

УДК. 004.658.2

ПЕРСПЕКТИВИ РОЗВИТКУ ІНФРАСТРУКТУРИ БАЗ ДАНИХ ДЛЯ ФІНАНСОВОГО СЕКТОРУ

DOI 10 36994 / 2788- 5518- 2022-02-04-20

Рожков С.М. Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. serj.rozhkov@gmail.com

Анотація: Метою цієї статті є дослідити та спрогнозувати етапи та шляхи розвитку інфраструктури для баз даних фінансових установ, що пов'язані із індустріальними регуляціями. Розглянуто основні тенденції в IT в цілому та у вибраному домені, наведені порівняльні розрахунки для безсерверних хмарних сервісів та самоінстальованих баз даних.

Ключові слова: бази даних, AWS, публічна хмара, безсерверні технології, RDS, EC2, фінансовий сектор.

PROSPECTS OF THE INFRASTRUCTURE DEVELOPMENT FOR FINTECH COMPANY'S DATABASES

Sergii Rozhkov. Open International University of Human Development "Ukraine", Kyiv, Ukraine. serj.rozhkov@gmail.com

Abstract: Technologies used in the development of applications are changing rapidly. On the other hand, one of the main parts of the backend is a database which does not undergo fundamental changes so often. The purpose of this work is to investigate and, if possible, to predict ways of fintech companies' infrastructural solutions for the nearest future.

One of the biggest users of relational databases is the financial sector. In order to predict something, one must first determine the current state of database infra. By 2020, almost 75 percent of the infrastructure was located on real servers in rackspaces. Despite all this, most fintech companies plan to move to public clouds in the future as it provides

many advantages for their business. Even 5-7 years ago, it was not possible to talk about the fact that highly loaded databases of serious financial companies could be deployed in the public cloud, since public clouds could not provide the necessary performance, primarily in the speed of working with the volumes[1][4]. But progress in hypervisor technologies has led to the fact that almost any database can be migrated or deployed in the cloud. Therefore, it is safe to say that the future deployment of database infrastructure are public or private clouds.

But the use of serverless technologies (for example, RDS in AWS) will still not gain such a large popularity. Since so far self-installed databases are cheaper not only in terms of hardware, but also in terms of licenses for paid versions of databases. In addition, self-installed versions provide greater flexibility and maneuverability in solving emerging issues during operation, for example, to work with a disk, you can use software RAID for disks of different types, distribute the load on different partitions (for example, database files on one partition, log files on another partition)[1] or use optimization due to the file system itself and its additional options, use combinations of instances with ephemeral storage and orchestrators and configuration management systems for read-only replications.

Key words: databases, AWS, public cloud, serverless services, RDS, EC2, fintech.

Вступ.

В сучасному бізнесі існує велика конкуренція, яка є рушієм прогресу, а для кінцевих користувачів виражена у тому, що сервіси постійно покращуються, не є виключенням і фінансовий сектор, де, наприклад, вже важко знайти людину, у якої буде менше двох мобільних банків на смартфоні. Тому, як ми бачимо, навіть тут, де кожна дія обмежена індустріальними та урядовими регуляціями, треба орієнтуватися на високопродуктивні фреймворки, безсерверні та хмарні технології – щоб бути швидшим, кращим і давати те, чого немає у конкурентів.

Технології що використовуються у розробці застосунків змінюються швидко, але, не зважаючи на це, однією із основних частин бекенду є бази даних де корінні зміни відбуваються не так часто, тому в даній статті основна увага буде приділена саме базам, так як окрім того що це є важливим елементом він, як правило, є ще і найдорожчою частиною архітектурного рішення.

Як видно із рис.1. одним із найбільших користувачів реляційних баз даних є саме фінансовий сектор.

Виконання досліджень

Для того щоб щось прогнозувати треба спочатку визначитися із поточним станом речей. На 2020 рік майже 75 відсотків інфраструктури знаходилися саме на «залізних» серверах(див рис.2.)

Але якщо подивитися на топових користувачів саме хмарних сервісів (рис.3.)[7] то можна побачити що саме фінансовий сектор є одним із лідерів по інтеграції своїх сервісів, хоча захист облікових даних – одна з найбільших проблем, пов'язаних із розташуванням даних у хмарі. Візьмемо, наприклад, AWS (Amazon Web Services). Оскільки з вашого кореневого облікового запису здійснюється керування пов'язаними ресурсами та сервісами AWS, потрібно надати повний доступ до всіх даних, включаючи ті, на які поширюються права адміністратора.

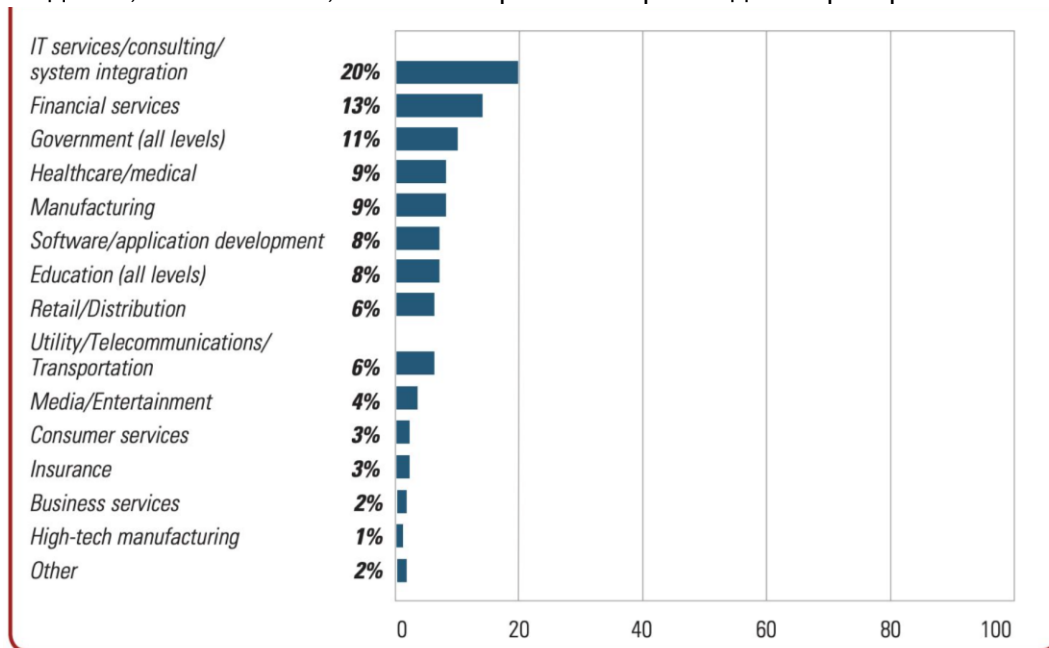


Рис.1. Статистика використання баз даних за індустрією.

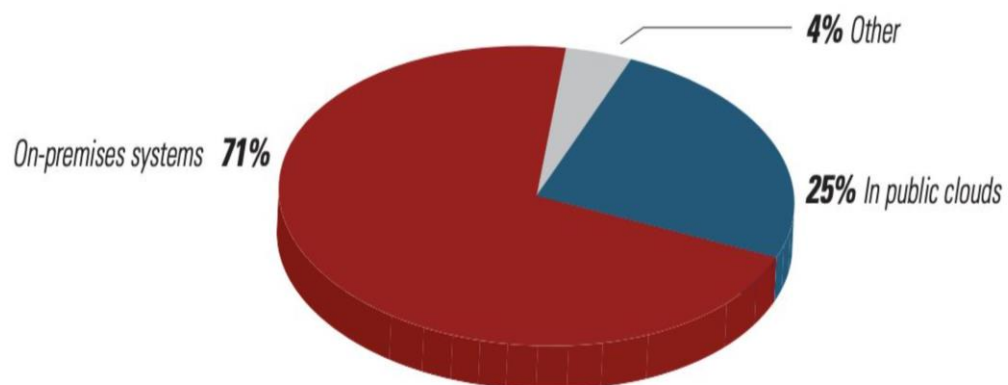


Рис.2. Розміщення інфраструктури оператора.



Рис.3. Лідери використання хмарних сервісів.

Незважаючи на все це, більшість фінтек компаній планує і надалі переїздити у публічні хмари(рис.4.) так як це надає безліч переваг[3]:

- більша гнучкість масштабування;
- менший кошторис;
- не потрібно витрачати час на підтримку серверної інфраструктури;
- швидший шлях сетапу від розробки до продакшену;
- менший рівень кваліфікації персоналу;
- більший вибір СУБД;
- інтеграція із іншими хмарними сервісами;
- швидкі бекапи;
- можливість переносити пікове навантаження із нарощуванням потужності;
- краще балансування навантаження.

Згідно із рис.5 основні зусилля європейських компаній витрачаються саме на міграцію у хмари і, як бачимо, далі на оптимізацію коштів, що і не дивно, так як тарифікація сервісів є багатовекторною і складною. Це навіть породило компанії які просто прораховують вартість майбутньої конфігурації інфраструктури, а потім контролюють рахунок і надають свої рекомендації щодо його оптимізації.

	<i>Past 12 months</i>	<i>Next 12 months</i>
<i>Decrease</i>	<i>2%</i>	<i>2%</i>
<i>No change</i>	<i>32%</i>	<i>18%</i>
<i>Increase 1–5%</i>	<i>18%</i>	<i>15%</i>
<i>Increase 6–10%</i>	<i>16%</i>	<i>17%</i>
<i>Increase 11–25%</i>	<i>13%</i>	<i>17%</i>
<i>Increase 26–50%</i>	<i>9%</i>	<i>11%</i>
<i>Increase by more than 50%</i>	<i>4%</i>	<i>10%</i>
<i>Total Increase</i>	<i>26%</i>	<i>38%</i>
<i>Don't know/unsure</i>	<i>7%</i>	<i>9%</i>

Рис.4. Найближчі плани стосовно міграції баз даних у хмарні сервіси.

Top 3 Cloud Initiatives in Europe



Рис.5. На які роботи у хмарі витрачають ресурси європейські компанії

Із наведеної вище статистики випливає лише один висновок – надалі більшість компанії будуть переносити свої як звичайні сервіси і застосунки там і бази даних саме у публічні хмари.

Розглянемо ефективність використання коштів, так як перш за все мета бізнесу -це більше заробляти та менше витрачати. А далі відштовхуватися від результатів. Публічна хмара, мабуть, найбільш економне рішення, особливо якщо ви працюєте з SaaS-додатком зі зростаючими вимогами до інфраструктури. Приватна хмара, навпаки, буде найкращим варіантом, якщо контроль та безпека для вас найважливіша. Так як найпопулярнішим публічним хмарним сервісом є AWS далі будемо розглядати само його(рис.6.).

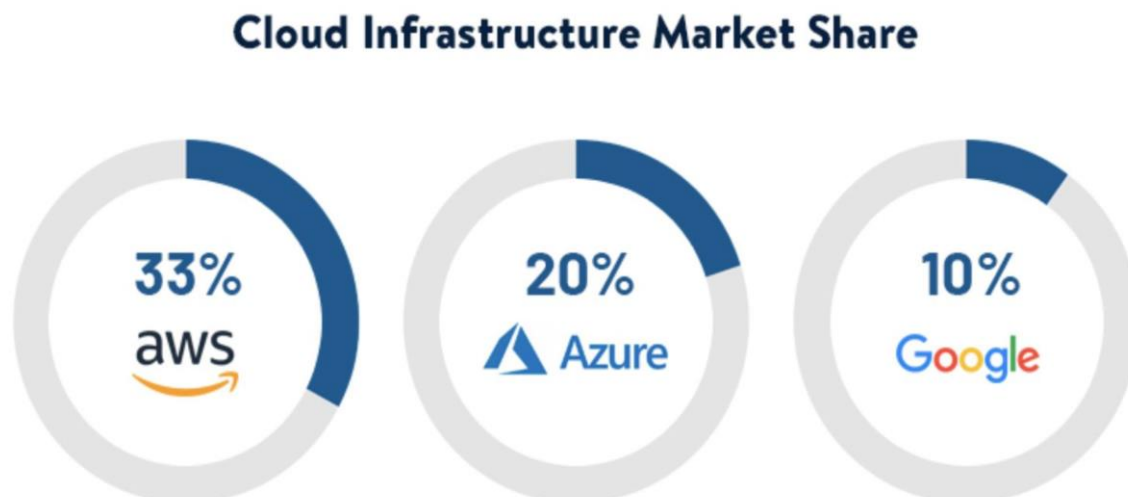


Рис.6. Статистика використання хмарних сервісів.

Далі постає питання – чи треба використовувати хмарні сервіси баз даних чи все ж використовувати самоінстальовану базу даних.

Розглянемо AWS RDS (Relational Database Service) – це база даних як сервіс. Вам надають доступ до бази даних без необхідності її адмініструвати, а всі турботи щодо підтримки резервного копіювання, налаштування реплікації та забезпечення високої доступності бере на себе постачальник послуги (тобто AWS)[5].

Простіше кажучи ви отримуєте один або кілька екземплярів БД, з якими можна одразу працювати – писати чи читати дані. При цьому немає необхідності самостійно розбиратися з встановленням та початковою конфігурацією самої БД. Але, насправді, є нюанси. Обмеження RDS можна загалом сформулювати так[2]:

- оскільки весь процес адміністрування бере на себе провайдер, ви як користувач отримуєте обмежену можливість конфігурування БД, що іноді буває критично, тому що при цьому відсутня можливість при необхідності зробити тонке налаштування;
- із точки зору безпеки ви навіть не маєте користувача із повним набором адміністративних привілеїв;
- якщо хтось робить за вас всю «брудну» роботу, то за це доведеться заплатити, а у випадку RDS рахунок буває чималим.

Зробимо невеличкий примірний підрахунок-порівняння між RDS та самоінстальованою базою на EC2. Для прикладу візьмемо базу середнього розміру із параметрами 32 CPU 128 RAM. Для простоти прикладу візьмемо звичайний SSD диск на 1Tb.

Для RDS:

Estimated monthly costs	
DB instance	2300.96 USD
Storage	127.00 USD
Total	2427.96 USD

Рис.7. Розрахунок вартості інстансу RDS із майстер ногою

Якщо додати ноду для репліки:

Estimated monthly costs	
DB instance	2300.96 USD
Multi-AZ standby instance	2300.96 USD
Storage	253.00 USD
Total	4854.92 USD

Рис.8. Розрахунок вартості інстансу RDS із майстер ногою та реплікою

Приклад для бази на EC2:

Estimated monthly costs

	instance type	hour	day	month	storage per Gb	1 Tb	Total
master	m6i.8xlarge	1,712	41,088	1232,64	0,11	110	1342,64
Total							1342,64

Рис.9. Розрахунок вартості інстансу EC2 із майстер ногою

із реплікою:

Estimated monthly costs

	instance type	hour	day	month	storage per Gb	1 Tb	Total
master	m6i.8xlarge	1,712	41,088	1232,64	0,11	110	1342,64
replica	m6i.8xlarge	1,712	41,088	1232,64	0,11	110	1342,64
Total							2685,28

Рис.10. Розрахунок вартості інстансу EC2 із майстер ногою та реплікою

Як впливає із розрахунків – самостійно інстальована база даних в два рази дешевше, тобто, якщо ми говоримо про кластерну архітектуру для високонавантажених баз даних, різниця може сягати десятків тисяч доларів на місяць.

Також можна розглянути ліцензовані бази даних і порівняти їх ціни, візьмемо до прикладу дуже популярний Microsoft SQL Server:

- RDS RI 32 cores price - 178116\$/рік
- Marketplace IAM 32 cores price - 137777\$/рік + 15894\$/рік за EC2 інстанс
- Own License 32 cores - 95632\$/рік + 15894\$/рік за EC2 інстанс

Це означає що із точки зору бізнесу повністю нівелюється головна перевага RDS – те що не треба мати окремого спеціаліста із баз даних, тому що для великих об'ємів вигідніше найняти команду DBA.

Часто буває, коли проект тільки починається, то вигідніше взяти RDS. Але в міру зростання та розвитку самого проекту зростає, як правило, і база даних, а значить і вартість за кожен екземпляр RDS. У якийсь момент приходить розуміння, що вигідніше мати self-hosted-рішення, навіть з урахуванням витрат на його розгортання та підтримку часу і ресурсів.

Висновок

Ще 5-7 років тому годі було говорити про те щоб високонавантажені бази даних серйозних фінансових установ могли бути розгорнуті у хмарі, так як публічні хмари не могли надати необхідну продуктивність насамперед у швидкості роботи із диском[1][4]. Але прогрес у технологіях роботи гіпервізора призвів до того що вже майже будь-яку базу можна мігрувати або розгорнути у хмарі. Тому можна впевнено говорити що майбутнє розгортання інфраструктури баз даних – це публічні або приватні хмари.

А ось використання безсерверних технологій(наприклад RDS у AWS) все ж не набуде такого великого масштабу. Так як поки що самоінстальовані бази даних окрім того що є дешевшими не тільки у плані заліза, а і у плані ліцензій на платні версії баз даних. Окрім того, самоінстальовані версії надають більшу гнучкість та маневреність по вирішенню виникаючих питань по мірі експлуатації, наприклад, для роботи із диском можна використовувати програмні RAID для дисків різних типів, розносити навантаження на різні розділи(наприклад на один розділ файли бази, на інший розділ файли журналів)[1] або ж використовувати оптимізацію за рахунок самої файлової системи та її додаткових опцій, використовувати комбінації із інстансів із ефемерним стораджем та оркестраторів і систем управління конфігураціями для реплікацій що використовуються лише для читання.

References

1. Vinicius M. Grippa, Sergey Kuzmichev «Learning MySQL: Get a Handle on Your Data»
2. Hans-Jurgen Scholing «Mastering PostgreSQL 13: Build, administer, and maintain database applications efficiently with PostgreSQL 13», 4th Edition.
3. Brendan Gregg's "Systems Performance: Enterprise and the Cloud", 2nd Edition.
4. M. Ma, Z. Yin, "Diagnosing root causes of intermittent slow queries in cloud databases".
5. Shared Responsibility Model aws.amazon.com/ru/compliance/shared-responsibility-model.
6. Database in a cloud survey <https://d1.awsstatic.com/whitepapers/RDS/IOUG-AWS-Databases-in-the-Cloud-Final-Report.pdf>
7. Cloud computing statistic <https://99firms.com/blog/cloud-computing-statistics/>