Regulatory Grade Company KYB & Financial Risk Assessment – 1-Page Summary

Overall Approach

The solution is an end-to-end Company KYB onboarding and financial risk assessment system built entirely within a private Azure subscription. It consists of a **FastAPI backend** (with JWT authentication securing all endpoints), a **Next.js frontend** for human-in-the-loop review, and a **PostgreSQL database** hosted on Azure. File uploads are stored in **Azure Blob Storage** (minimum tier), and all sensitive connections and credentials are retrieved securely from **Azure Key Vault**. The system emphasizes **explainability, auditability, and regulatory-aligned risk handling**, with deterministic, rule-based decision logic.

Document Ingestion & Classification

- Supports bulk uploads of synthetic/company documents.
- Automatic classification: Trade License, MOA/AOA, Board Resolution, IDs, Bank Letters, VAT/TRN, Financial Statements using deterministic rules and regex patterns.
- Extracts language, issue dates, and expiry dates.
- Flags unsupported or unrelated files and missing mandatory documents.
- Confidence scores assigned to all classifications.

Data Extraction & Financial Indicators

- Extracts and normalizes legal identity, jurisdiction, legal form, license authority, shareholders, UBOs, authorized signatories.
- Extracts high-level financial indicators: revenue, net profit/loss, total assets/liabilities, audit status.
- Conflicting/missing data flagged; unrelated files labeled.
- All fields traceable to source documents with confidence scores.

Financial Health Risk Scoring

- Deterministic, rule-based scoring system.
- Numeric score (0–100), risk band (Low/Medium/High), risk drivers listed.
- Conservative defaults applied for missing/ambiguous data.
- Example drivers: negative profitability, liabilities > assets, missing/unaudited statements.

Review & Confirm Interface

- Next.js frontend displays extracted data, risk scores, and drivers.
- Manual edits allowed with clear distinction from automated extraction.
- Regulatory confirmation statement included.

Auditability & Logging

- Structured logs for ingestion, extraction, risk computation, and manual edits.
- Captures old/new values, user actions, and company ID.
- Logs stored in `.log` files with sensitive data masked.
- Actions fully traceable and auditable.

Trade-offs and Assumptions

- Deterministic rules prioritized for explainability over ML.
- Minimal Azure tiers to reduce cost.
- Synthetic/public documents used to avoid PII.
- Designed for short-term technical evaluation; scalability optimizations not implemented.

Known Limitations

- Extraction/classification rely on deterministic rules; edge cases require manual review.
- Financial scoring is high-level; not a substitute for full audit or regulatory review.

High-Level Azure Cost Estimate

- **Blob Storage (Hot Tier, minimal usage):** ~$5/month
- **Azure PostgreSQL (Basic Tier):** ~$15–20/month
- **App Service / VM (FastAPI & Next.js):** ~$20/month
- **Azure Key Vault (minimal usage):** <$5/month
- **Total:** <$50/month (short-term exercise)