**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**ТАВРІЙСЬКИЙ ДЕРЖАВНИЙ АГРОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ ДМИТРА МОТОРНОГО**

**Факультет економіки та бізнесу**

**Кафедра маркетингу**

|  |  |
| --- | --- |

**УПРАВЛІННЯ ВЕЛИКИМИ ДАНИМИ**

**BIG DATA MANAGEMENT**

методичні вказівки до самостійної роботи

здобувачів ступеня вищої освіти «Магістр»

зі спеціальності ***075 «Маркетинг»***

**Мелітополь, 2022р.**

**Управління великими даними (Big Data Management).** Методичні вказівки до самостійної роботи здобувачів ступеня вищої освіти «Магістр» за спеціальністю 075 «Маркетинг». – Мелітополь, ТДАТУ, 2022 р. – 24 с.

**Розробники:**

*Кюрчев В.М.,*д.т.н., професор кафедри експлуатації та технічного сервісу машин, ТДАТУ

*Куліш Т.В.,* д.е.н., доцент кафедри маркетингу ТДАТУ;

*Шевчук С.В.,* к.е.н., доцент кафедри обліку і економічного аналізу, Національний університет кораблебудування імені адмірала Макарова

**Рецензенти:**

*Марина Тропманн-Фрік* -д.т.н., проф., Університет прикладних наук Гамбургу (Німеччина)

*Мироненко Д.С.* к.т.н., доцент кафедри інформатики ПДТУ

Розглянуто та рекомендовано до друку на засіданні кафедри маркетингу

Протокол № *10* від « *09*» травня 2022 року

Завідувач кафедри «Маркетинг»

д.е.н, професор \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Дар’я ЛЕГЕЗА

Схвалено методичною комісією факультету економіки та бізнесу для здобувачів освітнього рівня «Магістр» за спеціальності 075 «Маркетинг»

Протокол № 11 від *«16» червня* 2022 року

Голова, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_ Анна КОСТЯКОВА

Схвалено Навчально-методичною радою Таврійського державного агротехнологічного університету імені Дмитра Моторного для здобувачів освітнього рівня «Магістр» за спеціальності 075 «Маркетинг»

Протокол № *1* від «*26*»  *серпня* 2022 року

Голова, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_ Олександр ЛОМЕЙКО

©Шевчук С.В., 2022 рік

**ЗМІСТ**

ВСТУП

Тема 1. Основні положення концепції Big Data

Тема 2. Типи даних та їх вимірювання

Тема 3. Управління даними. База даних та СУБД

Тема 4. Дизайн бази даних. Розподілена система баз даних

Тема 5. Основні технології та інструменти роботи з великими даними

Тема 6. Особливості Apache Hadoop. Компонента MapReduce

**ВСТУП**

Курс «Управління великими даними» спрямований на формування у студентів компетентностей з питань отримання, обробки, аналізу та оцінювання інформації в різних суспільних сферах на основі управління великими даними та використання отриманих результатів для обґрунтування управлінських рішень. Мета курсу — ознайомлення з концепціями управління великими даними, інструментами організації управління даними, отримання доступ до баз даних, механізмами обробки інформаційних потоків.

Завдання курсу:

* ознайомлення із основними положеннями концепції Big Data;
* формування розуміння напрямів застосування великих даних в економіці, промисловості, суспільному житті, у науці та дослідженнях;
* засвоєння загальних методів обробки масиву даних;
* ознайомлення із напрямами управління даними;
* вивчення можливостей створення, збереження, оновлення та пошуку інформації в [базах даних](https://uk.wikipedia.org/wiki/%D0%91%D0%B0%D0%B7%D0%B0_%D0%B4%D0%B0%D0%BD%D0%B8%D1%85) з контролем доступу до них;
* формування розуміння щодо дизайну бази даних;
* здобуття знань про можливості застосування основних технологій та інструментів роботи з великими даними, структуру даних та хмарні технології;
* з’ясування значення інфраструктури збереження, обробки великих обсягів даних та використання простору даних для моделювання Big Data.

Мета самостійної роботи:

* поглиблення, систематизація та закріплення набутих теоретичних знань за допомогою самостійного опрацювання здобувачами вищої освіти навчальної та спеціальної наукової літератури, а також більш глибока розробка одного з конкретних питань;
* вміння узагальнювати теоретичні матеріали, обробляти і аналізувати зібрані дані;
* розвинення навички формулювання власних висновків, розробляти пропозиції з проблем, що досліджуються;
* напрацювання навичок застосування інструментарію аналізу даних, а також вміння аналізувати результати конкретних ситуацій поведінки виробників на ринку.

Самостійна робота – це важлива форма навчання, спосіб розвитку мислення, її виконання сприяє підвищенню вміння здобувача вищої освіти:

* самостійно формулювати проблему дослідження;
* здійснювати пошук необхідної наукової інформації;
* аналізувати наукову літературу та вивчати законодавчі матеріали з досліджуваних питань;
* досліджувати поведінку та механізм прийняття рішень окремими економічними суб’єктами у різних ринкових ситуаціях;
* моделювати процес погодження господарських цілей сукупності мікроекономічних суб’єктів за допомогою ринкового механізму;
* оформляти науково-дослідний матеріал;
* захищати підготовлений матеріал (робити наукові доповіді, відповідати на запитання, захищати свою точку зору).

У методичних рекомендація наведені питання, що підлягають опрацюванню, завдання для самоперевірки знань за кожною темою, що допоможе зрозуміти необхідність та практичні напрямами використання методів управління великими даними.

**ЗАВДАННЯ ДЛЯ САМОСТІЙНОГО ОПРАЦЮВАННЯ**

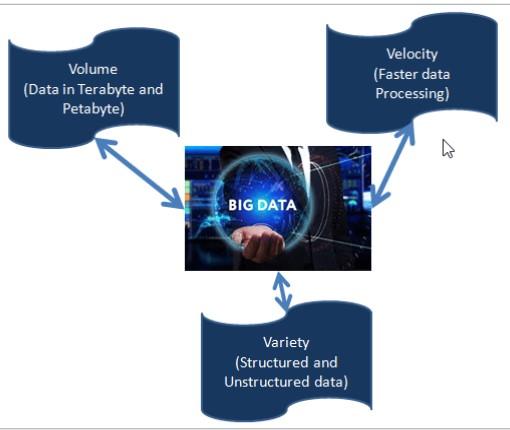
Завдання на самостійну роботу підготовка реферату з обраною тематикою та проробка тестових завдань.

**ТЕМАТИКА ЗА НАПРЯМАМИ**

**1. Основні характеристики Big Data**

У нас є три основні характеристики Big Data, і якщо будь-які дані задовольняють цим характеристикам, вони будуть розглядатися як Big Data. Це – комбінація трьох V, згаданих нижче:

* Гучність
* Швидкість
* Різноманітність



Гучність: Дані мають бути величезного обсягу. Big Data має рішення для підтримки великого обсягу даних, що знаходяться в терабайті або петабайті. Ми можемо виконувати операції CRUD (створення, читання, оновлення та видалення) на Big Data легко та ефективно.

Швидкість: Він відповідає за швидший доступ до даних. Сьогодні соціальні медіа потребують швидкого обміну даними протягом частки часу, і Big Data є найкращим рішенням для цього. Отже, швидкість - це ще одна характеристика, і це швидкість обробки даних.

Різноманітність: У соціальних мережах ми маємо справу з неструктурованими даними, такими як аудіо- чи відеозаписи, зображення тощо. Крім того, різні сектори, такі як банківський домен, потребують структурованих та напівструктурованих даних. Big Data – це рішення для підтримки обох типів даних в одному місці. Різноманітність означає різні типи даних, такі як структуровані / неструктуровані дані, що надходять з різних джерел.

**2. Проблеми традиційних баз даних**

Традиційна база даних не підтримує різноманітні дані, тобто вона не може обробляти неструктуровані та напівструктуровані дані. Традиційна база даних працює повільно, маючи справу з великим обсягом даних. У традиційних базах даних обробка або аналіз великого обсягу даних дуже важкий. Традиційна база даних здатна зберігати дані у терабайтах або петабайтах. Традиційна база даних не може обробляти історичні дані та звіти.

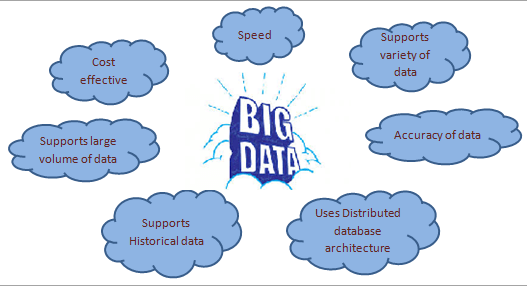
Через певний проміжок часу необхідне очищення даних від бази даних. Витрати на підтримку великого обсягу даних дуже високі за традиційної бази даних. Точність даних менша в традиційній базі даних, оскільки в ній не зберігаються повні історичні дані.

**3. Переваги Big Data порівняно з традиційною базою даних**

Великі дані відповідають за обробку, управління та обробку різних типів даних, таких як структуровані, напівструктуровані та неструктуровані.

Це рентабельно з точки зору збереження великого обсягу даних. Він працює на системі розподілених баз даних. Можемо довго зберігати великі обсяги даних, використовуючи методи Big Data. Тож легко обробляти історичні дані та створювати точні звіти. Швидкість обробки даних дуже висока, і тому соціальні медіа використовують методи великих даних.

Точність даних - велика перевага Big Data. Це дозволяє користувачам приймати ефективні рішення для свого бізнесу на основі поточних та історичних даних.



Обробка помилок, контроль версій та взаємодія з клієнтами дуже ефективні в Big Data.

**4. Виклики та ризики в Big Data**

Однією з основних проблем у роботі з великими даними є управління великими обсягами даних. На сьогоднішній день дані надходять у систему з різних джерел з різноманітністю. Тож для компаній є дуже великим випробуванням - правильно ним керувати. Наприклад, щоб сформувати звіт, що містить дані за останні 20 років, потрібно зберегти та підтримувати дані системи за останні 20 років. Для того, щоб надати точний звіт, необхідно вносити в систему лише відповідні дані. Він не повинен містити нерелевантних або непотрібних даних, інакше підтримка такої кількості даних буде великим викликом для компаній.

Ще однією проблемою цієї технології є синхронізація різних типів даних. Як ми всі знаємо, великі дані підтримують структуровані, неструктуровані та напівструктуровані дані, що надходять з різних джерел, синхронізувати їх та отримати узгодженість даних дуже складно.

Наступним викликом, з яким стикаються компанії, є розрив експертів, які можуть допомогти та реалізувати проблеми, з якими вони стикаються в системі. У цій галузі є великий розрив у талантах.

Обробка аспекту дотримання вимог є дорогою. Збір даних, агрегування, зберігання, аналіз та звітування Big Data має величезні витрати. Організація повинна мати можливість управляти всіма цими витратами.

Ризики: Він може обробляти різноманітні дані, але якщо компанії не можуть належним чином зрозуміти вимоги та контролювати джерело даних, це дасть помилкові результати. Як результат, для дослідження та виправлення результатів знадобиться багато часу та грошей.

Захист даних - ще один ризик для Big Data. При великому обсязі даних є більше шансів, що хтось їх вкраде. Хакери даних можуть викрадати та продавати важливу інформацію (включаючи історичні дані) компанії.

Крім того, конфіденційність даних - ще один ризик для Big Data. Якщо ми хочемо захистити особисті та конфіденційні дані від хакерів, вони повинні бути захищені та повинні відповідати всій політиці конфіденційності.

Технології великих даних:

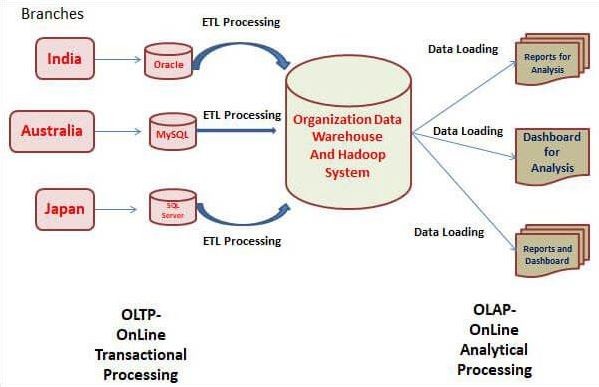
* Apache Hadoop
* Microsoft HDInsight
* NoSQL
* Sqoop
* Big Data в Excel

**5. Склад великих даних та даних**

Data Warehouse – це основна концепція, яку ми повинні зрозуміти перед обговоренням тестування Hadoop або Big Data. Сховище даних на прикладі в реальному часі. Наприклад , є компанія, яка створила свої філії в трьох різних країнах, припустимо філію в Індії, Австралії та Японії. У кожному відділенні всі дані про клієнтів зберігаються в Локальній базі даних. Ці локальні бази даних можуть бути звичайними класичними СУБД, такими як Oracle, MySQL чи SQL Server тощо, і всі дані клієнтів будуть зберігатися в них щодня.

Зараз щокварталу, півроку чи щороку організація хоче аналізувати ці дані для розвитку бізнесу. Щоб зробити те саме, організація буде збирати всі ці дані з кількох джерел, а потім складати їх в одному місці, і це місце викликається **«**Інформаційне сховище».

Сховище даних – це різновид бази даних, яка містить усі дані, отримані з декількох джерел або декількох типів баз даних через **“ETL”** (що є **Є** xtract, **Т** ransform і **L** oad) процес. Коли дані будуть готові в сховищі даних, ми зможемо використовувати їх для аналітичних цілей. Тож для аналізу ми можемо генерувати звіти з даних, доступних у сховищі даних. За допомогою інструментів бізнес-аналітики можна створити кілька діаграм та звітів. Для аналітичних цілей нам потрібен Data Warehouse для розвитку бізнесу та прийняття відповідних рішень для організацій.



У цьому процесі відбуваються три речі. По-перше, зібрано дані з кількох джерел і розмістили їх в одному місці, яке є сховищем даних. Тут використовуємо процес “ETL”, таким чином, завантажуючи дані з декількох джерел в одне місце, ми застосовуватимемо їх у коренях трансформації, а потім тут зможемо використовувати різні види інструментів ETL.

Після того, як дані будуть готові до сховища даних, можемо створювати різні звіти для аналізу ділових даних за допомогою інструментів бізнес-аналітики (BI) або ми також називаємо їх інструментами звітування. Такі інструменти, як Tableau або Cognos, можуть бути використані для створення звітів та панелей інструментів для аналізу даних для бізнесу.

**6. Що таке OLTP і що таке OLAP**

Викликаються бази даних, які ведуться локально і використовуються для транзакційних цілей OLTP, тобто обробка онлайн-транзакцій. Щоденні транзакції будуть зберігатися тут і негайно оновлюватися, і тому ми назвали їх системою OLTP.

Тут ми використовуємо традиційні бази даних, у нас є кілька таблиць і є взаємозв'язки, отже, все систематично планується відповідно до бази даних. Ми не використовуємо ці дані в аналітичних цілях. Тут ми можемо використовувати класичні бази даних RDMBS, такі як Oracle, MySQL, SQL Server тощо. Коли ми приходимо до частини Data Warehouse, ми використовуємо Teradata або Hadoop Systems, які також є своєрідною базою даних, але дані в DataWarehouse зазвичай використовуються для аналітичних цілей і називаються OLAP або Інтернет-аналітична обробка.

Тут дані можна оновлювати щокварталу, півроку чи щороку. Іноді дані оновлюються також “Offerly”, де “Offerly” означає, що дані оновлюються та отримуються для аналізу відповідно до вимог замовника.

Крім того, дані для аналізу не оновлюються щодня, оскільки ми отримуватимемо дані з кількох джерел, за розкладом, і ми можемо виконувати це завдання ETL. Ось як працює Інтернет-аналітична система обробки. Тут знову ж таки, BI Tools або Reporting Tools можуть генерувати звіти, а також інформаційні панелі, і на основі цього ділові люди прийматимуть рішення щодо вдосконалення свого бізнесу.

**7. Тестування Big Data та Hadoop**

Після зберігання та обробки розробником дані надходять на формування звіту. До цього нам потрібно перевірити оброблювані дані на точність і перевірити, чи дані точно завантажені та оброблені правильно чи ні. Отже, програму та / або сценарії, створені розробником, потрібно перевірити Hadoop або Big Data Tester. Тестувальник повинен знати базове програмування, таке як Mapper, Hive, Pig Scripts тощо, щоб перевірити сценарії та виконати команди.

Отже, перед тестуванням тестувальники повинні знати, які всі програми та сценарії працюють, як написати код, а потім подумати, як їх протестувати. Тестування може проводитися як вручну, так і за допомогою засобів автоматизації.

Hadoop має різні види тестування, такі як модульне тестування, тестування регресії, тестування системи та тестування продуктивності тощо. Отже, це загальні типи тестування, які ми використовуємо в нашому звичайному тестуванні, а також тестування Hadoop та Big Data. У нас є однакові терміни тестування, такі як стратегія тестування, сценарії тестування, тестові кейси тощо у тестуванні Hadoop та Big Data. Тільки середовище відрізняється, і існують різні види методів, які ми використовуємо для тестування системи Big Data та Hadoop, оскільки тут нам потрібно тестувати дані, а не додаток.

Тестування Big Data означає перевірку та перевірку даних при зберіганні та обробці їх у сховищі даних. Під час тестування Big Data нам потрібно перевірити обсяг та різноманітність даних, вилучених з різних баз даних та завантажених, а також оброблених у сховищі даних або системі Hadoop, це тестування підлягає функціональному тестуванню. Нам потрібно протестувати швидкість даних, завантажених з різних баз даних і завантажених до системи Hadoop, яка є частиною тестування продуктивності. Отже, як план або стратегія, нам потрібно зосередитись на функціональному, а також на тестуванні продуктивності тестування Big Data.

У тестуванні Big Data тестувальник повинен перевірити обробку величезної кількості даних за допомогою товарного обладнання та відносних компонентів. Отже, якість даних також відіграє важливу роль у тестуванні Big Data. Важливо перевірити та перевірити якість даних.

**8. Типи тестування для тестування Big Data**

Функціональне тестування та тестування продуктивності відіграють життєво важливу роль у тестуванні Big Data, окрім тестування Big Data, нам потрібно провести ще кілька видів тестування, таких як тестування баз даних, а також архітектурне тестування.

Ці типи тестування також настільки важливі, як тестування функціональності та продуктивності.

1) Архітектурне тестування

Це тестування проводиться для того, щоб переконатися, що обробка даних правильна та відповідає вимогам. Насправді, система Hadoop обробляє величезні обсяги даних і є всебічною.

Якщо архітектура неправильна, це може погіршити продуктивність, через що обробка даних може перерватись і може статися втрата даних.

2) Тестування бази даних

Тут перевірка процесу вкладається в картину, і нам потрібно перевірити дані з різних баз даних, тобто нам потрібно переконатися, що дані, отримані з вихідних баз даних або локальних баз даних, повинні бути правильними та належними. Крім того, нам потрібно перевірити, щоб дані, наявні у вихідних базах даних, відповідали даним, що вводяться в систему Hadoop.

 3) Тестування продуктивності

В рамках тестування продуктивності нам потрібно перевірити швидкість завантаження та обробки даних, тобто як IOPS (Input Output Per Second).

Потрібно перевірити швидкість введення даних або даних як вхідних даних з різних баз даних до сховища даних або системи Hadoop та від системи Hadoop або сховища даних до середовища клієнта.

Потрібно також перевірити швидкість даних, що надходять з різних баз даних та зі Складу даних, як вихідні дані. Це те, що ми називаємо вхідним виходом на секунду або IOPS.

Окрім цього, ще одним аспектом є перевірка ефективності поглинання та розподілу даних, а також наскільки швидко дані споживаються сховищем даних з різних баз даних та системою клієнта з системи Hadoop.

Також, як тестувальник, нам потрібно перевірити ефективність розподілу даних, наприклад, наскільки швидко дані розподіляються між різними файлами, доступними в системі Hadoop або в сховищі даних. Подібним чином той самий процес відбувається під час розподілу даних до клієнтських систем.

Система Hadoop або сховище даних складається з декількох компонентів, тому тестер повинен перевірити ефективність усіх цих компонентів, таких як Завдання MapReduce, вставка та споживання даних, час відгуку запитів та їх продуктивність, а також продуктивність пошуку операцій. Все це включено до тестування продуктивності.

4) Функціональне тестування

Функціональне тестування містить тестування всіх підкомпонентів, програм та скриптів, інструментів, що використовуються для виконання операцій зберігання або завантаження та обробки тощо.

Для тестера це чотири важливі типи та етапи, через які дані потрібно відфільтрувати, щоб клієнт отримав ідеальні дані без помилок.

Існують різні інструменти, які використовуються для тестування Big Data:

* Файлова система розподілу HDFS Hadoop для зберігання Big Data.
* Зниження карти HDFS для обробки Big Data.
* Для NoSQL або HQL Cassandra DB, ZooKeeper та HBase тощо.
* Хмарні серверні інструменти, такі як EC2.

**9. Ролі та обов'язки тестування Hadoop**

Як тестувальник Hadoop, ми несемо відповідальність за розуміння вимог, підготовку кошторисів випробувань, планування тестових кейсів, отримання деяких даних тестування для тестування деяких тестових кейсів, участь у створенні випробувального стенду, виконанні планів випробувань, звітуванні та повторному тестуванні дефектів. Крім того, ми повинні відповідати за щоденне звітування про стан та проходження тестів.

Перше, що ми збираємося обговорити, це Тестова стратегія. Після того, як ми запропонуємо рішення нашої проблеми, нам потрібно продовжити і спланувати або розробити стратегію нашого плану випробувань, ми можемо обговорити стратегію автоматизації, яку ми можемо використовувати там, план про графік випробувань, який залежить від наших термінів постачання, також ми може обговорювати планування ресурсів.

Стратегія автоматизації допоможе нам зменшити ручні зусилля, необхідні для тестування продукту. Графік випробувань важливий, оскільки він забезпечить своєчасну доставку товару.

Планування ресурсів буде мати вирішальне значення, оскільки нам потрібно спланувати, скільки людських годин нам потрібно на нашому тестуванні та скільки ресурсів Hadoop потрібно для виконання нашого планування тестів.

Після того, як ми розробимо стратегію тестування, нам потрібно продовжувати створювати плани розробки тестів, які включають створення планів тестування, створення тестових сценаріїв, які допоможуть нам автоматизувати тестування, а також визначити деякі дані тестування, які будуть використовуватися в планах тестування. і допомагає нам виконувати ці тестові плани. Коли закінчимо розробку тестів, що включає створення тестових планів, тестових сценаріїв та тестових даних, ми продовжуємо і починаємо виконувати ці тестові плани. Коли ми виконуємо тестові плани, можуть бути певні сценарії, коли фактичний результат не відповідає очікуваному, і ці речі називаються дефектами. Всякий раз, коли є дефект, нам також потрібно перевірити ці дефекти, і нам потрібно створити і підтримувати матриці для них. Всі ці речі підпадають під наступну категорію, яка є Управління дефектами.

**10. Що таке управління дефектами?**

Управління дефектами складається з відстеження помилок, виправлення помилок та перевірки помилок. Щоразу, коли план тестування виконується щодо будь-якого з наявних у нас продуктів, і як тільки виявляється конкретна помилка або виявляється дефект, про цей дефект потрібно повідомляти розробнику або призначати розробнику.

Тож розробник може розглянути це і почати над ним працювати. Як тестувальник, нам потрібно відстежувати хід помилки та відстежувати, чи помилку виправлено. Якщо помилка була виправлена, як повідомляється, тоді нам потрібно повторно протестувати її та перевірити, чи вона вирішена.

Як тільки всі помилки будуть виправлені, закриті та перевірені, нам потрібно продовжувати і пропонувати випробуваний продукт OKAY. Але перед тим, як ми доставимо товар, ми повинні переконатися, що UAT (Тест прийняття користувача) успішно завершено.

Ми переконуємось, що тестування встановлення та перевірка вимог зроблені належним чином, тобто продукт, який доставляється клієнту або кінцевому користувачеві, відповідає вимогам, зазначеним у Документі вимоги до програмного забезпечення.

Крок, який ми обговорили, ґрунтується на уяві, будь-який із сценаріїв тестування або будь-який з підходів до тестування, які ми будемо використовувати для цих кроків, або вимовляємо ці фрази для тестування нашого продукту та отримання кінцевого результату, що є OKAY Випробуваний продукт.

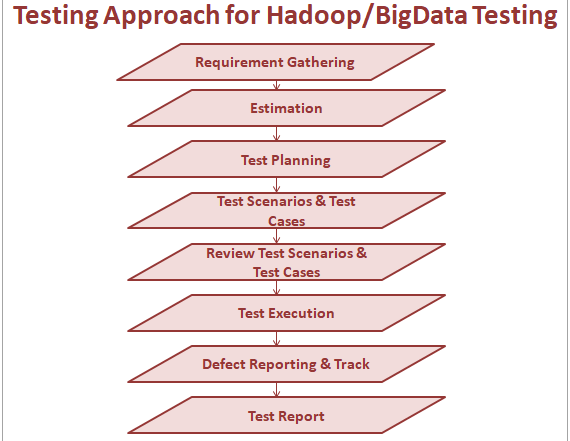
Ми можемо використовувати Sqoop або Flume або будь-який конкретний продукт для отримання Даних, записів або чого-небудь іншого як своїх джерел даних. Ми можемо використовувати ці інструменти для отримання даних із джерел даних до мого проміжного каталогу, який називається першим етапом нашого процесу.

**11. Підходи до тестування для тестування Hadoop / тестування Big Data**

Підхід до тестування є загальним для всіх видів тестування не тільки тому, що це тестування Big Data або Hadoop, коли ми переходимо до звичайного ручного тестування або тестування автоматизації чи тестування безпеки, тестування продуктивності, тому будь-яке тестування дотримується того самого підходу.

Як частина підходу до тестування, нам слід почати з «Вимоги», Вимоги – це основна річ, в основному містить моделі даних, цілі, джерела, а також які трансформації нам потрібно застосовувати, які інструменти ми повинні використовувати? Усі ці деталі будуть доступні в розділі «Вимоги». В основному це вимоги клієнта або вимоги клієнта. Виходячи з цієї вимоги, ми розпочнемо процес тестування.

Ще однією частиною Підходу є «Оцінка», Скільки часу нам потрібно витратити, щоб виконати всю діяльність як частину тестування. Ми робимо планування випробувань, готуємо сценарії випробувань, готуємо кейси для тестування та виконуємо їх, а також знаходимо дефекти та повідомляємо про них, а також готуємо звіти про випробування. Всі ці заходи займуть певний час, тому скільки часу нам потрібно для завершення всіх цих дій, і це в основному називається оцінкою. Нам потрібно дати грубу оцінку керівництву.



«Планування тестів»– це не що інше, як опис процесів, що тестувати, що не тестувати, який обсяг тестування, які графіки, скільки потрібно ресурсів, вимоги до апаратного та програмного забезпечення та які терміни, а також цикли тестування буде використовуватися, які рівні тестування ми вимагали тощо. Під час планування тестування вони виконуватимуть певний розподіл ресурсів для проекту та які різні моделі ми маємо, скільки ресурсів потрібно і які набори навичок потрібні тощо. Усі ці речі та аспекти будуть включені в тест Етап планування.

«Виконання тесту». Після того, як документ буде готовий, ми отримаємо підпис від верхньої команди, щоб розпочати процес виконання, який в основному називається тестовим виконанням. Якщо ми хочемо виконати наші тестові випадки під час виконання, нам потрібно перевірити, чи розробник повинен надіслати інформацію, якщо це нормальне функціональне тестування чи якесь інше тестування або тестування автоматизації, нам потрібна збірка.

**12. Звітування про дефекти та відстеження**

Вході виконання тестових заходів системи, ми виявляємо деякі дефекти, деякі очікувані, а деякі фактичні не дорівнюють очікуваним результатам, тому нам потрібно ввести їх у список і надати їх команді розробників для вирішення, і це в основному називається звітування про дефекти.

Припустимо, якщо ми виявимо дефект у завданні MapReduce, тоді ми повідомимо про це розробника, і вони знову відтворять завдання MapReduce, і вони внесуть деякі модифікації рівня коду, а потім знову нададуть останнє завдання MapReduce, яке нам потрібно протестувати випробування. Після того, як ми закінчили весь процес тестування та дефекти були усунені, нам потрібно створити наші Звіти про тестування. Звіти про тести – це все, що ми зробили для завершення процесу тестування до цього часу. Все планування, написання та виконання тестових випадків, який результат ми отримали тощо, все документується разом у формі тестових звітів.

Потрібно надсилати ці звіти щодня або щотижня або відповідно до потреб Клієнта. На сьогоднішній день організації використовують модель AGILE, тому кожен звіт про статус потрібно оновлювати під час щоденних випробувань.

# СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

# *Базова литература*

1. Кузьменко О., Козьменко О. Економіко-математичні методи і моделі. Економетрика : навч. посіб. К. : «Університетська книга», 2019. 406 с.

2. Pang-Ning Tan, Steinbach M., Karpatne A., Kumar V. Introduction to Data Mining. 2nd edition. New York: Pearson, 2018. 1450 p.

3 Zak Cameron. Data Mining Concepts and Techniques: Complete Guide to a Comprehensive Understanding of Data Mining. Independently published, 2020. 366 p.

4. Фрэнкс Б. «Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики» / Билл Фрэнкс. — Москва: Альпина Паблишер, 2017. — 320 c.

5. Глущенко Н. Большие данные большого города: как Big Data меняет жизнь Киева [Электронный ресурс] / Нина Глущенко // интернет-журнал AIN.UA. — Электронные данные. — [Киев: AIN.UA, 2017]. — Режим доступа: <https://ain.ua/special/big-data-in-kyiv/>

6. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.

7. Дэви С. Основы Data Science и Big Data. Python и наука о даннях // С. Дэви, М. Арно, А. Мохамед. – СПб.: Питер, 2017. – 336 c.: ил.

8. C. Chen, M. Lin, and X. Guo, “High-level modeling and synthesis of smart sensor networks for Industrial Internet of Things,” Computers & Electrical Engineering, vol. 61, pp. 48–66, 2017.

9. Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis und Florian Mansmann. „Visual Analytics“. 2010

10. Dimitri P. Bertsekas and John N. Tsitsiklis. Introduction to Probability. Charles Wheelan. Naked Statistics: Stripping the Dread from the Data. W. W. Norton and Company, 2013.

11. F. Liu, Y. Liu, D. Jin, X. Jia, and T. Wang, “Research on Workshop-Based Positioning Technology Based on Internet of Things in Big Data Background,” Complexity, vol. 2018, Article ID 875460, 11 pages, 2018.

12. H. Mora, M. Signes-Pont, D. Gil, and M. Johnsson, “Collaborative Working Architecture for IoT-Based Applications,” Sensors, vol. 18, no. 6, p. 1676, 2018.

13. H. Tahaei, R. Salleh, S. Khan, R. Izard, K.-K. R. Choo, and N. B. Anuar, “A multi-objective software defined network traffic measurement,” Measurement, vol. 95, pp. 317–327, 2017.

14. Hariri, R.H., Fredericks, E.M. & Bowers, K.M. Uncertainty in big data analytics: survey, opportunities, and challenges. J Big Data 6, 44 (2019). <https://doi.org/10.1186/s40537-019-0206-3>

# *Допоміжна література*

1. Barrow C., Barrow P., Brown R. The Business Plan Workbook: A Step-By-Step Guide to Creating and Developing a Successful Business. 9th Edition. Kogan Page, 2018. 407 p.
2. Behera H.S. (ed.) et al. Computational Intelligence in Data Mining. Springer, 2018. 895p.
3. Bhatia P. Data Mining and Data Warehousing: Principles and Practical Techniques. Cambridge University Press, 2019. 513 p.
4. Business Result Pre-Intermediate. Teacher’s book. 2nd edition. Rachel Appleby, Mark Bartan, David Grant. Oxford University Press, UK, 2017. 97 p.
5. Byrd M.J. Small Business Management: An Entrepreneur’s Guidebook. 8th Ed. McGraw-Hill, 2017. 494 p.
6. Carlberg C. Business Analysis with Microsoft Excel. 5th edition. Que Publishing, 2018. 576 p.
7. Cicala G. Project Management Using Microsoft Project 2019: A Training and Reference Guide for Project Managers Using Standard, Professional, Server, Web Application and Project Online for Office 365. Project Assistants Inc., 2019. 446 p.
8. Dr. Eckroth J. AI Blueprints: How to build and deploy AI business projects. Packt, 2018. 241p.
9. Esposito A., Esposito A.M., Jain L.C. (Eds.) Innovations in Big Data Mining and Embedded Knowledge. Springer, 2019. 286 p. (Intelligent Systems Reference Library 159).
10. Fitzgerald Brian, Stol Klaas-Jan. Scaling a Software Business. Springer, 2017. 264 p.
11. Harris Tom. Start-up: A Practical Guide to Starting and Running a New Business. Springer, 2018. 153 p.
12. Harvard Business Review – Entrepreneur’s Handbook. Boston: Harvard Business School Publishing Corporation, 2018. 303 p.
13. Hastie Trevor, Tibshirani Robert, Friedman Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009. Corrected 12th printing, 2017. 745 p.
14. Kreinovich V., Sriboonchitta S., Chakpitak N. (eds.) Predictive Econometrics and Big Data. Springer, 2018. 788 p.
15. Kryvinska N., Gregus M. (eds.) Data-Centric Business and Applications Evolvements in Business Information Processing and Management (Vol.2). Springer, 2020. 468 p.
16. Lee C.F., Chen H.Y., Lee J. Financial Econometrics, Mathematics and Statistics: Theory, Method and Application. Springer, 2019. 656 p.
17. Olson D.L., Wu D. Predictive Data Mining Models. 2nd.ed. Springer, 2020. 125 p. Computational Risk Management.
18. Olson David L. Descriptive Data Mining. 2nd.ed. Springer, 2019. 130 p. (Computational Risk Management).
19. Project Management Institute. A Guide to the Project Management Body of Knowledge: PMBOK Guide-Sixth Edition + Agile Practice Guide. 6th Edition. Project Management Institute, 2017. 800 p.
20. Project Management Institute. Agile: Practice Guide. Project Management Institute, 2017. 115 p.
21. Rojas I., Pomares H., Valenzuela O. (eds.) Advances in Time Series Analysis and Forecasting: Selected Contributions from ITISE 2016. Springer, 2017. 412 p.
22. Sul D. Panel Data Econometrics: Common Factor Analysis for Empirical Researchers. New York: Routledge, 2019. 165 p.
23. Wilcox J. Excel: The Ultimate Statistics Guide. Mobo Publications, 2017. 251p.