

Óbudai Egyetem Neumann János Informatikai Kar		Kiberfizikai Rendszerek Intézet		
Tantárgy neve és kódja: <i>Big data algoritmusok és programozás, NIEBD1GBNE</i> Kreditérték: 5				
<i>Mérnök Informatikus BSc szak</i>		<i>Nappali tagozat 2021/22 tanév I. félév</i>		
Tantárgy oktató(i): Nagy András, Dénes-Fazakas Lehel				
Előtanulmányi feltételek: (kóddal)		Korszerű adatbázisok, NIXKD1GBNE		
Heti óraszámok:	Előadás:2	Tantermi gyak.: 0	Laborgyakorlat: 2	Konzultáció: 0
Számonkérés módja:	Vizsga			
A tananyag				
Oktatási cél: A tárgy keretében a hallgatók megismerkednek a Big Data fogalomkörhöz kapcsolódó fogalmakkal, technológiákkal, paradigmákkal, komponensekkel, alkalmazási területekkel.				
Tematika: Bevezetés a Big Data világába, Hadoop alapok(HDFS, YARN), Big Data adattárolási módszerek(ORC, Arrow, Avro, Parquet), Adatbegyűjtési módszerek(Kafka), Adatfeldolgozás(MapReduce, Spark), Stream feldolgozás(Spark Stream), Feldolgozási folyamatok(Oozie vagy Airflow, Apache Nifi), Klaszter menedzsment és biztonság(Zookeeper, HCatalog, Ranger, LDAP / AD), Adathozzáférés(Impala, Hue, Tableau), Analitikai eszközök(Mahout, Spark Mlib), Felhő alapú megoldások(IaaS vs. PaaS Big Data használata Azure környezetben), Big Data esettanulmányok. Azure Databricks platform használata.				

Féléves ütemezés:	
Oktatási hét (konzultáció)	Témakör
1.	E: Bevezetés a Big Data világába. L: Hadoop telepítése, konfigurálása.
2.	E: Hadoop alapok(HDFS, YARN). L: HDFS.
3.	E: Big Data adattárolási módszerek(ORC, Arrow, Avro, Parquet). L: MapReduce.
4.	E: Adatbegyűjtési módszerek(Scoop, Kafka). L: YARN.
5.	E: Adatfeldolgozás(MapReduce, HIVE, PIG, Spark). L: Kafka.
6.	E: Stream feldolgozás(Spark Stream, Storm). L: Spark I.
7.	E: Feldolgozási folyamatok(Oozie vagy Airflow, Apache Nifi). L: Spark II.
8.	E: Klaszter menedzsment és biztonság(Zookeeper, HCatalog, Ranger, LDAP / AD). L: Azure Databricks.
9.	E: Adathozzáférés(Impala, Hue, Tableau). L: Azure Databricks.
10.	E: Analitikai eszközök(Mahout, Spark Mlib). L: Azure Databricks.
11.	E: Felhő alapú megoldások(IaaS vs. PaaS Big Data használata Azure környezetben). L: Azure Databricks.
12.	E: Big Data esettanulmányok. L: Konzultáció
13.	Elméleti ZH
Félévközi követelmények	

Az előadáson a részvétel kötelező. Nem kap aláírást az a hallgató, aki az órák 30%-ánál többet hiányzott (TVSZ-nek megfelelően).
Az aláírás megszerzésének feltétele, hogy az Elearning oldalra feltöltse az értékelhető beadandó feladatokat, és a félév végi tesztet legalább 50%-os eredménnyel írja meg.

A félévzáró érdemjegy (é) kialakításának módszere

A tantárgy teljesítését leíró érdemjegyet a vizsgán szerzett pontszám (max. 50 pont) és a labor feladataiból szerzett pontszám (max. 50 pont) összege határozza meg. Mindkettőből minimum 51%-ot kell elérni a tantárgy sikeres teljesítéséhez. A labor gyakorlatból elérhető 50 pont a következő 3 tétel összegéből áll össze:

1. Azure Databricks gyakorlati feladat megoldása (max. 20 pont)
2. Félévzáró labor ZH (teszt) megoldása (max. 20 pont)
3. 4 db kisebb beadandó feladat (max. 10 pont)

Az elégséges jegyhez 51, a közepeshez 62, a jóhoz 74, a jeleshez 85 pontot kell elérni.

Pótlás módja

Vizsga módja

Írásbeli vagy online szóbeli vizsga

Vizsgajegy kialakítása

Az elégséges jegyhez 51, a közepeshez 63, a jóhoz 74, a jeleshez 85 pontot kell elérni.

Irodalom

Kötelező:

A Moodle rendszerben közzétett jegyzet és kiegészítő tananyagok

Ajánlott:

Alex Holmes: Hadoop In Practice 2nd Edition

Egyéb segédletek:

-