УДК 004.6, 004.4

М. Альхаваша, стажист

Архітектура платформи відкритих даних для e-Government

Київський національний університет імені
Тараса Шевченка, Україна, 01601, м.Київ,
вул. Володимирська, 64/13
e-mail: mohammadalhawawsha@yahoo.com

M. Alhawawsha, intern

Architecture of the Platform for the E-Government Open Data

Taras Shevchenko National University of Kyiv,
64/13, Volodymyrska Street, Kyiv, Ukraine, 01601
e-mail: mohammadalhawawsha@yahoo.com

Електронний уряд - це сукупність поширених технологій та автоматизованих процесів. Відкриті дані відіграють вирішальну роль в успішній реалізації цієї концепції. Архітектура платформи відкритих даних (ODP) тут описана як основа для впровадження систем відкритого доступу до даних, включаючи конкретні вимоги. Пропонована архітектура та її компоненти були детально обговорені в цій роботі щодо її доступності, продуктивності та надійності. Підсистема відкритих даних на основі представленої тут архітектури була розроблена для уряду Йорданії та успішно впроваджена та протестована. Таким чином, ця архітектура показала свою життєздатність.

Основна увага в роботі - детальний аналіз запропонованої архітектури ODP та її характеристик. ODP є важливою системою для зрілого електронного уряду. Ми пропонуємо тут архітектуру для неї з перевіреними на використання характеристиками. Цей факт додає значення стабільності системи електронного урядування та значним характеристикам та покращує загальну якість системи.

Ключові слова: електронний уряд, відкриті дані, цифрова трансформація, оцифровка, архітектура системи, висока доступність, висока продуктивність, відкрита платформа даних.

E-Government is a set of pervasive technologies and automated processes now. The open data plays a crucial role in the successful implementation of this concept. The Open Data Platform (ODP) architecture is described here as the framework for the open data access systems implementation, including specific requirements. The proposed architecture and its components were discussed in this paper in detail for its availability, productivity, and reliability. The open data subsystem based on the architecture presented here was developed for the Jordan Government and was successfully implemented and tested. Thus, this architecture showed its viability.

The focus of the paper is the detailed analysis of the proposed ODP architecture and its characteristics. The ODP is a significant system for the mature e-Government. We propose here the architecture for it with usage-proven characteristics. This fact adds the value to the e-Government framework stability, and significant characteristics and improves the overall quality of the system

Key Words: e-government, open data, digital transformation, digitalization, system architecture, high availability, high performance, open data platform.

Статтю представила к.ф.-м.н., доц. Розора І.В.

Вступ

Цифрова трансформація суспільства, штатів, міст та громадян - це загальний процес на сьогодні. Це зачіпає майже всі сфери нашого життя. Оцифрування відбувається навіть у сфері

електронного здоров'я та електронного уряду, дані стають відкритими та широкодоступними. Усі ці зміни потребують зрілої інфраструктури, яка може впоратися з такими проблемами, як надійність, доступність, продуктивність, доступність та безпека.

Актуальність, підходи, архітектура, рішення та виклики електронного уряду та SMART-уряду були описані та обговорені в [1-12]. Потребу у відкритому електронному уряді розглянуто в [1-5]. Існує також багато публікацій, в яких висвітлюються різні аспекти електронного уряду, включаючи історії успіху та виклики під час розробки та впровадження [6-11].

Багато нетривіальних завдань, які повинні бути вирішені під час якісної реалізації концепції електронного уряду, повинні бути ретельно змальовані та вивчені заздалегідь – наприклад, питання безпеки при роботі з особистими даними та наявність самих даних належної якості. Однак одне з перших і важливих завдань – це концепція відкритих даних [12] та її відповідна реалізація. Це означає, що портал даних (або подібна система) повинен бути:

- достатньо надійний (включаючи відмовостійкість),
- висока доступність в різних умовах,
- продуктивність в сенсі відповідності вимогам до рівня обслуговування (SLA),
- зручність (зрозумілість для кінцевих споживачів, включаючи відповідні вимоги щодо доступності), та
- запобігання несанкціонованому доступу до чутливої частини даних (внутрішніх сервісів), а також до внутрішньої частини системи взагалі, від неуповноважених сутностей (включаючи користувачів).

Тут ми не згадуємо про безпеку та конфіденційність, оскільки тема "Відкриті дані", що означає, власне, відкритість для кінцевого користувача.

Дані повинні бути представлені у чіткій структурі. Ми маємо на увазі самі дані, а також правильні формати та форми подання, та систему каталогів, аби кінцевий користувач знаходив потрібні дані з меншими зусиллями. Проте, ці питання виходять за межі цього дослідження.

Отже, тут ми зосереджуємось на розробці Open Data Framework (ODF) та Open Data System (ODS), зокрема на їх архітектурі, блок-схемі та їх властивостях. Ми обговорюємо очікування, вимоги до такої платформи та її поведінку за різних умов.

Мета дослідження – розробити архітектуру Open Data Platform (ODP), яка відповідає вимогам, описаним вище. Методика є аналітично-порівняльною – вона спирається на сучасні

технології та кращі практики, адаптуючи їх для отримання найкращого результату.

Врахуємо вимоги до системи та, спираючись на поточну тенденцію до архітектури мікросервісів, запропоновано архітектуру платформи відкритих даних, обговорено її властивості та надано аналіз її переваг та недоліків.

Архітектура платформи відкритих даних для електронного урядування

На рис. 1 подано пропоновану структурну схему компонентів ODP: архітектура, робочі процеси, потоки даних та зв'язки.

ODP має бути веб-орієнтованою системою, оскільки користувачі матимуть доступ до відкритих даних через Інтернет. Отже, має бути Front-End частина із серверами доступу до даних та підготовкою та генерацією звітів. Цей блок включає службу кешування для прискорення обробки повторюваних запитів доступу до даних. Також використання кешу може покращити доступність, покращивши загальну продуктивність Back-End-сервісу, зводячи до мінімуму кількість запитів до цієї частини. Також він надає дані, якщо резервна система з якихось причин недоступна. Тут Back-End є сервери зберігання даних (Storage Engine) та доступу до даних (Querying Engine).

Front-End сервери можуть бути комбінацією SQL Server та/або OLAP (On-Line Analytical Processing – на відміну від On-Line Transactional Processing сервера, OLTP) для оптимізації продуктивності, особливо для запитів тривалої обробки з агрегатами над великими масивами даних. Ця частина всієї інфраструктури є основним сховищем вітрин даних (Data Marts), також відомим як склад даних (Data Warehouse) – єдине джерело достовірної інформації для всіх користувачів системи (ODS).

З іншого боку, ODP має багато різних джерел вхідних даних, тому потрібна колекція трансформаторів даних, які перетворюють дані з формату введення у внутрішнє представлення. Цей компонент ми називаємо «Збір вихідних даних» з різних форматів вхідних даних, що перетворюються і надалі спільно зберігаються у сховищі даних.

Як і в більшості систем обробки даних, нам потрібен Склад даних з такими компонентами:

- очищення даних, що означає пошук відхилень та помилок у даних, невідповідностей та колізій, ведення словників різних позначень одних і тих же сутностей та інших подібних завдань,
- попередня обробка даних, яка агрегує дані, перевіряючи їх цілісність та здійснюючи можливі перехресні перевірки,
- систематизація, приведення даних до належної форми, та
- зберігання отриманих очищених та систематизованих даних на складі даних SQL Server.

Усі ці компоненти (очищення, попередня обробка, ректифікація) мають важливе значення для підготовки якісних даних для подальшого використання громадою.

Зовнішні постачальники послуг – це, власне, видима зовні частина всієї Платформи, включаючи впровадження порталу відкритих даних. До таких послуг належать:

- інтерфейси прикладного програмування (API) для користувачів або “роботів” (інші програми, що запитують відкриті дані),
- адаптери даних, які являють собою джерела даних для зовнішнього програмного забезпечення, рядки підключення та інші засоби доступу до даних,
- послуги доступу до даних, які складаються з вітрин даних (у першу чергу – завантажувані файли, а також механізм запитів та інші засоби доступу), знімки даних (в першу чергу для версійності) та інші засоби.

Рис. 1. Структурна схема платформи відкритих даних (ODP).

API рекомендується будувати за допомогою сервісно-орієнтованої архітектури, заснованої на протоколі SOAP, використовуючи мову опису веб-служб WSDL для їх визначення та виявлення. Ця архітектура забезпечує високу доступність та масштабованість для цієї підсистеми та її користувачів.

Внутрішні постачальники послуг надають доступ до даних інших служб для внутрішнього використання. Це містить:

- підсистеми архівації (для історичних версій),
- аналітичні та звітні підсистеми – для інших державних установ, агентств та "внутрішніх" отримувачів даних, яким вони необхідні,

• інші системи, які забезпечують доступ до необхідних даних та відповідно до угод про сервіс.

Аналітична та звітна підсистема спрямована на пришвидшення доступу до складних вибірок даних, що вимагає агрегації багатьох первинних даних та, можливо, різних форматів подання даних.

Крім того, останньою, але не менш важливою складовою ODP є служби моніторингу стану, які включають ряд інфраструктурних рішень, спрямованих на:

- контроль функціональності усіх компонентів,
- попередження ситуації, коли спрацьовує певний тригер щодо критичної події,
- сигналізація про надмірне використання деяких ресурсів на ранніх етапах,
- запобігання проблемам шляхом інформування адміністраторів про критичні зміни ситуації (надмірне завантаження, рівень продуктивності підсистем, доступність),
- візуалізація стану всього комплексу,
- передбачення вузьких місць в продуктивності та інших крайових ситуаціях,
- автоматизація завдань технічного обслуговування (архівування, коригування параметрів віртуалізації обладнання, автоматизація масштабування, перезавантаження тощо)

Всі кроки реєструються, щоб відстежувати всі зміни, доступність та потенційно складні ситуації.

Як ми знаємо з великих проектів та компаній, таких як Google і Facebook, Grammarly та інших веб-орієнтованих продуктів, однією з найбільших команд є команда інфраструктури (розробка, підтримка та моніторинг).

Властивості платформ відкритих даних

Основні властивості впливають із запропонованої архітектури. Надамо короткий аналіз обговорених питань.

Пропонована архітектура ODP має такі позитивні характеристики:

- багаторівнева структура та мікросервісна архітектура окремих блоків підтримують високий рівень незалежності та масштабування,
- механізми кешування та запитів допомагають оптимізувати та збалансувати завантаженість системи,
- розділення внутрішніх та зовнішніх споживачів сервісів,

• підсистема моніторингу, яка допомагає виявити помилки та швидко вирішити критичні ситуації.

Розглянемо докладніше наступні аспекти.

Доступність. Дані залишатимуться доступними (деякий час згідно конфігурації) через кеш, навіть якщо двигун запитів та/або сховище даних стануть недоступними. Крім того, використання OLAP-запитів і механізмів багатовимірного зберігання допомагає поліпшити доступність даних за рахунок кращої швидкості обробки запитів.

Продуктивність. Високий рівень продуктивності (іншими словами, SLA – вимоги щодо рівня обслуговування) повинен бути забезпечений першочергово апаратно (або віртуальною, «хмарною» конфігурацією). Також наступні архітектурні рішення підтримують потрібний рівень продуктивності:

- механізм кешування та багатошарова структура вирішує ситуації з високим навантаженням,
- мікросервіси забезпечують масштабованість,
- поєднання двигунів запитів SQL та OLAP забезпечує гнучкість у створенні більш ефективного потоку даних всередині ODP,
- моніторинг допомагає рано виявити критичні відхилення у швидкодії.

Надійність. Надійність у нашому випадку базується на:

- архітектурі мікросервісів (таким чином досягається високий рівень незалежності компонентів ODP),
- моніторингу за допомогою реактивної підсистеми, яка запобігає системним катастрофам,
- кешах, які покращують доступність у деяких випадках,
- багаторівневій архітектурі, яка допомагає уникнути складних взаємозв'язків і, таким чином, швидко локалізувати потенційні проблеми (у першу чергу, щодо продуктивності),
- розділенні внутрішніх та зовнішніх клієнтів послуг,
- брандмауері, який допомагає фільтрувати трафік та уникати несанкціонованого доступу до системи за багатьма параметрами, наприклад, частотою та різноманітністю запитів.

Проте, безумовно, надійність даних в цілому залежить від якості вхідних даних.

Безпека. Безпека, як згадувалося раніше, не акцентована у цій роботі через відкритий характер ODP та даних всередині, проте коротко охарактеризуємо основні критерії:

- конфіденційність, яку слід враховувати лише на етапі попередньої обробки, оскільки будь-які особисті дані повинні бути виключені,
- цілісність повинна підтримуватись додатковими обмеженнями, застосовуваними до того, щоб обслуговуючий персонал не мав доступу до робочого процесу та потоку даних,
- доступність обговорювалась вище,
- підвітність стосується журналів та історичних знімків, описаних вище,
- достовірність можна гарантувати лише у випадку фізично обмеженого доступу до інфраструктури ODP.

Підсумок та аналіз

Пропонований ODP та його архітектура відрізняються від відомих раніше (наприклад, [6-11]). Описана Платформа слідує принципам архітектури мікросервісів для підвищення доступності та масштабованості, а також використовує ряд технологій (підсистеми OLAP, звітності та моніторингу) для підвищення продуктивності та надійності. Ця архітектура є новою для реалізації концепції відкритих даних для електронного урядування і публікується вперше.

Ми обговорили характеристики ODP запропонованої структури (архітектури). Ця дискусія не залежить від основних технологій впровадження та конкретних рішень – вона стосується концепції та властивостей системи.

Позитивний досвід розробки ODP, яку впроваджено, дає нам право зробити висновок про ефективність цієї архітектури.

Недоліком цієї архітектури є те, що вона складна та потенційно не дешева, поєднує широкий спектр технологій і прийомів (кешування, звітування, OLAP тощо) – але контролером над складністю виступає підсистема моніторингу. Окремі підсистеми можуть масштабуватися незалежно і бути стійкими до відмов (слід просто збільшити кількість ресурсів для них як для мікросервісів) – проте в цьому випадку впровадження та підтримка системи (в

першу чергу апаратні та "хмарні" ресурси) стануть ще дорожчими.

Експериментальна реалізація системи відкритих даних для уряду Йорданії підтверджує доцільність рішення та ефективність запропонованої архітектури. Дещо подібні архітектури можна побачити і в інфраструктурних проектах Facebook та Google, тому це підтримує запропонований метод і підходить до представленої архітектури.

Висновки

У роботі ми докладно дослідили запропоновану архітектуру платформи відкритих даних ODP як основу для відкритого доступу до даних для реалізації концепції електронного уряду. Проаналізовано її властивості, переваги та недоліки, показано, як можна досягти високих показників та як максимально використати інфраструктуру. Однак слід зазначити, що, звичайно, надійність даних (цінність ODP) в цілому залежить від якості вхідних даних.

Пропонована архітектура платформи відкритих даних має такі характеристики: доступність, продуктивність та надійність (відносна, стосовно вхідних даних) – та відповідає нефункціональним вимогам, описаним на початку дослідження.

Системи відкритих даних для уряду Йорданії була впроваджена з використанням запропонованої архітектури. Ця реалізація показала, що ODP є достатньо надійною, забезпечує високу доступність та продуктивність під час навантажувальних випробувань. Таким чином, можна зробити висновок, що запропонована архітектура для ODP є досить зрілою і її рекомендується використовувати в інших подібних системах та реалізаціях.

Запропонована архітектура є новою у описаній конфігурації. У цьому полягає новизна даної роботи. В результаті розробки системи за наведеною архітектурою вдалося розв'язати актуальну задачу для доступу до відкритих даних на прикладі державного проекту.

Список використаних джерел

1. *Alhawawsha M.* A comparison of the E-Government System Architecture in Jordan with the E-Government System of the United States / M. Alhawawsha // “Технологічний аудит та резерви виробництва”. – 2017. – No 2/2 (34). – С.18-24. DOI:10.15587/2312-8372.2017.100190
2. *Alhawawsha M.* Developing of the e-Government System based on Java for Online Voting, Scientific Journal / M. Alhawawsha // “Технологічний аудит та резерви виробництва”. – 2017. – No 3/2 (35). – С.9-13. DOI:10.15587/2312-8372.2017.104033
3. *Alhawawsha M.* E-Government Versus SMART Government: Jordan Versus The United States / M. Alhawawsha, A. Glybovets // Eureka journal. – 2017. – No 3. – С.3-11. DOI:10.21303/2504-5571.2017.00338
4. *Alhawawsha M.* E-Government Developed System Prototype about USA and Jordan / M. Alhawawsha, A. Glybovets // Геометричне моделювання та інформаційні технології. – 2017 (04). – N 1 (3). – С.42-26.
5. *Alhawawsha M.* Identification of the Issues in the E-Governance in Jordan and USA / M. Alhawawsha // Information Systems and Technologies. – 2017. – No 1/1(6). – С.37-40.
6. *Saleh Z.I.* A Framework for an E-government Based on Service-Oriented Architecture for Jordan / Z.I. Saleh, R.A. Obeidat, Y. Khamayseh // Information Journal of information engineering and electronic business. – 2013 (09). – Vol.5, No.3. – С.1-10. DOI:10.5815/ijieeb.2013.03.01
7. *Osho L.O.* Effective Networking Model for Efficient Implementation of E-Governance: A Case Study of Nigeria / L.O. Osho, M.B. Abdullahi, O. Osho, J.K. Alhassan // Information Journal of information engineering and electronic business. – 2015 (01). – Vol.7, No.1. – С.18-28. DOI:10.5815/ijieeb.2015.01.03

References

1. ALHAWAWSHA M. (2017): *A comparison of the E-Government System Architecture in Jordan with the E-Government System of the United States*, Scientific Journal “Technology Audit and Production Reserves”, No 2/2 (34), pp.18-24; DOI: 10.15587/2312-8372.2017.100190
2. ALHAWAWSHA M. (2017): *Developing of the e-Government System based on Java for Online Voting*, Scientific Journal “Technology Audit and Production Reserves”, No 3/2 (35), pp.9-13; DOI: 10.15587/2312-8372.2017.104033
3. ALHAWAWSHA M., GLYBOVETS A. (2017): *E-Government Versus SMART Government: Jordan Versus The United States*, Eureka journal, No3, pp.3-11; DOI:10.21303/2504-5571.2017.00338
4. ALHAWAWSHA M., GLYBOVETS A. (2017): *E-Government Developed System Prototype about USA and Jordan*, Geometry Modelling and Information Technologies, N1(3), 04/2017, pp.42-26
5. ALHAWAWSHA M.(2017): *Identification of the Issues in the E-Governance in Jordan and USA*, Information Systems and Technologies, No 1/1(6), pp.37-40
6. ZAKARIA I. SALEH, RAND A. OBEIDAT, YASER KHAMAYSEH (2013): *A Framework for an E-government Based on Service-Oriented Architecture for Jordan*, IJIEEB, MECS Press, Vol.5, No.3, Sep. 2013, pp.1-10; DOI: 10.5815/ijieeb.2013.03.01
7. LAURETTA O. OSHO, MUHAMMAD B. ABDULLAHI, OLUWAFEMI OSHO, JOHN K. ALHASSAN (2015): *Effective Networking Model for Efficient Implementation of E-Governance: A Case Study of Nigeria*, IJIEEB, MECS Press, Vol.7, No.1, pp.18-28; DOI: 10.5815/ijieeb.2015.01.03

8. *Balakumaran P.J.* Evolving An E-Governance System for Local Self-Government Institutions for Transparency and Accountability / P.J. Balakumaran, V. Ramamoorthy // *Information Journal of information engineering and electronic business.* – 2013 (12). – Vol.5, No.6. – C.40-46. DOI:10.5815/ijieeb.2013.06.05
9. *Olawale O.J.* Comparative Analysis of Personnel Distributions in the Local Government Service in Ekiti-State, Nigeria, for Service Delivery / O.J. Olawale, A.A. Justus // *International Journal of Mathematical Sciences and Computing.* – 2019 (04). – Vol.5, No.2. – C.44-53. DOI:10.5815/ijmsc.2019.02.04
10. *Mohammad H.* E-Government in Jordan / H. Mohammad, T. Almarabeh, A.A. Ali // *European Journal of Scientific Research.* – 2009. – Vol.35, No.2. – C.188-197.
11. *Majdalawi Y.Kh.* E-Government Strategy and Plans in Jordan / Y.Kh. Majdalawi, T. Almarabeh, H. Mohammad, W. Quteshate // *Journal of Software Engineering and Applications.* – 2015. – #8. – C.211-223. DOI:10.4236/jsea.2015.84022
12. *AlSukhayri A.M.* Leveraging the Saudi Linked Open Government Data: A Framework and Potential Benefits / A.M. AlSukhayri, M.A. Aslam, S. Arafat, N.R. Aljohani // *International Journal of Modern Education and Computer Science.* – 2019 (07). – Vol.11, No.7. – C.14-22. DOI:10.5815/ijmecs.2019.07.02
8. BALAKUMARAN P.J, VIGNESH RAMMAMORTHY. H (2013): *Evolving An E-Governance System for Local Self-Government Institutions for Transparency and Accountability*, IJIEEB, MECS Press, Vol.5, No.6, Dec. 2013, pp.40-46; DOI: 10.5815/ijieeb.2013.06.05
9. OLANIYAN JOSEPH OLAWALE, ADEMUYIWA A. JUSTUS (2019): *Comparative Analysis of Personnel Distributions in the Local Government Service in Ekiti-State, Nigeria, for Service Delivery*, IJMSC, MECS Press, Vol.5, No.2, pp.44-53; DOI: 10.5815/ijmsc.2019.02.04
10. HIBA MOHAMMAD, TAMARA ALMARABEH, AMER ABU ALI (2009): *E-Government in Jordan*, European Journal of Scientific Research, Vol.35, No.2, pp.188-197
11. YOUSEF KH. MAJDALAWI, TAMARA ALMARABEH, HIBA MOHAMMED, WALA QUTESHTE (2015): *E-Government Strategy and Plans in Jordan*, Journal of Software Engineering and Applications, 8, pp.211-223; DOI: 10.4236/jsea.2015.84022
12. AFNAN M. ALSUKHAYRI, MUHAMMED AHTISHAM ASLAM, SACHI ARAFAT, NAIF RADI ALJOHANI (2019): *Leveraging the Saudi Linked Open Government Data: A Framework and Potential Benefits*, IJMECS, MECS Press, Vol.11, No.7, pp.14-22; DOI: 10.5815/ijmecs.2019.07.02

Надійшла до редколегії 15.11.2018