Monitoring&alerting model

Метрики можуть збиратися комплексно. Для бізнес-логіки та JVM: Spring Boot Actuator з використанням Micrometer. Для інфраструктури: агенти Node Exporter/cAdvisor на хостах. Усі дані збираються, зберігаються та візуалізуються через Prometheus та Grafana.

- Бізнес-метрики (recipe.generation.time, recipe.favorites.added.count): Збираються безпосередньо в коді Recipe Generation Module та Recipe Management Module за допомогою Micrometer Timers та Counters.
- Системні метрики (http.server.requests, jvm.memory.used): Збираються автоматично Spring Boot Actuator для всіх HTTP-запитів та JVM.
- Інфраструктурні метрики (system.cpu.usage, system.disk.usage): Збираються агентами Node Exporter, встановленими на віртуальних машинах/контейнерах, та передаються до Prometheus.

Метрика	**Виміри**	**Зв'язок з інфраструктурними ресурсами**	**Призначення**
1. `recipe.generation.time`	Мілісекунди (мс)	CPU, Network I/O (зовнішній API)	Час відповіді OpenAI.
2. `http.server.requests` **(Response Time p95)**	Мілісекунди (мс)	CPU, RAM, Network I/O	Час відповіді на API-запити (SLO: < 500 мс).
3. `circuitbreaker.state`	Стан (OPEN/CLOSED/HALF_OPEN)	N/A	Стан Circuit Breaker для OpenAI API.
4. `recipe.generation.failure.count`	Кількість	N/A	Кількість збоїв викликів ШІ (5xx).
5. `http.server.requests` **(Status 5xx Rate)**	Відсоток (%)	N/A	Частота помилок на сервері.

6. `jvm.memory.used`	Мегабайти (МБ)	RAM	Використання пам'яті сервісом Spring Boot.
7. `system.cpu.usage`	Відсоток (%)	CPU	Загальне навантаження на процесор хоста.
8. `db.connection.active`	Кількість	RAM, Disk I/O	Кількість активних підключень до PostgreSQL.
9. `user.registration.count`	Кількість	Disk I/O (DB writes)	Кількість нових реєстрацій (бізнес-метрика).
10. `recipe.favorites.added.count`	Кількість	Disk I/O (DB writes)	Популярність функції "Улюблене".
11. `pdf.export.success.count`	Кількість	CPU (генерація), Disk I/O	Успішність роботи **PDF Export Module**.
12. `active.users.concurrent`	Кількість	CPU, RAM, DB Connections	Кількість одночасних користувачів (SLO: 1000, max 2000).
13. `process.uptime`	Секунди	N/A	Час безперебій

			ної роботи сервісу.
14. `system.disk.usage`	Відсоток (%)	Disk Space	Використання дискового простору резервні копії).

Alerts

Метрика	**Мін/Макс Допустиме Значення**	**Тип**	**Критичність**	**Mitigation Plan**
1. `recipe.generation.time`	Макс: 15 с (p95)	Threshold	**Критична**	1. Масштабування Spring Boot Backend (Recipe Generation Module). 2. Перевірка завантаження та лімітів OpenAI API. 3. Аналіз якості `prompt` для зменшення часу генерації.
2. `http.server.requests` **(Response Time p95)**	Макс: 500 мс	Threshold	**Висока**	1. Горизонтальне масштабування Spring Boot Backend. 2. Оптимізація повільних DB-запитів (R5.1). 3. Перевірка CPU/RAM на хостах.
3. `circuitbreaker.state`	Стан = OPEN	State Change	**Критична**	1. Перевірка доступності

	протягом 1 хв			OpenAI API. 2. Автоматичне/ручне скидання Circuit Breaker після підтвердження відновлення.
4. `http.server.requests` (Status 5xx Rate)	Середнє > 1% від загальної кількості запитів	Trend	**Висока**	1. Аналіз логів для виявлення причини 5xx (напр., R4.1 - помилка експорту PDF). 2. Можливий відкат (rollback) останнього розгортання.
5. `db.connection.active`	Макс: 90% від Max Pool Size	Threshold	**Критична**	1. Збільшення пулу підключень (Connection Pool Size). 2. Масштабування PostgreSQL (додавання Read Replicas). 3. Виявлення та завершення довготривалих транзакцій.
6. `system.cpu.usage`	Макс: 85% протягом 5 хв	Threshold	**Висока**	1. Горизонтальне масштабування Backend (автоскейлінг). 2. Виконання профілювання для виявлення "гарячих" точок коду.
7. `jvm.memory.used`	Макс: 80% від Heap	Threshold	**Середня**	1. Профілювання для виявлення Memory Leak. 2.

				Збільшення розміру Heap.