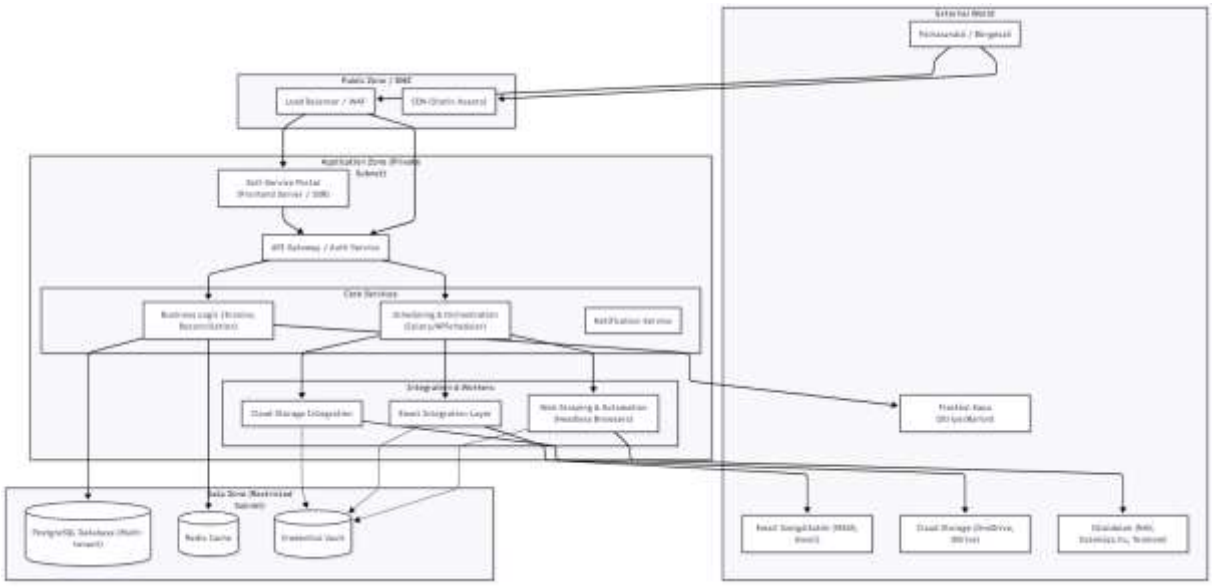
Rendszer Architektúra és Topológia Terv

Ez a dokumentum a tervezett SaaS rendszer magas szintű architektúráját és topológiáját tartalmazza a biztonsági tervezés elősegítése érdekében.

Architektúra Diagram



Komponens Részletek és Biztonsági Megfontolások

A rendszer rétegekre bontva, a fenti topológia alapján:

1. Public Zone (Nyilvános Zóna)

A belépési pont a rendszerbe.

- **Load Balancer & WAF (Web Application Firewall):** Védi az alkalmazást a DDoS támadások, SQL injection és egyéb gyakori fenyegetések ellen. Itt történik a HTTPS terminálás.
- **CDN:** A statikus tartalmak (képek, JS, CSS) kiszolgálása a gyorsabb betöltés érdekében.

2. Application Zone (Alkalmazás Zóna - Private Subnet)

Az üzleti logika itt fut, közvetlenül nem érhető el az internetről, csak a Load Balancer-en keresztül.

- **Self-Service Portal:** A felhasználói felületet kiszolgáló szerverek (React/Vue.js hoszting vagy SSR).
- **API Gateway / Auth Service:** Központi hitelesítés és felhatalmazás (Multi-tenancy enforcement). Minden kérést validál, mielőtt a belső szolgáltatásokhoz érne.

- **Core Services:**
    - **Business Logic:** Számlafeldolgozás, egyeztetés (Reconciliation), analitika, cashflow engine.
    - **Notification Service:** Email és SMS értesítések kiküldése.
- **Integration & Workers (Aszinkron feldolgozók):**
    - **Scraper Engine:** A legérzékenyebb pont a kimenő forgalom szempontjából. Headless böngészőket futtat (izolált konténerekben ajánlott, pl. Selenium Grid vagy Playwright) a külső oldalak (NAV, Telekom, stb.) eléréséhez.
    - **Email & Storage Workers:** API kapcsolatokat kezelnek a külső szolgáltatókkal (M365, Gmail, OneDrive).

3. Data Zone (Adat Zóna - Restricted Subnet)

A legszigorúbban védett zóna. Közvetlen internetelléréssel nem rendelkezik.

- **PostgreSQL:** A fő adatbázis. A multi-tenancy adatbiztonság itt fizikailag (adatbázis particionálás/schema separation) vagy logikailag (Row Level Security) biztosított.
- **Credential Vault: KRITIKUS ELEM.** Itt tároljuk titkosítva az ügyfelek külső szolgáltatásokhoz (pl. Telekom bejelentkezéshez) tartozó belépési adatait. Szigorú audit naplózás és hozzáférés-vezérlés szükséges.
- **Redis:** Gyorsítótár a munkamenetek, job queue és átmeneti adatok számára.

Biztonsági Tervezési Fókusz (Protection Planning)

A "védelem" tervezéséhez (ARLITECH biztonsági fókusz) ezeket a területeket javasolt kiemelni a topológián:

1. **Credential Vault Védelme:** Mivel a rendszer jelszavakat tárol (pl. Telekom bejelentkezéshez), ennek a komponensnek a védelme a legkritikusabb. Javasolt ipari standard megoldás használata (pl. HashiCorp Vault, Azure Key Vault), nem saját fejlesztésű tároló.
2. **Scraper / Headless Browser Izoláció:** A web scraper-ek idegen kódokat futtathatnak vagy sérülékeny weboldalakat látogatnak. Ezeket érdemes rövid életű, eldobjató konténerekben futtatni (ephemeral containers) egy külön hálózati szegmensben, hogy egy esetleges böngésző-szintű exploit ne terjedjen át az adatbázisra vagy a kódra.
3. **Tenancy Izoláció:** Az Auth rétegnek garantálnia kell, hogy a Tenant ID minden adatbázis lekérdezéshez hozzá legyen csatolva ("Logical Separation"), hibátlanul.
4. **Network Policies / Egress Filtering:** A Scraper Engine-nek engedélyezni kell a kimenő forgalmat a céloldalak felé, de a Data Zone felé csak a szükséges minimumot (eredmény vissz írása).

Kritikus Komponens Specifikációk

Ez a dokumentum a rendszer két legkritikusabb biztonsági és üzemeltetési komponensének technikai részleteit tartalmazza.

1. Credential Vault (Biztonságos Adattároló)

A rendszernek ügyfelek nevében kell bejelentkeznie külső szolgáltatókhoz (pl. Telekom, Számlázz.hu, Google). Ezeknek a jelszavaknak és API kulcsoknak a tárolása a legmagasabb biztonsági kockázatú pont.

Tervezett Megoldás: HashiCorp Vault (vagy Cloud Provider Key Management)

Saját fejlesztésű titkosítás helyett ipari szabvány megoldást használunk.

Architektúra

- **Storage Backend:** Encrypted storage (pl. AWS KMS / Azure Key Vault integration).
- **Access Control:** AppRole authentication. A Backend szolgáltatás csak a számára szükséges titkokat olvashatja.
- **Encryption:** AES-256-GCM. A kulcsok (Master Key) soha nem tárolódnak az alkalmazás kódban.

Adatmodell & Útvonalak

A Vault-on belüli útvonalstruktúra a Multi-tenancy-t tükrözi: secret/data/tenants/{tenant_id}/{provider_name}

Példa JSON payload egy Telekom fiókhoz:

```
{
  "username": "user@example.com",
  "password": "encrypted_stored_value",
  "2fa_seed": "otp_seed_if_applicable",
  "last_rotated": "2024-01-01"
}
```

Biztonsági Szabályok

1. **Write-Only UI:** A Frontend csak *írhat* a Vault-ba (jelszó megadása). Soha nem olvashatja vissza a jelszót.
2. **Ephemeral Access:** A Scraper Engine csak a futás idejére kap hozzáférést a konkrét tenant_id titkaihoz.
3. **Audit Logs:** Minden hozzáférés (olvasás/írás) naplózva van (ki, mikor, melyik titkot).

---

2. Web Scraper Engine & Automation Layer

A külső weboldalak (pl. NAV, közműszolgáltatók) adatainak kinyerésére szolgáló réteg.

Technológiai Stack

- **Core:** Python + Playwright (könnyebb karbantarthatóság és modernebb mint a Selenium).
- **Browser:** Chromium (Headless).

Izolációs Stratégia (VÉDELEM)

Mivel a böngészők idegen kódot (JS) futtatnak, és a céloldalak kompromittálódhatnak, a Scraper Engine-t karanténban kell tartani.

1. **Ephemeral Containers:** Minden scrape job (pl. "Telekom számla letöltés") egy teljesen új, tiszta Docker konténerben indul el.
    - A munka végeztével a konténer megsemmisül.
    - Nincs perzisztens state, így malware nem tud megmaradni.
2. **Network Policies (Egress):**
    - **Allow:** *.telekom.hu, *.google.com, *.szamlazz.hu (Whitelist alapú).
    - **Block:** Minden más internetes forgalom.
    - **Internal:** Csak a Message Queue (job request) és az Object Store (PDF feltöltés) felé kommunikálhat. Az adatbázist NEM érheti el közvetlenül.

Anti-Bot & Megbízhatóság

- **Stealth Mode:** playwright-stealth plugin használata a bot detektálás elkerülésére.
- **Retry Logic:** Exponential backoff hálózati hibák vagy karbantartás esetén.
- **Human Simulation:** Véletlenszerű késleltetések és egérmozgás emuláció.
- **Session Handling:** Sikeres bejelentkezés után a cookie-kat (ha biztonságos) rövid ideig Redis-ben tároljuk a következő futás gyorsítására, de a Vault mesterjelszó nélkül ezek értéktelenek.

CAPTCHA Kezelés

- **Szolgáltató:** 3rd party CAPTCHA solving service (pl. 2Captcha / Anti-Captcha) integráció API-n keresztül.
- **Emberi beavatkozás:** Opcionálisan, ha az automata nem boldogul, a job "Failed / User Action Required" státuszba kerül, és értesíti az ügyfelet.