

KONUMSAL BÜYÜK VERİ: ÖZELLİKLERİ VE AVANTAJLARI

Hacer Kübra SEVİNÇ, Prof. Dr. İsmail Rakıp KARAŞ², Huda İBRAHİM³

¹Öğr. Gör., Sinop Üniversitesi, Bilgisayar Teknolojileri Bölümü, 57400, Ayancık, Sinop, hacerkubra@gmail.com

²Prof. Dr., Karabük Üniversitesi, Bilgisayar Mühendisliği Bölümü, 78050, Karabük, ismail.karas@karabuk.edu.tr

³ Professor, Dept.of Information System, School of Computing, Universiti Utara Malaysia, Sintok 06010, Kedah, Malaysia, huda753@uum.edu.my

ÖZET

Konumsal büyük veri, mevcut bilgi sistemlerinde yer alan büyük verinin bir parçasını oluşturmaktadır. Her gün üretilen yaklaşık 2.5 kilyon byte verinin içerisinde konum verileri de önemli bir yer oluşturmaktadır. GNNS veya RFID ile özellikle mobil ve IoT cihazlar ile üretilen bu verinin işlenmesi ve elde edilen sonuçların değerlendirilmesi önem arz etmektedir. Özellikle gerçek zamanlı trafik hareketleri, hava durumu verisi, alternatif rota üreten sürücü asistan uygulamaları bu önem için örnek oluşturmaktadır. Toplanan konumsal büyük verinin hacmi, hızı, doğruluğu, değeri ve çeşitliliği yani büyük verinin 5V'si konumsal açıdan da incelenmelidir. Bu yazıda konumsal büyük verinin özellikleri anlatılırken, sağladığı avantajlardan (zaman ve iş gücü karı, kent planlama, sağlık hizmetleri vb.) bahsedilerek, bu alandaki çalışmalar incelenecektir. Ayrıca bu alanda çalışan genç araştırmacılar için yol göstermesi amaçlanmaktadır.

Anahtar Sözcükler: Konumsal büyük veri, büyük veri.

ABSTRACT

GEOSPATIAL BIG DATA: FEATURES AND AVANTAGES

Geospatial big data forms part of the big data in existing information systems. Location data also has an important place among the approximately 2.5 quintillion bytes of data produced every day. It is essential to process this data, produced with GNNS or RFID, especially with mobile and IoT devices, and evaluate the results obtained. Especially real-time traffic movements, weather data, and driver assistant applications that generate alternative routes are examples of this importance. The volume, speed, accuracy, value, and diversity of the geospatial big data collected, that is, the 5V of the big data should also examine from a spatial perspective. In this article, while explaining the features of geospatial big data, the advantages it provides (time and labor profit, urban planning, health services) will be mentioned, and the studies in this field will examine. It also intends to guide young researchers working in this field.

Keywords: Geospatial big data, big data,.

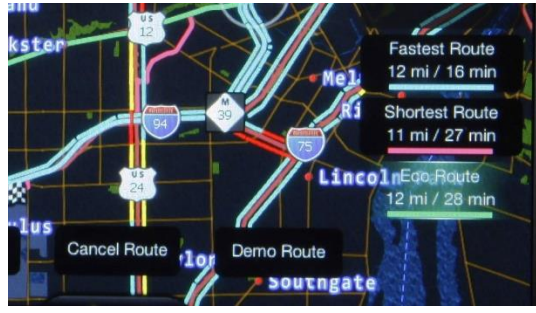
1. GİRİŞ

Günümüzde teknolojinin geldiği noktada, üretilen veri miktarı, günlük teknolojik alet kullanımı, günlük veri üretimi gibi bilgilerden bahsetmek yersizdir. İnsanlık geri dönüşü olmayan bir sensör, veri, bilgi üretimi ve tüketimi sağlayan teknoloji kullanımında çok yol kat etmiştir. Büyük veri, bulut bilişim, siber güvenlik, yapay zeka, otonom sistemler, akıllı cihazlar, nesnelerin interneti, sanal gerçeklik, blok zinciri (Duggal, 2022) gibi kavramlar her sektörde kullanılmakta ve her yöneticinin, araştırmacının bildiği temel kavramlar haline gelmiştir.

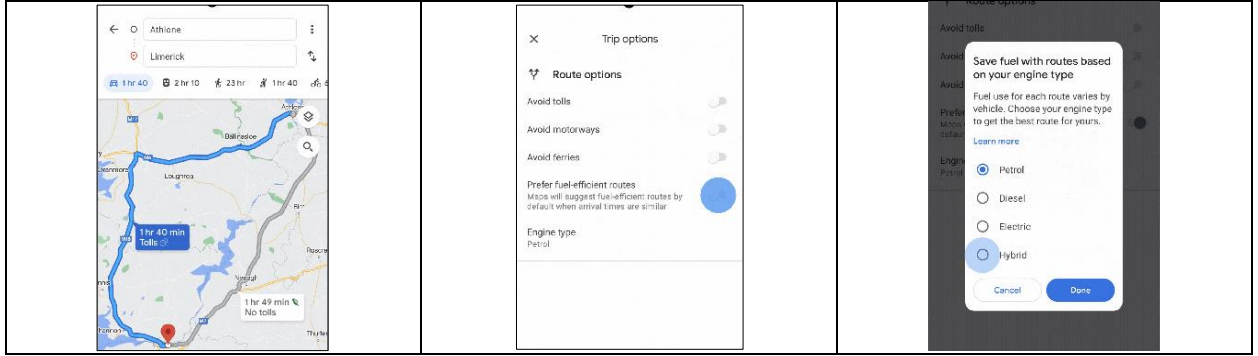
Büyük veri daha fazla çeşitlilik içeren ve hızla artan büyük hacimli veridir. Büyük veri özellikle kaynaklarından gelen daha büyük, daha karmaşık veri kümeleridir. Bu veri kümeleri o kadar hacimlidir ki, geleneksel veri işleme yazılımı onları yönetemez. Ancak bu büyük hacimli veriler, kaynak yetersizliği, klasik yöntemler ile veri işleme süresinin uzunluğu gibi daha önce üstesinden gelinemeyen problemleri çözmek için kullanılabilir. (Oracle, 2022)

Konumsal veri, genellikle haritalama için kullanılan topoloji ve koordinatları içeren coğrafi konum bilgilerini içeren verilerdir. (Lenka, 2016) Konumsal Büyük Veri (Geospatial Big Data) geniş çapta mevcut donanım, yazılım ve/veya insan kaynaklarının kapasitesini aşan ve konum bilgilerini içeren veri kümeleri olarak tanımlanabilir. Yaygın olarak bulunan kaynakların (yani donanım, yazılım, insan kaynakları) kapasitesini aşan ve üzerinde çalışılması özel çaba gerektiren coğrafi konumsal verilerdir. (McCoy, 2017)

Konumsal büyük veri (KBV) ile gerçek zamanlı trafik hareketleri, hava durumu verisi, alternatif rota üreten sürücü asistan uygulamaları gerçekleştirilebilir. Enerji tasarrufu konusunda Ford mühendisleri tarafından araç rotalama probleminde "En hızlı rota" seçeneğine ek olarak "Eko-Rota (Eco Route)" geliştirilmiştir (Şekil 1). Eko-rota seçeneği bazı araçlarda yaklaşık olarak %15 yakıt tasarrufu sağlamaktadır. (Lee & Kang, 2015) Ayrıca Google yakın zamanda 40 ülke için çevre dostu rotalar seçeneğini ekleyeceğini duyurmuştur. Bu özelliği öncelikle İngiltere'de yaşayan bazı kullanıcılar için aktif etmiştir. Kullanıcılar, yaprak etiketle işaretlenmiş çevre dostu rotayı görebilecekler (Şekil 2). (Metha, 2022)



Şekil 1. Eko-Rota seçeneği olan navigasyon sistemi



Şekil 2. Google Haritalar'da çevre dostu rotalar için yakıt türü seçeneği.

2. BÜYÜK VERİ

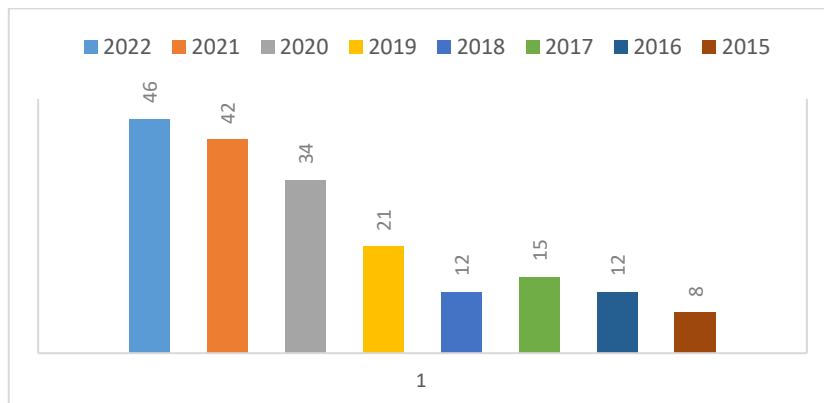
Büyük veri, geleneksel veri işleme uygulama yazılımları ile yönetilemeyecek kadar büyük ve karmaşık veri kümelerini analiz etme, sistematik olarak çalıştırma gibi süreçleri ele alan veri yönetim sistemleridir. Büyük veri, son yıllarda kuruluşların dijital dönüşüm sürecinde büyük yankı uyandıran, en yaygın bilgi ve araştırma alanlarından biridir. (BilgeAdam, 2022)

Büyük verinin 5V'si, Hacim (Volume), Hız (Velocity), Doğruluk (Veracity), Çeşitlilik (Variety), Değer (Value) olarak bilinen özelliklerdir. Hacim; veri büyüklüğünü, Hız; verinin üretilme hızını, Doğruluk; verinin güvenilirliğini, Çeşitlilik; verinin farklı formatlardan veya ortamlardan olmasını ve Değer; verinin işlendikten sonra ne anlam taşıdığını ifade etmektedir.

3. KONUMSAL BÜYÜK VERİ

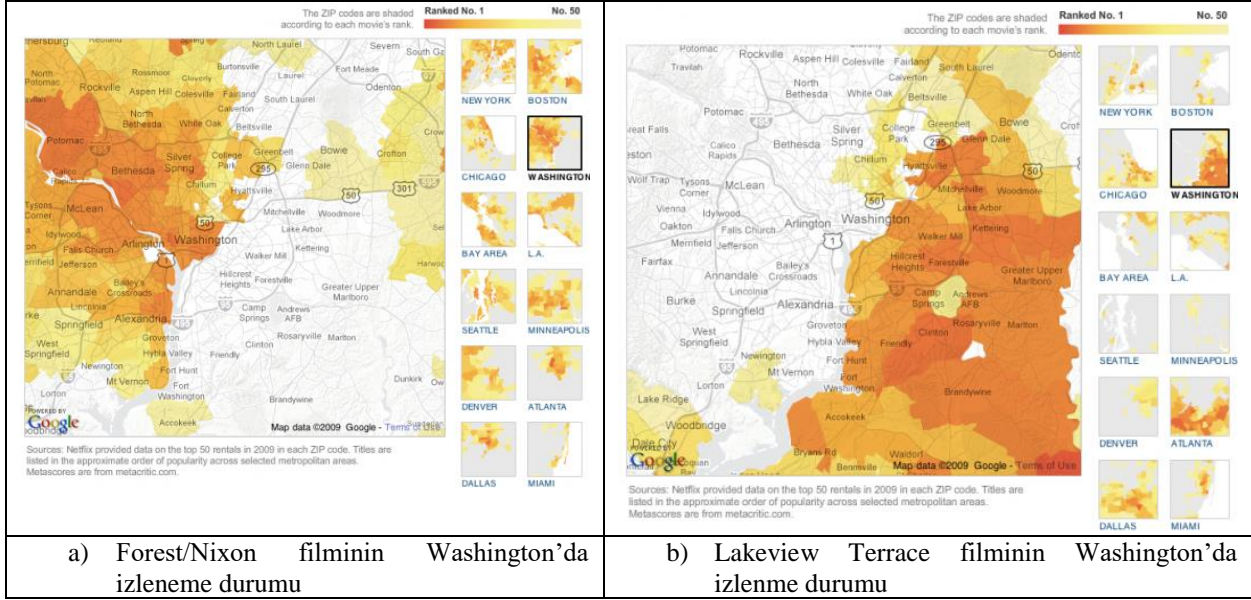
Konumsal büyük veri, büyük verinin bir parçasıdır. Konumsal Coğrafi Bilgi, çoğunlukla çevresel sensörler, kameralar, web kameraları, sosyal medya ve hatta insanların günlük aktiviteleri gibi sabit ve mobil sensörler tarafından her gün üretilmekte ve toplanmaktadır. Verinin gün geçtikçe büyümesi, CBS sistemlerini de etkilemektedir. Mevcut masaüstü yazılımları (ArcGIS ve QGIS gibi) birkaç gigabayt veriyi işleyebilmektedir, fakat günümüzde üretilen coğrafi veri miktarı terabayt hatta petabaytları bulmaktadır. Bu durumda coğrafi veri de büyük veri teknikleri ile işlenmelidir.

Hakem denetimli akademik literatür platformlarından birisi olan Scienccedirect üzerinde "geospatial big data" terimi arandığında son yıllarda bu alanda yapılan akademik çalışmalarında attığı görülmektedir (Şekil 3).



Şekil 3. Scienccedirect yayın sayıları

Konumsal veri her alanda değerlidir ki en büyük ücretli aboneliğe dayalı video akış şirketlerinden birisi olan Netflix üyelerinin posta koduna göre hangi filmin hangi bölgelerde popüler olduğunu görselleştirmiştir. (Şekil 4) Yapay zeka, makine öğrenmesi, büyük veri gibi konulardan bahsedildiğinde Netflix örneğine değinmemek imkansızdır. Netflix algoritması öncelikle benzer yapımları beğenen üyeleri eşleştirerek, öneride bulunma üzerine kurulu bir yapay zeka üzerine kurulmuş olmasına rağmen, kullanıcıların coğrafi bilgilerine göre de yapımların izlenme oranlarını ortaya koymuştur. Bu tür analizler pazarlama stratejilerinde önemli yer tutmaktadır. Bazı filmler tüm coğrafyalarda çok izlense de bazı filmler bazı bölgelerce daha fazla izlenmektedir. Bu da şirketin hedeflediği kitleye göre üretmesi gereken yapımlar konusunda önsezi geliştirmesini desteklemektedir.

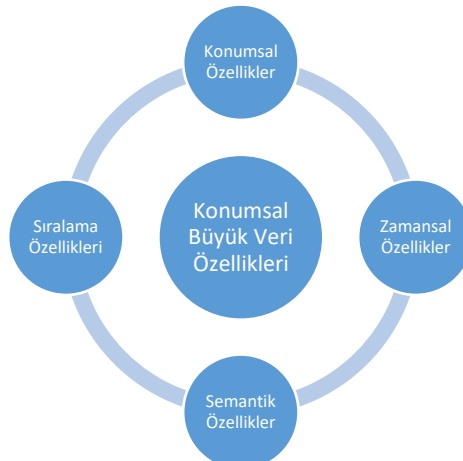


Şekil 4. Netflix bölgelere göre izlenme oranları ısı haritası (FlowingData, 2010)

Li ve diğerleri (2016)'ne göre, Konumsal büyük veri ISPRS TC II'nin çeşitli Çalışma Gruplarının odak noktası coğrafi veri işleme ve teorileri ile ilgili olan hedeflerinden bazılarını ele alır. Ayrıca yaptıkları çalışmada Konumsal büyük veriyi şu başlıklarda incelemişlerdir: Konumsal coğrafi verinin toplanması, Kalite değerlendirme, Veri modellenmesi ve yapılandırılması, Veri görselleştirme ve görsel analiz, Veri madenciliği ve bilgi keşfidir.

Yin ve diğerleri (2021), kentsel arazi kullanımı haritalaması için uzaktan algılama ve konumsal büyük verinin bütünleştirilmesi konulu çalışmalarında konumsal büyük verini özelliklerini şöyle sıralamışlardır (Şekil 5):

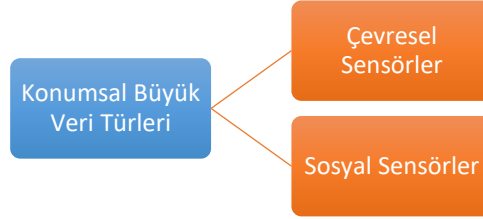
- Konumsal özellikler (OpenStreetMap ile sağlanan bina, yol park bilgileri gibi)
- Zamansal özellikler (İnsan faaliyetlerinin hareketlilik modellerini ortaya çıkarabilecek zamansal özelliklere sahip birçok veri kaynağı örneği (örneğin, cep telefonu verileri, trafik verileri, sosyal medya verileri ve akıllı kart verileri) vardır.
- Semantik özellikler (Fotoğraflar, sokak görünümleri, kitle kaynaklı coğrafi etiketli fotoğraflar, sosyal medya fotoğrafları)
- Sıralama özellikleri (Sosyal medya ve arama motorları verileri)



Şekil 5. Konumsal Büyük Veri Özellikleri

Huang v.d. (2021), akıllı şehirler için konum tabanlı büyük veri konulu çalışmalarında konumsal büyük verinin iki ana türü olduğundan bahsetmektedir (Şekil 6):

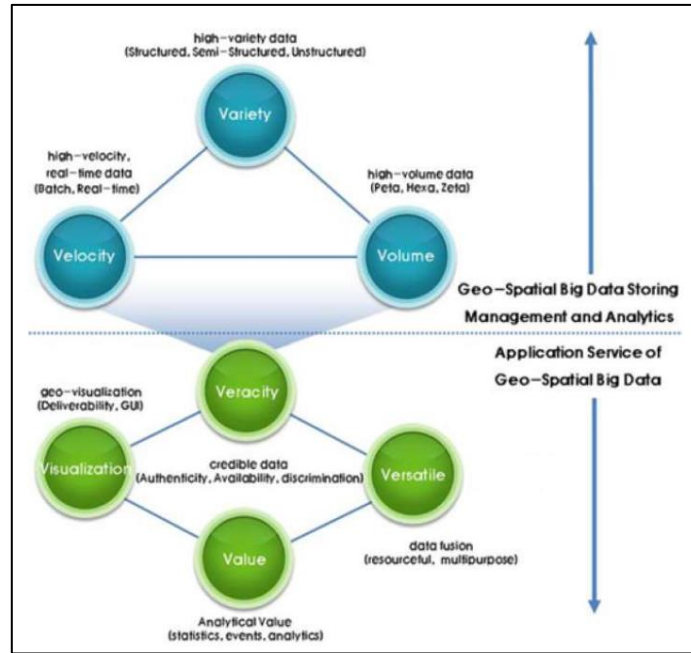
- Büyük yersel gözlem verisi (çevresel sensörler) : fiziksel çevre özellikleri ve yer kürenin karakteristiği
- Büyük insan davranışları verisi(sosyal sensörler): sosyal medya verileri, akıllı kart seyahat verileri, GNNS sistemleri içeren mobil cihazların verileri



Şekil 6. Konumsal Büyük Veri Türleri

Coğrafi veri gerçek dünya özelliklerini ve bilgilerini kullanarak üretilir. Bu özellikle coğrafi koordinat sistemi ile bağlantılıdır. Üretilen bu veri, nüfus, eğitim, su kaynakları, elektrik hatları, tedarik zinciri gibi konularda olabilir. CBS bu verilerin analiz edilmesini ve bu veriler arasındaki örüntü ve ilişkileri yorumlamaya yarar. Fakat örneğin Küresel Biyoçeşitlilik Danışma Tesisi (Global Biodiversity Information Facility) 400milyon tür için enlem ve boylam bilgi çifti kaydetmiştir ki bu verileri klasik yöntemler ile işlemek yeterli olmayacaktır. (Lenka, 2016)

Büyük veri için özellikle 3V (bugünlerde 5V) üzerinde durulurken, Konumsal Büyük Veri ile 2V daha eklenmektedir. Bunlar Visualization (görselleştirme) ve Versalite (çok yönlü) (Loukili & et al., 2020). (Şekil 7'de gösterilen Value ve Veracity büyük verinin 5V'i içerisine alınmıştır.) Görselleştirme, analiz sonucunun kullanıcılara ne kadar etkili bir şekilde iletilebileceğini ölçen bir özelliktir ve Versatile, konumsal büyük verinin çok yönlü kullanım olasılığı ile ilgilenir.

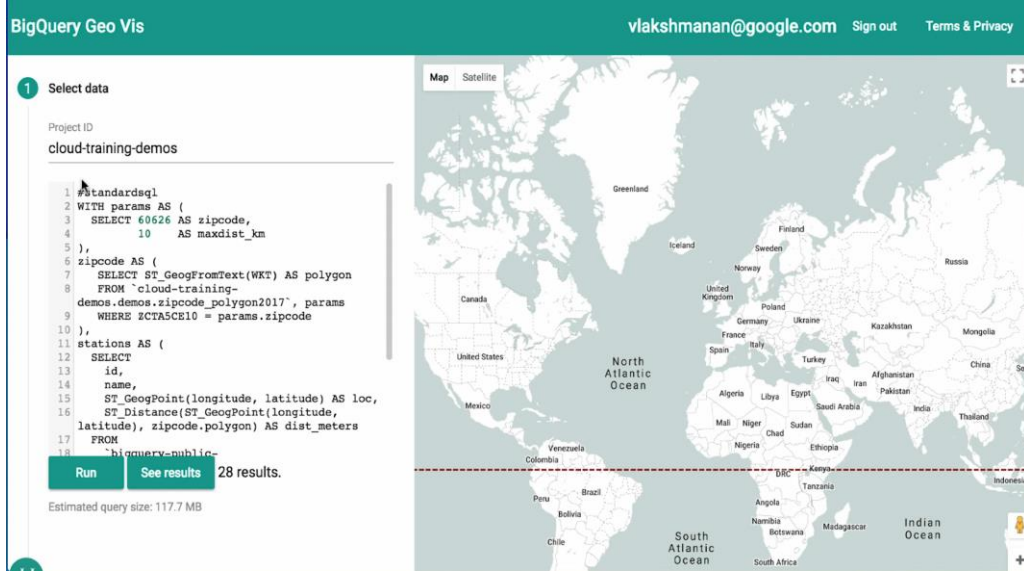


Şekil 7. Konumsal Büyük Verinin 7V'si

Büyük veri ile çalışabilmek için bazı çatı (framework) ve araçlar bulunmaktadır. Hadoop, mongoDB, Cassandra, Apache Spark, DynamoDB bunlardan bazılarıdır. Konumsal büyük veri ile çalışabilmek için kullanılacak araçlar ise şunlardır: SpatialHadoop, Google BigQuery GIS, GIS Tools for Hadoop (Esri), GeoSpark. MondoDB, konumsal veriler ile GeoJSON nesneleri ile çalışabilir. (MongoDB, 2022) Bu araçlar her ne kadar kullanım amaçlarına göre farklı performansları verseler de küme boyutuna göre SpatialHadoop ve GeoSpark karşılaştırıldığında Geospark'ın büyük veri analitiğinde daha hızlı sonuçlar verdiği görülmektedir. (Lenka, 2016)

3.1. Google BigQuery GIS

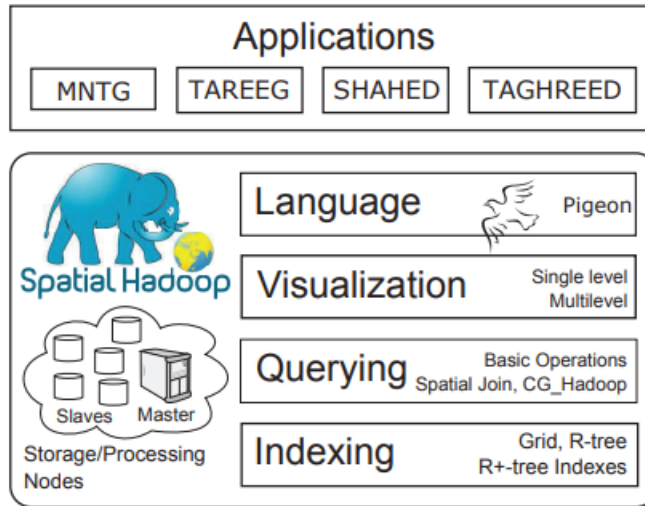
Google BigQuery, büyük veri setlerinin analizini sağlayan ve Google tarafından geliştirilen bulut üzerinde çalışan bir araç sunmaktadır (Şekil 8). BigQuery, verilerin sütunlar ile saklanması sağlar. Bunun sebebi, klasik veri tabanlarındaki indis oluşturma zorunluluğunu ortadan kaldırmaktır. Bu aracın sınırlılıkları ise sadece Google Standart SQL üzerinde sorgulama yapılabilir olmasıdır. Ayrıca Python için geliştirilen GEOGRAPHY veri tipini içeren BigQuery Kütüphanesi bulunmaktadır (Google, 2022). Diğer dilleri henüz desteklememektedir.



Şekil 8. Google BigQuery

3.2. SpatialHadoop

SpatialHadoop, Apache Hadoop'ta konumsal büyük veri kümelerini işlemek için özel olarak tasarlanmış açık kaynaklı bir MapReduce uzantısıdır. (Apache, 2022) Şekil 9'da SpatialHadoop'a genel bir bakış görülmektedir.



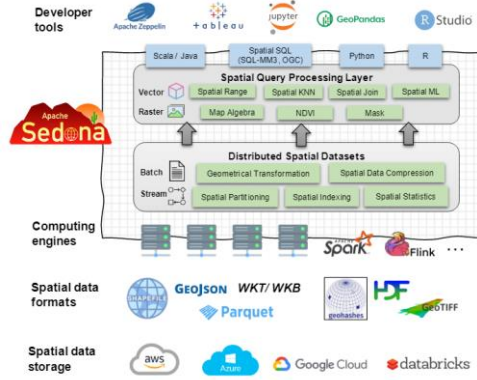
Şekil 9. SpatialHadoop genel bakış

SpatialHadoop, konumsal büyük veri içeren birçok gerçek zamanlı uygulama olan MNTG (web tabanlı trafik oluşturma aracı) veya TAREEG (OpenStreetMap verisi için MapReduce üretici) gibi, birlikçe çalışabilmektedir. (Eldawy & Mokbel, 2014)

3.3. Apache Sedona (GeoSpark)

Apache Sedona, büyük konumsal verileri işlemek için tasarlanmış bir küme bilgi işlem sistemidir. Sedona ile karmaşık konumsal nesnelere (vektör, raster ve diğer çeşitli formatlardaki veriler) ile çalışılabilir. R-Tree, Quad-Tree ya da K-

En yakın komşu algoritması ile dağıtık konumsal sorgular gerçekleştirebilmektedir. Scala, Java, Python, R programlama dillerini desteklemektedir. (Şekil 10)



Şekil 10. Apache Sedona Sistem Mimarisi

3.4. AWS Athena

Amazon Athena, standart SQL kullanarak verileri doğrudan Amazon Simple Storage Service (Amazon S3) içinde analiz etmeyi kolaylaştıran etkileşimli bir sorgu hizmetidir. Bu aynı zamanda Geospatial veri türlerini destekler ve Geospatial aramalar ve işlemler sağlamaktadır. (AWS, 2022)

4. SONUÇLAR

Konumsal büyük veri ile hacmi yüksek verilerle çalışıldığı için birçok araştırmacı verilerin güvenilirliği ve geçerliğinden şüphe duymaktadır. Fakat veri kalitesi ile ilgili farklı kaynaklardan toplanan verilerin karşılaştırılması veya ek doğrulama yöntemleri ile bu sorunun üstesinden gelinebilir. Ayrıca konumsal büyük verinin kent planlaması, tedarik zinciri veya akıllı şehirler konseptlerinde getireceği büyük faydalar yadsınamaz. Özellikle şehirlerin karşılaştığı trafik sorunları gibi gerçek zamanlı verilerin işlenerek çözülebileceği sorunlar akıllı şehirlerin daha ilerlemesini olmasını sağlayacaktır.

Büyük verinin 5V'sinden birisi olan çeşitlilik özelliği ile çok farklı veri kaynağından veri kümesi beslenmektedir. Bu verilerin arasında konumsal verilerde büyük yer tutmaktadır. Bu konumsal büyük veriyi doğru işlemek ilgili alana büyük getiriler sağlayacaktır. Netflix örneğinde görüldüğü gibi, çevrimiçi çalışan tüm sistemlerin sahip olduğu konumsal veri değerlendirilmeli analiz edilmeli ve işlenmelidir. Pazarlamadan, sağlığa, eğitimden üretime her alanda yapılacak analizler daha verimli sonuçlar elde edilmesini sağlayacaktır. Büyük verinin insan yaşamı içerisinde gün geçtikçe daha fazla yer alması ile artık bu alandaki çalışmaların artacağı gerçeği araştırmacıların dikkatini çekmektedir. Gelecek çalışmalarda konumsal büyük verinin analiz edilecek insan hayatına getireceği katkılar değerlendirilmeli ve sektörün bu analizleri yapabilmesi için teşvik edilmesi gerekmektedir.

KAYNAKLAR

Apache. (2022, 10 10). *SpatialHadoop*. SpatialHadoop: <http://spatialhadoop.cs.umn.edu/> adresinden alındı

AWS. (2022, 10 10). *AWS Supported geospatial functions*. <https://docs.aws.amazon.com/athena/latest/ug/geospatial-functions-list.html> adresinden alındı

BilgeAdam. (2022, 08 25). *Big Data*. Bilge Adam Akademi: <https://akademi.bilgeadam.com/courses/big-data/> adresinden alındı

Duggal, N. (2022, 09 12). *Top 18 New Technology Trends for 2022*. simplelearn: <https://www.simplilearn.com/top-technology-trends-and-jobs-article> adresinden alındı

Eldawy, A., & Mokbel, M. (2014). The Ecosystem of SpatialHadoop. *SIGSPATIAL Special*. doi:<https://doi.org/10.1145/2766196.2766198>

FlowingData. (2010, 01 11). *The Geography of Netflix Rentals*. FlowingData: <https://flowingdata.com/2010/01/11/the-geography-of-netflix-rentals/> adresinden alındı

- Google.** (2022, 10 10). *Introduction to geospatial analytics*. Google Cloud BİgQuery: <https://cloud.google.com/bigquery/docs/geospatial-intro> adresinden alındı
- Huang, H., Yao, X., Krisp, J., & Jiang, B.** (2021). Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Computers Environment and Urban Systems*, 90. doi:10.1016/j.compenvurbsys.2021.101712
- Lee, J.-G., & Kang, M.** (2015). Geospatial Big Data: Challenges and Opportunities. *Big Data Research*, 74-81. doi:<https://doi.org/10.1016/j.bdr.2015.01.003>
- Lenka, R. K.** (2016). Comparative analysis of SpatialHadoop and GeoSpark for geospatial big data analytics. *2nd International Conference on Contemporary Computing and Informatics (IC3I)* (s. 484-488). Greater Noida, India: IEEE. doi:10.1109/IC3I.2016.7918013
- Loukili, Y., & et al.** (2020). Analysis and exploitation of Geospatial Big Data: State of art. *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, (s. 1-6). doi:10.1109/AICT50176.2020.9368706.
- McCoy, M. D.** (2017). Geospatial Big Data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science*, 74-94. doi:<https://doi.org/10.1016/j.jas.2017.06.003>.
- Metha, I.** (2022, 09 6). *Google Maps is expanding its eco-friendly navigation feature to Europe*. techcrunch: <https://techcrunch.com/2022/09/06/google-maps-is-expanding-its-eco-friendly-navigation-feature-to-40-more-countries/> adresinden alındı
- MongoDB.** (2022). *Geospatial Queries*. 8 29, 2022 tarihinde MongoDB Docs: <https://www.mongodb.com/docs/manual/geospatial-queries/#std-label-geospatial-geojson> adresinden alındı
- Oracle.** (2022, 08 30). *Büyük Veri nedir?* Oracle: <https://www.oracle.com/tr/big-data/what-is-big-data/> adresinden alındı
- Songnian Li, S. D.** (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119-133. doi:<https://doi.org/10.1016/j.isprsjprs.2015.10.012>
- Yin, J., Dong, J., Hamm, N., Li, Z., & Wang, J.** (2021). Integrating remote sensing and geospatial big data for urban land use mapping: A review. *International Journal of Applied Earth Observation and Geoinformation*. doi:10.1016/j.jag.2021.102514