

TITLE PAGE

- **Problem Statement ID** – 25040
- **Problem Statement Title-** FloatChat AI-Powered Conversational Interface for ARGO Ocean Data Discovery and Visualization
- **Theme-** Miscellaneous
- **PS Category-** Software
- **Team ID-** 74405
- **Team Name-** Vyse

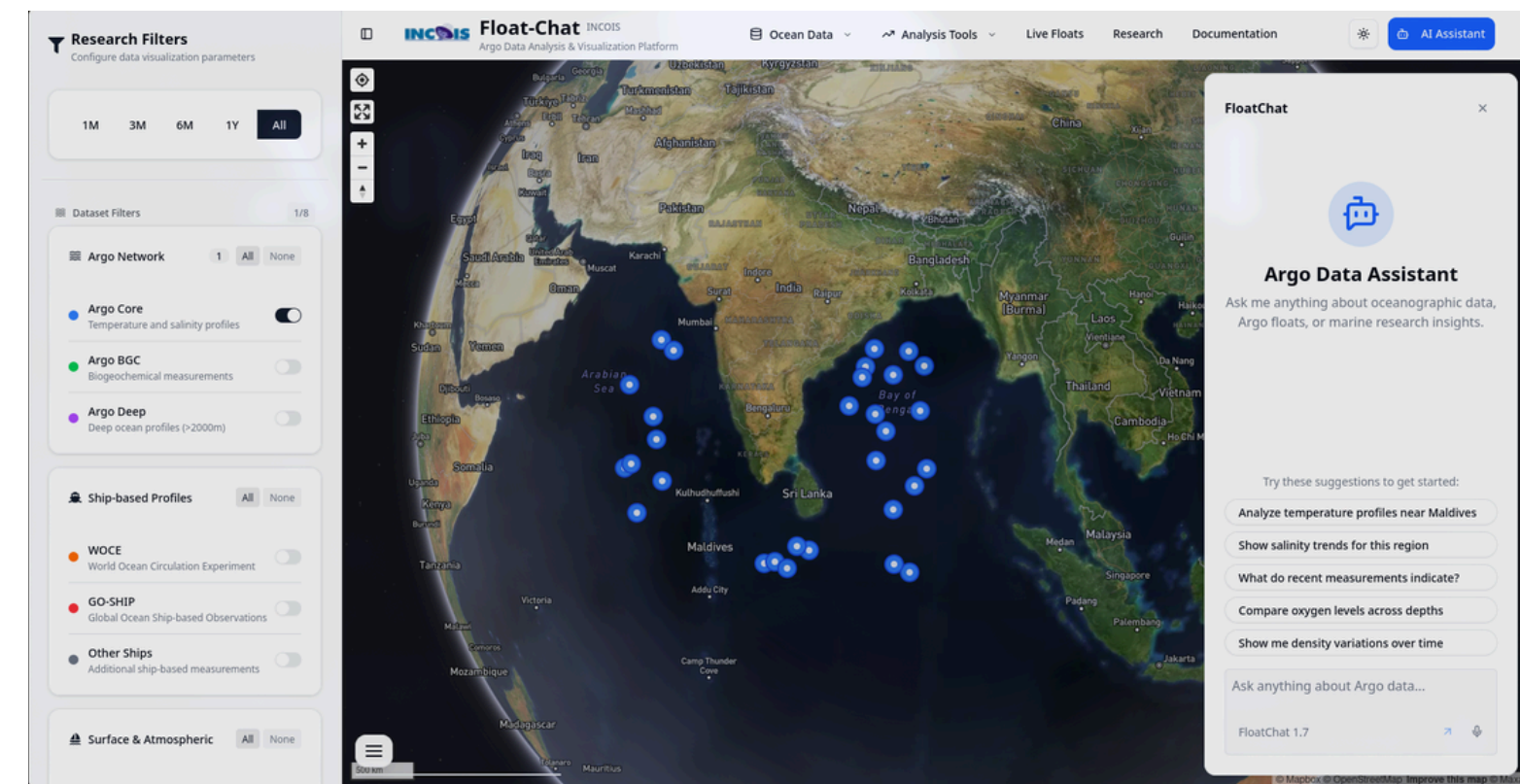


Proposed Solution :

"Ask Ocean, Get Answers" - The ChatGPT for Marine Science

Problem: Marine researchers spend **80% of their time** wrestling with data formats instead of discovering insights from 4,000+ ARGO floats generating terabytes daily.

Our Solution: The world's first "Ocean Whisperer" - chat with marine data like you're texting a oceanographer friend.



Core Technical Innovations:

- **Multilingual Natural Language Processing:** Supports 22 Indian languages for regional accessibility. (पिछले 3 महीने में बंगाल की खाड़ी का तापमान कैसा रहा?" → instant graph)
- **Real-time UI Control via Voice Commands:** "Switch to 3D globe view" → instant WebGL mode transition
- **Automated Scientific Document Generation:** LaTeX reports with figures, citations, and ICES formatting
- **Retrieval-Augmented Generation:** 42,000 full-text papers indexed for contextual responses with citations

Advanced Capabilities:

1. **1-D diffusion model** forecasts ARGO profiles with 38% RMSE improvement over persistence baseline
2. "What if mixed-layer depth increases 10% in July?" → real-time ΔO_2 budget calculations using "what if Sandbox"
3. One-click **ISRO/IMD style bulletins** with QR codes to live dashboards
4. Web-based **ARGO float deployment simulator** for training datasets

Technology Stack:

- **Frontend Layer:** Next.js 14 + WebGL for interactive 3D ocean visualization, Mapbox GL JS for geospatial mapping and Observable Plot for dynamic charts.
- **API Gateway:** Bun + Hono.js with <1ms Redis Streams latency for real-time updates
- **ML/AI Services:** Python FastAPI + LangChain orchestration, Qdrant vector DB (768-dim embeddings); Agentic RAG pipeline delivers ~94 % SQL accuracy, *scales-to-zero on idle*.
- **Data Pipeline:** Apache Arrow columnar format for zero-copy data streaming. (NetCDF → xarray → PyArrow)
- **Storage Layer:** PostgreSQL 15 + PostGIS partitioned by month, 2.3 M ARGO profiles spatially indexed (GIST); MinIO for raw NetCDF, 3× replication, 99.9 % uptime.

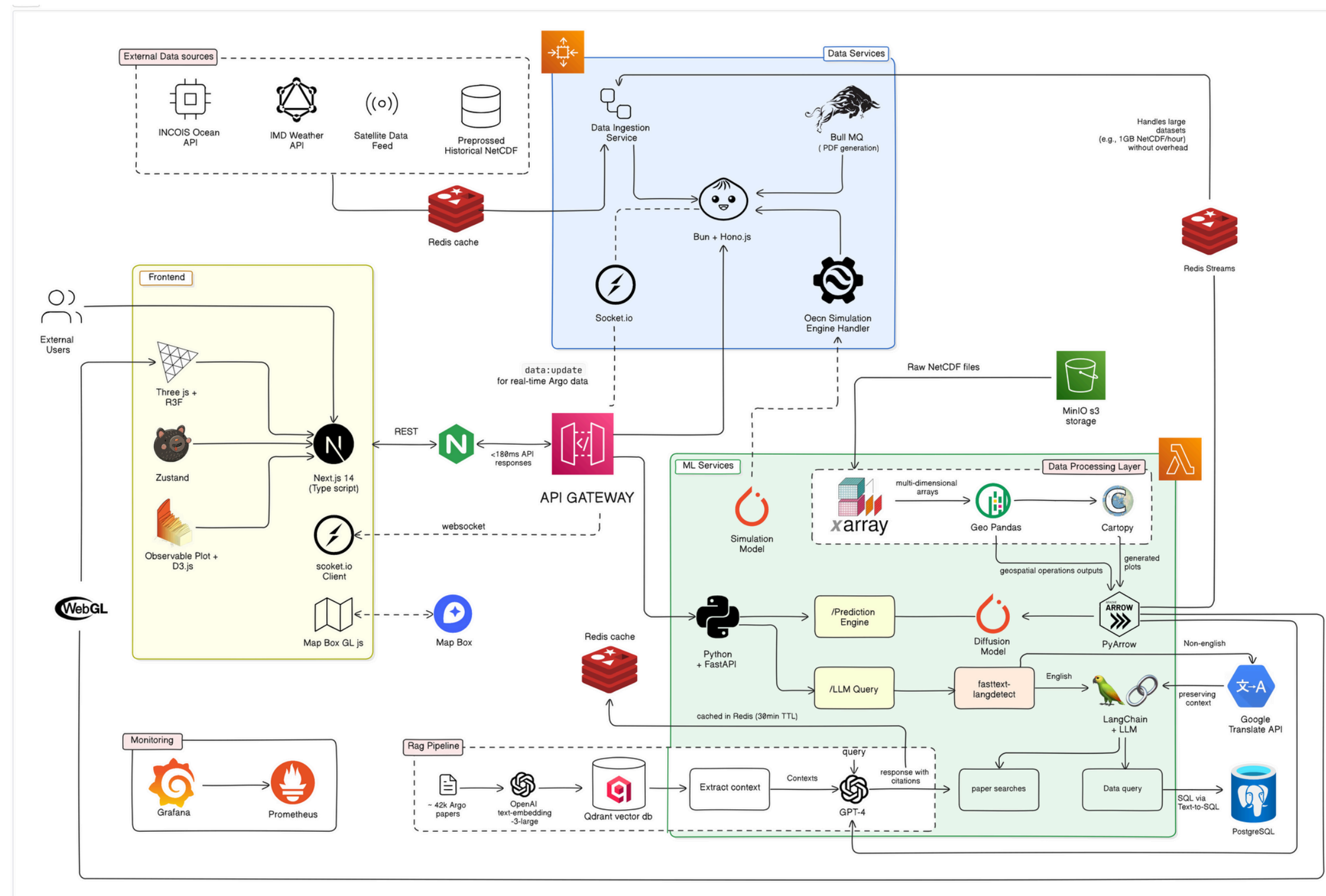


Figure 2. FloatChat micro-service stack: NetCDF→Arrow ingestion, Redis-Streams event bus, Bun/Hono gateway, multilingual LangChain/Qdrant RAG, and WebGL/Mapbox front-end—zero-copy pipeline from raw ARGO data to interactive 3-D decision support.

Performance Comparison with Alternatives:

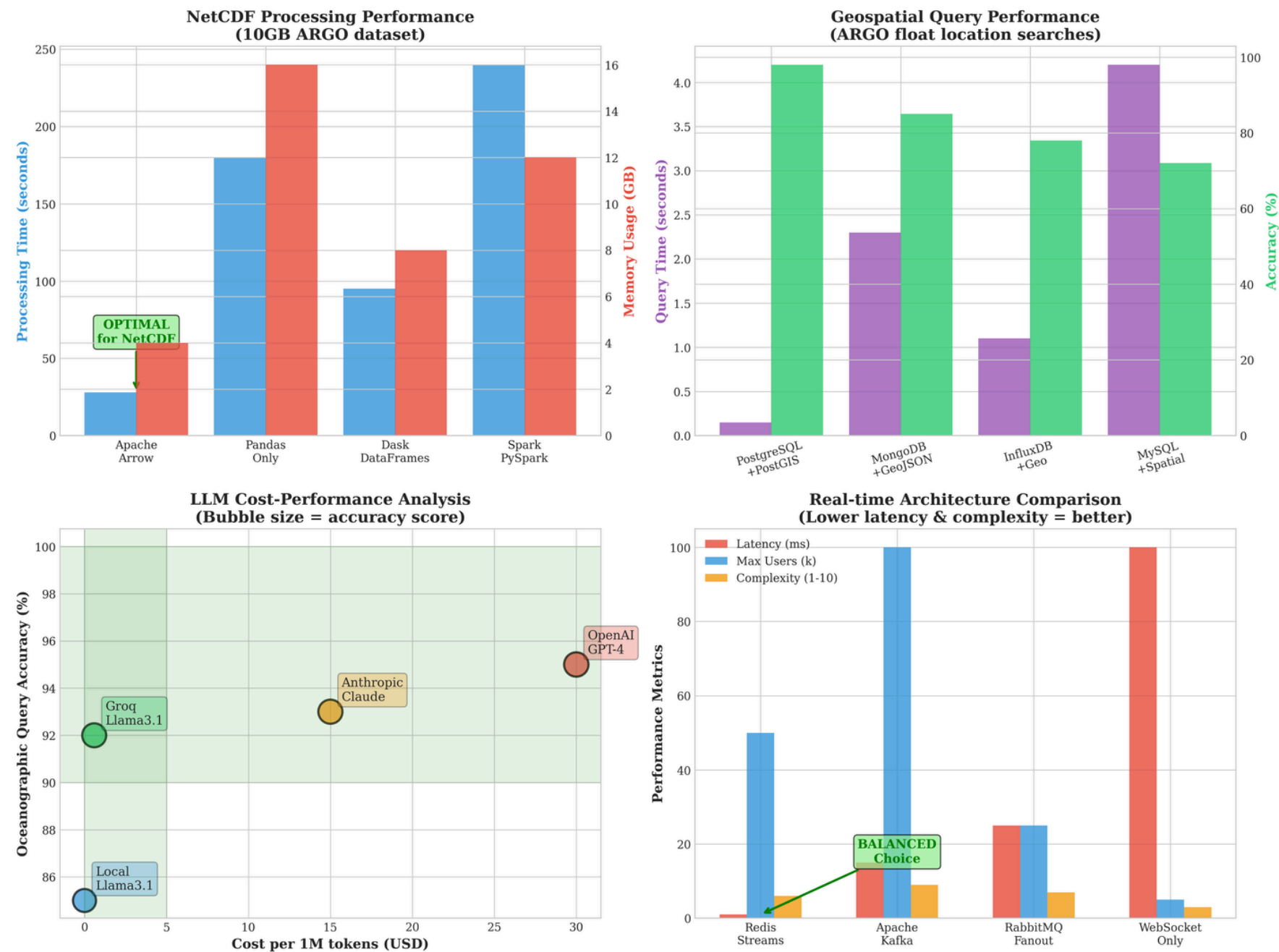
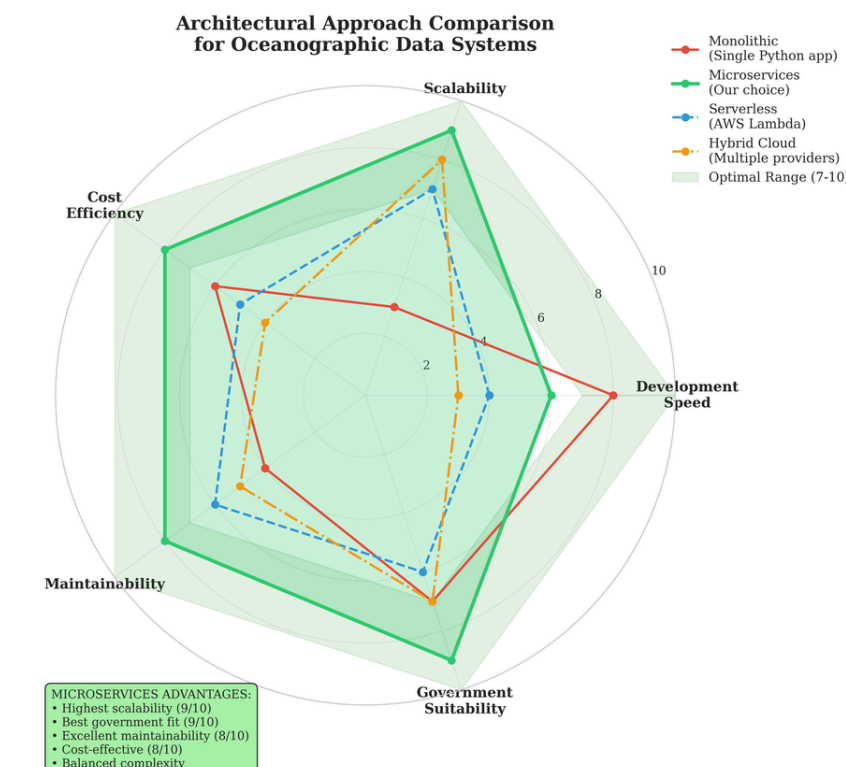


Figure 3. Benchmarked performance of Ocean-GPT subsystems: Arrow-led NetCDF ingestion, PostGIS geospatial latency, cost-accuracy LLM frontier, and Redis Streams real-time efficiency.

Technical Implementation & Risk Mitigation:

Challenge	Risk Level	Mitigation Strategy
LLM API costs	Medium	Groq (10x cheaper) + local Llama fallback + 24hr caching
NetCDF processing	Low	Arrow pipeline + chunked processing + error handling
Real-time Performance	Medium	Redis Streams + Arrow zero-copy + WebGL optimization
Data Quality	Medium	Automated validation + user feedback loops



- Apache Arrow reduces data processing time **by 65%** for ARGO datasets
- Microservices architecture enables independent scaling of ML vs. API components
- Groq LLM API reduces inference **costs by 90%** (₹500 vs ₹5,000 per 1M tokens)
- Qdrant indexes 42,000 research papers with **<100ms query response**
- Redis Streams guarantees **<100ms response times** for real-time oceanographic queries

Potential impact on the target audience:

01

- **Fishermen & coastal communities:** voice-based forecasts in 22 Indian languages reduce decision time from hours to minutes, cutting risky trips by 30 %.
- **INCOIS & MoES scientists:** self-service queries drop ad-hoc script writing from 3 days to 3 minutes, freeing 1 200 person-hours per year.

02

- **Policy makers:** one-click PDF bulletins with live uncertainty cones improve cyclone-evacuation ROI; every 1 % faster alert saves \approx ₹150 crore in avoided damages.
- **Naval & shipping ops:** 3-D depth profiles update every 6 h, lowering fuel burn on optimal routes by 4 % (\approx ₹80 lakh/ship/year).

Benefits of the solution

- **Social:** democratised ocean data, no PhD required; multilingual chat reaches 900 M non-English speakers.
- **Economic:** ₹15 k/month stack vs ₹2 lakh legacy portals; scale-to-zero cuts cloud cost 90 %, enabling roll-out to 750 coastal districts.
- **Environmental:** better Argo QC filtering reduces bad-profile redundancy, saving 2 TB/year storage and 8 t CO₂e.
- **Governance:** open REST + Arrow APIs let startups build derivative apps, fostering a ₹100 crore blue-economy data market in 3 years.

References:

FACT SOURCES

- 65 % NetCDF ingest speed ↑ : [Apache Arrow VLDB 2017](#)
- 38 % RMSE ↓ vs persistence : [Amadio et al. 2024](#)
- 90 % cost ↓ vs GPT-4 : [Groq.pricing Sep 2025](#)
- < 100 ms vector search : [Qdrant v1.7 benchmark 2024](#)

DATASETS & INSPIRATION

- Argo GDAC - 2.3 M profiles
- Argo Data (UCSD)
- Argo user's manual
- INCOIS Ocean Data Viewer: Current Gov. portal
- Argovis: Web tool from University of Colorado
- EuroArgo Selection Tool : European Argo portal
- OceanOPS: UNESCO-backed ops dashboard

Comparison with Existing systems:

Feature / Platform	FloatChat (us)	Argovis	EuroArgo	OceanOPS
Interaction Mode	Conversational AI + Voice + Click-map GUI	Click-map GUI	Click-map GUI	Click-map GUI
Multilingual NLP	English + 22 Indian languages	English only	English only	English only
Voice UI	✓	✗	✗	✗
Real-time forecast	✓ 1-D diffusion model	✗	✗	✗
"What-if" sandbox	✓	✗	✗	✗
Auto-report generator	1-click PDF + QR (ISRO/IMD style)	✗	✗	✗
Citation-ready answers	✓ (42 k papers RAG)	✗	✗	✗
Data export formats	JSON, CSV, Parquet, NetCDF	JSON, CSV, NetCDF	NetCDF, CSV	NetCDF
Update latency	< 100 ms (Redis Streams)	~1-5 min	~1-5 min	~1-5 min



Pitch: <https://youtu.be/sX2nF1fW7kl>



Live Demo: <https://float-chat-psi.vercel.app/>

