Network Architecture Overview

Key Components:

- Client-side (User Interaction) - Web Extension (Chrome).
- Backend Infrastructure - Servers, API layer, AI models.
- Data Storage - Databases for scraped data.
- AI Integration - For data validation, enrichment, and lead scoring.
- Payment Gateway - For API subscriptions, licensing, and monetization.
- External Integrations - 3rd party APIs like Clearbit, OpenAI, Google Vertex, etc.

Client-Side (User Interaction)

1. Web Scraping Chrome Extension:
   - Purpose: Extract data (emails, phone numbers, etc.) from websites in real time.
   - Components: Browser API, Content Script, Popup UI.
   - Data Flow: User visits a website -> Data is collected -> Sent to Backend API -> Processed by AI -> Stored in database.

Backend Infrastructure (API and Business Logic)

1. API Server Layer:
   - Purpose: Interface between frontend and backend.
   - Technology Stack: Node.js / Python (FastAPI / Flask), GraphQL, NGINX/API Gateway.
   - Data Flow: Chrome Extension -> REST API -> AI/Database -> Response back to UI.
   - Key Services: Authentication, API Rate Limiting.

AI Integration (Data Processing & Validation)

1. AI Layer (GPT, Clearbit, Vertex, etc.):
   - Purpose: Data validation, enrichment, and lead scoring.
   - Technology: OpenAI API, Clearbit API, Google Vertex AI.
   - Data Flow: Raw scraped data -> AI models process -> Validate/Enrich/Score -> Return enhanced data.

Data Storage & Database

1. Database Layer:

- Purpose: Store scraped, validated, and enriched data.
- Technology: SQL (PostgreSQL, MySQL), NoSQL (MongoDB), Data Warehouses (BigQuery).
- Data Flow: Data stored -> Queried when needed.

Payment Gateway Integration

1. Subscription & Billing:

- Purpose: Manage API access and handle payments.
- Technology: Stripe/PayPal for payments, API key generation for access control.
- Data Flow: Subscription -> Payment processed -> API key generated.

External Integrations (CRM, Email Services, 3rd Party APIs)

1. CRM Integrations (Lead Management):

- Purpose: Sync leads with CRM tools (Salesforce, HubSpot, etc.).
- Technology: Zapier, Webhooks.
- External APIs: Clearbit for data enrichment, Google Vertex AI for lead classification.

Scalability & Security Considerations

- Load balancers for handling traffic spikes.
- Microservices for modular architecture.
- Auto-scaling (AWS/GCP) for dynamic resource allocation.
- SSL Encryption, JWT authentication, API Rate Limiting for security.

Conclusion

This architecture ensures a scalable, secure, and AI-powered web scraping tool optimized for lead generation and monetization.