

Dynamic Dashboard for Indian Higher-Ed Data & NAAC Accreditation Analytics

End-to-end technical roadmap — modules, front end, back end, algorithms and analytics

1. Vision & Scope

A single, self-updating dashboard that ingests data from every Higher Education Institution (HEI) in India — Universities, Colleges, IITs/IIMs/NITs, Standalone Institutions — and continuously scores them against NAAC's seven-criterion framework, producing predicted grades (A++ to C), gap analysis, peer benchmarking, and drill-downs from national → state → district → institution → department level.

2. Authoritative Data Sources

Source	What it provides	Access
AISHE (MoE)	Enrolment, faculty, programmes, infrastructure	Annual CSV / API
NAAC public portal	SSR, AQAR, IIQA, grades, CGPA scores	HTML scrape / PDF
NIRF (nirfindia.org)	Rankings, TLR, RPC, GO, OI, Perception scores	JSON / PDF
UGC / AICTE	Recognition status, approvals, sanctioned intake	REST / CSV
Vidya Samiksha Kendra	Real-time HEI performance feed	API (OAuth)
Scopus / Web of Science	Publications, citations, h-index	API (keyed)
Institution ERPs	Attendance, results, placements, finance	SFTP / REST
Samarth DigiLocker	eGov, Verified degrees, student records	API (OAuth)

3. NAAC Framework Mapped to Metrics

Criterion (Weight)	Key Quantitative Indicators (KIs)
C1 Curricular Aspects (100/150)	New programmes, CBCS adoption, syllabus revision cycle, feedback loop
C2 Teaching-Learning & Evaluation (350/200)	Student-teacher ratio, fill rate, PhD faculty %, results, ICT-enabled classrooms
C3 Research, Innovations & Extension (120/250)	Publications, citations, h-index, patents, grants, MoUs, extension activities
C4 Infrastructure & Learning Resources (100)	Classrooms, labs, library spend, e-resources, sports
C5 Student Support & Progression (130)	Scholarships, placements, higher-studies, competitive exam success
C6 Governance, Leadership & Management (100)	Strategic plan, e-governance, faculty development, financial audit
C7 Institutional Values & Best Practices (100)	Gender/DEI, environment, human values, two best practices, institutional distinctiveness

Final grade = $0.7 \times \text{KI (quantitative)}$ + $0.3 \times \text{QIM (peer-reviewed qualitative)}$. The engine must reproduce this weighting on the fly.

4. Product Modules

M1 · Data Ingestion Hub

Connectors for AISHE, NAAC, NIRF, UGC, Scopus, ERP SFTP; scheduled + on-demand pulls; schema registry; PII masking.

M2 · HEI Master Registry

Golden record per institution (AISHE code + PAN + DISE) with de-duplication and hierarchy (Uni → affiliated colleges → depts).

M3 · NAAC Scoring Engine

Rule engine that maps raw metrics to Key-Indicator scores, applies criterion weights, outputs predicted CGPA + Grade.

M4 · Gap & Recommendation Engine

Diff between current score and target grade; prescriptive actions per KI; effort/impact matrix.

M5 · Benchmarking & Peer Analytics

Cluster HEIs by type/size/geo; percentile rank per criterion; radar comparisons.

M6 · Prediction & Forecasting

Grade trajectory, dropout risk, placement forecast, research-output projection.

M7 · Geo & Equity Dashboard

State/district heatmaps; gender, SC/ST/OBC, PwD, rural-urban splits.

M8 · Report Studio

Auto-generate SSR/AQAR-ready PDFs, board presentations, ministry briefs.

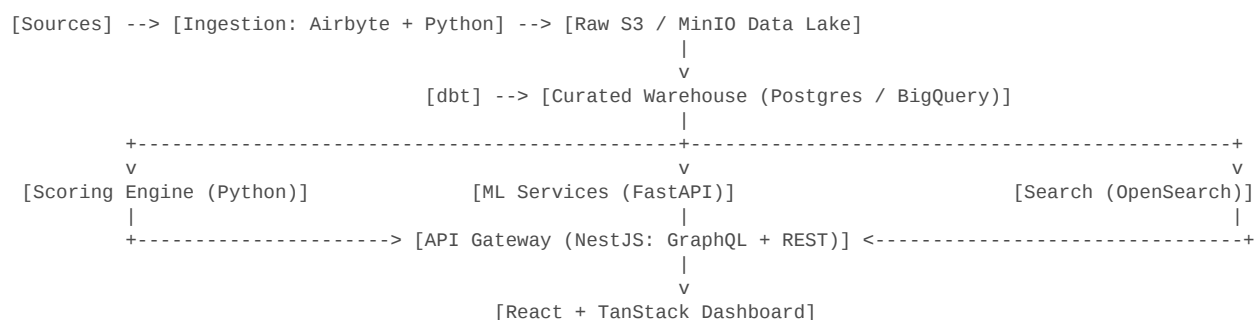
M9 · Admin, RBAC & Audit

Ministry, State, University, College, Auditor, Public roles; every edit versioned.

M10 · Public Portal (read-only)

Citizen-facing rankings, transparency data, RTI-friendly downloads.

5. Reference Architecture



6. Front-End Stack

- **Framework:** React 19 + TanStack Start (SSR, file-based routes) with TypeScript.
- **UI system:** Tailwind v4 + shadcn/ui; semantic tokens for light/dark; role-aware themes (Ministry vs Institution).
- **Data-viz:** Recharts + Visx for radars/heatmaps, deck.gl / MapLibre for India choropleth, AG-Grid for 50k-row tables.

- **State:** TanStack Query for server cache, Zustand for UI state, URL as source of truth for filters.
- **Auth:** OAuth via NAD/DigiLocker for institution logins; SSO (SAML) for ministry users.
- **Accessibility:** WCAG 2.2 AA, Hindi + 10 regional languages via i18next.
- **Performance:** Route-level code-splitting, virtualised lists, server-side pagination, edge caching on stable filters.

7. Back-End Stack

- **API layer:** NestJS (TypeScript) exposing REST + GraphQL; Zod / class-validator for input validation.
- **Warehouse:** Postgres 16 with Citus for sharding, or BigQuery for >10 TB; partition by academic-year.
- **Ingestion:** Airbyte for standard sources, Python (httpx + BeautifulSoup + Playwright) for NAAC/NIRF scraping; Prefect/Airflow orchestrator.
- **Streaming:** Kafka topics for ERP webhooks and real-time attendance/result feeds.
- **Scoring & ML services:** FastAPI micro-services (Python 3.12) containerised on Kubernetes.
- **Storage:** S3/MinIO for raw docs and SSR PDFs; Redis for hot caches; OpenSearch for text/faculty/publication search.
- **Observability:** OpenTelemetry → Grafana, Loki, Tempo; SLOs per endpoint.
- **Security:** Row-level security in Postgres, Vault for secrets, PII tokenisation, DPDP Act 2023 compliance, MeitY audit-ready logs.

8. Algorithms & Analytical Models

8.1 Data quality & entity resolution

- Fuzzy matching (Jaro–Winkler + Levenshtein) + blocking on AISHE code to deduplicate institutions across sources.
- Isolation Forest to flag outlier submissions (e.g. 100% placement claims).
- Bayesian imputation for missing KIs using peer-cluster priors.

8.2 NAAC scoring engine

- Deterministic rule engine encoding NAAC's KI formulas (e.g. Full-Time-Teacher ratio, Publications/Teacher).
- Weighted sum per criterion → CGPA on 4.0 scale; grade bands A++ (≥ 3.76) ... C (1.51–2.00).
- Sensitivity analysis: SHAP-style attribution to show which KI moved the grade.

8.3 Peer clustering & benchmarking

- K-Means / HDBSCAN over standardised features (size, discipline mix, geography, funding).
- Data Envelopment Analysis (DEA) to compute technical-efficiency frontier per cluster.
- Percentile ranking within cluster shown on radar + bullet charts.

8.4 Predictive models

- Gradient-boosted trees (XGBoost / LightGBM) to predict next-cycle NAAC grade — features: 3-yr KI history, faculty churn, R&D grants.
- Survival analysis (Cox PH) for student dropout risk.
- Time-series (Prophet / N-BEATS) for enrolment and placement forecasts.
- LLM-assisted qualitative scoring: extract evidence spans from SSR PDFs, cross-check against KI claims.

8.5 Recommendation engine

- Constraint-optimisation (OR-Tools) — maximise Δ CGPA subject to budget, faculty-hire, and infrastructure constraints.
- Retrieval-augmented action library: past AQARs of A++ peers as playbooks.

9. Analytical Deliverables per HEI

- Predicted NAAC CGPA + Grade with 90% confidence interval.
- Criterion-wise scorecard vs national and cluster median.
- Top-10 gaps with prescriptive actions and estimated CGPA lift.
- Faculty productivity heatmap (publications, citations, patents, projects).
- Student outcome funnel: intake → progression → results → placement / higher-studies.
- Equity view: gender, category, PwD, rural-urban parity indices.
- Financial-health snapshot: unit cost / student, grant utilisation, audit flags.

10. Delivery Roadmap (12 months)

Phase	Weeks	Key Deliverables
P0 Discovery	1–3	Stakeholder workshops (MoE, NAAC, UGC); data-sharing MoUs; KI dictionary v1

P1 Foundation	4–10	Data lake, HEI registry, AISHE + NAAC + NIRF connectors, RBAC skeleton
P2 Scoring MVP	11–18	Rule-based scoring engine, criterion dashboards, 500-college pilot in one state
P3 Analytics	19–26	Peer clustering, DEA, gap engine, report studio, mobile-responsive UI
P4 Predictive	27–34	ML models, LLM SSR parser, recommendation engine, public portal beta
P5 Scale & Certify	35–44	All-India rollout, DPDP audit, ISO 27001, load test @ 50k concurrent
P6 Continuous	45–52	Feedback loop, annual AQAR automation, quarterly model retraining

11. Team & Governance

- Core: 1 PM, 1 Product Designer, 3 Front-end, 4 Back-end, 2 Data Engineers, 2 ML Engineers, 1 QA Lead, 1 DevOps, 1 Security & Compliance.
- Advisory: NAAC assessor, UGC policy expert, Vice-Chancellor panel (4), student-representative panel.
- Governance: Bi-weekly Steering Committee with MoE; open data-standards board; quarterly public transparency report.

12. Risks & Mitigations

- **Data heterogeneity** — enforce a common KI dictionary; auto-map with LLM + human-in-the-loop review.
- **Institutional resistance** — start with opt-in benefits (free gap analysis) before mandate.
- **Privacy (DPDP Act)** — tokenise student PII, publish only aggregate views on the public portal.
- **Gaming of metrics** — anomaly detection, cross-source triangulation, on-site audit sampling.
- **Model drift** — quarterly retraining, drift alerts on feature distributions.

Outcome: a living, evidence-backed system that lets every HEI see — and improve — its NAAC standing continuously, and lets the Ministry steer Indian higher education with real-time data.