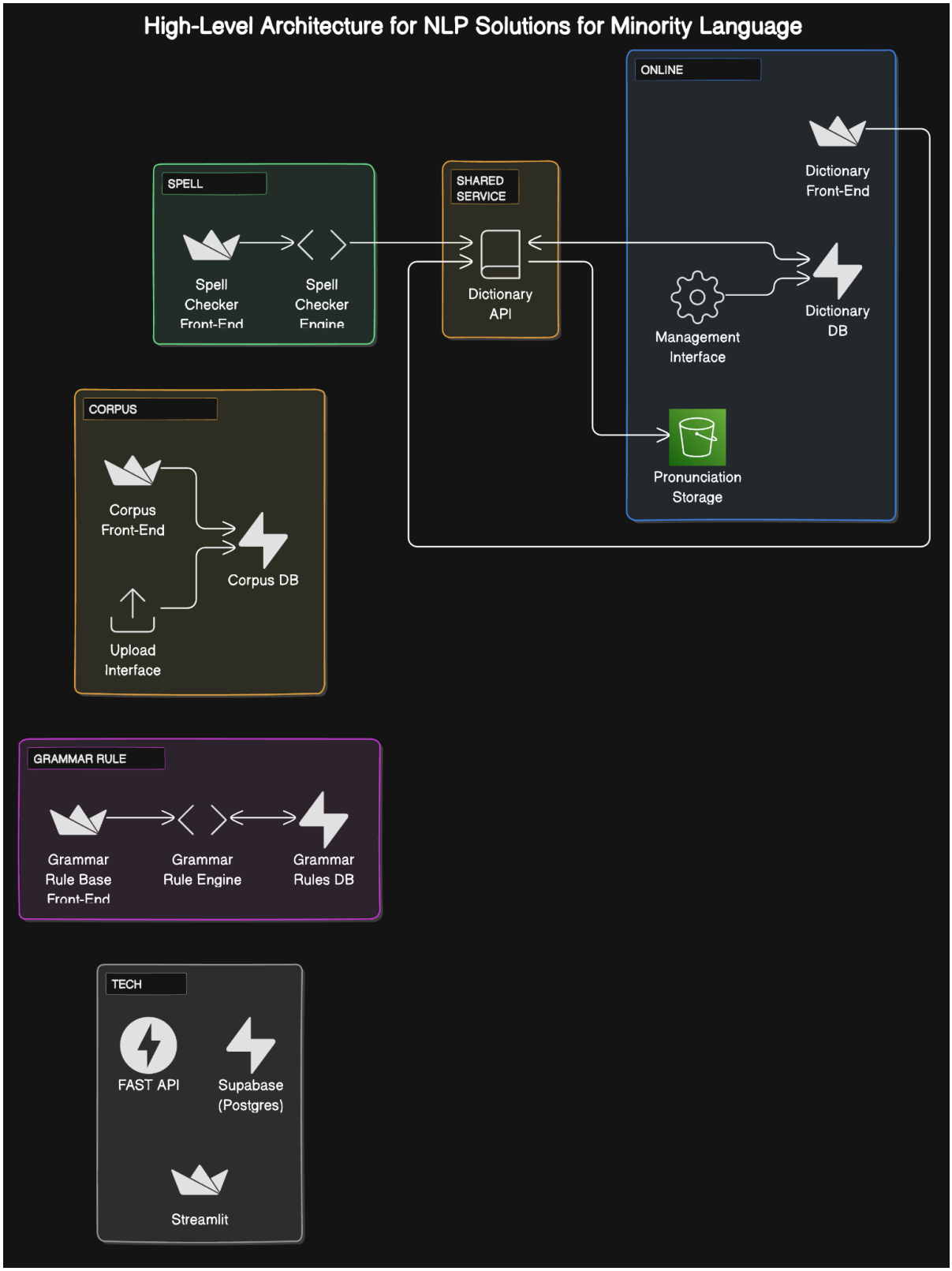
NLP Solutions for a Minority Language



1. Estimates per Work Package

Work Package 1: Online Dictionary

- **Effort Estimate**: 4-5 days
- **Cost**: 1000usd
- **Tasks**:
    - Develop UI, relational database, API endpoints, and bulk upload functionality.
- **Assumptions**: Data is pre-processed and ready for use.
- **Tools**: Supabase (self-hosted), Streamlit.

Work Package 2: Spell Checker

- **Effort Estimate**: 4-5 days
- **Cost**: 1000usd
- **Tasks**:
    - Tokenization, lemmatization, spell-checking engine, UI.
- **Assumptions**: The rule-based approach suffices for spelling corrections.
- **Tools**: SpaCy/ NLTK/ Custom Library, Streamlit.

Work Package 3: Corpus Collection

- **Effort Estimate**: 2-3 days
- **Cost**: 600usd
- **Tasks**:
    - Build UI for search/upload, and implement metadata handling.
- **Assumptions**: Corpus resources are readily available.
- **Tools**: Supabase.

Work Package 4: Grammar Rule Base

- **Effort Estimate**: 2-3 days
- **Cost**: 600ud
- **Tasks**:
    - UI for managing rules, and rule engine development.
- **Assumptions**: Rules are pre-defined.
- **Tools**: PostgreSQL, Streamlit.

2. Technical Recommendations & Assumptions

- **Supabase** (self-hosted) for cost-effective database and API management.
- **SpaCy/ NLTK/ Custom Library** for NLP tasks with custom tokenization and spell-checking.
- **Streamlit** for front-end development.
- **PostgreSQL** for storing dictionary, grammar, and corpus data.

Assumptions: Good data quality, minimal dialect complexity, and no significant delays in sourcing resources. Given these conditions, the project timeline of 2-3 weeks is feasible.