

# ВИМІРЮВАННЯ ТА АНАЛІЗ ДАНИХ СІЛЬСЬКОГОСПОДАРСЬКОГО ПОЛЯ ЗА ДОПОМОГОЮ ХМАРНОГО КОНВЕРБОВОДУ ОПРАЦЮВАННЯ ДАНИХ

*Д.Шутка , магістр , Р.Продан, магістр, В.Татарин , к.т.н., доц.  
Національний університет «Львівська політехніка», Україна;  
e-mail: denys.shutka.mknuo.2022@lpnu.ua*

## Анотація

Зростання попиту на точне землеробство спонукало до інтеграції передових технологій для оптимізації сільськогосподарської практики. У цій статті представлено підхід до обробки даних про сільськогосподарські поля за допомогою хмарного конвеєра даних. Система використовує дані з різних датчиків, розміщених на полях, для збору інформації в режимі реального часу про ключові параметри, такі як вологість ґрунту, температура, вологість тощо. Зібрані дані передаються в хмару, де вони проходять ряд етапів обробки та аналізу даних. .  
Стаття демонструє ефективність хмарного конвеєра даних у підвищенні стійкості сільського господарства. Це сприяє швидкому ухваленню рішень фермерами та зацікавленими сторонами на основі аналізу даних у реальному часі. Крім того, система пропонує цінний інструмент для моніторингу та оптимізації стратегій зрошення, розподілу ресурсів і практик управління врожаєм. Це дослідження підкреслює потенціал хмарних конвеєрів даних у революції точного землеробства. Здатність точно й ефективно вимірювати й аналізувати дані про сільськогосподарські поля відкриває нові шляхи для сталого ведення сільського господарства та пом'якшення ризиків, пов'язаних із лісовими пожежами та посухами.

## Ключові слова

Обробка даних, ETL, аналіз даних, моніторинг і вимірювання, Інтернет речей, сільське господарство, хмарні технології.

## 1. Введення

У міру зростання економіки попит на харчові продукти зростає [1], що робить контроль над сільськогосподарською продукцією дедалі важливішим. Сільськогосподарський сектор потребує поєднання традиційних методів ведення сільського господарства, сучасної науки і техніки та ефективних методів управління для підвищення продуктивності та покращення якості та кількості сільськогосподарського виробництва.

Поля та ферми, які охоплюють великі площі, піддаються ризику пожежі з різних причин, наприклад, від ударів блискавки під час грози або необережного поводження з вогнем. Зважаючи на постійну загрозу бомбардування інфраструктури в Україні, потреба в ефективному сільськогосподарському моніторингу стала як ніколи актуальною.

Україна займає помітне місце як провідний виробник чорноземів, охоплюючи 27,8 млн га або 8,7% світових чорноземних ресурсів [2]. Цей родючий ґрунт займає 67,7% сільськогосподарських угідь України і за належного використання забезпечує найвищий урожай для різних сільськогосподарських культур з достатньою вологістю. Завдяки розробці прототипу конвеєра обробки сільськогосподарських даних для моніторингу полів у реальному часі можна отримати дані різних польових датчиків із будь-якої площі поля в будь-який час. Використання хмарних технологій для обробки системи відкриває численні можливості для розширення її функціональності.

Наприклад, інтеграція системи штучного інтелекту [3] для безперервного аналізу даних шляхом виявлення аномалій є сучасним підходом, якому сприяє джерело даних, надане конвеєром хмарної обробки. Крім того, система сповіщень може надсилати електронні листи зацікавленим сторонам у разі системних збоїв або критичних показників. Доступ у режимі реального часу до інформації з різних датчиків відкриває двері до різних рішень, які можуть підвищити продуктивність сільського господарства.

Використання великих даних [4], Інтернету речей [5], штучного інтелекту, систем обробки подій, сповіщень у хмарі, нереляційних баз даних, технологій візуалізації даних тощо є важливими для успіху сучасного бізнесу. Застосування цих технологій у сільськогосподарській практиці може прокласти шлях до підвищення продуктивності та присутності на світовому ринку. І поточна відсутність таких систем призводить до необхідності ручного моніторингу, який не є таким швидким і ефективним і вимагає більших витрат з часом.

Конвеєр обробки сільськогосподарських даних – це автоматизована система, призначена для оперативного сповіщення користувачів про пожежі на полі або вказівки оптимального часу внесення добрив для рослин. Попит на таку систему є високим, оскільки вона усуває потребу у великих групах, які перевіряють великі сільськогосподарські поля, завдяки доступності даних вимірювань через Інтернет.

Обробка даних є новою та процвітаючою сферою ІТ, яка використовується в різних галузях, від банківської справи до розумних будинків. Оскільки операції з даними стають майбутнім кожного бізнесу,

сільськогосподарський сектор може використовувати ці оброблені результати для підвищення продуктивності та посилення своїх позицій на українському ринку.

## 2. Недоліки

Ручний моніторинг сільськогосподарських полів створює значні недоліки, які перешкоджають ефективному веденню сільського господарства. Одним із ключових обмежень є відсутність моніторингу в режимі реального часу, оскільки дані, зібрані за допомогою перевірок вручну, часто застаріють до моменту, коли вони потрапляють до осіб, які приймають рішення. Ця затримка в отриманні важливої інформації може призвести до втрачених можливостей і затримки реагування на нові виклики.

Крім того, застосування ручних методів робить процес моніторингу повільним і трудомістким. Фахівці повинні фізично відвідати кожне поле, охоплюючи величезні території, які можуть охоплювати десятки чи сотні гектарів, що призводить до значних витрат часу та зусиль. Такий повільний темп збору даних може перешкодити своєчасному прийняттю рішень і завадити швидкому втручанням в критичних ситуаціях.

Іншим недоліком ручного моніторингу є можливість непослідовності в зборі даних. Різний персонал може по-різному інтерпретувати та записувати інформацію, що призводить до розбіжностей у зібраних даних. Такі невідповідності можуть створити неточності в аналізі та процесі прийняття рішень, вносячи невизначеність у сільськогосподарське планування.

Відсутність моніторингу в реальному часі та суперечливі дані можуть призвести до втрат продуктивності сільського господарства. Без своєчасної оцінки здоров'я врожаю, стану ґрунту та потенційних ризиків фермери можуть упустити важливі можливості для швидкого вирішення проблем, що призведе до зниження врожайності та зниження загальної прибутковості.

Крім того, ручний моніторинг сприйнятливий до людських помилок і недоглядів, що ще більше ускладнює точність збору даних. Ці помилки можуть бути результатом втоми, відволікань або різного рівня досвіду серед моніторингового персоналу, що додає ще один рівень невизначеності в процес прийняття рішень.

## 3. Мета

Метою цієї статті є представити та продемонструвати використання автоматизованого хмарного конвеєра даних для вимірювання та аналізу даних сільськогосподарських полів, а також продемонструвати ефективність і результативність цієї технології для вдосконалення точного землеробства шляхом використання даних від датчиків, розміщених на полях, обробляючи їх у хмарі та дозволяючи точні дані в режимі реального часу про стан сільськогосподарських полів.

## 4. Вимірювання та аналіз даних сільськогосподарських полів

У прагненні генерувати цінну інформацію на основі зростаючого обсягу та різноманітності даних важливість інженерії даних та її ролі в аналітиці, науці про дані та машинному навчанні зростає. Команди розробників даних стикаються зі зростаючим тиском, щоб перетворити необроблені неструктуровані дані в чистий і надійний формат, що є вирішальним кроком, який передусє їх ефективному використанню для вирішення бізнес-завдань. Невід'ємною складовою конвеєра обробки даних є побудова моделі Extract - Transform - Load (ETL) [6]. ETL служить процесом, який інженери обробки даних використовують для отримання даних із різних джерел, перетворення їх у корисний і надійний формат і подальшого завантаження в системи, доступні кінцевим користувачам, для вирішення проблем і аналізу даних.

На рис. 1 представлена архітектурна схема конвеєра обробки сільськогосподарських даних з використанням Google Cloud Platform [7]. Кожен компонент рішення зображено як покроковий процес, що ілюструє, як він обробляє дані в потоковому режимі [8].

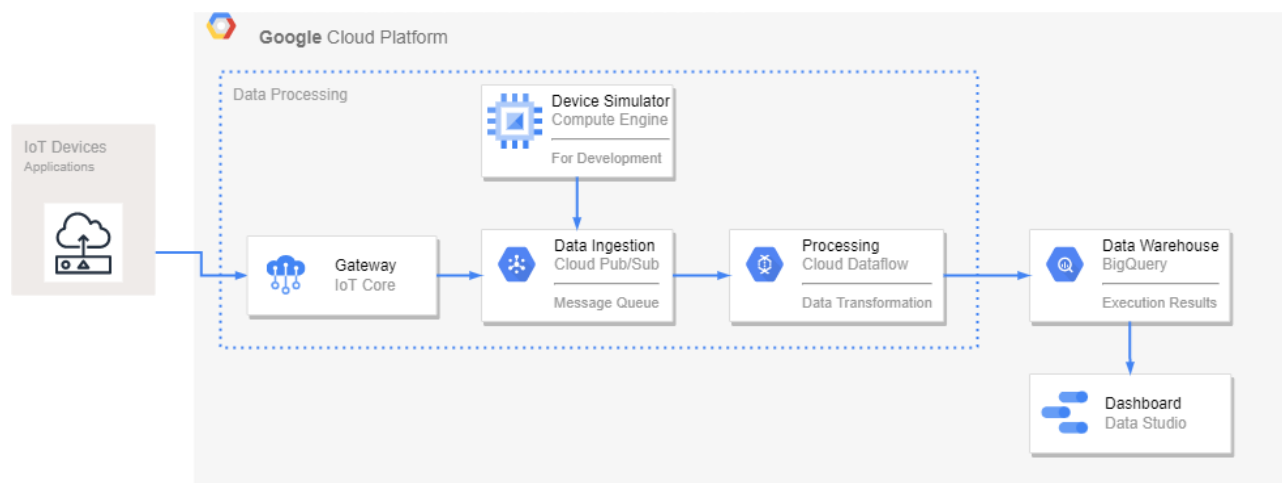


Рисунок 1. Архітектурна схема конвеєра даних

Конвеєр складається з кількох компонентів із різними функціями:

- Пристрої IoT: вони охоплюють датчики Інтернету речей, які збирають важливі дані про різні умови, такі як температура, вологість і сонячна радіація. Крім того, вони передають такі важливі метадані, як ідентифікатори датчиків, час спостереження та геодані.
  - Шлюз: як частина основної служби Google IoT [9], шлюз дозволяє керувати віддаленими датчиками за допомогою інструментів реєстрації, забезпечуючи безпечний зв'язок і передачу даних через мережу.
  - Симулятор пристрою: щоб перевірити розвиток хмарної системи, сценарій Python [10] під назвою Device Simulator генерує змодельовані дані датчиків. Сценарій виконується у віртуальному середовищі Python 3 на віртуальній машині Linux, налаштованій за допомогою служби Google Compute Engine [11]. Весь доступ здійснюється через канал SSH [12], підключений через Cloud Console, оскільки віртуальна машина не має графічного інтерфейсу. Симулятор використовується для прискорення розробки та полегшення впровадження нових змін і вдосконалень.
  - Приймання даних: цей компонент, частина служби Pub/Sub [13], отримує повідомлення даних від датчиків і зберігає їх у черзі. Він діє як проміжний елемент у конвеєрі, забезпечуючи стабільність системи під час роботи в режимі реального часу. Навіть якщо виникають затримки або помилки в наступних компонентах обробки, дані з датчиків зберігаються в цій черзі.
  - Обробка: ключовий компонент конвеєра ETL, розміщений у службі Dataflow платформи GCP [14], безпосередньо обробляє отримані дані. Він виконує перетворення даних із теми Pub/Sub і записує оброблені дані до таблиць зберігання.
  - Data Warehouse: репозиторій для всіх конвеєрних даних, реалізований за допомогою служби Google BigQuery [15]. Зберігання організовано в набори даних, кожен з яких містить таблиці, де зберігаються оброблені дані з попереднього компонента.
  - Інформаційна панель: панель візуалізації даних, створена за допомогою служби Data Studio [16]. Це слугує як варіант для відображення отриманих даних і створення важливих звітів про стан системи.
- Після зберігання даних у таблицях можна отримати цінні відомості, побудувавши відповідні графіки та застосувавши відповідні агрегації. Ці різноманітні діаграми пропонують уявлення про функціонування системи, надаючи параметри окремих станів датчиків, комбіновані показники, а також залежності та розподіли даних. Зразковий графік на рисунку 2 ілюструє агрегацію даних температури від усіх датчиків. Середнє значення дає загальну характеристику системи, тоді як максимальне та мінімальне значення виділяють критичні параметри. Швидкі коливання температури, що утворюють стрибки даних, можуть свідчити про зміни в навколишньому середовищі, як-от займання поля, де дані температури датчика можуть різко підскочити або призвести до збою в параметрах.

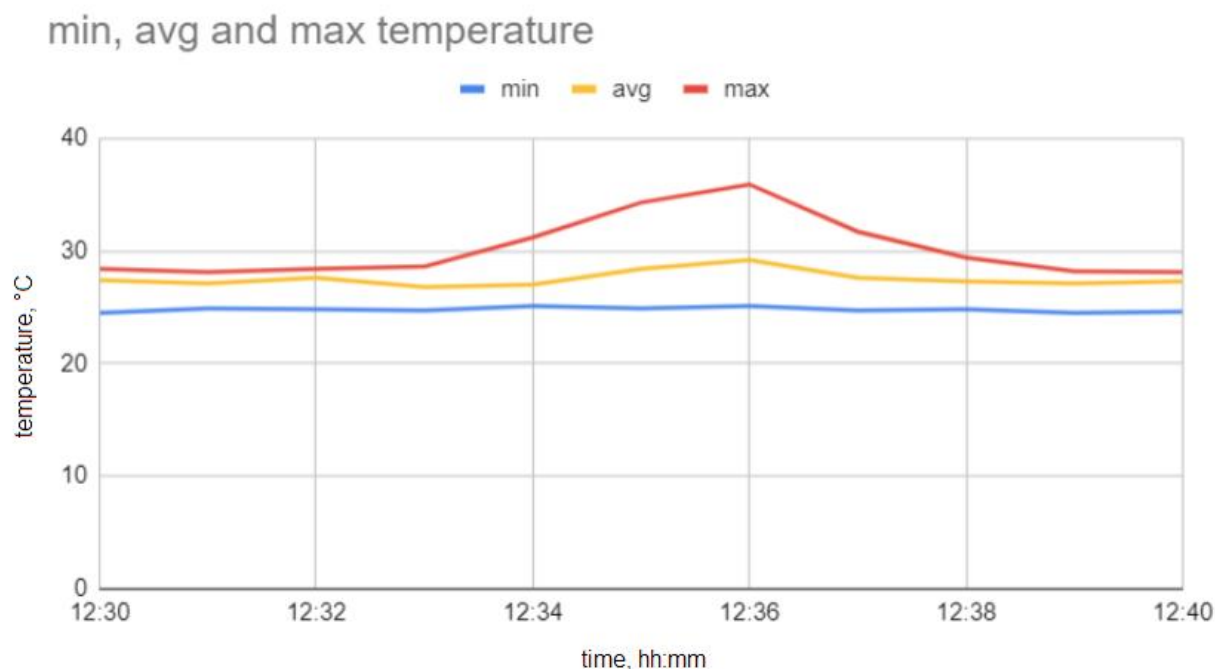


Рис. 2. Зміни середнього, мінімального та максимального значень температури в часі за даними датчиків

На малюнку 3 кругова діаграма відображає рівень освітленості на датчиках, що дозволяє статистику природних умов на всіх полях або окремих частинах системи за допомогою різних фільтрів. Такі дані можуть виступати в якості метрики для розуміння загального розподілу світла протягом дня, дозволяючи прийняти відповідні рішення, наприклад вибрати правильні види культур і оцінити їх зростання. Визначення рівнів можна встановити вручну для конкретних культур, щоб представити точні значення для кожного типу рослин.

### lighting level

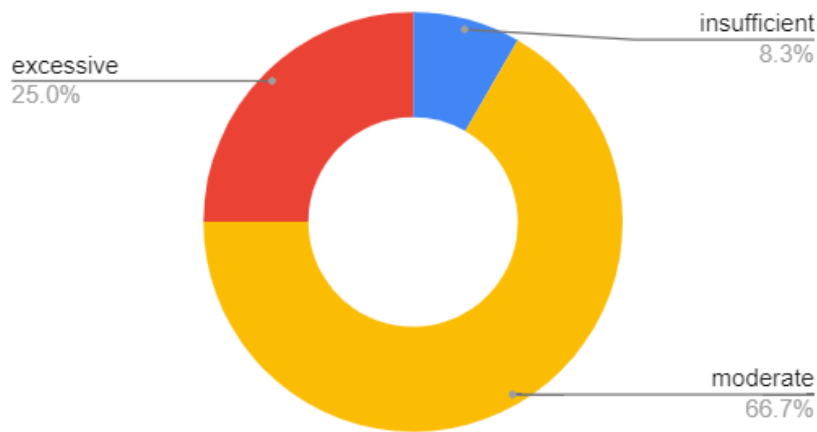


Рис. 3. Кругова діаграма рівня освітленості датчика для вітрини. Рівні освітленості відповідають таким значенням: недостатня – нижче 2000 лк; помірна – від 2000 до 5000 лк; надмірна – понад 5000 лк.

Ці типи агрегації дозволяють переглядати статистику про природні умови по всіх полях в цілому або по окремих частинах системи, використовуючи різні фільтри. Розробка більш складних типів графіків здійснюється з використанням інших методів аналізу даних [17] та розробки засобів фільтрації [18]. Дані також можна обробляти через більші часові інтервали, щоб отримати довгострокову статистику та планувати посіви на наступні сезони. На рисунку 4 показана гістограма розподілу вологості, яку можна отримати шляхом збору даних протягом тривалого періоду часу. Візуалізації такого типу допомагають краще зрозуміти умови конкретних земельних ділянок, щоб спеціалісти мали достатню інформацію про погодні умови на конкретних полях у певні сезони, щоб вибрати правильні види добрив і зробити точний вибір урожаю.

### average humidity

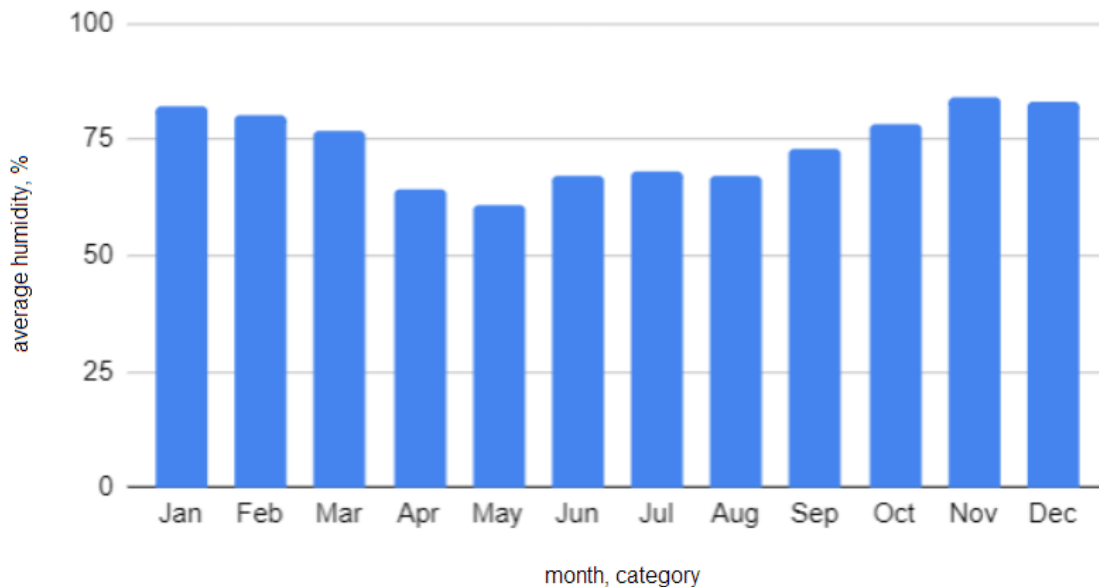


Рис. 4. Гістограма розподілу рівнів вологості протягом року

Окрім даних про загальні характеристики середовища, багато аналітичних операцій можна виконувати з метаданими – даними, які надають інформацію не про показники, а про стан або розташування датчиків. Приклад використання таких метаданих можна побачити на малюнку 5, де доступність датчиків відображається за допомогою секторної діаграми. Якщо певні датчики вийшли з ладу або втратили з'єднання з мережею, це може свідчити або про їхню несправність, або про якісь зовнішні чинники, які могли їх пошкодити. І залежно від розподілу недоступних датчиків у певний час можна зробити прогноз потенційної першопричини.

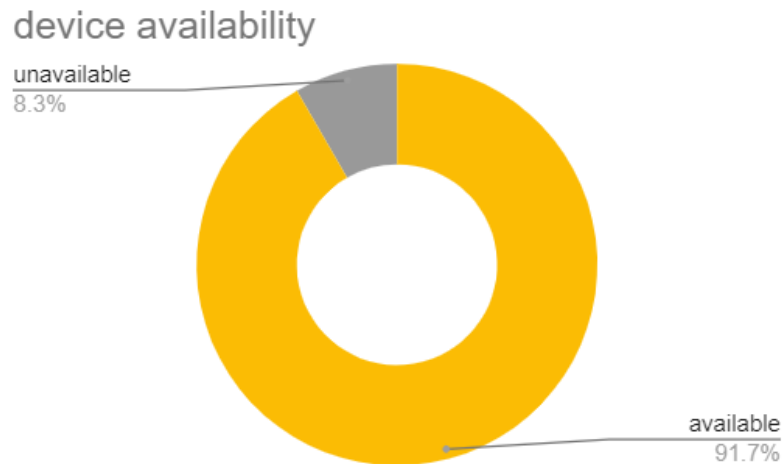


Рис. 5. Кругова діаграма доступності датчика

Ці підходи, керовані даними, пропонують цінні відповіді на різноманітні бізнес-запитання, надаючи лише короткий огляд безлічі прикладів, які можна вивести з цього, здавалося б, невеликого набору інформації. Успішний результат конвеєра хмарної обробки даних полягає в його здатності надавати такі дані та надавати інформацію про численні запити, що робить його безцінним інструментом для використання хмарних технологій для покращення сільськогосподарської практики.

Основна мета цього конвеєра — забезпечити точні вимірювання та аналіз даних сільськогосподарських полів. Використовуючи потужність хмарних обчислень, конвеєр даних забезпечує масштабовані та ефективні можливості обробки даних, що дозволяє проводити швидкий і своєчасний аналіз. Потім оброблені дані використовуються для отримання інформації про стан сільськогосподарських полів, включаючи такі фактори, як можливі лісові пожежі та посухи.

Потенціал розуміння на основі даних величезний, пропонуючи численні рішення для вирішення питань сільськогосподарського бізнесу. Приклади, представлені в статті, дають лише уявлення про можливості, які можна отримати з таких наборів даних. Можливість доступу до даних у реальному часі та аналізу різноманітних запитів є свідченням успіху конвеєрної системи обробки даних із використанням хмарних технологій.

Запровадження конвеєрної обробки сільськогосподарських даних за допомогою хмарних технологій знаменує собою ключовий крок до майбутнього підвищення врожайності та стабільності для сільськогосподарських компаній. Ці системи набувають популярності серед провідних компаній у багатьох країнах. Натхнення успішними рішеннями таких компаній у Європі та Америці, охоплення родючих земель України та величезний потенціал використання передових технологій дозволять значно збільшити поставки продуктів харчування як усередині країни, так і в світі.

Кілька країн уже спостерігали значне щорічне збільшення продуктивності сільськогосподарських угідь на 15% завдяки активному впровадженню технологій Інтернету речей і великих даних, використання даних, зібраних під час посівних сезонів, дає змогу фахівцям краще планувати використання добрив, визначати родючі території та оптимізувати сільськогосподарську діяльність. практики.

Сильна сторона моніторингу даних полягає в його здатності оцінювати різноманітні вимірювання на великих полях площею десятки чи сотні гектарів без необхідності перевірки вручну. У разі лісових пожеж або надзвичайних ситуацій системи обробки даних можуть миттєво сповістити керівників, зменшуючи бізнес-ризик та підвищуючи операційну стабільність у складні часи.

Розробляючи конвеєр даних, користувачі отримують можливість вирішувати важливі питання планування розширення бізнесу та проводити моніторинг сільськогосподарських угідь у реальному часі для забезпечення оптимальної продуктивності. Технології обробки даних представляють майбутнє і активно розвиваються. Багато світових компаній уже впроваджують їх, і українські сільськогосподарські фірми мають потенціал стати наступними бенефіціарами цих трансформаційних досягнень. Застосування цих технологій, безсумнівно, відкриє нові горизонти для зростання та розширення сільськогосподарського сектора.

## 5. Висновки

Впроваджено та успішно використано прототип хмарного конвеєра обробки даних для аграрного сектора для обробки даних із датчиків. Сховище даних — це сховище вихідних даних, яке можна використовувати для отримання інформації з оброблених записів і створення візуалізацій, як показано в статті.

Представлені показники є лише верхівкою айсберга, коли йдеться про безліч способів інтерпретації даних. У наданих візуалізаціях такі показники, як значення температури, вологості та освітлення, а також доступність пристрою, використовуються для демонстрації деяких ключових параметрів, які можна отримати шляхом аналізу даних у реальному часі. Подібні вимірювання корисні фахівцям з точного землеробства для подальшого планування та оцінки врожайності конкретного виду культури в аграрному бізнесі.

## Список літератури

- [1] Е. Фукасе, В. Мартін, «Економічне зростання, конвергенція та світовий попит і пропозиція продовольства», Світовий розвиток, Том 132, 2020. DOI: <https://doi.org/10.1016/j.worlddev.2020.104954>
- [2] Позняк С.П. Чорноземи України: географія, генезис та сучасний стан, 2016. DOI: <https://doi.org/10.15407/ugz2016.01.009>.
- [3] MS Alkathairi, «Штучний інтелект покращив взаємодію між людиною та комп'ютером для комп'ютерних систем», Комп'ютери та електротехніка, Том 101, 2022. DOI: <https://doi.org/10.1016/j.compeleceng.2022.107950>
- [4] Г.Е. Пенс, «Що таке великі дані і чому це важливо?», Journal of Educational Technology Systems, 43(2), 159–171, 2014. DOI: <https://doi.org/10.2190/ET.43.2.d>
- [5] М. Б. Хой (2015), «Інтернет речей»: що це таке і що це означає для бібліотек», Щоквартальник медичних довідкових служб, 34:3, 353-358. DOI: 10.1080/02763869.2015.1052699
- [6] Н. Бісвас, «Новий підхід до концептуального моделювання процесу вилучення-перетворення-завантаження», International Journal of Ambient Computing and Intelligence 10.1, 30-45, 2019. DOI: DOI: 10.4018/IJACI.2019010102
- [7] А. Гупта, П. Госвами, Н. Чаудхарі, Р. Бансал, «Розгортання програми за допомогою Google Cloud Platform», 2020 2-га Міжнародна конференція з інноваційних механізмів для галузевих додатків (ICIMIA), 2020, стор. 236-239 . DOI: <https://doi.org/10.1109/ICIMIA48430.2020.9074911>
- [8] G. van Dongen, D. van den Poel, "Evaluation of Stream Processing Frameworks", IEEE Transactions on Parallel and Distributed Systems, vol. 31, вип. 8, стор. 1845-1858, 2020. DOI: <https://doi.org/10.1109/TPDS.2020.2978480>
- [9] Б. Градос, Х. Бедон. «Програмні компоненти платформи моніторингу IoT у хмарній платформі Google: описове дослідження та архітектурна пропозиція», Комунікації в комп'ютерних та інформаційних науках, том 1193. 2019. DOI: [https://doi.org/10.1007/978-3-030-42517-3\\_12](https://doi.org/10.1007/978-3-030-42517-3_12)
- [10] З. Добесова, «Мова програмування Python для обробки даних», 2011 Міжнародна конференція з електротехніки та керування, 2011, стор. 4866-4869, DOI: 10.1109/ICCEENG.2011.6057428
- [11] Дж. Шах і Д. Дубарія, «Створення сучасних хмар: використання Docker, Kubernetes і Google Cloud Platform», 2019 IEEE 9th Annual Computing and Communication Workshop and Conference, 2019, pp. 0184-0189, DOI: <https://doi.org/10.1109/CCWC.2019.8666479>
- [12] Ф. Бергсма, Б. Дюлінг, Ф. Колар, Дж. Швенк, Д. Стебіла, «Безпека протоколу Secure Shell (SSH) з багатьма шифрами», Конференція ACM SIGSAC з комп'ютерної та комунікаційної безпеки, 2014 р., 369– 381, 2014. DOI <https://doi.org/10.1145/2660267.2660286>
- [13] Ж. Гаскон-Самсон, Ф.-П. Garcia, В. Kemme, J. Kienzle, "Dynamoth: A Scalable Pub/Sub Middleware for Latency-Constrained Applications in the Cloud", 2015 IEEE 35th International Conference on Distributed Computing Systems, 2015, pp. 486-496, DOI: <https://doi.org/10.1109/ICDCS.2015.56>
- [14] J. Sreemathy, R. Brindha, M. Selva Nagalakshmi, N. Suvexha, N. Karthick Ragul and M. Praveennandha, "Overview of ETL Tools and Talend-Data Integration", 2021 7th International Conference on Advanced Computing and Communication Системи, 2021, стор. 1650-1654, DOI: <https://doi.org/10.1109/ICACCS51430.2021.9441984>
- [15] D. Dzulhikam, ME Rana, «A Critical Review of Cloud Computing Environment for Big Data Analytics», 2022 International Conference on Decision Aid Sciences and Applications, 2022, pp. 76-81, DOI: <https://doi.org/10.1109/DASA54658.2022.9765168>
- [16] С. М. Алі, Н. Гупта, Г. К. Наяк, Р. К. Ленка, «Візуалізація великих даних: інструменти та проблеми», 2016 2-га Міжнародна конференція з сучасних обчислювальних машин та інформатики, 2016 р., стор. 656-660, DOI: <https://doi.org/10.1109/IC3I.2016.7918044>
- [17] Б. Ся, П. Гонг, «Огляд бізнес-розвідки через аналіз даних», Бенчмаркінг: міжнародний журнал. 21. 300-311, 2014. DOI: <https://doi.org/10.1108/BIJ-08-2012-0050>.
- [18] MS Gounder, VV Iyer, A. Al Mazyad, "A survey on business intelligence tools for university dashboard development", 2016 3rd MEC International Conference on Big Data and Smart City, 2016, pp. 1-7, DOI: <https://doi.org/10.1109/ICBDSC.2016.7460347>

## Дані про авторів

Денис Шутка

Магістр

Кафедра комп'ютеризованих систем автоматизації

НУ «Львівська політехніка».

Бандери, 12, Львів, Україна, 79013

Електронна адреса: [denys.shutka.mknuo.2022@lpnu.ua](mailto:denys.shutka.mknuo.2022@lpnu.ua)

ORCID: <https://orcid.org/0009-0008-9759-279X>

Роман Продан

Магістр

Кафедра комп'ютеризованих систем автоматизації  
НУ «Львівська політехніка».

Бандери, 12, Львів, Україна, 79013

Електронна адреса: roman.prodan.mknuo.2022@lpnu.ua

ORCID: <https://orcid.org/0009-0000-9725-8582>

Василь Татарин

Кандидат технічних наук, доцент

Кафедра комп'ютеризованих систем автоматизації  
НУ «Львівська політехніка».

Бандери, 12, Львів, Україна, 79013

E-mail: vasyly.tataryn@lpnu.ua

ORCID: <https://orcid.org/0000-0001-9740-1924>