
BLM433 Dağıtık Sistemlere Giriş

Aktivite 3

Dr. Öğr. Üyesi Süleyman Eken

BAŞLANGIÇ: 27 Aralık 2020 Pazar

BITİŞ : 10 Ocak 2021 Pazar

(50 p) Kısım 1 (Apache Spark Uygulaması Geliştirme)

1- Apache Zeppelin notebook kurulmalıdır (Jupyter Notebook'a benzemekle beraber farklılıkları vardır).

<http://zeppelin.apache.org/>

2- Apache Spark SQL ve MLlib kütüphanelerini kullanarak Zeppelin üzerinde iki farklı uygulama gerçekleyiniz. Aşağıdaki linklerde çeşitli örnekler mevcuttur.

3- Zeppelin notebook'unuzu ogrenci_numarasi.json formatında gönderiniz.

Çeşitli Zeppelin Notebookları için bakınız:

<https://github.com/hortonworks-gallery/zeppelin-notebooks>

<https://www.zepl.com/explore>

Faydalı videolar:

https://www.youtube.com/watch?v=pQbVH_aO5E

<https://www.youtube.com/watch?v=raHmOCAPWg0>

<https://www.youtube.com/watch?v=CfhYFqNyjGc>

<https://www.youtube.com/watch?v=J6Ei1RMG5Xo>

(50 p) Kısım 2 (Hadoop MapReduce ile Seyrek Matris Çarpımı)

Sayısal analiz ve bilimsel hesaplamada, seyrek bir matris veya seyrek dizi, öğelerin çoğunun sıfır olduğu bir matristir. Bir matrisin seyrek olarak değerlendirilmesi için kaç öğenin sıfır olması gerektiğine dair kesin bir tanım yoktur; ancak ortak bir kriter, sıfır olmayan öğelerin sayısının kabaca satır veya sütun sayısı kadar olmasıdır¹.

Bu bölümde Apache MapReduce programlama paradigması kullanılarak seyrek matris çarpımı yapan bir program geliştirilecektir. Sonuçları tek makine üzerinde ve fiziksel gerçek dağıtılmış küme üzerinde ayrıca veriniz. Seyrek matrislere şuradan erişebilirsiniz: <https://sparse.tamu.edu/> Bu matrislerden farklı büyüklükteki en az 3 matrisi kullanınız.

	Tek makine (sn) (pseudo-distributed)	Cluster (sn)
1. çalıştırma		
2. çalıştırma		
3. çalıştırma		
4. çalıştırma		
5. çalıştırma		
Ortalama		

Not: Projeler 2 kişilik olup aktivite 2'deki grup arkadaşınızla devam edeceksiniz.

¹https://en.wikipedia.org/wiki/Sparse_matrix