

1. Büyük Veri nedir? Kısa bir tanım veriniz.

Büyük Veri, geleneksel veri işleme yöntemleriyle yönetilemeyecek kadar büyük, hızlı ve çeşitli veri kümelerini ifade eder.

2. Büyük Verinin geleneksel 3 Vs'si nelerdir? Her birini kısaca tanımlayın.

- **Hacim (Volume):** Büyük miktarda veri üretilmesi.
- **Hız (Velocity):** Verinin sürekli ve hızlı bir şekilde üretilip işlenmesi.
- **Çeşitlilik (Variety):** Yapısal, yarı yapılandırılmış ve yapılandırılmamış farklı veri türleri içermesi.

3. Google ve Amazon gibi şirketlerin neden Büyük Veri sorununu ele alan ilk şirketler arasında yer aldığını açıklayın.

Bu şirketler, büyük ölçekli veri işlemesi ve depolaması gerektiren arama motorları, e-ticaret ve bulut hizmetleri gibi sistemler geliştirdikleri için Büyük Veri çözümlerine öncülük etmiştir.

4. Ölçek büyütme ile ölçek genişletme arasındaki farkı açıklayın.

- **Ölçek büyütme (Scaling Up):** Daha güçlü donanım ekleyerek mevcut sunucuyu geliştirme.
- **Ölçek genişletme (Scaling Out):** Daha fazla sunucu ekleyerek yatay olarak büyütme.

5. Akış işleme nedir ve neden bazen gereklidir?

Verinin anlık olarak işlenmesi ve analiz edilmesidir. Gerçek zamanlı kararlar almak için gereklidir (örneğin, finansal işlemler ve IoT sistemleri).

6. Akış işlemenin geri bildirim döngüsü işlemeden farkı nedir?

Akış işleme verileri anında işlerken, geri bildirim döngüsü geçmiş verileri analiz edip yeni stratejiler geliştirmek için kullanılır.

7. Doğruluk, değer ve görselleştirmenin neden ilişkisel veritabanlarının yanı sıra Büyük Veri için geçerli olduğunu açıklayınız.

Büyük Veri'nin anlamlı hale gelmesi için verinin doğru, değerli ve anlaşılır bir şekilde görselleştirilmesi gerekir; bu durum ilişkisel veritabanları için de geçerlidir.

8. Çoklu kalıcılık nedir ve neden yeni bir yaklaşım olarak kabul edilir?

Verilerin farklı türde veri depolama sistemlerinde (NoSQL, ilişkisel vb.) aynı anda saklanmasıdır. Büyük Veri'nin gerektirdiği esneklik nedeniyle yeni bir yaklaşımdır.

9. Hadoop Dağıtılmış Dosya Sistemi yaklaşımı tarafından yapılan temel varsayımlar nelerdir?

- Büyük veri kümeleri işlenir.
- Veri daha çok yazılır, az değiştirilir.

- Ölçeklenebilirlik ve hata toleransı önceliklidir.

10. HDFS'de bir isim düğümü ile bir veri düğümü arasındaki fark nedir?

- **İsim Düğümü:** Dosya sisteminin meta verilerini yönetir.
- **Veri Düğümü:** Verinin kendisini saklar ve işler.

11. MapReduce işlemedeki temel adımları açıklayın.

1. **Map:** Veriyi bölüp işler.
2. **Shuffle & Sort:** Ara sonuçları gruplar.
3. **Reduce:** Gruplanmış verileri özetleyerek nihai sonucu üretir.

12. HDFS ve MapReduce'un birbirini nasıl kısaca açıklayınız.

HDFS, veriyi saklarken MapReduce, bu veriyi işlemek için kullanılır.

13. NoSQL veritabanlarının dört temel kategorisi nelerdir?

1. **Anahtar-Değer**
2. **Belge**
3. **Sütun Ailesi**
4. **Grafik**

14. Bir anahtar-değer veritabanı ile bir belge veritabanının değer bileşenleri nasıl farklıdır?

Anahtar-değer veritabanında değerler genellikle basit saklanırken, belge veritabanlarında değerler daha karmaşık ve yapılandırılmıştır (örneğin, JSON).

15. Satır merkezli ve sütun merkezli veri depolama arasındaki farkı kısaca açıklayınız.

- **Satır merkezli:** Veriler satır bazında saklanır, OLTP için uygundur.
- **Sütun merkezli:** Veriler sütun bazında saklanır, analiz iş yükleri için uygundur.

16. Sütun ailesi veritabanında bir sütun ile bir süper sütun arasındaki fark nedir?

- **Sütun:** Bir anahtar-değer çiftidir.
- **Süper Sütun:** Birden fazla sütunu içeren yapıdır.

17. Grafik veritabanlarının neden ölçek büyütme ile mücadele etme eğiliminde olduğunu açıklayın.

Bağlantılar yoğun olduğu için geleneksel yatay ölçeklendirme yöntemleri verimli çalışmaz.

18. Bir veritabanının toplu farkındalığa sahip olmasının ne anlama geldiğini açıklayın.

Veritabanının, işlemleri gerçekleştirdiği sırada diğer verilerin ve işlemlerin bağlamını anlamasıdır (örneğin, graf veritabanları ilişkileri değerlendirirken toplu farkındalık gerektirir).