

Назва. Високопродуктивні системи обробки великих даних

Тип. Вибіркова

Рік навчання. 2019–2020.

Семестр. I.

Кількість кредитів ЄКТС. 5.

ШБ лектора, науковий ступінь, посада. Мінухін С.В., докт. техн. н., проф.

Результати навчання.

здатність визначення великих даних та їх використання для побудови розподілених інформаційних систем;

здатність використовувати технології та особливості обробки великих даних у розподілених системах;

здатність обирати та обґрунтовувати тип програмного забезпечення та відповідних інструментів для пакетної та потокової обробки даних;

здатність розв'язувати проблеми масштабованості, проектування та експлуатації систем розподілених сховищ даних;

володіти інструментами аналізу великих даних з використанням мови запитів та роботи з різними типами організації даних.

Обов'язкові попередні навчальні дисципліни. "Бази даних", "Комп'ютерні мережі", "Програмування", Операційні системи.

Зміст.

Розподілені файлові системи. Модель обчислень MapReduce. Архітектура Apache Hadoop. Потокова та пакетна обробка даних. Принципи та архітектура функціонування Apache Spark. Архітектури систем обробки великих даних. Принципи побудови архітектури системи Apache Storm. BigTable-подібні сховища даних. Технологія Apache HBase.

Рекомендовані джерела.

1. Book on “Hadoop: The Definitive Guide” by Tom White by O’Reilly Publication, 2010.
2. M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, “Spark: Cluster computing with working sets,” ser. HotCloud’10.
3. M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi, et al. Spark sql: Relational data processing in spark. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1383–1394. ACM, 2015.
4. V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth, B. Saha, C. Curino, O. O’Malley, S. Radia, B. Reed, and E. Baldeschwieler. Apache hadoop yarn: Yet another resource negotia- tor. In *Proceedings of the 4th Annual Symposium on Cloud Computing, SOCC ’13*, pages 5:1–5:16, New York, NY, USA, 2013. ACM.
5. M. Zaharia, M. Chowd4hury, M. J. Franklin, S. Shenker, and I. Stoica. Spark: Cluster computing with working sets. In *Proceedings of the 2Nd*

USENIX Conference on Hot Topics in Cloud Computing, HotCloud'10, pages 10–10, Berkeley, CA, USA, 2010. USENIX Association.

6. A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony, H. Liu, P. Wyck-off, and R. Murthy. Hive: a warehousing solution over a mapreduce framework. Proceedings of the VLDB Endowment, 2(2):1626–1629, 2009.

7. Hadoop. <http://hadoop.apache.org/>.

8. Spark <http://spark.apache.org/>

9. N. J. Gunther, P. Puglia, and K. Tomasette, “Hadoop superlinear scalability,” Commun. ACM, vol. 58, no. 4, pp. 46–55, 2015.

10. Spark QuickStart Guide

11.

Методи навчання.

Лекції та лабораторні заняття з використанням інформаційних технологій.

Методи оцінювання:

- поточний контроль (опитування);
- підсумкова контрольна робота
- підсумковий контроль (іспит).

Мова навчання. Українська.