Project Performance Phase

Date	18 February 2026
Team ID	LTVIP2026TMIDS81723
Project Name	Civil Engineering Insight Studio
Maximum Marks	5 Marks

Project Performance Specifications

The following performance criteria are designed to ensure that the AI-powered audit process does not become a bottleneck at the construction site.

Metric	**Requirement**	**Design Consideration / Technology**
Inference Speed	AI analysis of structural members must be completed within **15–30 seconds** of image upload. +2	Use of **Google Gemini 2.5 Flash**, which is optimized for low-latency multimodal reasoning. +1
Response Time	The web dashboard should load and become interactive in under **3 seconds**. +1	Implementation of a **CDN (Content Delivery Network)** to serve static assets and UI components. +2
Concurrent Users	The system must support at least **50 concurrent site audits** without a degradation in processing speed. +1	**Scalable Architecture** using a 3-tier model and cloud-managed load balancers to distribute traffic. +2
Data Throughput	High-resolution images (up to 10MB) must be uploaded and validated in under **5 seconds** on a standard 4G connection. +1	Use of **Google Cloud Storage (GCS)** for fast, parallelized file ingestion. +1

Metric	Requirement	Design Consideration / Technology
Report Generation	Automated PDF/Markdown reports must be compiled and ready for download within **10 seconds** of analysis completion. +1	Background task processing using **Python** logic to separate report compilation from the main UI thread. +2

Implementation Strategy for High Performance

- To meet the **NFR-4 (Performance)** requirement , the project utilizes the following strategies as outlined in the **Technology Stack**:
- **Use of Cache**: Implementing **Redis Cache** to store frequently accessed project metadata, reducing the number of direct hits to the **PostgreSQL** database.
- **Image Optimization**: Before sending photos to the **Gemini API**, the **Python** backend performs client-side or edge-server resizing to reduce payload size without losing structural detail.
- **Asynchronous Processing**: The UI remains responsive while the heavy "Visual Reasoning" takes place in the background, providing the user with a real-time progress bar.
- **Distributed Infrastructure**: The application is deployed on **Google Cloud Platform (GCP)**, allowing for regional edge locations that bring the service closer to the field engineer's physical location.