

Óbudai Egyetem Neumann János Informatikai Kar		Kiberfizikai Rendszerek Intézet		
Tantárgy neve és kódja: Big data algoritmusok és programozás, NIEBD1GBNE Kreditérték: 5				
Mérnök Informatikus BSc szak		Nappali tagozat 2020/21 tanév II. félév		
Tantárgy oktató(i): Fleiner Rita, Nagy András				
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(Scoop, Kafka), Adatfeldolgozás(MapReduce, HIVE, PIG, Spark), Stream feldolgozás(Spark Stream, Storm), 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 keretrendszer általános ismertetése.
2.	E: Hadoop alapok(HDFS, YARN). L: HDFS fájl rendszer.
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: Fájl formátumok.
6.	E: Stream feldolgozás(Spark Stream, Storm). L: Impala. Azure Databricks.
7.	E: Feldolgozási folyamatok(Oozie vagy Airflow, Apache Nifi). L: Hue. Azure Databricks.
8.	E: Klaszter menedzsment és biztonság(Zookeeper, HCatalog, Ranger, LDAP / AD). L: Hive. Azure Databricks.
9.	E: Adathozzáférés(Impala, Hue, Tableau). L: Sqoop. Azure Databricks.
10.	E: Analitikai eszközök(Mahout, Spark Mlib). L: Flume.
11.	E: Felhő alapú megoldások(IaaS vs. PaaS Big Data használata Azure környezetben). L: Spark.
12.	E: Big Data esettanulmányok. L: HBase.
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, a 12 darab gyakorlati feladat feltöltése a moodle rendszeren keresztül, illetve az 51%-os teljesítés a Databricks feladat és a félévzáró moodle teszt során.	

Zárthelyi dolgozatok	
Oktatási hét (konzultáció)	Témakör
14.	Labor ZH
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. A labor gyakorlatból elérhető 50 pont a következő 2 tétel összegéből áll össze: 1. Azure Databricks gyakorlati feladat megoldása (max. 24 pont) 2. Félévzáró labor ZH (teszt) megoldása (max. 26 pont) Mindhárom megmérettetésen (vizsga, labor ZH, Databricks gyakorlati feladat) minimum 51%-ot kell elérni a tantárgy sikeres teljesítéséhez. Az elégséges jegyhez 51, a közepeshez 62, a jóhoz 74, a jeleshez 85 pontot kell elérni.	
Pótlás módja	
A labor zárthelyi az utolsó héten vagy a vizsgaidőszakban pótolható vagy javítható.	
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:	
-	