

РЕФЕРАТ

Дисертацію виконано на 73 аркушах, вона містить 2 додатки та перелік посилань на використані джерела з 15 найменувань. У роботі наведено 13 рисунки та 5 таблицю.

Актуальність теми. Сьогодні люди значною мірою залежать від комп'ютерів. Саме технічні пристрої допомагають в обрахунках, обробці та зберіганні важливих даних. Так як дані постійно доповнюються та зростають в об'ємі, важливо щоб комп'ютерні програми виконували обробку даних як можна швидше. Для зберігання та обробки великої кількості даних популярність отримала Big Data. Але важливим є дослідження та виявлення саме найшвидшого, найзручнішого компоненту платформи та компоненту, що дасть змогу мінімізувати витрати. Важливо дослідити метод мінімізації часу обробки запиту на платформі Hadoop. Даний метод дає змогу дослідити поведінку вхідних даних великого об'єму та саму швидкість обробки даних на платформі. Також наразі Hadoop не має зручного інтерфейсу для користувача.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно з планом науково-дослідних робіт кафедри прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Мета і задачі дослідження. Метою дисертаційної роботи є розробка методу мінімізації часу обробки запиту на платформі HADOOP.

Для досягнення вказаної мети було розв'язано такі задачі:

- систематизувати існуючі методи для мінімізації часу обробки даних;
- розробити модель для мінімізації часу обробки;
- розробити методи розв'язання задачі мінімізації часу обробки;
- провести експериментальні дослідження з використанням отриманих даних.

Об'єктом дослідження є платформа Hadoop для обробки Big Data; виконання запитів у об'єктах екосистеми Hadoop, такі як Cloudera Impala, Apache Drill, Apache

Hive, Apache Tez, Apache Spark, Cloudera Hue; різні типи зтиснення, такі як parquet, snappy, gzip, bzip2.

Предметом дослідження є метод мінімізації часу обробки запиту на платформі Hadoop для прискорення результату дії запиту з використанням Big Data.

Методи дослідження. Для розв'язання поставленої задачі використовувалися такі методи: методи мінімізації часу (для розроблення моделі обробки); методи прогнозування (для розроблення методів розв'язання задачі прогнозу швидкості роботи компонентів); методи теорії алгоритмів та програмування (для програмної реалізації розроблених алгоритмів).

Наукова новизна одержаних результатів складається з таких положень:

- уперше поставлено задачу мінімізації часу обробки даних на платформі Hadoop, яка, на відміну від існуючих, передбачає опираючись на прогноз обрати найшвидший компонент для обробки потрібних даних;
- удосконалено методи прогнозування, які, на відміну від існуючих, враховують нечіткість даних, що дає змогу забезпечувати кращий та точніший прогноз;
- удосконалено інтерфейс користувача платформи, який відрізняються від існуючого тим, що має інтуїтивне управління.

Практичне значення одержаних результатів. Запропоновано методи, які може бути використано під час мінімізації часу обробки даних. Розроблені методи, математичне й програмне забезпечення для мінімізації часу обробки даних спрощують роботу програмного забезпечення та час роботи, сприяють забезпеченню швидких вирішень задач.

Апробація результатів дисертації. Основні положення і результати роботи були представлені та опубліковані :

- Кулик Р. Ю. Дослідження та порівняльний аналіз компонентів платформи Hadoop Р. Ю. Кулик // «Современные тенденции развития науки» (г. Харьков, 26-27 мая 2017 г.). — Херсон : Издательский дом "Гельветика", 2017.

– Кулик Р. Ю. Методологічні аспекти прогнозування часу обробки запиту на платформі Hadoop Р. Ю. Кулик // «Современные тенденции развития науки» (г. Харьков, 26-27 мая 2017 г.). — Херсон : Издательский дом "Гельветика", 2017.

Ключові слова: Big Data, Hadoop, база даних, Cloudera Impala, Apache Drill, Apache Hive, Apache Tez, Apache Spark, Cloudera Hue, прогнозування, регресія.