop çok daha iyi performans göstermiştir. Grep uygulaması içinse Hadoop ve spark arasında kayda değer bir fark görülememiştir.

Salur (2016), yaptığı tez çalışmasına Hadoop ortamında veri madenciliği teknikleri kullanarak duygu analizi yapmaya çalışmıştır. Çalışma için 105.00 haber başlığı, veri olarak kullanılmıştır. Analiz işlemleri hem tek düğüm, hem de 4 düğümlü Hadoop sisteminde gerçekleştirilmiştir. Çalışma sonunda sınıflandırma işleminde yaklaşık olarak %80 başarı elde edildiği gözlemlenmiştir.

Çiftçi (2016), yaptığı tez çalışmasında bir sağlıklı gıda tüketimi arama hizmeti olan Mobile APP Search Service (MSS) uygulamasını anlatmıştır. Uygulamanın en önemli özelliği Hadoop ve Mapreduce programlama modelini kullanmasıdır.

Bu özelliğinden dolayı bu uygulamanın çok daha hızlı çalışabildiği çalışmada belirtilmiştir.

Toros (1993), yaptığı tez çalışmasında Türkiye genelini temsilen 18 istasyona ait yağış ve sıcaklık verilerini almıştır. Alınan bu veriler üzerinde Mann-Kendall ve Spearman Rho trend analizi tekniklerini kullanılmış ve bölgesel olarak iklim değişikliğinin olup olmadığını tespit edilmeye çalışılmıştır. Çalışma sonunda Türkiye genelinde yağışlar konusunda genel olarak bir trend eğilimine rastlanılmamıştır. Ancak Türkiye genelinde gece sıcaklıklarında önemli artışlar tespit edilmiştir. Özellikle bu artışlar ilkbahar mevsiminde daha belirgin şekilde ortaya çıkmıştır. Ayrıca verilerde görülen trend değişimleri 1950'li yıllarda başlamaktadır. Bu durumun Türkiye'nin endüstrileşmeye başlaması, hızlı şehirleşme ve artan nüfus yoğunluğu ile ilişkili olabileceği düşünülmüştür.

Partal (2003), yaptığı tez çalışmasında Türkiye'deki iklim değişiklerini incelemek için 96 istasyona ait yağış verilerini incelenmiştir. Yağış verileri üzerinde Mann-Kendall ve Sen'in T testi yöntemleri kullanılmıştır. Çalışma sonunda Türkiye'nin batı ve güneyi ile Karadeniz kıyı şeridinde ocak, şubat ve eylül aylarında azalma eğilimleri görülmüştür. Ancak bu bilgiler iklim değişikliğinin olduğunu onaylamak için yeterli olmamıştır.

Gürkan (2005), yaptığı tez çalışmasında Seyhan Havzası üzerinde bir hidrolik model oluşturmuştur. Oluşturulan model sayesinde iklim değişiminin yüzey suyu kaynaklarına olan etkisi incelenmiştir. Çalışma kapsamında yağışın azalması ile sıcaklık ve buharlaşmanın artması dikkate alınarak 2 farklı senaryo oluşturulmuştur. Oluşturulan senaryolara göre Seyhan havzasının iklim değişimlerine karşı oldukça duyarlı olduğu tespit edilmiştir.

Em (2005), yaptığı tez çalışmasında Küresel ısınmanın GAP bölgesindeki etkilerini incelemiştir. Güneydoğu Anadolu Bölgesi Türkiye'nin önemli bir tarım bölgesi olduğundan büyük projeler yapılmıştır. Bunlardan en büyüğü ise GAP'tır. Çalışmada 1970-2013 yılları arası 15 istasyondan günlük yağış verileri toplanmış ve bu veriler üzerinden trend analizi tekniklerinden t testi ve Mann-

Kendall testi uygulanarak kuraklığın gidişatı tespit edilmeye çalışılmıştır. Elde edilen verilerin sonuçlarına göre yıllık toplam yağış verilerinde 10 istasyonda artış görülürken 5 istasyonda azalma tespit edilmiştir. Ayrıca GAP sayesinde inşa edile barajlar sayesinde tutulan su miktarı arttığından bölgedeki nemliliğin arttığı tespit edilmiştir.

Önol (2007), yaptığı tez çalışmasında sera gazlarının iklim değişimlerini belirleyebilmek için bir iklim senaryosu oluşturmuştur. Bu amaçla 1961-1990 ve 2071-2100 arası yıllar için olası senaryolar hazırlanmıştır. Çalışma sonunda 2071-2100 yılları arasında Balkan ülkelerinde yaz sıcaklığında 7 °C ye varan artışların olacağı görülmüştür. Ülkemizde ise genel olarak 3 ila 4 °C arasında bir değişim beklenmektedir. Ege Bölgesi'nde 5 °C'den fazla sıcaklık artışının olabileceği belirlenmiştir.

Özfidaner (2007), tez çalışmasında 1932-2002 yılları arasındaki Türkiye genelindeki gözlem istasyonlarından alınan aylık ve yıllık yağış verilerinin eğilimlerini belirlemek için Mann-Kendall ve t-testi uyguladı. Çalışmada elde edilen değerler, 1968-1997 yılları arasında daha önceden akım verileri ile karşılaştırılmıştır. Çalışma sonunda noktasal ölçekte 7 bölge için yağışları kış aylarında azalma eğiliminde olduğu, buna karşın ilkbahar ve yaz aylarında ise artma eğiliminde olduğu görülmüştür. Bölgesel ölçekte ise Güneydoğu Anadolu Bölgesi'ndeki yağış verilerinde azalma eğilimi görülmektedir.

Ecer (2009), yaptığı tez çalışmasında Fırat Havzası için Elektrik İşleri Etüt İdaresi (EİEİ) ve Devlet Meteoroloji İşleri (DMİ) tarafından işletilen istasyonların verilerinden faydalanmıştır. Çalışmada EİEİ ait 21 akım gözlem istasyonlarının yıllık minimum, yıllık maksimum ve yıllık ortalama değerleri ile DMİ ait 38 gözlem istasyonunun yıllık toplam yağış, yıllık maksimum yağış, yıllık ortalama sıcaklık, yıllık maksimum sıcaklık, yıllık minimum sıcaklık, yıllık ortalama nem, yıllık maksimum nem, yıllık minimum nem değerleri kullanılmıştır. Bu değerler üzerinde Mann-Kendall testi ve Spearman'ın Rho testleri ile Sen'in Trend Eğim metodu kullanılmıştır. Daha sonra elde edilen sonuçlar üzerinde "harita bindirme" (overlay mapping) tekniği kullanılarak

parametrelerdeki deęişiklikler incelenip sebep-sonuç iliřki mantığı yorumlanmıřtır. alıřma sonunda havzanın gney, batı ve gneybatı kesimlerinde yıllık minimum akım deęerlerinde anlamlı bir azalan trend tespit edilmiřtir. Sıcaklık ve nem deęerlerinde ise artan bir eęilim söz konusudur. GAP kapsamında gerekleřtirilen barajların nem deęerlerindeki artıřa etken olduęu da belirtilmiřtir.

Bahadır (2011), Gneydoęu Anadolu Proje (GAP) alanında sıcaklık ve yaęıř verileri zerinde trend analizi alıřması yaptı. Yaptığı alıřmanın sonucuna gre kresek ısınmanın Gneydoęu Anadolu Blgesi'ni etkileyeceęi belirtilmiřtir. alıřmada 1992 yılından sonra sıcaklıklarda ykselme eęilimini olduęu belirtilmiřtir. Buna karřın aynı blgede yaęıřlarda azalma eęiliminin olduęu grlmektedir. alıřma sonunda bu durumun l alanlarının geniřlemesine etki edeceęi vurgulanmıřtır.

Ghali (2011), yaptığı tez alıřmasında iklim deęiřiklięinin bilimsel temelini ve olası etkilerini incelemiřtir. Ayrıca iklim deęiřiklięinin Finlandiya ve Fas orman endstrisine etkilerini belirtmiřtir. Yapılan tez alıřmasında iki farklı teknik kullanılmıřtır. Birincisi konuyla alakalı kaynaklar arařtırılıp incelenmiřtir. İkinci yntem ise uzmanlardan oluřan bir mlakat yapılımasıdır. Katılanların oęunluęu Fin veya Fas'tandır. alıřma sonunda kaęıt endstrisinin sera gazı salınımlarında etkisinin byk olduęu belirlenmiřtir. Ayrıca hem Finlandiya'da hem de Fas'ta gelecek 10 yıl iinde yarı derece sıcaklık artıřı olacaęı belirtilmiřtir.

Yılmaz (2012), yaptığı tez alıřmasında, Trkiye ve blgesi iin u iklim olaylarının daęılımını incelemiřtir. Yapılan alıřma iin istasyon ve dinamik olarak klmř iklim projeksiyon verileri kullanılmıřtır. Yapılan analiz alıřmaları genellikle u yaęıřlar ve u sıcaklıklar zerine yoęunlařmıřtır. U deęer istatistiklerinin hesaplanması iin U Deęer Analizi (UDA) teknięi kullanılmıřtır. 30 yıllık verilerin analizi sonucunda u deęerler perspektifinden bakıldıęında minimum ve maksimum yaęıř u deęerlerinin zamanla azaldığı buna karřın sıcaklık u deęerlerinin ise ykseldięi grlmřtir.

Azadani (2012), yaptığı tez çalışmasında, iklim değişikliklerinin Colorda' daki en büyük nehirlerden biri olan Arkansas Nehrine ve bu nehir havzasına etkilerini araştırmıştır. Çalışmada 2013-2040 yılları arasında Arkansas Nehir Havzasındaki gelecek yıllara ait sıcaklık ve yağış tahminlerinde bulunulmaya çalışılmıştır. Bu amaçla 3 farklı senaryo oluşturulmuştur. Çalışma sonunda, her üç senaryoya göre de iklim değişikliklerinin Arkansas Nehir Havzası üzerinde olumsuz etkileri olduğu görülmüştür.

Macana (2014), yaptığı tez çalışmasında Büyük Menderes Havzası bölgesindeki 8 meteorolojik istasyonun 1970-2011 yıllarına ait aylık ortalama sıcaklık, aylık ortalama minimum ve aylık ortalama maksimum verilerini kullanmıştır. Elde edilen verilerin zamansal değişimleri iklim değişimi bağlamında incelenmiştir. Bu amaçla Mann-Kendall ve Sen eğilim testleri uygulanmıştır. Çalışma sonunda istasyonların çoğunda artış eğilimleri gözlemlenmiştir. Ortalama sıcaklıkta en yüksek artış 0,050°C /yıl ile Denizlide olmuştur. Ortalama maksimum sıcaklıkta en yüksek artış 0,062°C/yıl ile Yatağan'da olurken, ortalama minimum sıcaklıkta artış 0,067°C/yıl ile Denizli'de tespit edilmiştir.

Baltacı (2014), yaptığı tez çalışmasında güneş aktivitelerindeki değişimle iklim değişimleri arasında bir ilişki olup olmadığını araştırmıştır. Bu çalışma için Meteoroloji Genel Müdürlüğü'nden 19 istasyona ait ortalama yüzey sıcaklık bilgileri ve Nasa'dan güneş radyasyonu göstergesi olarak güneş lekesi verileri alınmıştır. Bu veriler karşılaştırılarak analizler yapılmıştır. Yapılan analiz sonuçlarına göre sıcaklık ve güneş aktivitesi arasında %80 oranında korelasyon mevcuttur.

Katzenbergeri (2014), yaptığı tez çalışmasında, küresel ısınmanın canlı türleri üzerindeki etkisini araştırmıştır. Bu çalışma için farklı iklim bölgelerinden 71 amfibi türü incelenmiştir. Her bir tür için Isınma Toleransı ve Termal Emniyet Marjı belirlenmiştir. Çalışma sonucunda, ılıman bölgedeki bazı türlerin küresel ısınmadan olumsuz etkilendiği belirlenmiştir.

Slagle (2014), yaptığı tez çalışmasında iklim değişikliklerinin Myanmar üzerinde ki etkilerini ve Myanmar hükümetinin bu konu için aldığı önlemleri incelemiştir.

Myanmar IPCC raporuna göre oldukça hassas bölgelere sahiptir. Sıcaklıkların yükselmesi, yağışlardaki değişkenlikler ve deniz seviyesinin yükselmesi tarım ürünlerinde elde edilen verimi azaltmaktadır. Çalışma sonunda, hükümet önlemler olsa da çok yavaş şekilde ilerleme kaydettiğinden Myanmar'ı iklim değişikliğine karşı oldukça savunmasız bıraktığı belirtilmiştir.

Çeribaşı ve Doğan (2015), yaptıkları çalışma da Batı Karadeniz Havzası, Doğu Karadeniz Havzası ve Sakarya Havzası için yıllık ortalama yağışların trend analizini yaptı. Analiz sonuçlarına göre; Batı Karadeniz Havzasının bazı bölgelerinde azalan yönde trendler olduğu görülmüştür. Doğu Karadeniz Havzası trend analizi sonuçlarına bakıldığında havzanın bazı bölgelerinde artan yönde trendlere rastlanılmaktadır. Sakarya Havzası trend analizi sonuçlarına bakıldığında ise havzanın çoğu bölgelerinde azalan yönde trendler görülmüştür. Artan yada azalan yönde trend gösteren bölgelerde değişime neden en büyük sebep ülkemizde son yıllarda meydana gelen iklim değişikliğidir. Oluşan iklim değişikliği, ülkemizin önceki yıllara göre bazı bölgelerinde sıcaklıkların artmasına ve daha az yağış almasına neden olurken bazı bölgelerin ise daha fazla yağış almasına sebep olmaktadır.

Bulut (2015), yaptığı tez çalışmasında 1985 - 2014 yılları arasında, 13 il, 50 ilçe için haşhaş ekimi yapılan bölgelerdeki haşhaş verimini ve haşhaş bitkisinin gelişimini en çok etkileyen faktörler incelenmiştir. Bu amaçla toprak sıcaklığı, minimum sıcaklık, ortalama sıcaklık, maksimum sıcaklık, ortalama nem, donlu günler sayısı, toplam yağış ve toplam güneşlenme süresi verileri elde edilmiştir. Bu verilerin arasındaki ilişkiler korelasyon Analizi yöntemiyle incelenmiştir. Ayrıca bu çalışmada, HadGEM2-ES küresel iklim modeli kullanılarak, 2016 - 2099 yılları arasındaki iklim değişimlerinin haşhaş verimine olası etkileri açıklanmıştır. Çalışma sonunda iklim faktörlerinin haşhaş verimi üzerinde önemli etkileri olduğu belirlenmiştir. HadGEM2-ES küresel modeli kullanılarak gerçekleştirilen senaryoya göre; haşhaş tarımı yapılan iller gelecek dönemlerde yaşanması muhtemel iklim değişikliklerinden etkilenecektir.

Çaylak (2015), yaptığı tez çalışmasında olası iklim değişikliği senaryolarının Kırıkkale'de yetiştirilen buğday bitkisine etkisini araştırmıştır. Bu amaçla

DSSAT bitki iklim benzetim modellerinden Ceres-Wheat modeli kullanılmıştır. Model için gerekli veriler Kırklareli Toprak Su ve Tarımsal Meteoroloji Araştırma İstasyonu'ndan temin edilmiştir. Çalışma sonunda sıcaklık artışının ve yağış miktarındaki azalış buğday verimine olumsuz etki ettiği belirtilmiştir. Ayrıca 2013-2100 yılları için yapılan senaryoya göre buğday veriminin %3 ile 6 arasında azalacağı gözlemlenmiştir.

Lahera (2017), yaptığı tez çalışmasında iklim değişikliğinin Amerika'nın doğu kesimlerindeki ürünlere etkisini belirtip 2 farklı çalışma yapmıştır. İlk çalışmada Orta Atlantik'te uygulanan yöntemler için görüşmeler ve anketler yapılmıştır. İkinci çalışmada ise gelecekte Amerika'nın doğu kesimlerindeki mısır ve soya ürünlerine iklim değişikliğinin etkilerini ve bu ürünlerin adaptasyonlarını değerlendirmek için APEX (Agricultural Policy/Environmental Extender) modeli kullanılmıştır. Çalışma sonunda iklim değişiklikleri zaman içerisinde ürünler üzerinde artan bir olumsuz etkiye sahip olduğu görülmüştür. Özellikle mısır ürününün, soya ürününe göre daha fazla etkilendiği belirtilmiştir.

Yukarıdaki tez ve taramalarından ülkemizde Hadoop ve iklim değişiklikleri üzerine çeşitli çalışmalar yapıldığı görülmektedir. İklim değişikliği ile ilgili çalışmalar eski tarihlere uzansa da yapılan çalışmalar sadece sıcaklık ve yağış verileriyle sınırlıdır. Bu tez çalışmasında ise çok daha fazla veri üzerinde analiz yapılmıştır. Bu durum yapılan tez çalışmasını diğer çalışmalardan ayıran en önemli özelliktir. Elde edilen büyük miktardaki verilerden anlamlı sonuçlar çıkarmak Hadoop'tan faydalanılmıştır. Hadoop'un büyük verileri analiz etmek için kullanılması ve son yıllarda kullanımının yaygınlaşması tercih edilmesindeki en önemli sebeptir.

### 3. MATERYAL VE YÖNTEM

Bu bölümde, çalışmada kullanılan veriler, Hadoop yapısı ile Spearman'ın Rho ve Mann Kendall trend analizi yöntemleri açıklanmıştır.

#### 3.1. Veri Seti

Meteoroloji; kısaca atmosfer bilimidir. Yunanca "meteoron" kelimesinden adını almıştır ve gökyüzünde olan olaylar anlamına gelmektedir. Meteorolojiye ilgi çok eski tarihlere dayanmaktadır. Eski Yunanlılar bulutları, rüzgârları ve yağmuru anlamak ve birbirleriyle ilişkilerini tespit etmek için rasat yapmışlardır. Ülkemizde ise yapılan çalışmalar Selçuklular zamanına kadar uzanmaktadır. Ancak bunlar daha ziyade astronomik gözlemlerde bulunmak içindi. 1500'lü yılların son çeyreğinde Osmanlı İmparatorluğu'nda astronomi ve meteoroloji alanında önemli çalışmalar yapılmıştır. Tanzimattan sonra da bu çalışmalar devam etmiştir. Osmanlı İmparatorluğu'nda meteorolojinin kurumsallaşma çalışmaları 1867 yılında Kandilli Rasathanesi'nin kurulması ile başlamış ve bu kurumsallaşma 1937 yılında Cumhuriyet Türkiye'sinde tamamlanmıştır (MGM, 2018). Günümüzde Meteoroloji Genel Müdürlüğü, bünyesindeki 15 bölge müdürlüğü ile Orman ve Su İşleri Bakanlığı'na bağlı olarak çalışmaktadır.

Tez çalışması kapsamında gerekli veriler Meteoroloji 10. Bölge Müdürlüğünden elde edilmiştir. Tez çalışması Düzce, Bolu, Zonguldak, Bartın, Karabük, Kastamonu, Sinop, Samsun, Çorum, Amasya, Tokat, Ordu, Giresun, Gümüşhane, Trabzon, Bayburt, Rize ve Artvin illerini kapsamaktadır. Kapsama alanı Şekil 3.1.'de harita üzerinde gösterilmiştir.



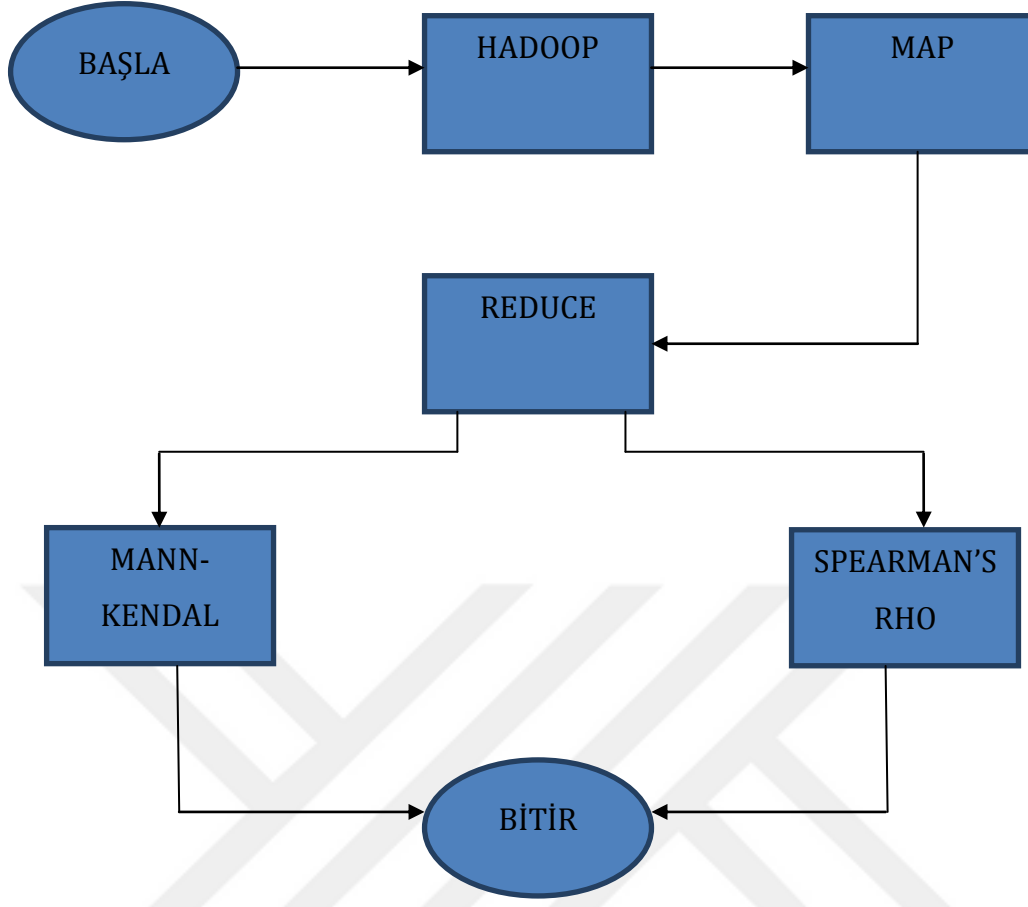
Şekil 3.1. Tez çalışmasının kapsadığı iller

Bu çalışma için Meteoroloji 10. Bölge Müdürlüğünden günlük yağış, sıcaklık, rüzgâr, nem, basınç, 20 cm toprak altı sıcaklığı, 50 cm toprak altı sıcaklığı, açık yüzey buharlaşma miktarı, global güneş radyasyonu, güneşlenme süresi ve güneş şiddeti verileri alınmıştır. Alınan bu veriler ve özellikleri Çizelge 3.1’de verilmiştir.

Çizelge 3.1. 10. Bölge Müdürlüğünden alınan veriler

Veri Adı	Verinin Süresi	Veri Adedi	İstasyon Adedi
Günlük Toplam Yağış (mm=kg÷m <sup>2</sup> )	1927-2017	1.840.092	433
Günlük Ortalama Sıcaklık (°C)	1926-2017	1.733.686	448
Günlük Güneşlenme Süresi (saat)	1947-2017	482.977	162
Günlük Ortalama Toprak Sıcaklığı 20 cm için (°C)	1929-2017	18.913	114
Günlük Ortalama Toprak Sıcaklığı 50 cm için (°C)	1927-2017	18.913	111
Günlük Ortalama Aktüel Basınç (hPa)	1927-2017	798.674	135
Günlük Ortalama Nispi Nem (%)	1927-2017	1.115.931	318
Günlük Ortalama Rüzgâr Hızı (m÷sn)	1927-2017	1.653.821	355
Günlük Toplam Açık Yüzey Buharlaşma (mm)	1936-2017	189.130	30
Günlük Toplam Global Güneş Radyasyonu (kwsaat÷m <sup>2</sup> )	2005-2017	61.334	84
Günlük Toplam Güneş Şiddeti(cal÷cm <sup>2</sup> )	1967-2017	64.720	132
<b>Toplam</b>		7.978.191	

Çizelge 3.1’ de görüleceği gibi günlük yağış, sıcaklık, rüzgâr, nem, basınç, 20 cm toprak altı sıcaklığı, 50 cm toprak altı sıcaklığı, açık yüzey buharlaşma miktarı, global güneş radyasyonu, güneşlenme süresi ve güneş şiddeti bilgilerini içeren 11 farklı parametre mevcuttur. Tez çalışmasında 11 farklı parametrede 1926-2017 tarihleri arasına ait toplam 7.978.191 adet veri üzerinde çalışılmıştır. Verilerin Hadoop ortamında işlenip, trend analizleri ile yorumlanması aşamaları Şekil 3.2.’de gösterilmiştir.



Şekil 3.2. Verilerin Hadoop ve trend analiz teknikleri ile işlenmesinin şematik gösterimi

### 3.2. Hadoop Büyük Veri İşleme Aracı

Hadoop, Mapreduce yöntemini kullanarak büyük veri dosyalarının paralel şekilde işlenmesini sağlayan açık kaynak kodlu bir yapıdır (Lam, 2010). Bilgisayar kümeleri üzerinde çalışan dağıtık programlar için alt yapı desteği sunmaktadır (Er, 2013).

Hadoop, kullanıcı etkileşimine ihtiyaç duymadan verilerin ilgili düğüme taşınmasını ve olası bir donanımsal sorunda kurtarma işlemlerinin otomatik olarak gerçekleştirilmesini sağlar. Kullanılan Mapreduce tekniğinde, yazılan program bilgisayar kümesinde bulunan bir düğümde (node) çalışabilecek küçük parçalara ayrılır. Ayrıca Hadoop bilgisayar kümesi için yüksek bant genişliğine olanak sağlayan dağıtık dosya sistemini (Hadoop Distributed File System -

HDFS) kullanıcıya sunar (White, 2009). Hem Mapreduce hem de HDFS dosya sisteminin birlikte kullanımıyla düğüm hataları otomatik olarak düzeltilir. Böylece kullanıcıya çok daha yüksek güvenilirlikli bir sistem sunulmuş olunur.

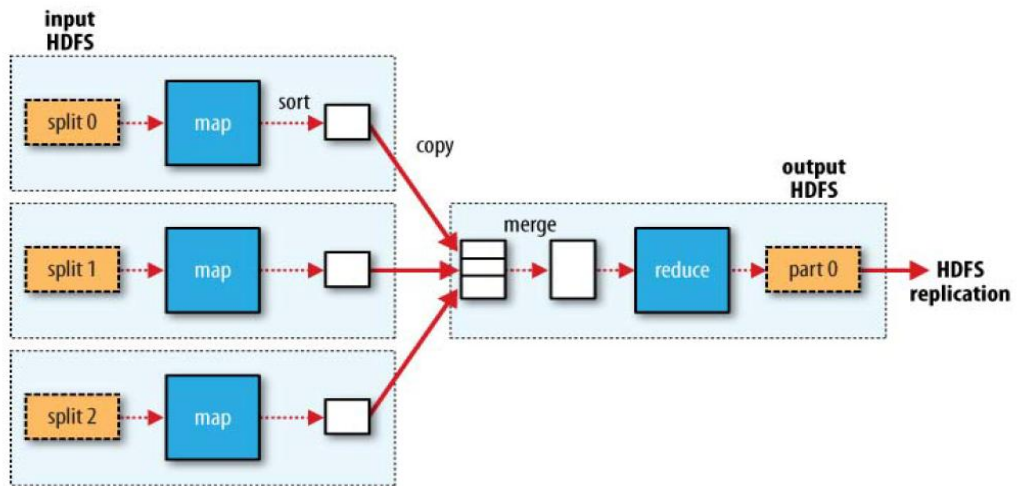
Hadoop yatay olarak büyüyerek pahalı ve güçlü bilgisayarları zorunlu tutmadan büyük veriye alt yapı sağlamaktadır. Yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış veriler Hadoop ile tutulabilmektedir. Hadoop'un bir diğer özelliği de veriyi saklamaya ve dağıtık hesaplamaya aynı anada izin veriyor olmasıdır.

Dünyada birçok firma Hadoop'tan yararlanmaktadır. Adobe, Alibaba, Ebay, Facebook bu firmalardan bazılarıdır (Wang, 2018).

### **3.2.1. Mapreduce mimarisi**

Mapreduce büyük veri kümelerindeki verilere en hızlı şekilde erişip işlemeyi sağlayan bir programlama modelidir (Dean ve Ghemawat, 2008). İlk olarak Jeffrey Dean ve Sanjay Ghemawat tarafından ortaya atılan bu model, Google sunucularında kayıtlı verilerin paralel işlenmesi amacıyla geliştirilmiştir. Günlük olarak kaydedilen web sayfalarının kütüğe yazılması sonucu oluşan veri kümesi içerisinde çok kısa sürede sonuç üretecek şekilde arama yapmak için geliştirilmiştir. Binlerce paralel çalışan bilgisayarı kullanabilecek şekilde tasarlanmıştır. Bu noktada dağıtık sistemlerdeki verinin parçalanması, dağıtılması ve işlemin paralelleştirilmesi gibi problemler ortaya çıkmıştır. Bu problemleri aşabilmek için yük dengeleme (load balancing), verinin dağıtılması ve hata toleransı işlemleri için bir soyutlama modeli tasarlanmıştır. Geliştirdikleri Mapreduce yöntemi bu nedenlerle kullanıcı tarafından kolayca kullanılabilir ve dağıtık sistem üzerinde kolayca işlem yapmaya izin veren bir ara yüze sahiptir. Basit ve güçlü olan bu sistem sayesinde kullanıcı sadece map ve reduce fonksiyonlarını yazarak paralelleştirilmesini istediği işi tanımlar, sistem bu iş için gerekli veri bölütlenmesini, paralelleştirmeyi ve hata yönetimini kendi yönetir. Mapreduce kullanarak paralelleştirilebilen farklı uygulamalar mevcuttur. Bunlara örnek olarak web sayfası erişim frekansı

tarama ve bulma, arama motorlarında kelime veya dosya arama (Ghemawat vd., 2003), büyük boyutlu astronomik imgeler üzerinde işlem yapmak (Wiley K vd., 2010), dağıtık olarak sıralama algoritmaları çalıştırmak ve paralel görüntü işlemek (Krishna vd., 2010) gibi paralel hesaplama gerektiren bazı işlemler verilebilir. Mapreduce modeli daha sonra Apache yazılım topluluğu tarafından Hadoop adında açık kaynak kodlu bir proje olarak yeniden geliştirilmiştir. Hadoop ücretsiz bir sistem olduğundan çok geniş bir geliştirici topluluğu tarafından kullanılmaktadır. Hadoop kullanıcıları arasında Yahoo gibi büyük kurumlar da mevcuttur. Şekil 3.3.'deki veri akış diyagramında görülen Hadoop'ta paralelleştirilecek hesaplama bir görev (job) olarak tanımlanır. Bu görev geliştirici tarafından kodlanan map ve reduce fonksiyonlarını içerir. Bu soyutlamada ilham kaynağı Lisp gibi fonksiyonel dillerdeki benzer fonksiyonlar olmuştur. Mapreduce sistemi bu fonksiyonların birçok kopyasını paralel olarak farklı makinelerdeki işleyicilerde çalıştırır. Map fonksiyonu anahtar/kilit (key/value) ikilisinden oluşan bir kayıt bilgisini girdi olarak alır ve yine anahtar/kilit ikilisinden oluşan bir çıktı verisini oluşturur. Çok sayıda paralel çalışan bu Map fonksiyonları Hadoop sisteminde çıktılarını ortanca çıktı (intermediate output) olarak tutarlar ve bu çıktılar reduce fonksiyonlarına girdi olarak gönderilir. Tanımlı reduce fonksiyonlarını çalıştıran işleyiciler bu ortanca çıktıları girdi olarak kabul edip bunları tekrardan düzenleyip birleştirerek nihai çıktıyı oluşturur. Bu yeniden düzenleme sırasında aynı anahtar değerine sahip olan ortanca çıktıların değerleri birleştirilir.



Şekil 3.3. Mapreduce veri akış diyagramı (White,2009)

Mapreduce çalışma şeklini biraz daha detaylı incelemeden önce bu tez içerisinde kullanılan bazı kavramların açıklanması gerekmektedir.

**Makine:** Dağıtık sistemi oluşturan fiziksel bilgisayarlardan her birine denir.

**İşleyici (Task):** Her makine de çalışan paralelleşmiş her bir Mapreduce iş parçacığı. Map işleyicisi ve Reduce işleyicisi olarak tanımlanır.

**Düğüm (Node):** Hadoop tarafından her bir makineye verilen dağıtık sistemdeki rol veya tanım. Her makine aslında bir düğüm olarak görev almaktadır.

**Girdi Parçacığı (Input Split):** İşleyicilerin dosya sisteminden okudukları ve işleyecekleri veri parçacığıdır.

**Ortanca Çıktı (Intermediate Output):** Map fonksiyonları tarafında oluşturulan ara çıktı verisi. Bu veri anahtar/değer ikilisi şeklinde olup Reduce fonksiyonuna girdi olarak verilir.

Mapreduce görevinin çalışma şekli aşağıdaki adımlar şeklinde incelenebilir:

Hadoop Mapreduce sisteminde kayıtlı olan büyük boyutlu dosyalar öncelikle dosya sistemi varsayılan blok boyutu olan 64 megabyte boyutundaki parçacıklara ayrılır ve bu parçacıklar farklı düğümlerde kaydedilir. Bir parçanın birden fazla kopyası sistemde olabilir (replication), bunun sebebi hata oluşan düğümdeki verinin kaybolmasını engellemektir. Ayrıca replication yapılan veri paralel işleme sırasında daha fazla düğümde bulunacağından daha hızlı bir şekilde işlenecektir. Bu çalışmada 3 düğüm kullanılmıştır.

HDFS sistemindeki düğümlerden bir tanesi yönetici (master) düğümdür. Yönetici düğüm diğer işçi düğümlere işleyici olarak görevler verir. Bir işleyici Map işleyicisi veya Reduce işleyicisi olarak farklı görev alabilir. Sistemdeki Map işleyicisi sayısı işlenecek veri boyutlarına göre otomatik olarak belirlenir. Reduce sayısı kullanıcı tarafından görev ayarları yapılırken belirlenebilir.

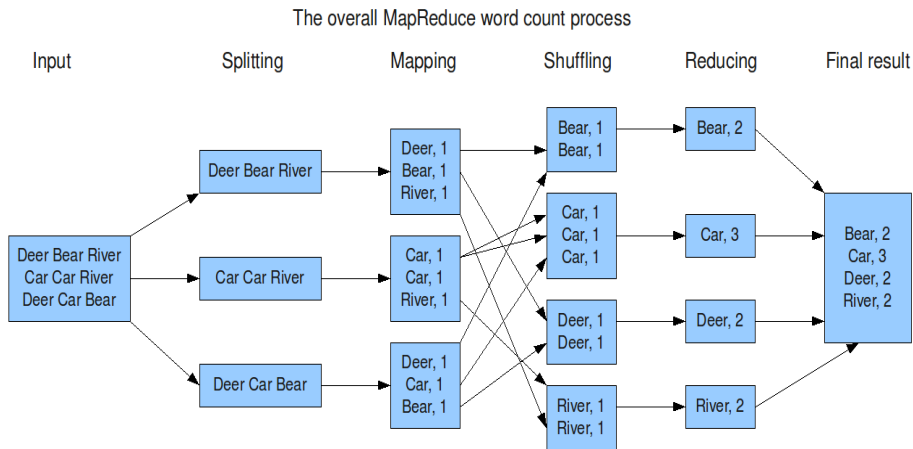
Map işleyicisi girdi parçacığı verisini sistemden okuyarak bundan map fonksiyonu için anahtar/değer ikilisini içeren kayıt verilerini oluşturur. Kayıt girdileri ile kullanıcı tarafından oluşturulmuş olan map fonksiyonları çağrılır. Map fonksiyonlarının oluşturduğu ortanca çıktı ilgili makine tarafından reduce fonksiyonuna gönderilmek üzere geçici hafızaya ve sonra da yerel diske kaydedilir. Reduce fonksiyonuna gönderildikten sonra bu veri silinir.

Oluşturulan ortanca çıktı verileri Reduce işleyicilerine paylaşılır ve Reduce işleyicileri bu verileri okuyarak ortanca anahtar verilerine göre sıralar ve guruplar. Böylelikle aynı anahtar değerine sahip kayıtların değerleri gruplandırılmış olur. Kullanıcı isterse bu gruplama ve sıralama şeklini değiştirebilir.

Reduce fonksiyonu içerisine gruplandırılmış yeni anahtar/değer ikilileri girdi olarak verilir. Kullanıcı tarafından oluşturulmuş reduce fonksiyonu nihai işleme tabi tutup oluşturduğu çıktıları çıktı dosyasına ekler.

Yönetici düğüm, görevi çalıştıran kullanıcıya görevin tamamlanma bilgilerini sunar ve görevi sonlandırır (Demir, 2012).

Şekil 3.3.'deki Hadoop çalışma mantığını daha iyi anlayabilmek için bir text dosyasındaki her bir kelimenin kaç defa kullanıldığını bulan Mapreduce programının çalışma şekli Şekil 3.4.'de gösterilmiştir.



Şekil 3.4. Örnek programın çalışma mantığı (Kumar, 2014)

**Input:** Veriler bir .txt dosyasına yazılır. Bunu için “Not Defteri” veya “Notepad++” programları kullanılabilir.

**Splitting:** Verileri kullanıcının isteğine göre belli büyüklüklere ayırır. Kullanıcı herhangi bir işlem yapmazsa varsayılan olarak veriler 64 MB lık parçalara otomatik ayrılır.

**Mapping:** Dosyada bulunan bütün kelimeler key(word) veya value(1) şeklinde bölümlere ayrılır .

**Shuffling :** Map fonksiyonu çalıştıktan sonra elde edilen çıktıları Reduce fonksiyonuna yönlendirir . Programımız kelime hesaplama(word-count) programı olduğundan aynı olan kelime grupları aynı Reducer’a iletilir.

**Reducing:** Map kısmından elde edilen çıktılar üzerinden toplama işlemi uygulanarak (farklı bir işlemde yapılabilir) sonuçlar bulunur. Elde edilen çıktılar istenen hedefe yazılır (HDFS , SQL , NoSQL)

**Final Result:** Reducing aşamasıyla elde edilen sonuçlar dosyaya yazılır. Oluşturulan bu dosyanın belli bir uzantısı yoktur. Ancak “Not Defteri” veya “Notepad++” programlarıyla bu dosya açılabilir.

### 3.2.1.1. Map fonksiyonu

Map fonksiyonunu girdi olarak gelen verileri gruplayıp, sıralamakla görevlidir. Ayrıca elde edilen çıktıların reduce fonksiyonuna aktarılmasından sorumludur (Lam, 2010). Hadoop yapısında, girdi olarak verilen dosya yapısı HDFS dosya sisteminde olmalıdır. Dosyanın her satırı anahtar/değer (key/value) şeklinde bir kayıt olarak okunmaktadır. Anahtar değeri okuma tipine göre byte cinsinden satırın başlangıç adresi, değer ise satırın metin kısmıdır. Okunan kayıtlar HDFS dosya sisteminin daha önceden belirlenen blok boyutuna göre gruplanır ve gruplanan bu kayıtlar “split” olarak adlandırılmaktadır.

Şekil 3.5.'deki örnek programda görüldüğü gibi Map fonksiyonu girdi anahtar/değer (key/value) olarak long/text tipinde değer aldığı görülmektedir. Örnekte Map fonksiyonu değer olarak ise satırı alır, anahtar değeri ile ilgili bir işlem yapmamaktadır. Satırı kelimelerine ayırır. Ayrıştırılan her bir kelime ve kelimenin sayısını çıktıya aktarır. Çıktı anahtar/değer ikilisi olarak text/int (bulunan kelime/kelimenin sayısı) dışarıya verilir.

```
package kelime_hesaplama;
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
public class hesap_map extends Mapper<LongWritable, Text, Text, IntWritable>{

    private final static IntWritable bir = new IntWritable(1);
    private Text kelime = new Text();

    public void map(LongWritable key, Text value, Context context
        ) throws IOException, InterruptedException {
        StringTokenizer ifade = new StringTokenizer(value.toString());
        while (ifade.hasMoreTokens()) {
            kelime.set(ifade.nextToken().toUpperCase());
            context.write(kelime, bir);
        }
    }
}
```

Şekil 3.5. Map Fonksiyonu

### 3.2.1.2. Reduce fonksiyonu

Reduce fonksiyonunun girdileri, map fonksiyonlarının çıktılarıdır. Bu yüzden map fonksiyonunun çıktı anahtar/değer (key/value) ikilisi ile reduce fonksiyonunun girdi anahtar/değer (key/value) ikilisinin veri tipleri aynı olmalıdır (Lam, 2010). Reduce fonksiyonu anahtar düzeyinde sıralanmış ve gruplanmış girdi değerlerini alır ve ilgili işlemler çözümü gerçekleştirir.

Reduce fonksiyonu, Şekil 3.6.'de gösterildiği üzere girdi anahtar/değer ikilisi olarak text/int almaktadır. Bu girdi değerler map fonksiyonunun çıktı tipi ile aynıdır. Map fonksiyonunda çıktı değeri olarak sadece int kullanılırken reduce

fonksiyonu değer olarak int tipinde bir dizi oluşturmaktadır. Çünkü aynı anahtar değerine sahip birden çok değer olabilir. Reduce fonksiyonu girdi olarak aldığı anahtarı (ilgili kelime) direkt çıktı olarak yazar. Girdi olarak aldığı değer parametresini ise "sum" değişkenine ekler. Burada amaç aynı anahtar değerine sahip olan değerlerin sayısını bulmaktır.

```
package kelime_hesaplama;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class hesap_red extends Reducer<Text,IntWritable,Text,IntWritable> {
    private IntWritable sonuc = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values,
        Context context
    ) throws IOException, InterruptedException {
        int toplam = 0;
        for (IntWritable deger : values) {
            toplam += deger.get();
        }
        sonuc.set(toplam);
        context.write(key, sonuc);
    }
}
```

Şekil 3.6. Reduce Fonksiyonu

### 3.2.1.3. Mapreduce görevi tanımlama

Map ve reduce fonksiyonlarını içeren Mapper ve Reducer sınıflarını tanımladıktan sonra işleyicileri oluşturmak için öncelikle bir Hadoop görevi tanımlanması gerekmektedir. Şekil 3.7.'daki örnek kod parçasığında görülebileceği gibi JobConf sınıfından oluşturulan nesne, göreve ait özelliklerin atandığı ve görevin nasıl çalıştırılması gerektiğinin ayarlarının yapılabileceği nesnedir. Hadoop'ta görev çalıştırıldığı zaman, görevin tanımlandığı bu sınıf, Mapper ve Reducer sınıfları ile birlikte bir java çalıştırılabilir dosyasına (Java Archive (JAR) dönüştürülür. Bu dosya dağıtık sistemde görevi çalıştıracak tüm düğümlere gönderilir. JobConf sınıfı bu görevin düğümlerde nasıl çalıştırılacağı bilgilerini içerdiği için çok önemlidir. Bu tanımlama yapılırken öncelikle girdi ve çıktı dosyalarının nereden okunacağı ve nereye yazılacağına dair adres bilgileri

FileInputFormat ve FileOutputFormat sınıfları içinde tanımlanır. "AddInputPath" metodu kullanılarak birden fazla girdi klasörü veya dosyası da görev için tanımlanabilir. Çıktı dosyasının adresi tanımlandıktan sonra görev çalıştırıldığında eğer HDFS'de bu isimde bir dosya mevcutsa görev hata vererek çalışmaz. Bu klasörün sistemde olmaması görev çalıştığında görevin bu klasörü yaratması açısından gereklidir. "SetOutputKeyClass" ve "SetOutputValueClass" fonksiyonları ile çıktı dosyasına yazılan kayıtların anahtar ve değer ikililerinin formatları belirlenir. Tüm yapılandırma bittikten sonra "runJob" fonksiyonu çağrılarak görev çalıştırılır ve görevin sonlanması beklenir.

```
1 package kelime_hesaplama;
2
3 import org.apache.hadoop.conf.Configuration;
4 import org.apache.hadoop.fs.Path;
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Job;
8 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
9 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
10 public class kelime_hesaplama {
11     public static void main(String[] args) throws Exception {
12
13
14         Configuration conf = new Configuration();
15         Job job = Job.getInstance(conf, "kelime_hesaplama");
16         job.setJarByClass(kelime_hesaplama.class);
17         job.setMapperClass(hesap_map.class);
18         job.setCombinerClass(hesap_red.class);
19         job.setReducerClass(hesap_red.class);
20         job.setOutputKeyClass(Text.class);
21         job.setOutputValueClass(IntWritable.class);
22         FileInputFormat.addInputPath(job, new Path(args[0]));
23         FileOutputFormat.setOutputPath(job, new Path(args[1]));
24         System.exit(job.waitForCompletion(true) ? 0 : 1);
25     }
26 }
```

Şekil 3.7. Mapreduce görevi tanımlama program kodu

### 3.2.2. HDFS (Hadoop Dağıtık Dosya Sistemi)

HDFS çok büyük miktarda veya büyük boyutta dosyaların dağıtık bir sistemde saklanması ve bu dosyalara en hızlı ve güvenli bir şekilde erişilebilmesi için tasarlanmıştır (Demir, 2012). Mapreduce prensiplerine uygun olarak verilerin kolay işlenebilmesine uygun bir mimari kullanılmıştır. Normal dosya sistemlerinde tutulabilen boyutların çok üzerinde, petabyte boyutlarında dosyaların saklanıp işlenebilmesi için; HDFS saklanacak dosyaları blok

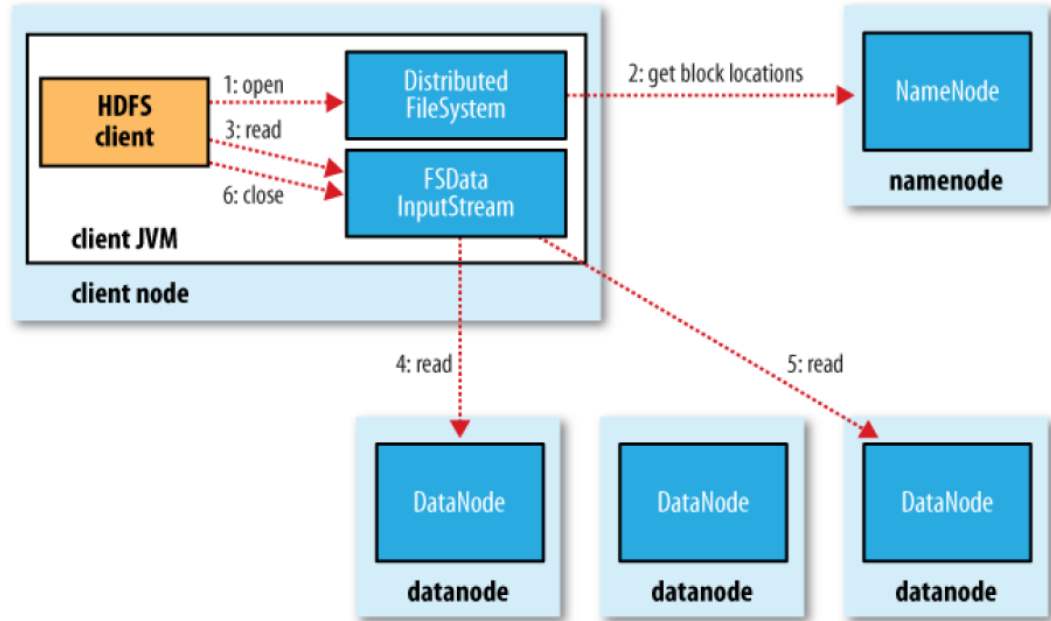
boyutunda (örn: 64 MB) parçalara ayırır ve parçaları farklı düğümlerde saklar. Böylelikle binlerce düğüm üzerinde petabyte boyutunda dosyaların saklanması mümkün olabilmektedir. Dosyaların parçalar halinde dağıtık olarak saklanması, bu dosyalar üzerinde çalışacak olan işleyicilerin de kendi buldukları düğümdeki dosyaya ait parçayı işlemesine imkân verdiğinden, bu şekilde çalışması Mapreduce işleyicilerinin de veriye daha çabuk ulaşarak daha hızlı işlem yapmasını sağlamıştır. Veri yerelliği (Data Locality) (Demir, 2012) prensibi denen bu yapı HDFS'nin paralel işleme performansını da ciddi anlamda artırmaktadır. Veri yerelliği prensibinde, işlenecek verinin sürekli olarak işleyici olan düğümlere ağ üzerinden transferi gerekmemektedir, dolayısıyla ağ gecikmesinden kaynaklanan yavaşlamaları azaltmaktadır.

### **3.2.2.1. HDFS namenode ve datanode yazılımları**

HDFS dosya sistemi iki tür düğümden oluşur. İlki dosya sistemi yönetici düğümünde bulunan NameNode ve diğeri işçi düğümlerde bulunan DataNode yazılımlarıdır. NameNode yazılımı, dosya sisteminde kayıtlı bulunan tüm dosya ve klasörlere ait dosya sistem ağacını tutar ve yönetir. Her dosyaya ait kayıt bilgisi (Metadata) NameNode tarafından yerel diskte tutulur. NameNode ayrıca bir dosyaya ait blokların bazılarının tutulduğu DataNode'lara ait bilgileri de tutar. Fakat yerel blok adreslerini yerel olarak tutmaz sadece hangi bloğun hangi DataNode tarafından yönetildiği bilgisini tutar. DataNode ise kendinde bulunan blokların saklanması, eklenmesi, silinmesi ve adreslenmesi gibi işleri kontrol eder. Dolayısıyla DataNode yazılımları işçi düğümlerde bulunan blok işlemlerinin yürütüldüğü ve yapılan işlemlerin blok listelerini NameNode yazılımına raporlayan bir yazılımdır. NameNode yazılımı tüm dosya sistemini yönettiğinden bozulması durumunda tüm dosya sistemi ve dosyalar kaybolacaktır. Hadoop bunun önüne geçmek için iki mekanizma geliştirmiştir. Bunlardan ilki: NameNode'un tüm dosya sistemindeki ağaç verisinin belli aralıklarla imajını almasıdır, ikincisi ise paralelde çalışan ikincil bir NameNode oluşturmaktır. Bu ikincil (Secondary) NameNode bilgileri yedekleme amaçlı çalışır ve genellikle ayrı bir fiziksel makinede bulunur.

### 3.2.2.2. HDFS girdi çıktı mekanizması (HDFS I/O)

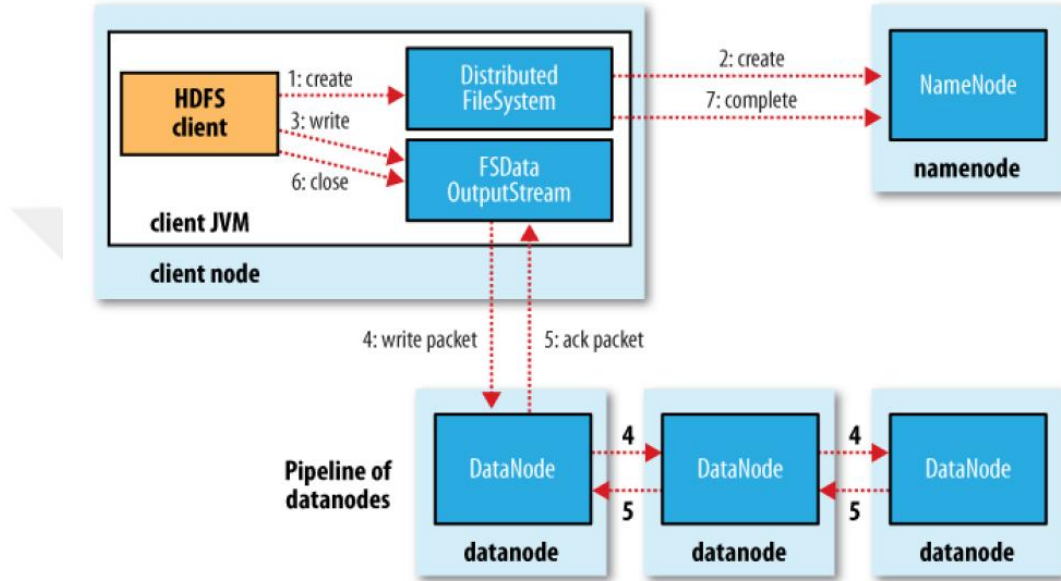
Kullanıcı bir dosyayı HDFS'de okumak istediği zaman bununla ilgili istekte bulunduktan sonra HDFS'nin bunu gerçekleştirme mekanizması Şekil 3.8.'de sergilenmiştir.



Şekil 3.8. HDFS'de istemci tarafından dosya okunma şeması (White, 2009)

Kullanıcı HDFS'de open() ile DistributedFileSystem nesnesinin NameNode ile haberleşmesini başlatır. Bunu gerçekleştirmek için Remote Procedure Call (RPC) (Birrell ve Nelson, 1983) mekanizmasını kullanır. Böylelikle istenilen dosyaya ait blokların bulunduğu DataNode'lara ait adresleri NameNode geri döndürür. Bu gelen DataNode'lardaki blokları adreslerinden okumak için FSDDataInputStream üzerinden read() fonksiyonu çağrılır. Bunu her blok bilgisini okumak için tekrar yapar. Tüm blok bilgileri okunduktan sonra da FSDDataInputStream üzerinde close() fonksiyonu kullanarak veri akışı durdurulur. Eğer veriyi DataNode'lardan okuma sırasında problem oluşursa okunacak blok bilgisini içeren diğer node üzerinden bu blok tekrardan okunur. Eğer blok verisinin kontrol-toplam (checksum) bilgisi kontrol edilerek bozuk olduğu anlaşılırsa, bu durum NameNode'a bildirilir ve yedeklenmiş blok bilgisi

(Replication) ile bozuk blok verisi güncellenir. NameNode her zaman istemciye en yakın DataNode adresine döner bu da erişim sürelerini kısaltan performansı artıran bir mekanizmadır. Dosyaların kullanıcı tarafından HDFS'ye yazılması da aşağıda Şekil 3.9.'de sergilenmiştir. Burada izlenen ana sıra, önce dosyanın yaratılması, sonra verisinin dosyaya yazılması, ardından da dosyanın kapatılmasıdır.



Şekil 3.9. HDFS'ye istemci tarafından dosya yazılma şeması (White, 2009)

Kullanıcı tarafından dosya DistributedFileSystem nesnesi üzerinden create() fonksiyonu çağrılarak yaratılır. DistributedFileSystem NameNode'a bir RPC (Birrell ve Nelson, 1983) mesajı ile dosya sisteminde yeni bir dosya yaratması komutunu gönderir. İlk oluşturulan bu dosya bilgisi hiç bir blok verisi içermez. NameNode'da eğer aynı isimde dosya yoksa istemcinin yazma haklarını kontrol ederek bu isteği gerçekleştirir. DistributedFileSystem dosya içeriğinin yazılması için istemciye bir FSDDataOutputStream nesnesi oluşturur. Bu kendi içinde yazılacak olan veriyi paketlere ayırır. Bu parçalar daha sonra bloklar halinde sistemde bulunan DataNode'lara kaydedilir. Bu blokların çoğaltılarak kopyalanması işlemi bir boru hattı (pipeline) mekanizmasına göre olur. İlk DataNode bloğu kaydeder ve ikinci DataNode'a gönderir o da kaydeder, daha sonra ikinci kaydedilmesi için üçüncüye gönderir. Tüm yazma işlemi

tamamlandıktan sonra tekrardan `FSDataOutputStream` nesnesinin `close()` fonksiyonu çağrılarak kapatılır ve yazma sonlandırılmış olur.

### 3.3 Trend Analiz Yöntemleri

Trend analizi, uzun dönemlere ait veriler ile mevcut durum veya geleceğe yönelik değişimlerinin yorumlanmasına yardımcı olan bir istatistiksel tekniktir. Bu değişimler artma, azalma ya da durgunluk şeklinde olabilir. Herhangi bir veri setinin trend analizlerini belirlemek için farklı testler uygulanmaktadır. Bu çalışmada Mann-Kendall ve Spearman Rho trend analiz teknikleri kullanılmıştır.

Kendall'ın Tau'su ve Spearman'ın Rho'su şeklinde adlandırılan katsayı ve yöntemler Sosyal Bilimler alanındaki araştırmalarda (analizlerde) sıklıkla kullanılırlar (Gibbons, 1993). Değişkenlerin bağımlı veya bağımsız olması dikkate alınmaksızın aralarındaki ilişkinin derecesini ve yönünü belirlemek amacıyla kullanılan istatistik yöntemleri korelasyon olarak tanımlanır. Bu bağlamda Parametrik Pearson Korelasyon testinin uygulanamadığı durumlarda Kendall ve Spearman testleri kullanılır (Sipahi vd., 2010).

#### 3.3.1. Mann-Kendall Testi

Mann-Kendall testi, hidroloji ve klimatoloji gibi alanlardaki zaman serilerinde yer alan trendin belirlenmesinde yaygın olarak kullanılan ve Dünya Meteoroloji Örgütü (WMO) tarafından da önerilen istatistiksel bir test yöntemidir. Mann-Kendall testi, birçok çalışmada trend analizinin yapılmasında üstünlüğünü ortaya koymuştur. Mann-Kendall test istatistiği ve S istatistiği hesaplamaları şu şekilde yapılmaktadır (Bulut vd. 2006).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (3.1)$$

Denklemden yıl uzunluğu  $n$  ifade edilmiştir.  $S$  değeri ise  $n \geq 8$  olduğunda normal dağılım göstermektedir. Eğer  $n \geq 30$  durumu olursa  $z$  testi,  $t$ -testine yaklaşmış demektir.  $\text{sgn}$  işaret fonksiyonu olup, bu gidiş testi  $i = 1, \dots, n-1$ 'e kadar sıralanmış olan bir  $x_i$  veri setine ve  $j = i + 1, \dots, n$ 'e kadar sıralanmış olan bir  $x_j$  veri setine uygulanır. Her bir sıralanmış rakam  $x_i$  bir referans noktası olarak kabul edilir ve diğer sıralanmış veri grubu  $x_j$  ile aşağıda gösterilen denklemde verildiği gibi karşılaştırılır (Özfidaner vd., 2007).

$$\text{sgn}(x_j - x_i) = \begin{cases} 1; & x_j > x_i \\ 0; & x_j = x_i \\ -1; & x_j < x_i \end{cases} \quad (3.2)$$

eşitliği ile hesaplanır.  $S$ 'nin varyansı aşağıdaki eşitlikle hesaplanır:

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^p t_i(t_i-1)(2t_i+5)}{18} \quad (3.3)$$

Burada,  $p$  veri setindeki bağıl grupların sayıları,  $t_i$  değeri  $i$  uzunluğundaki bir seride bağıl gözlemleri göstermektedir. Eşitlikteki toplama terimi sadece veride bağıl gözlem olduğunda kullanılır. Standartlaştırılmış Mann-Kendall istatistiği  $Z$  ise Denklem (4)'de verildiği gibi hesaplanabilmekte ve seride gidiş yoktur sıfır hipotezi ( $H_0$ ) varsayımı altında ortalaması sıfır, varyansı bir olan standart normal dağılım göstermektedir.

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}}; & S > 0 \\ 0 & ; S = 0 \\ \frac{S+1}{\sqrt{\text{Var}(S)}}; & S < 0 \end{cases} \quad (3.4)$$

$Z$  değerinin pozitif yani artı olması artışı gösterirken, negatif yani eksi  $Z$  durumunda olması azalışa işaret etmektedir. Burada iki yönlü %95 güven aralığında ( $Z_{1-0.05/2} = Z_{0.975}$ ) karşılığı 1,96 olarak alınmıştır (Ek 1).

### 3.3.2. Sperman'ın Rho Testi

İki gözlem serisi arasında korelasyon olup olmadığını gözlemleyebilmek amacıyla kullanılan Sperman'ın Rho Testi, lineer trend varlığının olup olmadığının araştırılmasında basit ve hızlı bir testtir. Sıra istatistiği olan  $Rx_i$  verilerin büyükten küçüğe veya küçükten büyüğe sıralanması ile belirlenir. Gözlem serisi  $X=(x_1, x_2, \dots, x_n)$  vektörü olmak üzere; iki yönlü test ile tanımlanan  $H_0$  hipotezine göre  $x_i$  ( $i=1, 2, 3, \dots, n$ ) değerleri eş olasılıklı dağılımlardır,  $H_1$  hipotezine göre ise  $x_i$  ( $i=1, 2, 3, \dots, n$ ) değerleri zamanla artar veya azalır. Spearman'ın Rho testi istatistiği ( $r_s$ ), (5) bağıntısı ile hesaplanır (Gümüş, 2006).

$$r_s = 1 - 6 \frac{[\sum_{i=1}^n (R(X_i - i)^2)]}{n^3 - n} \quad (3.5)$$

$n > 30$  için  $r_s$  dağılımı normale yaklaşacağından normal dağılım tabloları kullanılır. Bunun için  $r_s$ 'nin test istatistiği ( $Z$ ), hesaplanması denklem (6)'da belirtilen şekilde yapılmaktadır:

$$Z = r_s \sqrt{n - 1} \quad (3.6)$$

Eğer  $Z$  değeri,  $\alpha$  önem seviyesinde standart normal dağılım tablolarından (Ek 2) tespit edilen  $Z_{\alpha/2}$  değerinden büyük ise,  $H_0$  hipotezi reddedilir.  $H_0$  hipotezi değişkenler arasında bir ilişkinin olmadığını belirtir. Aksi ispat edilmediği sürece doğru kabul edilir.

## 4. BULGULAR

Bu bölümde çalışma kapsamındaki Hadoop ortamının hazırlanması, veri setinin Hadoop ile işlenmesi ve verilerin analizi konuları detaylandırılmıştır.

### 4.1. Hadoop Ortamının Hazırlanması

Verilerin analizinin yapılabilmesi için öncelikle ağ üzerindeki bilgisayarlara gerekli olan Hadoop bileşenlerinin ve gerekli diğer yazılımların kurulması gerekmektedir. Bu amaçla Java SDK, Cygwin, Eclipse ve Hadoop programları bilgisayarlara kurulur.

#### ➤ Java SDK

Hadoop sistemindeki bütün bileşenler farklı bir jvm (java sanal makinesi) üzerinde çalışmaktadır. Bu yüzden Hadoop sisteminin kullanılacağı bütün makinelerde java SDK programının kurulmasına ihtiyaç vardır. Bu çalışmada Java versiyonu olarak 1.7.02 versiyonu kullanılmıştır. Bilgisayara Java kurulumu yapıldıktan sonra Java konumu Hadoop sistemine belirtilir. Bu belirtme işi JAVA\_HOME ortam değişkeni oluşturularak yapılabilir.

#### ➤ Cygwin

Hadoop sisteminin ilk olarak Linux işletim sistemleri için tasarlanmıştı. Bu yüzden windows işletim sistemlerinde Hadoop sistemini daha iyi kullanabilmek için linux işletim sistemine ait hizmetlerin ve komutların kullanılması gerekmektedir. Cygwin programı bu işlemi gerçekleştirebildiğinden kurulması gerekmektedir.

## ➤ Eclipse

Tez çalışması için oluşturulan algoritma java programlama dili ile yazılmıştır. Eclipse derleyicisi de Java için en iyi geliştirme ortamlarından biridir. Bu tezde Eclipse Neon.2 versiyonu kullanılmıştır.

Mapreduce programı geliştirmek için farklı programlama dilleri ve derleyiciler de kullanılabilir. Hadoop tüm diller için gerekli entegrasyonu sağlamıştır. Ancak Java asıl gerçekleştirme dili olarak sunulmaktadır. Çünkü Hadoop yapısı java programlama dili ile yazılmıştır.

## ➤ Hadoop

Hadoop kurulumu için <http://www.hadoop.apache.org/> sitesinden istenilen Hadoop versiyonu .zip dosyası olarak indirilir. İndirilen dosya, istenilen bir dizine çıkartılır. Tez için Hadoop 2.7.1 versiyonu kullanılmaktadır. Hadoop üç farklı konfigürasyonda çalışabilmektedir. Bunlar:

**Tekil mod (Standalone mode):** Bu modda geliştirilen program tek makinede ve aynı jvm üzerinde çalışır. Yeni başlayanlar ve geliştirme için ideal bir moddur (Venner , 2009; White, 2009; Lam, 2010).

**Sahte-dağıtık mod (Pseudo-distributed mode):** Bu modda geliştirilen program aynı bilgisayarda birden fazla java sanal makineleri üzerinde yerel olarak çalıştırılır. Esasen sahte-dağıtık mod tam dağıtık sistemi birebir simüle etmektedir.

**Tam dağıtık mod (Fully-distributed mode):** Bir program önce sahte-dağıtık mod ile hatalarından ayıklanır. Daha sonra hatalarından ayıklanan programın Hadoop sisteminin sunduğu hizmetlerden yararlanabilmesi ve daha iyi performans göstermesi için tam dağıtık modda çalıştırılmalıdır. Tam dağıtık modda Hadoop sisteminin çalıştırabilmek için ağ üzerinde en az iki makineye ihtiyaç duyulmaktadır. Ağ üzerindeki bu makinelerden bir tanesi master makine

olarak kullanılacak olup üzerinde NameNode, Secondary NameNode ve JobTraker bileşenleri çalıştırılır. Kalan diğer makineler ise slave olarak kullanılır. Bu makinelerde veriyi depolayan ve Mapreduce görevlerini işleyen sırasıyla DataNode ve TaskTracker bileşenleri çalıştırılır.

Hadoop sistemini tekil-modda çalıştırmak diğer modlarda çalıştırmaktan çok daha kolay olmasına rağmen Hadoop'u gerçek amacına uygun şekilde kullanabilmek için tam-dağıtık moda çalıştırmak gerekir. Bu tez çalışmasında Hadoop tam-dağıtık moda çalıştırılmıştır. Hadoop ortamını sağlamak için 3 adet çift çekirdekli 2.4 Ghz saat hızı işlemciye sahip bilgisayar kullanıldı. Bu bilgisayarlarda Windows 7 Professional 32 bit işletim sistemi bulunmaktadır. Her bilgisayarın RAM kapasitesi 2 GB boyutundadır. 3 bilgisayar için görev dağılımı şu şekilde yapılmıştır:

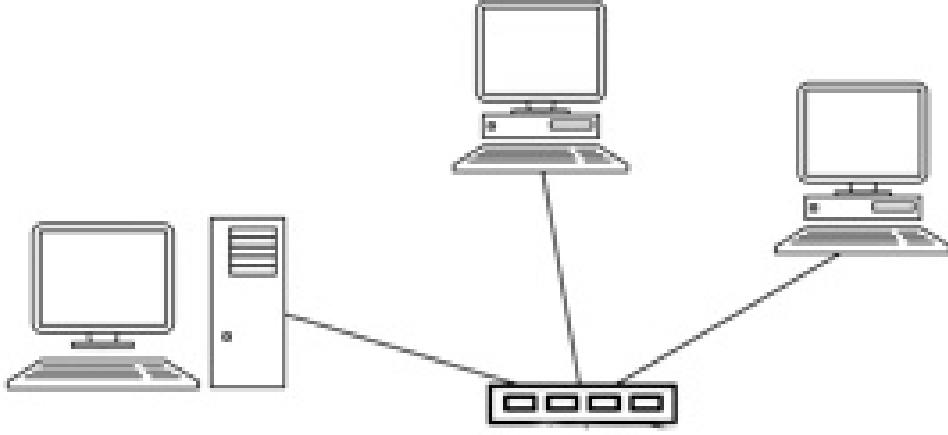
- Düğümlerden biri (Master)

1. Hem ana düğüm hem de veri düğüm
2. İş denetleyici
3. Görev denetleyici

- Diğer 2 düğüm (Slave)

1. Veri düğümler
2. Görev denetleyiciler

Bu bilgisayarlar birbirlerine switch aracılığıyla bağlanmıştır. Bu 3 bilgisayar için Şekil 4.1.'deki gibi bir star topolojisi oluşturulmuştur.



Şekil 4.1. Hadoop ortamı için Tam dağıtık mod yapısı

Bilgisayarlara bahsedilen yazılımlar kurulduktan sonra, Hadoop ortamında tam-dağıtık modu istenilen şekilde çalıştırmak için bazı ayarlamalar yapılmasına ihtiyaç duyulmuştur. Çünkü birden fazla makinede ortak bir çalışma gerçekleştirilmiştir. Programların kurulumu ve yapılan konfigürasyon ayarları 15 adımlık bir süreçte tanımlanmıştır.

**1-**Elimizdeki makinelerle cygwin programı kurulur. Bu program tüm linux işletim sistemine ait özellikleri ve komutları çalıştırmamıza yardımcı olmaktadır. Bunlar arasında en önemlisi ağ üzerinde bilgisayarlar arasında haberleşmeyi sağlayan ssh hizmetidir.

**2-**Bilgisayara cygwin kurulumu yaptıktan sonra üzerinden path ortam değişkenine şu satırlar eklenmelidir:

```
cygwin/bin
```

```
cygwin/usr/sbin
```

Böylece Hadoop komutları sorunsuz bir şekilde çalıştırılacaktır.

**3-**Hadoop sistemindeki master bilgisayarın diğer slave bilgisayarlar ile sorunsuz bir şekilde iletişim kurabilmesi için şifresiz giriş ayarlarının yapılmasına ihtiyaç duyulmaktadır. Bunun için şu adımlar gerçekleştirilir:

- Öncelikle bir cygwin terminali açılır ve “\$ ssh-host-config” komutu yazılır.
- Komut çalıştıktan sonra ekrana gelen “privilege seperation” isteği için “no” yazılıp devam edilir.
- Ekrana “sshd” kurulum isteği geldiğinde “yes” yazılıp devam edilir.
- Son olarak “cygwin enviroment variable” ifadesi için “ntsec” girişlerinin yapılması gerekmektedir.

**4-Bilgisayarlara “ssh” (Secure Shell-Güvenli Kabuk) servisinin kurulumu yapıldıktan sonra bu servisin başlatılması gerekmektedir. Bu servis ağ üzerinde başka bilgisayarlara bağlanmak, uzaktaki bir bilgisayarda komut çalıştırmak ve bilgisayarlar arasında dosya transferi sağlamak için kullanılan Linux tabanlı bir protokoldür. Bu servisin aktifleştirilmesi için “Bilgisayarım ->Yönet -> Hizmetler” sekmesine gidilir. Bu sekme içersindeki “CYGWIN ssh” hizmetine sağ tıklanıp “Başlat” seçeneği seçilir. Böylece servis başlatılmış olur. Servis bir kere başlatmak yeterlidir daha sonraki işlemlerde servis otomatik başlayacak ve herhangi bir işlem yapmaya ihtiyaç duyulmayacaktır.**

**5-Bilgisayarlara, şifresiz girişi sağlamak için cygwin terminali açılır ve burada “\$ ssh-keygen” komutu çalıştırılır. Ekrana gelen tüm istekler “ENTER” ile geçilir. Burada gerçekleştirilen adım ile her bilgisayar kendisine ait id\_rsa ve id\_rsa.pub isminde özel ve ortak anahtar dosyalarını oluşturur.**

**6-Bir önceki adımda oluşturulan ortak anahtar dosyalarının isimleri bilgisayar ismi olarak belirlenir. Oluşturulan dosyalar diğer bilgisayarlara da dağıtılır. Böylece bütün bilgisayarlarda her dosya için, “\$ cat xxx.pub >> authorized\_keys” komutu yürütülerek, public anahtar dosyaları diğer bilgisayarlara dağıtılmış olur. Bunun sonucunda ağ üzerindeki herhangi bir bilgisayar başka bir bilgisayara şifresiz olarak erişim sağlayabilmektedir.**

**7-Bilgisayarlar arasındaki bağlantılardan sonra bütün bilgisayarlara java kurulumu yapılır. Java kurulumu varsayılan olarak yapılabileceği gibi, istenilen bir dizine de kurulabilir. Örneğin; C:/Java.**

**8-**Sistemimizde gerekli java kodlarını yazmak ve derleyebilmek için master bilgisayara Eclipse derleyicisi kuruldu. Eclipse Java için en iyi geliştirme ortamlarından biri ve Hadoop ile uyumlu çalıştığından dolayı tercih edildi. Bilgisayara Eclipse Neon.2 versiyonu kurulduktan sonra Hadoop dosyalarını çalıştırmak için “Hadoop-eclipse-plugin-2.6.0” dosyası Eclipse içinde bulunan “plugins” klasörüne kopyalanmıştır. Böylece java kodlarının derlenmesi ve çalıştırılması çok daha kolay olmuştur.

**9-**Hadoop sistemi içi gerekli programları kurup ayarları yaptıktan sonra indirilen Hadoop.zip dosyası cygwin/home/<kullanıcı-adı> dizini altına çıkartılır.

**10-**Hadoop sistemini çalıştıracak olan ağdaki bütün bilgisayarlarda c/windows/system32/drivers/ dizini altındaki host dosyası düzenlenir. master ve slave olarak çalışacak olan bilgisayarlar ve bu bilgisayarların ip adresleri bu dosyaya eklenir. (192.168.0.1 m1, 192.168.0.2 slave2, 192.168.0.3 slave3)

**11-**Hadoop-2.7.1\etc\Hadoop klasörü içinde bulunan Hadoop-env.sh dosyası “Not Defteri” veya “Notpad++” programı aracılığıyla açılır. Açılan dosyaya JAVA\_HOME isminde bir ortam değişkeni tanımlanıp şu satırlar eklenir.

```
export JAVA_HOME=/cygdrive/c/Java  
export HADOOP_IDENT_STRING = HADOOPEE)
```

**12-** Tüm düğümlerdeki HDFS url adresini “hdfs://m1:9000” olarak değiştirilir. Bu işlem için Hadoop-2.7.1\etc\Hadoop klasörü içinde bulunan core-site.xml dosyası “Not Defteri” veya “Notpad++” programı aracılığıyla açılır. Açılan dosya içeriği Şekil 4.2.’deki gibi düzenlenir.

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://m1:9000</value>
  </property>
</configuration>
```

Şekil 4.2. Makinenin çalışacağı portu belirleme

**13-** Hadoop-2.7.1\etc\Hadoop dizininde hdfs-site.xml “Not Defteri” veya “Notpad++” programı aracılığıyla açılır ve Şekil 4.3.’teki gibi düzenlenir.

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
</configuration>
```

Şekil 4.3. NameNode konfigürasyonu

**14-** Hadoop-2.7.1\etc\Hadoop dizininde mapred-site.xml “Not Defteri” veya “Notpad++” programı aracılığıyla açılır. Açılan dosya içeriği Şekil 4.4.’deki gibi düzenlenir.

```
<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>localhost:9001</value>
  </property>
</configuration>
```

Şekil 4.4. Mapreduce konfigürasyonu

**15-**Hadoop ortamının Windows işletim sistemimiz de çalıştırmak için; Başlat -> Çalıştır sekmeleri tıklanır. Ekranı “cmd” yazılır. Açılan komut işlemi sayfasında ekrana Hadoop’u kurduğumuz yol yazılıp “Enter” tuşuna basılır (“cd *bulunduğu klasör*/Hadoop-2.7.1\sbin”). Daha sonra alt satıra: “start-all.cmd” yazılarak Hadoop çalışır hale gelecektir.

#### **4.2. Veri Setinin Hadoop İle İşlenmesi**

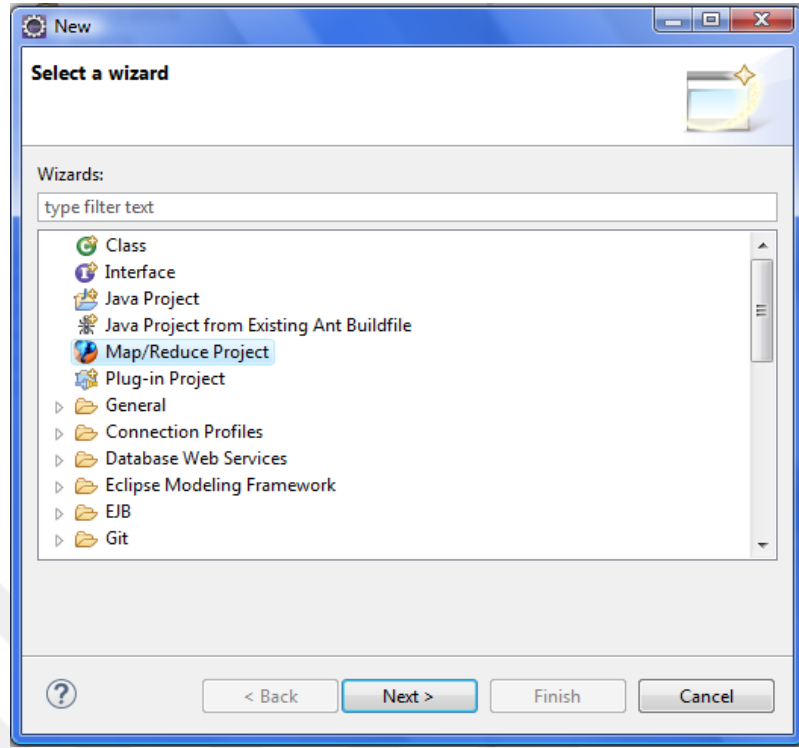
Meteoroloji 10. Bölge Müdürlüğü’nden temin edilen 448 istasyona ait 11 farklı 7.978.191 adet veriden oluşan veri setinin analizi için Java programlama dilinde bir program yazılmıştır. Sıcaklık verileri Şekil 4.5.’de görüldüğü üzere .txt dosyası olarak temin edilmiştir. Diğer veri setlerinde ayrı ayrı .txt dosyası olarak temin edilmiştir.

İstasyon No	İstasyon Adı	Yıl	Ay	Gün	Ortalama Sıcaklık (°C)
1	ABANA	1963	6	14	17.5
2	ABANA	1963	6	15	18.5
3	ABANA	1963	6	16	23.3
4	ABANA	1963	6	17	24.8
5	ABANA	1963	6	18	19.4
6	ABANA	1963	6	19	19.0
7	ABANA	1963	6	20	19.6
8	ABANA	1963	6	21	19.3
9	ABANA	1963	6	22	21.3
10	ABANA	1963	6	23	22.3
11	ABANA	1963	6	24	23.6
12	ABANA	1963	6	25	21.8
13	ABANA	1963	6	26	18.8
14	ABANA	1963	6	27	21.6
15	ABANA	1963	6	28	22.1
16	ABANA	1963	6	29	19.5
17	ABANA	1963	6	30	18.8
18	ABANA	1963	7	1	22.7
19	ABANA	1963	7	2	22.2
20	ABANA	1963	7	3	23.2
21	ABANA	1963	7	3	23.2

Şekil 4.5. Meteoroloji Genel Müdürlüğü tarafından gönderilen sıcaklık verileri

Bu veriler üzerinde incelemeler ve düzenlemeler yapılarak, oluşturulan programın algoritması belirlenip ardından kod yazma işlemine geçilmiştir.

Eclipse derleyicisinde Mapreduce işlemleri için bir çalışma alanı oluşturmak için önce File->New->Other sekmeleri tıklanır. Şekil 4.6.'da görülen pencere açılır.



Şekil 4.6. Mapreduce projesi oluşturma

Şekil 4.6.'da açılan pencereden "Map/Reduce Project" sekmesi tıklanarak yeni bir çalışma sayfası açılır. "Map/Reduce Project" sekmesi Eclipse sürümlerinde olmayan bir çalışma sayfasıdır. Eclipse derleyicisi içinde bulunan plugins klasörüne kopyaladığımız "Hadoop-eclipse-plugin-2.6.0.jar" eklentisi sayesinde bu çalışma alanı oluşturulmuştur. Açılan sayfa içinde map işlemleri, reduce işlemleri ve Mapreduce görevi tanımlama için 3 farklı class (sınıf) oluşturuldu. Şekil 4.7.'da map işlemlerini gerçekleştiren kodlar belirtilmiştir.

```

import java.io.IOException;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
public class Gunluk_Ortalama_Sicaklik_map extends MapReduceBase implements
Mapper<LongWritable, Text, Text, IntWritable>
{
    public void map(LongWritable key, Text value, OutputCollector<Text,
IntWritable> output, Reporter reporter) throws IOException
    {
        int yagis;
        String line = value.toString();
        int yil=0,sayac=0;
        for(int i=0; i<line.length(); i++)
        {
            if(line.substring(i,i+1).equals("|")) //VERİLERDEKİ YIL DEĞİŞKENİNE
ERİŞEBİLMEK İÇİN YAZILAN KOD BLOĞU
            {
                yil=i+1;
                sayac++;
            }
            if(sayac==2)
                break;
        }
        String year = line.substring(yil, yil+4);

        int b=0;
        for(int i=0; i<line.length(); i++) //HER YILIN SICAKLIK DEĞERİNE
ERŞEBİLMEK İÇİN YAZILAN KOD BLOĞU
        {
            if(line.substring(i,i+1).equals("|"))
                b=i+1;
        }
        yagis=Math.round(Float.parseFloat(line.substring(b))*10);
        output.collect(new Text(year),new IntWritable(yagis));

    }
}

```

Şekil 4.7. Sıcaklık verilerini işleyen map fonksiyonu için gerekli kodlar

Map fonksiyonunun iki temel değişkeni vardır. Birincisi anahtar kelimeler (key), diğeri ise bu anahtar kelimelere ait değer veya değerlerdir (value). Bu çalışmada her yıla ait günlük ortalama sıcaklık değerleri bulunacağından anahtar kelimeler olarak yıl verileri seçilmiştir. Bu sayede her yıla ait sıcaklık değerleri o yıla ilişkilendirilmiş olunur. Bu program çalıştıktan sonra aynı olan kelime grupları aynı Reducer'a iletilir. Reduce fonksiyonuna iletilen değerler. Şekil 4.8.'deki kodlar yazılır.

```
import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
public abstract class Gunluk_Ortalama_Sicaklik_red extends MapReduceBase
implements Reducer<Text, IntWritable, Text, IntWritable>
{
    public void reduce(Text key, Iterator<IntWritable> values,
OutputCollector<Text, IntWritable> output, Reporter reporter) throws
IOException
    {int sayac=0;
        int toplam = 0;

        while (values.hasNext()) //HER BİR YILLA AIT SICAKLIK DEĞERLERİ
        TOPLANIP, ORTALAMASI HESAPLANIYOR
        {
            int current=values.next().get();
                toplam=toplam+current;
                sayac++;
            }
        output.collect(key, new IntWritable((toplam/sayac)));
    }
}
```

Şekil 4.8. Sıcaklık verilerini işleyen reduce fonksiyonu için gerekli kodlar

Şekil 4.8.'de map fonksiyonundan gelen aynı olan kelime gruplarını birleştiren reduce fonksiyonuna ait kodlar görülmektedir. Aynı anahtar değere sahip gruplar üzerinde bir while döngüsü kullanılarak toplama işlemi gerçekleştirilmiştir. While döngüsündeki toplama işlemiyle beraber bir "sayac"

değişkeni yerleştirilmiştir. Böylece aynı anahtar değere sahip kaç tane değişkenin olduğu hesaplanabilmiştir. Toplama işleminin sonucu "sayac" değişkeninin değerine bölünerek ortalama hesaplanmıştır.

Bu iki fonksiyonun beraber çalışabilmesi için Mapreduce görev tanımlamasının yapılması gerekmektedir. Şekil 4.9.'da bu iki fonksiyon için oluşturulmuş görev tanımlaması görülmektedir.

```
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;
public class Gunluk_Ortalama_Sicaklik extends Configured implements Tool{
    public int run(String[] args) throws Exception
    {
        JobConf conf = new JobConf(getConf(), Gunluk_Ortalama_Sicaklik.class);
        conf.setJobName("Ortalama_Sicaklik_Hesaplama");
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(Gunluk_Ortalama_Sicaklik_map.class); //PROGRAM İÇİN
        MAP SINIFI DAHİL EDİLİYOR.
        conf.setReducerClass(Gunluk_Ortalama_Sicaklik_red.class); //PROGRAM İÇİN
        REDUCE SINIFI DAHİL EDİLİYOR.
        Path inp = new Path(args[0]); //VERİLERİN ALIYANCAĞI DOSYA BELİRTİLİYOR
        Path out = new Path(args[1]); //SONUÇLARI YAZILACAĞI DOSYA BELİRTİLİYOR
        FileInputFormat.addInputPath(conf, inp);
        FileOutputFormat.setOutputPath(conf, out);
        JobClient.runJob(conf);
        return 0;    }
    public static void main(String[] args) throws Exception
    {
        int res = ToolRunner.run(new Configuration(), new
        Gunluk_Ortalama_Sicaklik(),args);
        System.exit(res);
    } }
}
```

Şekil 4.9. Map ve reduce fonksiyonları için görev tanımlama

Şekil 4.9.'da görüldüğü üzere çalıştırılacak map ve reduce fonksiyonları çağırılıyor. Kaç farklı dosyadan verilerin alınacağı ve işlem bittikten sonra yazma işleminin yapılması için gerekli kodlar bu sayfaya yazılır. Program çıktı olarak “part-00000” isimli bir dosya oluşturur. Bu dosyanın içindeki sonuçlara “Notpad++” ile erişmek mümkündür.

Gerekli kod işlemleri yazıldıktan sonra program çalıştırılır. Program çalıştıktan sonra Şekil 4.10. ve Şekil 4.11.'deki ekran görüntüleri elde edilir.

## Summary

Security is off.  
 Safemode is off.  
 1 files and directories, 0 blocks = 1 total filesystem object(s).  
 Heap Memory used 50.43 MB of 90.28 MB Heap Memory. Max Heap Memory is 966.69 MB.  
 Non Heap Memory used 24.7 MB of 25.59 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	153 GB
DFS Used:	310 B (0%)
Non DFS Used:	76.38 GB
DFS Remaining:	76.62 GB (50.08%)
Block Pool Used:	310 B (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	3 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	26.04.2018 23:55:31

Şekil 4.10. Hadoop kümesi çalıştırdıktan sonra sistemin genel özellikleri

The screenshot shows the Hadoop cluster management interface. The main heading is "Nodes of the cluster". On the left, there is a navigation menu with options like "Cluster", "About Nodes", "Node Labels", "Applications", "NEW", "NEW\_SAVING", "SUBMITTED", "ACCEPTED", "RUNNING", "FINISHED", "FAILED", "KILLED", "Scheduler", and "Tools". The main content area displays "Cluster Metrics" and "Scheduler Metrics".

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Node
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	3	0	0	0

**Scheduler Metrics**

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

**Node Details Table**

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail	VCores Used	VC Avail
/default-rack		RUNNING	192.168.1.49:2417	192.168.1.49:8042	Cum Nis 27 16:11:34 +0300 2018		0	0 B	8 GB	0	8

Showing 1 to 1 of 1 entries

Şekil 4.11. Hadoop kümesi çalıştırıldıktan sonra izleyicinin özet bilgileri

Program çalıştıktan sonra yıllık ortalama sıcaklık değerleri bulunulmuş olur. Çizelge 4.1.'de sonuçlar görülmektedir.

Çizelge 4.1. Karadeniz Bölgesi yıllık ortalama sıcaklık sonuçları

Yıl	Ortalama	Yıl	Ortalama	Yıl	Ortalama	Yıl	Ortalama
1926	13,4	1949	12	1972	11,4	1995	12
1927	16,3	1950	12,9	1973	11,3	1996	12,7
1928	13,8	1951	13,3	1974	11,8	1997	11,8
1929	13,1	1952	13,7	1975	12,2	1998	13
1930	13,6	1953	12,2	1976	10,9	1999	13,1
1931	12,7	1954	13,2	1977	11,6	2000	12,2
1932	12,4	1955	14	1978	11,6	2001	13,5
1933	11,7	1956	11,9	1979	12,5	2002	12,7
1934	13	1957	13,4	1980	11,6	2003	12,4
1935	12,3	1958	13,3	1981	12,4	2004	12,5
1936	12,8	1959	11,1	1982	11	2005	12,9
1937	13,2	1960	11,9	1983	11,1	2006	12,3
1938	13,3	1961	11,8	1984	11,6	2007	13,1
1939	13,7	1962	12,6	1985	11,6	2008	12,6
1940	12,9	1963	12,5	1986	11,9	2009	12,9
1941	12,6	1964	11,6	1987	11,3	2010	14
1942	12,1	1965	11,8	1988	11,6	2011	11,5
1943	12,3	1966	13,7	1989	11,8	2012	12,3
1944	13	1967	11,4	1990	11,4	2013	12,4
1945	12,2	1968	12	1991	11,4	2014	13,6
1946	13	1969	11,9	1992	10,2	2015	12
1947	13,5	1970	12,6	1993	10,9	2016	11,2
1948	12,4	1971	12	1994	12,5		

Aynı işlemler yıllık toplam yağış, günlük güneşlenme süresi, günlük ortalama toprak sıcaklığı (20 cm için), günlük ortalama toprak sıcaklığı (50 cm için) günlük aktüel toplam basınç, günlük ortalama nispi nem, günlük ortalama

rüzgar hızı, günlük açık yüzey buharlaşma, günlük toplam güneş şiddeti ve günlük global güneş radyasyonu veri setleri içinde gerçekleştirilmiştir.

### 4.3. Verilerin Analizi

Elde edilen yıllık ortalama verilerin yorumlanabilmesi için Mann-Kendall ve Spearman Rho trend analizleri yapılmıştır. Analiz işlemleri sonucunda sağlıklı sonuçlar alınabilmesi için 1967 yılından itibaren ortalama sıcaklık değerleri alınmıştır. Karadeniz Bölgesi'nin bazı istasyonların kuruluşu 1926 yılına kadar gidiyorken bazıları daha yenidir. Her bölgede istasyon olmadığından, eski verileri tüm bölge analizi yaparken kullanmak yanlış sonuçlar ortaya çıkarabilir. Bu sebepten 1967 yılından itibaren değerlendirme yapılmıştır. Analiz işlemleri için Çizelge 4.2.'deki değerler kullanılmıştır. Yıllık ortalama sıcaklık verileri için yapılan analiz işlemlerinin sonuçları şu şekildedir:

### Günlük Ortalama Sıcaklık (°C)

Çizelge 4.2. Yıllara göre sıcaklık değerleri

Yıl	Ortalama	Yıl	Ortalama	Yıl	Ortalama	Yıl	Ortalama
1967	11,4	1980	11,6	1993	10,9	2006	12,3
1968	12	1981	12,4	1994	12,5	2007	13,1
1969	11,9	1982	11	1995	12	2008	12,6
1970	12,6	1983	11,1	1996	12,7	2009	12,9
1971	12	1984	11,6	1997	11,8	2010	14
1972	11,4	1985	11,6	1998	13	2011	11,5
1973	11,3	1986	11,9	1999	13,1	2012	12,3
1974	11,8	1987	11,3	2000	12,2	2013	12,4
1975	12,2	1988	11,6	2001	13,5	2014	13,6
1976	10,9	1989	11,8	2002	12,7	2015	12
1977	11,6	1990	11,4	2003	12,4	2016	11,2
1978	11,6	1991	11,4	2004	12,5		
1979	12,5	1992	10,2	2005	12,9		

Çizelge 4.2.'deki yıllık ortalama sıcaklık verilerine trend analiz testlerinden Mann-Kendall ve Spearman Rho trend analizleri uygulanmıştır. Mann-Kendall testi sonucunda  $u(t)$  ve  $u'(t)$  değerleri de hesaplanmıştır. Bu değerler sayesinde trendin zamanla değişimi hesaplanır.  $u(t)$  ve  $u'(t)$  yöntemi bir trendin başladığı yeri belirlemek ve trendin zaman içerisindeki seyrini izleyebilmek için kullanılır. Bu yöntem Mann-Kendall testi için bulunan  $z$  değerinin baştan itibaren her değer için bulunması ile elde edilir. (Sneyers, 1990).  $(X_1, X_2, \dots, X_n)$  zamana göre sıralanmış veri ve onların zamana göre sıraları  $j=(1, 2, \dots, n)$  olmak üzere her  $j$  için  $S$  istatistiği ve  $z$  değeri bulunur. Bulunan  $z$  değerleri  $u(t)$  fonksiyonunu meydana getirir.  $\alpha= 0,05$  olarak kabul edilerek  $u(t)$  fonksiyonunun herhangi bir trende sahip olup olmadığına belirlenir.  $u(t)$  değerinin (+) yönde 1,96'nın üzerinde olması artan yönde bir eğilim olduğunu, (-) yönde 1,96'nın altında olması ise azalan yönde bir eğilim olduğunu gösterir.  $u'(t)$  değeri de  $u(t)$ 'ye benzer şekilde seri içinde ters yönde hesaplanarak bulunur.

$u(t)-u'(t)$  grafikleri incelendiğinde ise değerler değişimin başladığı yerde birbirine yaklaşır. Daha sonra birbirlerinden uzaklaşarak trendin başladığı yer ile kuvveti gösterirler. Eğer seri içerisinde herhangi bir trend yoksa  $u(t)$  ve  $u'(t)$  birbirlerine çok defa yaklaşarak yakın salınımlar yapacaklardır (Partal, 2003).

Çizelge 4.3.'de Sıcaklık verilerinin analizi için yapılan Mann\_Kendall Korelasyon testi sonucu bulunan  $u(t)$  değerleri, Çizelge 4.4.'de ise  $u'(t)$  değerleri gösterilmiştir.

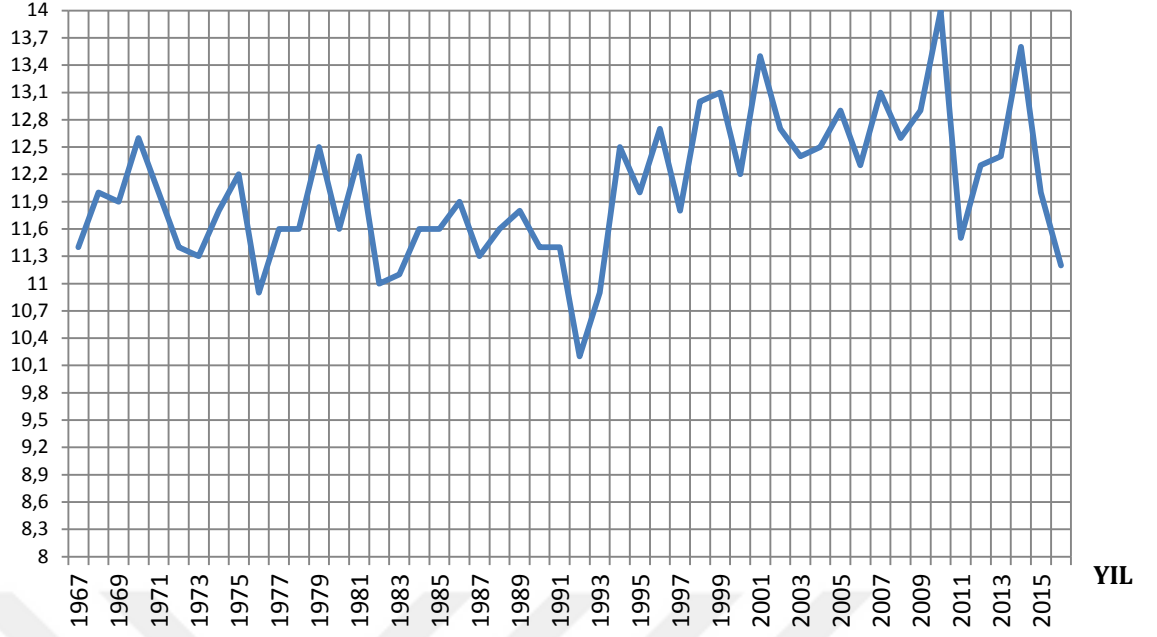
Çizelge 4.3. Sıcaklık için Mann-Kendall Korelasyon Testine Göre u(t) Çizelgesi

SIRA	YIL	DEĞER	n	t	Var(t)	E(t)	u(t)
1	1967	11,4	0	0	0	0	0
2	1968	12	1	1	0,25	0,5	1
3	1969	11,9	1	2	0,92	1,5	0,52
4	1970	12,6	3	5	2,17	3	1,36
5	1971	12	2	7	4,17	5	0,98
6	1972	11,4	0	7	7,08	7,5	-0,19
7	1973	11,3	0	7	11,08	10,5	-1,05
8	1974	11,8	3	10	16,33	14	-0,99
9	1975	12,2	7	17	23	18	-0,21
10	1976	10,9	0	17	31,25	22,5	-0,98
11	1977	11,6	4	21	41,25	27,5	-1,01
12	1978	11,6	4	25	53,17	33	-1,1
13	1979	12,5	11	36	67,17	39	-0,37
14	1980	11,6	4	40	83,42	45,5	-0,6
15	1981	12,4	12	52	102,08	52,5	-0,5
16	1982	11	1	53	123,33	60	-0,63
17	1983	11,1	2	55	147,33	68	-1,07
18	1984	11,6	6	61	174,25	76,5	-1,17
19	1985	11,6	6	67	204,25	85,5	-1,29
20	1986	11,9	12	79	237,5	95	-1,4
21	1987	11,3	3	82	274,17	105	-1,39
22	1988	11,6	7	89	314,42	115,5	-1,49
23	1989	11,8	13	102	358,42	126,5	-1,29
24	1990	11,4	5	107	406,33	138	-1,54
25	1991	11,4	5	112	458,33	150	-1,77
26	1992	10,2	0	112	514,58	162,5	-2,23
27	1993	10,9	1	113	575,25	175,5	-2,61
28	1994	12,5	25	138	640,5	189	-2,2
29	1995	12	21	159	710,5	203	-1,65
30	1996	12,7	29	188	785,42	217,5	-1,05
31	1997	11,8	17	205	865,42	232,5	-0,93
32	1998	13	31	236	950,67	248	-0,39
33	1999	13,1	32	268	1041,33	264	0,12
34	2000	12,2	25	293	1137,58	280,5	0,37
35	2001	13,5	34	327	1239,58	297,5	0,84
36	2002	12,7	31	358	1347,5	315	1,17
37	2003	12,4	27	385	1461,5	333	1,36
38	2004	12,5	29	414	1581,75	351,5	1,57
39	2005	12,9	35	449	1708,42	370,5	1,9
40	2006	12,3	27	476	1841,67	390	2
41	2007	13,1	38	514	1981,67	410	2,34
42	2008	12,6	33	547	2128,58	430,5	2,53
43	2009	12,9	37	584	2282,58	451,5	2,77
44	2010	14	43	627	2443,83	473	3,12
45	2011	11,5	11	638	2612,5	495	2,8
46	2012	12,3	28	666	2788,75	517,5	2,81
47	2013	12,4	30	696	2972,75	540,5	2,85
48	2014	13,6	46	742	3164,67	564	3,16
49	2015	12	23	765	3364,67	588	3,05
50	2016	11,2	5	770	3572,92	612,5	2,63

Çizelge 4.4. Sıcaklık için Mann-Kendall Korelasyon Testine Göre  $u'(t)$  Çizelgesi

SIRA	YIL	DEĞER	n'	t'	Var'(t)	E'(t)	$u'(t)$
50	2016	11,2	0	0	0	0	0
49	2015	12	1	1	0,25	0,5	-1
48	2014	13,6	2	3	0,92	1,5	-1,57
47	2013	12,4	2	5	2,17	3	-1,36
46	2012	12,3	2	7	4,17	5	-0,98
45	2011	11,5	1	8	7,08	7,5	-0,19
44	2010	14	6	14	11,08	10,5	-1,05
43	2009	12,9	5	19	16,33	14	-1,24
42	2008	12,6	5	24	23	18	-1,25
41	2007	13,1	7	31	31,25	22,5	-1,52
40	2006	12,3	3	34	41,25	27,5	-1,01
39	2005	12,9	7	41	53,17	33	-1,1
38	2004	12,5	6	47	67,17	39	-0,98
37	2003	12,4	5	52	83,42	45,5	-0,71
36	2002	12,7	9	61	102,08	52,5	-0,84
35	2001	13,5	13	74	123,33	60	-1,26
34	2000	12,2	3	77	147,33	68	-0,74
33	1999	13,1	13	90	174,25	76,5	-1,02
32	1998	13	13	103	204,25	85,5	-1,22
31	1997	11,8	2	105	237,5	95	-0,65
30	1996	12,7	11	116	274,17	105	-0,66
29	1995	12	3	119	314,42	115,5	-0,2
28	1994	12,5	10	129	358,42	126,5	-0,13
27	1993	10,9	0	129	406,33	138	0,45
26	1992	10,2	0	129	458,33	150	0,98
25	1991	11,4	3	132	514,58	162,5	1,34
24	1990	11,4	3	135	575,25	175,5	1,69
23	1989	11,8	6	141	640,5	189	1,9
22	1988	11,6	6	147	710,5	203	2,1
21	1987	11,3	3	150	785,42	217,5	2,41
20	1986	11,9	10	160	865,42	232,5	2,46
19	1985	11,6	7	167	950,67	248	2,63
18	1984	11,6	7	174	1041,33	264	2,79
17	1983	11,1	2	176	1137,58	280,5	3,1
16	1982	11	2	178	1239,58	297,5	3,39
15	1981	12,4	20	198	1347,5	315	3,19
14	1980	11,6	9	207	1461,5	333	3,3
13	1979	12,5	24	231	1581,75	351,5	3,3
12	1978	11,6	9	240	1708,42	370,5	3,16
11	1977	11,6	9	249	1841,67	390	3,29
10	1976	10,9	1	250	1981,67	410	3,59
9	1975	12,2	21	271	2128,58	430,5	3,46
8	1974	11,8	16	287	2282,58	451,5	3,44
7	1973	11,3	6	293	2443,83	473	3,64
6	1972	11,4	8	301	2612,5	495	3,8
5	1971	12	22	323	2788,75	517,5	3,68
4	1970	12,6	35	358	2972,75	540,5	3,35
3	1969	11,9	21	379	3164,67	564	3,29
2	1968	12	23	402	3364,67	588	3,21
1	1967	11,4	8	410	3572,92	612,5	3,39

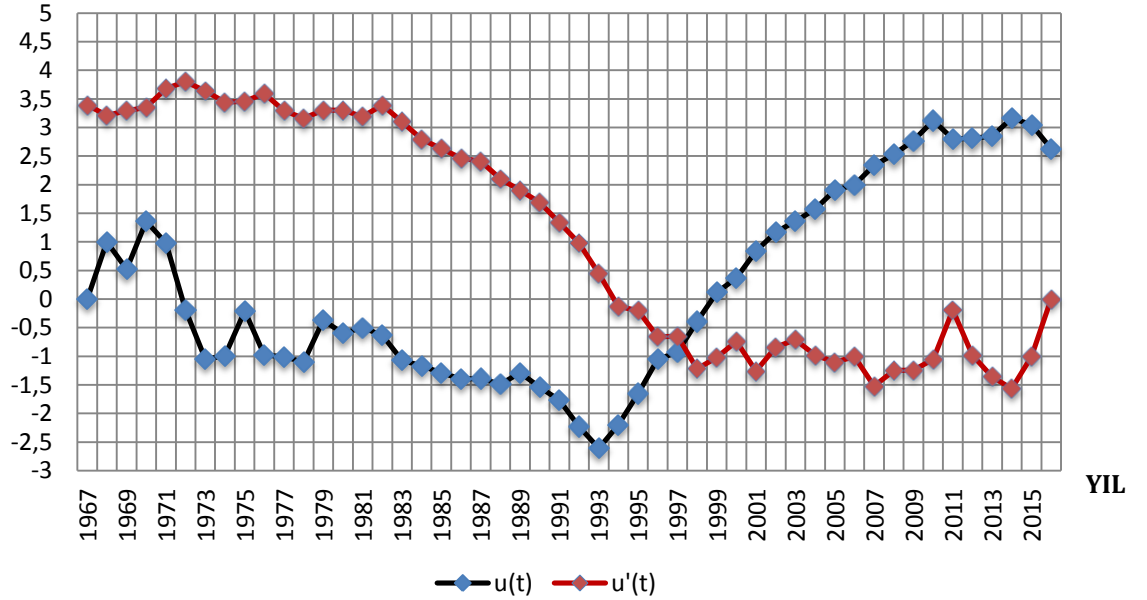
## SICAKLIK



Şekil 4.12. Sıcaklık verilerinin yıllara göre değişim grafiği

Şekil 4.12.'de, Çizelge 4.2.'de verilen sıcaklık ortalama değerlerinin grafikte gösterimi verilmiştir.

## $u(t), u'(t)$



Şekil 4.13. Mann-Kendall korelasyon testi sonucunda  $u(t)-u'(t)$  grafiği

Şekil 4.13.'de, Çizelge 4.3.'de gösterilen  $u(t)$  değerleri ile, Çizelge 4.4.'de gösterilen  $u'(t)$  değerlerinin grafikte gösterimi verilmiştir.

Karadeniz Bölgesi için yıllık ortalama sıcaklık analiz sonuçları şu şekilde bulunmuştur:

#### **Mann-Kendall Testi Sonuçları**

<b>Za/2</b>	1,96
<b>S</b>	405
<b>Sigma S</b>	119,55
<b>Kendal Kor. Kat.</b>	0,33
<b>Z</b>	3,38
<b>Sonuç</b>	İncelenen zaman serisinde artan yönde anlamlı bir trend olduğu sonucuna varılır.

#### **Spearman'in Rho Testi Sonuçları**

<b>Za/2</b>	1,96
<b>Rho Test İst. (rs)</b>	0,45
<b>Z</b>	3,15
<b>Sonuç</b>	İncelenen zaman serisinde artan yönde anlamlı bir trend olduğu sonucuna varılır..

Karadeniz Bölgesi'nin son 50 yıllık ortalama sıcaklık verilerinin analizi sonucunda hem Mann-Kendall ve hem de Spearman Rho testlerine göre %95 güven aralığında Z değeri 1.96'dan büyük olduğundan dolayı artan yönde bir eğilim olduğu görülmektedir. Elde edilen diğer meteorolojik verilere de Mann-Kendall ve Spearman Rho trend analiz testleri uygulanarak herhangi bir yönde eğilimlerinin olup olmadığı gözlemlenmiştir. Elde edilen Z değerleri ve trend sonuçları Çizelge 4.5. 'de gösterilmiştir.

Çizelge 4.5. Mann-Kendall ve Spearman'ın Rho test sonuçları

Veri Adı	Gözlem Yılı	Mann-Kendal Testi		Spearman Rho Testi	
		Z	Trend	Z	Trend
Yıllık Toplam Yağış (mm=kg÷m <sup>2</sup> )	1967-2016	2,1	Artan Trend	1,86	Trend Yok
Günlük Ortalama Sıcaklık (°C)	1967-2016	3,38	Artan Trend	3,15	Artan Trend
Günlük Güneşlenme Süresi (saat)	1967-2016	-1,36	Trend Yok	-1,98	Azalan Trend
Günlük Ortalama Toprak Sıcaklığı 20 cm için (°C)	1967-2016	1,99	Artan Trend	1,56	Trend Yok
Günlük Ortalama Toprak Sıcaklığı 50 cm için (°C)	1967-2016	2,31	Artan Trend	2,8	Artan Trend
Günlük Ortalama Aktüel Basınç (hPa)	1967-2016	0,9	Trend Yok	0	Trend Yok
Günlük Ortalama Nispi Nem (%)	1967-2016	4,67	Artan Trend	4,34	Artan Trend
Günlük Ortalama Rüzgâr Hızı (m÷sn)	1967-2016	1,74	Trend Yok	0,74	Trend Yok
Günlük Toplam Açık Yüzey Buharlaşıma (mm)	1967-2016	4,85	Artan Trend	3,34	Artan Trend
Günlük Toplam Güneş Şiddeti(cal÷cm <sup>2</sup> )	1967-2016	-0,57	Trend Yok	-0,61	Trend Yok

Elde Z değeri -1,96'dan küçük olduğunda azalan yönde bir eğilim, +1,96'dan büyük olduğunda ise artan yönde bir eğilim olduğunu gösterir. Z değeri -1,96 ile +1,96 değerleri arasında olduğunda ise herhangi bir eğilimden söz edilemez.

Çizelge 4.5.'de görüldüğü gibi Mann-Kendall ve Spearman Rho testleri günlük ortalama sıcaklık, 50 cm için günlük ortalama toprak altı sıcaklığı, günlük ortalama basınç, günlük ortalama nispi nem, günlük ortalama rüzgar hızı,

günlük toplam açık yüzey buharlaşma ve günlük toplam güneş şiddeti verileri için eğilimleri aynı yönde hesaplamıştır. Ancak yıllık toplam yağış, günlük güneşlenme süresi ve 20 cm için günlük toprak altı sıcaklığı için eğilimler farklılıklar göstermiştir. Yıllık toplam yağış ve 20 cm için toprak altı sıcaklığı verileri için Mann-Kendall testinde artan bir eğilim olmasına karşın, Spearman Rho testinde herhangi bir eğilime rastlanılmamıştır. Günlük güneşlenme süresi incelendiğinde ise Mann-Kendall testinde artan veya azalan bir eğilim görülmezken, Spearman Rho testinde azalan yönde bir eğilim vardır.



## 5. SONUÇ VE ÖNERİ

Teknolojinin hızla gelişmesine paralel olarak bilgisayar yazılımlarının sayısı ve kullanım alanları da artmıştır. Giderek gelişen bu yapı sonucu büyüyen veriler, hayatımıza “büyük veri” kavramının girmesine neden olmuş ve beraberinde iş süreçlerini olumsuz etkileyebilecek problemler getirmiştir. Geleneksel veri tabanı yöntemleri yerine alternatif özel yöntemlere yönelme ihtiyacını ortaya çıkarmıştır.

Veri madenciliği, verileri işleyerek büyük veriler içerisinde bilgisayar programları vasıtası ile anlamlı verilere ulaşmamızı sağlayan işlemler bütünüdür. Bu kavram ışığında büyük veriler yönetimsel açıdan sınırlanabilir hale getirilmiş ve anlamlı veriye ulaşmamız kolaylaşmıştır. Büyük verilerin işlenmesi için farklı çözüm yöntemleri vardır. Bunlardan bir tanesi de Hadoop teknolojisidir. Hadoop’un mapreduce mimarisi büyük verilerin paralel işlenmesi amacıyla geliştirilmiştir. Ayrıca verilerin saklanması için kullanılan HDFS sistemi verilere daha hızlı erişim sağlamakla beraber veri kaybının da önüne geçmektedir. Mapreduce mimarisinin yeni sayılabilecek bir teknoloji olmasına rağmen yaygınlığı hızla artmaktadır. Esas olarak Hadoop’un Linux ortamında geliştirilmiş olmasına rağmen Hadoop diğer işletim sistemlerine de entegre olabilmektedir. Ancak yine de Hadoop’u Windows işletim sistemi üzerinde çalıştırmak Linux ortamında kullanmaktan daha çok çaba gerektirmiştir.

Yapılan bu çalışma ile Hadoop mapreduce programlama modeli açıklanmış, HDFS ile Mapreduce arasındaki ilişki ortaya konmuştur. Çalışma kapsamında yapılan uygulamalar neticesinde Hadoop kullanılarak büyük veriler üzerinde daha iyi performanslar elde edilebileceği görülmüştür. Ayrıca Hadoop ağına daha fazla makine eklenerek, çok daha büyük verilerin analizi daha hızlı gerçekleştirilebilir.

Bu tez için 10. Meteoroloji Bölge Müdürlüğünden alınan Karadeniz Bölgesindeki aktif bütün istasyonların günlük yağış, sıcaklık, rüzgâr, nem, basınç, 20 cm toprak altı sıcaklığı, 50 cm toprak altı sıcaklığı, açık yüzey buharlaşma miktarı,

güneşlenme süresi ve güneş şiddeti bilgileri Hadoop/mapreduce yöntemi kullanılarak başarıyla işlenebilir hale getirildi. İstasyondaki verilerin başlangıç tarihleri değişkenlik göstermekle beraber 1927 yılına kadar uzanan veriler mevcuttur. Elde edilen verilerin yorumlanması için trend analizi yapılması gerekiyordu. Bu çalışmada Spearman'ın Rho ve Mann Kendall trend testleri kullanılmıştır. Spearman'ın Rho ve Mann Kendall trend testlerinin çıktıları incelendiğinde şu sonuçlara ulaşılmıştır:

- Günlük Ortalama Sıcaklık ( $^{\circ}\text{C}$ ) değerleri incelendiğinde her iki teste göre artan yönde bir eğilim mevcuttur.
- Günlük Ortalama Nispi Nem (%) değerleri incelendiğinde her iki teste göre artan yönde bir eğilim mevcuttur.
- Günlük Toplam Açık Yüzey Buharlaşma (mm) değerleri incelendiğinde her iki teste göre artan yönde bir eğilim mevcuttur.
- Günlük Ortalama Toprak Sıcaklığı 50 cm için ( $^{\circ}\text{C}$ ) değerler incelendiğinde her iki teste göre artan yönde bir eğilim mevcuttur.
- Günlük Ortalama Toprak Sıcaklığı 20 cm için ( $^{\circ}\text{C}$ ) değerler incelendiğinde Mann Kendall testinde artan bir eğilim göstermiştir.
- Günlük Ortalama Güneşlenme Süresi (saat) değerleri incelendiğinde Spearman'ın Rho testinde azalan yönde eğilim göstermiştir.
- Günlük Toplam Yağış ( $\text{mm}=\text{kg}\div\text{m}^2$ ) değerler incelendiğinde Mann Kendall testinde artan bir eğilim göstermiştir.
- Geriye kalan Günlük Ortalama Aktüel Basınç (hPa), Günlük Ortalama Rüzgâr Hızı ( $\text{m}\div\text{sn}$ ) ve Günlük Toplam Güneş Şiddeti( $\text{cal}\div\text{cm}^2$ ) değerleri incelendiğinde %95 anlamlılık seviyesinde azalan veya artan yönde bir trend tespit edilmemiştir.

Elde edilen analiz sonuçları son 50 yıllık verilere göre değerlendirilmiştir. Çalışmamızda elde edilen analiz sonuçları günlük ortalama sıcaklık ve sıcaklıkla alakalı 50 cm için toprak altı sıcaklığı, günlük nem ve günlük açık yüzey buharlaşma verilerinde artış yönünde bir eğilime işaret etmektedir. Sadece 20 cm toprak altı sıcaklığında durum farklılık göstermektedir. Mann-Kendall testine göre artış eğilimi gösteren 20 cm toprak altı sıcaklığı verileri, Spearman

Rho testinde Z deęeri  $-1,96$  ile  $+1,96$  arasında bir deęer aldıęından herhangi bir eęilim göstermemektedir. Toplam yaęıř verileri incelendięinde de durum benzerlik göstermektedir. Toplam yıllık yaęıř verilerinde Mann-Kendall testine gre artıř eęilimi olmasına karřın Spearman Rho testine gre eęilim grlmemektedir. Bazı istasyonların kurulum tarihleri 1926'lı yıllara giderken bazıları ok daha yenidir. Bu farklılıklarından dolayı Karadeniz Blgesi'ndeki genel meteorolojik deęiřimler incelenirken ok eski veriler saęlıklı sonular elde edilebilmesi iin gz ardı edilmiřtir. Bu sebepten tek bir istasyon zerinde yapılacak alıřmalar daha uzun yıllara ait verilerin incelenmesine olanak saęlayacaęından yerel blgedeki deęiřimler ok daha iyi yorumlanabilir.



## KAYNAKLAR

Alabay, N., 2014. Veri madenciliği ve İşletmelerde Uygulama Alanları.

Erişim Tarihi: 22.11.2017

<https://dralabay.wordpress.com/2014/03/07/veri-madenciligi-ve-isletmelerde-uygulama-alanlari/>

Albayrak, M., 2008. EEG Sinyallerindeki Epileptiform Aktivitenin Veri

Madenciliği Süreci ile Tespiti. Sakarya Üniversitesi, Fen Bilimleri

Enstitüsü, Doktora Tezi, 143s, Sakarya.

Alewiwi, M., 2015. Efficient And Secure Document Similarity Search Over Cloud

Utilizing Mapreduce. Sabancı Üniversitesi, Mühendislik Ve Doğa Bilimleri

Enstitüsü, Doktora Tezi, 133s, İstanbul.

Ali, A. M. A., 2016. A Comparative Study Of Hadoop And Spark Frameworks.

Atılım Üniversitesi, Doğal Ve Uygulamalı Bilimler Enstitüsü, Yüksek

Lisans Tezi, 75s, Ankara.

Akkoyun, E., 2011. Parallelization Of Functional Flow To Predict Protein

Functions. Middle East Technical University, Fen Bilimleri Enstitüsü,

Yüksek Lisans Tezi, 51s, Ankara.

Alkan, S., 2010. A Distributed Graph Mining Framework Based On Mapreduce,

Middle East Technical University, Fen Bilimleri Enstitüsü, Yüksek Lisans

Tezi, 47s, Ankara.

An Enterprise Architect's Guide to Big Data Reference Architecture Overview,

2016. Oracle Enterprise Architecture White Paper, 49p.

Akbaş, A., 2017. Büyük Veri (Big Data) Bileşenleri. Erişim Tarihi: 23.11.2017

<https://www.linkedin.com/pulse/b%C3%BCy%C3%BCk-veri-big-data-bile%C5%9Fenleri-alper-akba%C5%9F?published=t>

- Akın, G., 2006. Küresel Isınma, Nedenleri ve Sonuçları. Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi, 46(2), 29-43.
- Apak, G., Ubay, B. ,(2007). Türkiye İklim Değişikliği Birinci Ulusal Bildirimi, [www.meteor.gov.tr](http://www.meteor.gov.tr)
- Astronomical Data Analysis Software and Systems XX. ASP Conference Proceedings, Vol. 442, Massachusetts, USA, 93p.
- Azadani, F., N., 2012. Modeling The Impact Of Climate Change On Water Resources Case Study: Arkansas River Basin In Colorada. Colorado State University, Department Of Civil And Environmental Engineering, Degree Of Master Of Science, 94p, Colorado.
- Bahadır, M. (2011) Güneydoğu Anadolu Proje (GAP) Alanında Sıcaklık ve Yağışın Trend Analizi, Uluslararası Sosyal Araştırmalar Dergisi, The Journal of International Social Research, 4(6): 46-59
- Baltacı, U., 2014. Güneş Aktivitesi İle İklim Değişimi Arasındaki İlişkiler ve Türkiye'deki Olası Etkileri, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 98s, Ankara.
- Bayhan, Y., 2015. Classification Of Search Keywords And Implementation Of A Recommender System. Dokuz Eylül Üniversitesi, Doğal ve Uygulamalı Bilimler Enstitüsü, Yüksek Lisans Tezi, 61s, İzmir.
- Beşel, C., Kayıkçı, E.T, (2016). Meteorolojik Verilerin Zaman Serisi ve Tanımlayıcı İstatistikler ile Yorumlanması; Karadeniz Bölgesi Örneği. TÜCAUM Uluslararası Coğrafya Sempozyumu 13-14 Ekim 2016, Ankara, 2-3.

- Birrell, A. D. Ve Nelson, B. J., 1984. Implementing Remote Procedure Calls. ACM Transactions on Computer Systems, Vol. 2(1), Xerox Corporation Palo Alto Research Center California.
- Bulut, H., 2015. Türkiye’de iklim faktörlerinin ve İklim Değişikliğinin Haşhaş (Papaver Somniferum) Bitkisinin Verimi Üzerine Etkisi. Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 151s, Ankara.
- Cattell, R., 2011, Scalable SQL and NoSQL data stores. ACM SIGMOD Record, 39(3), 12-27.
- Chen, C. L. P., and Zhang, C. Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314-347.
- Cisco Foundation, Cisco Visual Networking Index: Forecast and Methodology, 2016–2021. Erişim tarihi: 22.011.2017  
<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>
- Çakır, F., Akgöbek Ö., 2009. Veri Madenciliğinde Bir Uzman Sistem Tasarımı, Akademik Bilişim, 11-13 Şubat Harran Üniversitesi. Şanlıurfa, 801-806.
- Çaylak, O., 2015. İklim Değişikliğinin Buğday Bitkisinin Gelişimi Ve Verimine Olası Etkilerinin Bitki-İklim Simulasyon Modeli ile İncelenmesi. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 92s, İstanbul.
- Çeribaşı, G., Doğan, E., (2015). Karadeniz ve Sakarya Havzalarında Yıllık Ortalama Yağışların Trend Analizi. SDU International Technologic Science, 7(1), 1-7.

- Çetin, Y., 2014. Mapreduce Kullanarak RDFS Üzerinde Dağıtık Çıkarsama. TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Bilimleri Enstitüsü, 68s, Ankara.
- Çetinkaya, S., 2016. Hadoop/Mapreduce Teknolojisi Kullanılarak Hızlı Tüketim Sektöründe Büyük Veri Analizi. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 82s, İstanbul.
- Çiftçi, M. A., 2016. Hadoop Ve Mapreduce Teknolojisi Aracılığıyla Gıda Tabanlı Mobil Uygulamalar İçin Bir Arama Hizmeti. İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 90s, İstanbul.
- Dean, J. and Ghemawat, 2008. S., "MapReduce: Simplified data processing on large clusters," Communications of the ACM, 51, 107-113.
- Demir, İ., 2012, Hadoop Tabanlı Büyük Ölçekli Görüntü İşleme Altyapısı. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, 87s, Kocaeli
- Demir, Ö., Atay, H., Eskiöğlü, O., Tuvan, A., Demircan, M., Akçakaya, A., 2013. RCP4.5 Senaryosuna Göre Türkiye’de Sıcaklık ve Yağış Projeksiyonları. III. Türkiye İklim Değişikliği Kongresi, 3-5 Haziran, İstanbul, 2-6.
- Dong, X. L., and Srivastava, D., 2013. Big data integration. Data Engineering (ICDE), 2013 IEEE 29th International Conference on, IEEE, 1245-1248.
- Duffy, D. Q., Schnase, J. L., Thompson, J. H., Freeman, S. M., Clune, T. L., 2011. Preliminary Evaluation of MapReduce for High-Performance Climate Data Analysis. Erişim Tarihi:10.04.2017. <http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120009187.pdf>
- Ecer, R., 2009. İklim Değişikliği Ve GAP Bölgesindeki Su Kaynaklarına Etkisi. Harran Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 82s, Şanlıurfa.

Em, A., 2005. Hidrolojik Değişken Yağışa Göre Gap Bölgesindeki Kuraklığın Trend Analizi Bakımından İncelenmesi. Dicle Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 103s, Diyarbakır.

Er, H. R., 2013. Gezgin Satıcı Probleminin Hadoop Üzerinde Çalışan Paralel Genetik Algoritma İle Çözümü. Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 83s, Kocaeli.

Ghali, E., O., 2011. The Effects Of Climate Change On Forest Industry And Environment: Finland And Morocco. Saimaa University Of Applied Sciences, Faculty Of Technology Imatra, Degree Program In Chemical Engineering, 81p, Imatra, Finland .

Ghemawat, S., Gobioff, H., Leung, S., 2003. "The Google File System," Proc. of the 19th ACM Symp. on Operating System Principles, 29–43.

GIBBONS, J. D., (1993), Nonparametric Measures of Association. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–091. Newbury Park, CA: Sage.

Göksu, C., 2011. "Big Data nedir? Geleneksel Veri Yönetimi"ne etkisi ne olabilir?". Erişim Tarihi: 22.11.2017  
<http://www.bthaber.com/gorus/%E2%80%9Cbig-data%E2%80%9D-nedir-geleneksel-%E2%80%9Cveri-yonetimi%E2%80%9Dne-etkisi-ne-olabilir/1/4580>

Gürkan, D.,2005. Seyhan Havzasında İklim Değişimlerinin Yüzeysel Su Kaynaklarına Olan Etkilerinin Araştırılması. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 85s, Ankara.

GÜMÜŞ, V., 2006. Fırat Havzası Akımlarının Trend Analizi ile Değerlendirilmesi. Harran Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 83s, Şanlıurfa.

- Hallaç, İ.R., 2014. Büyük Veri Analizinde Dağıtık Makine Öğrenmesi Algoritmalarının Kullanılması. Fırat üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 86s, Elazığ.
- Hand, D. J., 1998. Data Mining: Statistics and More?, The American Statistician, 52, 112-118.
- Helsel, D.R. ve Hirsch, R.M., 2002. Statistical Methods in Water Resources, Techniques of Water-Resources Investigations of the United States Geological Survey, U.S. Geological Survey, 524p.
- Hilbert, M., and López, P., 2011. The world's technological capacity to store, communicate, and compute information science. Science, 332: 60-65.
- İnan, O., 2003. Veri Madenciliği. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, 83s, Konya.
- Jacobs, P., 1999. Data Mining: What General Managers Need to Know. Harvard Management Update., 4(10), 2p.
- Lam, C., 2010. Hadoop in Action. Manning Publications Co., 301p, Stamford, USA.
- Katzenberger, M., 2014. Impact Of Global Warming In Holarctic And Neotropical Communities of Amphibians. Universidad De Sevilla, Estación Biológica de Doñana Consejo Superior de Investigaciones Científicas, Doctoral Thesis, 203p, Sevilla.
- Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., and Gani, A., 2014. Big data: survey, technologies, opportunities, and challenges, The Scientific World Journal, 4-8.
- Kitler, R. ve Wang, W., 1998. The Emerging Role of Data Mining, Solid State Technology, 42(11), 45.

Koca, M., 2014. The Costs And Benefits Of Turning Data Into Information Using Big Data Systems. Özyeğin Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 86s, İstanbul.

Krishna, M., Kannan, B., Ramani, A., Sathish, S. J., 2010. Implementation and Performance Evaluation of a Hybrid Distributed System for Storing and Processing Images from the Web. Cloudcom, 2010 IEEE Second International Conference on Cloud Computing Technology and Science, 762-767.

Kumar, A., 2014. Basics Of Big Data Analytics Hadoop. Erişim Tarihi: 30.04.2018. <https://www.slideshare.net/AmbujKumar4/basics-of-big-data-analytics-hadoop>.

Macana, E., 2014. Büyük Menderes Havzasında Ortalama, Maksimum ve Minimum Sıcaklıklarda 1970 Sonrasında Gözlenen Eğilimler. Adnan Menderes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 218s, Aydın.

Meteoroloji Genel Müdürlüğü, 2018. Türk Meteoroloji Tarihi. Erişim Tarihi: 03.04.2009. <https://www.mgm.gov.tr/genel/meteorolojitarihi.aspx>.

Oracle, 2016. An Enterprise Architect's Guide to Big Data, 49p.

Önol, B., 2007. Downscaling Climate Change Scenario using Regional Climate Model Over Eastern Mediterranean. Istanbul Technical University, Institute Of Science Technology, Ph.D. THESIS, 106s, İstanbul.

Özfidaner, M. (2007) Türkiye Yağış Verilerinin Trend Analizi ve Nehir Akımları Üzerine Etkisi, Çukurova Üniversitesi, Yüksek Lisans Tez, 82s.

Özekes, S., 2003. Data Mining Models and Application Areas, İstanbul Commerce University Journal of Science, 3, 65-82.

- Padhy, R. P., 2013. Big Data Processing With Hadoop-MapReduce In Cloud Systems. International Journal of Cloud Computing and Services Science, 2(1), 16-27.
- Partal, T., 2003. Türkiye Yağış Verilerinin Trend Analizi. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 101s, İstanbul.
- Redman, T. C., and Blanton, A., 1997. Data Quality For The Information Age. Artech House, 303p, Inc. Norwood, MA, USA.
- Salazar Lahera, N., 2017. Climate Change Impacts And Adaptations In Eastern Us Crop Production. Faculty of the Graduate School of the University of Maryland, Master of Science, 140p, College Park-Washington DC.
- Salur, M. U., 2016. Büyük Veri Araçlarından Hadoop Kullanarak Veri Madenciliği. Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 102s, Denizli.
- Selçuk, A., 2015. Büyük Veri Üzerinde Dağıtık Dosya Sistemi Ve Paralel İşleme Kullanarak Mahremiyet Korunmalı Arama. Sabancı Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 63s, İstanbul.
- Shearer, C., 2000. The Crisp-DM Model: The New Blueprint for Data Mining, Journal of Data Warehousing, 5(4), 13-23.
- Slagle, J., T., 2014. Climate Change In Myanmar: Impacts And Adaptation. Naval PostGraduate School. Master's Thesis, 97p, California.
- Sneyers, R., 1990. On The Statistical Analysis Of Series Of Observations. World Meteorological Organization Technical Note N° 143, 218s.
- Söylev, A., 2013. Biyolojik Veritabanlarında Etkin Benzerlik Hesaplama, TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 72s, Ankara.

Şarkışla, H., 2015. Hadoop MapReduce Algoritmasının Analizi İle Performansa Etki Eden Parametrelerin Tespiti Ve Hadoop Üzerinde Başarım Artımı, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 66s, Sakarya.

Topaloğlu, F., 2007. Veri Madenciliği ile Meteorolojik Parametrelerin analizi ve Zirai Meteoroloji Haritasının Çıkarılması. Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 63s, Elazığ.

Toplaoğlu, N., Yılmaz, M., Savaş, S., 2012. Veri Madenciliği ve Türkiye'deki uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 21, 1-23.

Toros, H., 1993. Klimatolojik Serilerden Türkiye ikliminde Trend Analizi. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 191s, İstanbul.

Türkeş, M., Şen, Ö. L., Kurnaz, L., Madra, Ö., Şahin, Ü., 2013. İklim Değişikliğinde Son Gelişmeler: IPCC 2013 Raporu, 40s.

Türkeş, M., Sümer M. U., Çetiner, G.,( 2000). Küresel İklim Değişikliği ve Olası Etkileri. Çevre Bakanlığı, Birleşmiş Milletler İklim Değişikliği Çerçeve Sözleşmesi Seminer Notları, s. 7-24.

Türkoğlu,İ., Doğan Ş, "Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees ," Doğu Anadolu Bölgesi Araştırmaları Dergisi, 5(2), 163-169., 2007.

Uzun, M., 2016. Map Ve Reduce İle Mobil Ağlarda Uçtan Uca İnternet Hız Analizi. TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans tezi, 83s, Ankara.

Venner, J., 2009. Pro Hadoop. Apress Media LLC., 442p, New York, USA.

Wang, R.Y. and Strong, D.M, 1996. Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, 12(4), 5-33.

White, T., 2009. Hadoop: The Definitive Guide. O'Reilly Media, Inc., 625p, Sebastopol, USA.

Wiley K., Connolly A., Krugho S., Gardner J., Balazinska M., Howe B., Kwon Y. and Bu Y., 2011. Astronomical Image Processing with Hadoop.

Xing, W., 2018. Dünyada Hadoop Kullanan Şirketler. Erişim Tarihi: 15.04.2018. <https://wiki.apache.org/hadoop/PoweredBy>.

Yamaoğlu Çılgın, G., 2006. Türkiye'de Küresel Isınmaya Yol Açan Sera Gazı Emisyonlarındaki Artış İle Mücadelede İktisadi Araçların Rolü. Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 151s.

Yılboğa, H., 2015. Büyük Veritabanlarında Özniteliklerin Etiketlere İndirgenmesine Dayalı İçerik Tabanlı Görüntü Erişimi. Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 109s, İstanbul.

Yılmaz, R. A., 2015. Düşük Maliyet İle Mikro Süper Bilgisayar Oluşturma Ve Apache Hadoop Entegrasyonu. Maltepe Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 106s, İstanbul.

Yılmaz, Y., 2012. Estimating Climate Extremes For Turkey And Its Region. Istanbul Technical University, Informatics Institute, M.Sc. THESIS, 65s, Istanbul.

Zantinge D., Adriaans P., 1997. Data Mining, MA, USA Addison Wesley Longman Publishing, 158p, Boston.

## EKLER

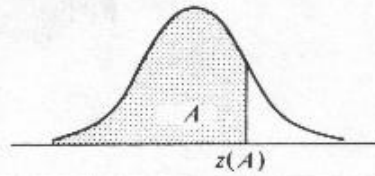
### EK 1: Z Normal Dağılım Çizelgesi

<b>Z</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0.0</b>	0.0000	0.0040	0.0080	0.0080	0.0080	0.0199	0.0239	0.0279	0.0319	0.0359
<b>0.1</b>	0.0398	0.0438	0.0478	0.0478	0.0478	0.0596	0.0636	0.0675	0.0714	0.0753
<b>0.2</b>	0.0793	0.0832	0.0871	0.0871	0.0871	0.0987	0.1026	0.1064	0.1103	0.1141
<b>0.3</b>	0.1179	0.1217	0.1255	0.1255	0.1255	0.1368	0.1406	0.1443	0.1480	0.1517
<b>0.4</b>	0.1554	0.1591	0.1628	0.1628	0.1628	0.1736	0.1772	0.1808	0.1844	0.1879
<b>0.5</b>	0.1915	0.1950	0.1985	0.1985	0.1985	0.2088	0.2123	0.2157	0.2190	0.2224
<b>0.6</b>	0.2257	0.2291	0.2324	0.2324	0.2324	0.2422	0.2454	0.2486	0.2517	0.2549
<b>0.7</b>	0.2580	0.2611	0.2642	0.2642	0.2642	0.2734	0.2764	0.2794	0.2823	0.2852
<b>0.8</b>	0.2881	0.2910	0.2939	0.2939	0.2939	0.3023	0.3051	0.3078	0.3106	0.3133
<b>0.9</b>	0.3159	0.3186	0.3212	0.3212	0.3212	0.3289	0.3315	0.3340	0.3365	0.3389
<b>1.0</b>	0.3413	0.3438	0.3461	0.3461	0.3461	0.3531	0.3554	0.3577	0.3599	0.3621
<b>1.1</b>	0.3643	0.3665	0.3686	0.3686	0.3686	0.3749	0.3770	0.3790	0.3810	0.3830
<b>1.2</b>	0.3849	0.3869	0.3888	0.3888	0.3888	0.3944	0.3962	0.3980	0.3997	0.4015
<b>1.3</b>	0.4032	0.4049	0.4066	0.4066	0.4066	0.4115	0.4131	0.4147	0.4162	0.4177
<b>1.4</b>	0.4192	0.4207	0.4222	0.4222	0.4222	0.4265	0.4279	0.4292	0.4306	0.4319
<b>1.5</b>	0.4332	0.4345	0.4357	0.4357	0.4357	0.4394	0.4406	0.4418	0.4429	0.4441
<b>1.6</b>	0.4452	0.4463	0.4474	0.4474	0.4474	0.4505	0.4515	0.4525	0.4535	0.4545
<b>1.7</b>	0.4554	0.4564	0.4573	0.4573	0.4573	0.4599	0.4608	0.4616	0.4625	0.4633
<b>1.8</b>	0.4641	0.4649	0.4656	0.4656	0.4656	0.4678	0.4686	0.4693	0.4699	0.4706
<b>1.9</b>	0.4713	0.4719	0.4726	0.4726	0.4726	0.4744	0.4750	0.4756	0.4761	0.4767
<b>2.0</b>	0.4772	0.4778	0.4783	0.4783	0.4783	0.4798	0.4803	0.4808	0.4812	0.4817
<b>2.1</b>	0.4821	0.4826	0.4830	0.4830	0.4830	0.4842	0.4846	0.4850	0.4854	0.4857
<b>2.2</b>	0.4861	0.4864	0.4868	0.4868	0.4868	0.4878	0.4881	0.4884	0.4887	0.4890
<b>2.3</b>	0.4893	0.4896	0.4898	0.4898	0.4898	0.4906	0.4909	0.4911	0.4913	0.4916
<b>2.4</b>	0.4918	0.4920	0.4922	0.4922	0.4922	0.4929	0.4931	0.4932	0.4934	0.4936
<b>2.5</b>	0.4938	0.4940	0.4941	0.4941	0.4941	0.4946	0.4948	0.4949	0.4951	0.4952
<b>2.6</b>	0.4953	0.4955	0.4956	0.4956	0.4956	0.4960	0.4961	0.4962	0.4963	0.4964
<b>2.7</b>	0.4965	0.4966	0.4967	0.4967	0.4967	0.4970	0.4971	0.4972	0.4973	0.4974
<b>2.8</b>	0.4974	0.4975	0.4976	0.4976	0.4976	0.4978	0.4979	0.4979	0.4980	0.4981
<b>2.9</b>	0.4981	0.4982	0.4982	0.4982	0.4982	0.4984	0.4985	0.4985	0.4986	0.4986
<b>3.0</b>	0.4987	0.4987	0.4987	0.4987	0.4987	0.4989	0.4989	0.4989	0.4990	0.4990

EK 2: Cumulative probabilities of the standart normal distribution

**Cumulative Probabilities of the Standard Normal Distribution.**

Entry is area  $A$  under the standard normal curve from  $-\infty$  to  $z(A)$



$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Selected Percentiles	
Cumulative probability $A$ :	.90    .95    .975    .98    .99    .995    .999
$z(A)$ :	1.282    1.645    1.960    2.054    2.326    2.576    3.090

## ÖZGEÇMİŞ

Adı Soyadı : Mustafa KAYA  
Doğum Yeri ve Yılı : Merzifon, 1986  
Medeni Hali : Bekâr  
Yabancı Dili : İngilizce  
E-posta : mustafa.kaya@omu.edu.tr



## Eğitim Durumu

Lise : Selçuklu Cumhuriyet Yabancı Dil Ağırlıklı Lisesi, 2005  
Lisans : Selçuk Üniversitesi, Teknik Eğitim Fakültesi, Elektronik-  
Bilgisayar Öğretmenliği, 2009

## Mesleki Deneyim

Bolvadin Teknik Ve Endüstri Meslek Lisesi	2009-2010
Seydişehir Leman - Kadir Parıltı Mes. ve Tek. And. Lisesi	2010-2015
OMU Bafra Meslek Yüksekokulu	2015-..... (halen)