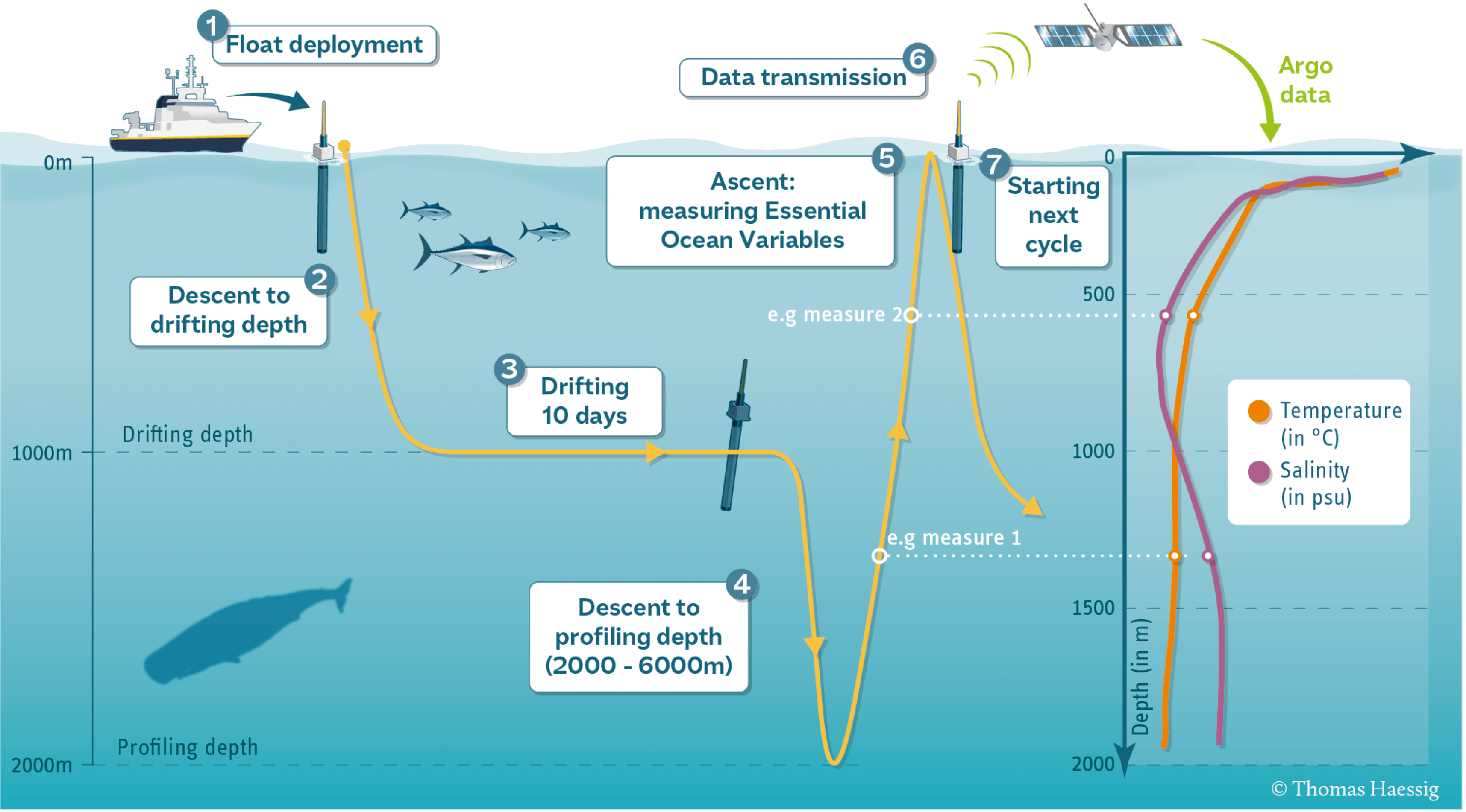
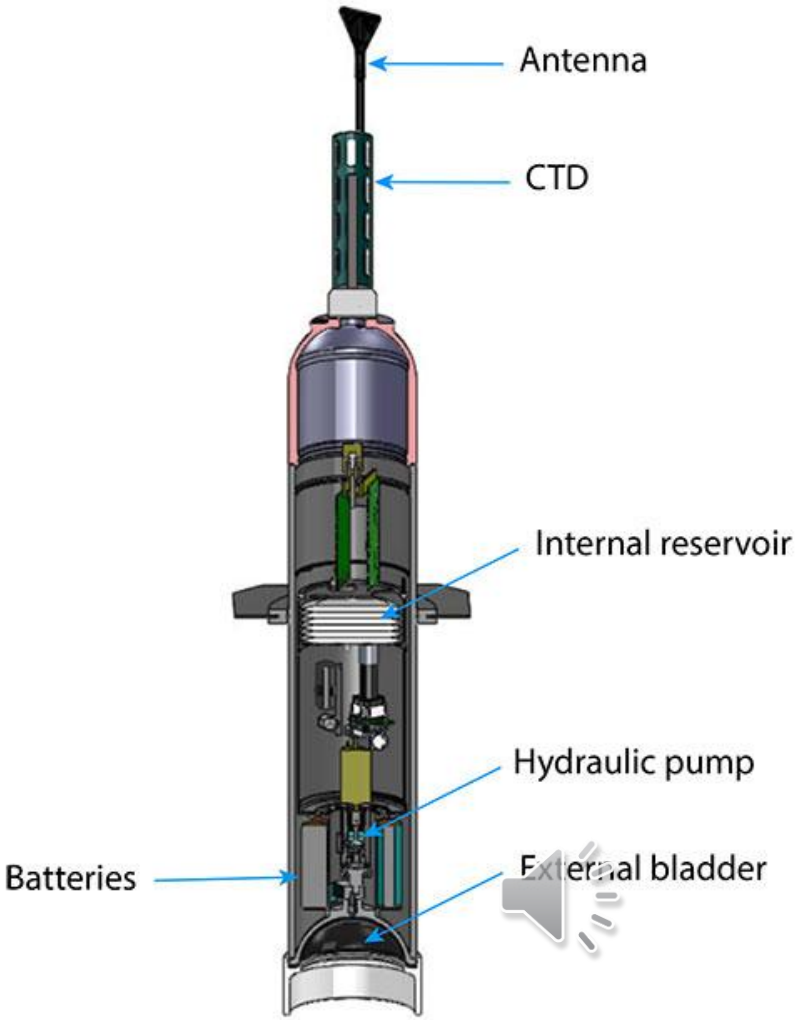
SMART INDIA HACKATHON 2025

SMART INDIA HACKATHON 2025

- **Problem Statement ID –** SIH25040
- **Organization Name** : Ministry of Earth Sciences (MoES)
- **Problem Statement Title:** **ARGO ocean data is vast but locked behind complexity, leaving non-experts without an easy way to explore and gain insights.**
- **Team Name :** Epic Innovators
- **Team Leader Name :** KIRIT P S
- **PS Category :** Software
- **Team Name :** Epic Innovators
- **Theme Name :** AI-Powered Software for Ocean Data Discovery



1 Float deployment
6 Data transmission
Argo data
0m
5 Ascent: measuring Essential Ocean Variables
7 Starting next cycle
2 Descent to drifting depth
3 Drifting 10 days
Drifting depth
1000m
e.g measure 2
e.g measure 1
4 Descent to profiling depth (2000 - 6000m)
Profiling depth
2000m
0
500
1000
1500
2000
Depth (in m)
Temperature (in °C)
Salinity (in psu)
© Thomas Haessig



Antenna
CTD
Internal reservoir
Hydraulic pump
Batteries
External bladder

EPIC INNOVATORS

Jal VaaniAI

SMART INDIA HACKATHON 2025

Proposed Solution:

1. Data Pipeline:

Ingest ARGO NetCDF files and convert them into MongoDB, storing Vector DB and summaries for efficient querying.

2. AI Backend:

Use RAG pipelines with multimodal LLMs to translate natural language queries into database queries, supporting multilingual input and output.

3. Visualization:

Interactive dashboards with maps, depth-time plots, and profile comparisons; enable exporting summaries to ASCII or NetCDF.

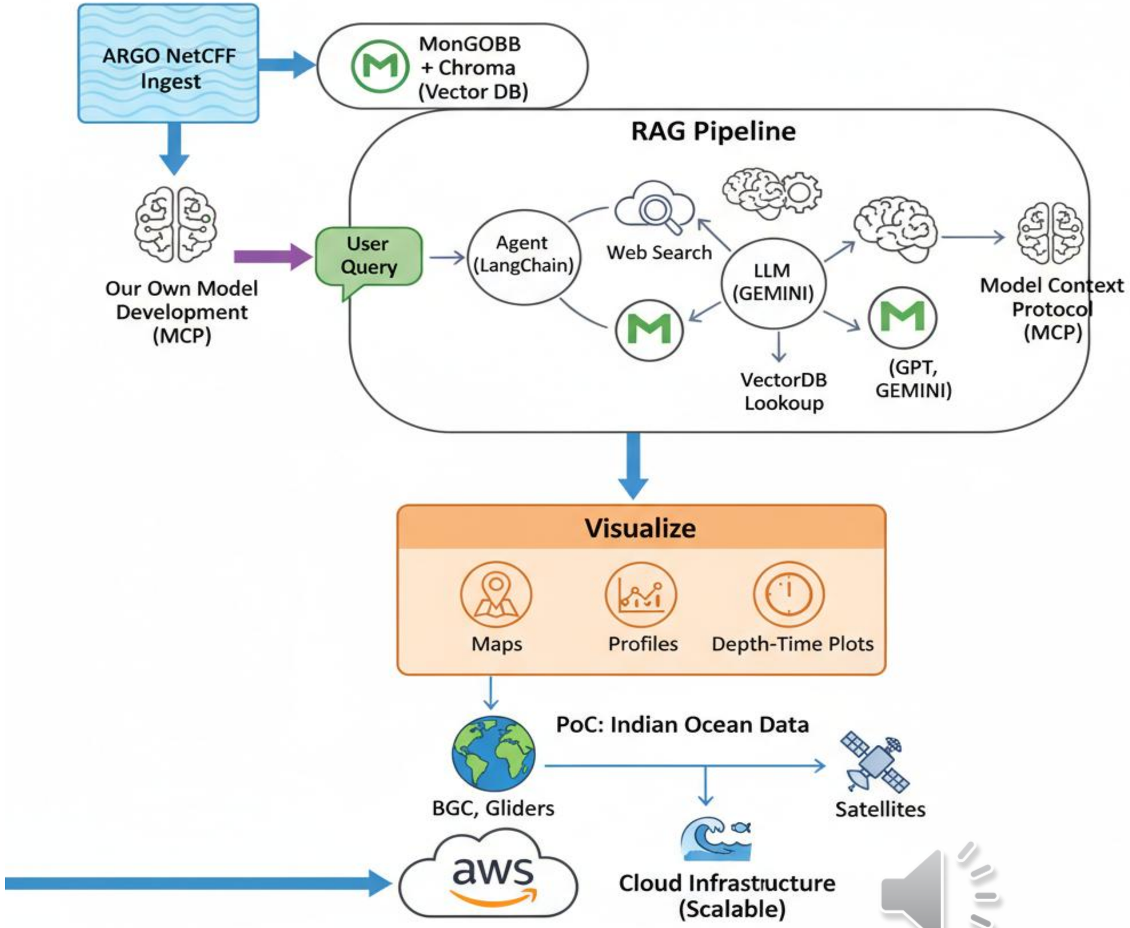
4. Chat Interface:

Conversational AI allows users to query floats, salinity, or BGC parameters, guiding discovery in multiple languages.

5. PoC Scope:

Demonstrate Indian Ocean ARGO dataset functionality, scalable to BGC floats, gliders, satellites, and future datasets.

Link : https://drive.google.com/file/d/1LQ0DB-JRFfgnwdxrhIncKqCjyC5Qii1e/view?usp=sharing



EPIC INNOVATORS

TECHNICAL APPROACH

SMART INDIA HACKATHON 2025

Technologies:

- **Programming & Databases:** Python, MongoDB
- **Vector Search & RAG:** Chroma, LangChain
- **Frontend & Visualization:** Flask, Dash, Plotly, React
- **AI & Cloud:** GPT, GEMINI, Custom LLM(**Developing Model**), AWS
- **Integration Protocol:** MCP (Model Context Protocol

Methodology:)

- **Data Ingestion:** ARGO NetCDF → MongoDB + Vector Database for structured storage.
- **AI Query Engine:** RAG pipeline converts user queries → LLM → Vector DB + Web Search for precise insights.
- **Interactive Visualization:** Maps, profiles, depth-time plots for intuitive exploration.
- **PoC & Scalability:** Indian Ocean ARGO datasets demonstrated; easily extendable to BGC floats, gliders, and satellite data.

EPIC INNOVATORS

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2025

Feasibility & Scalability:

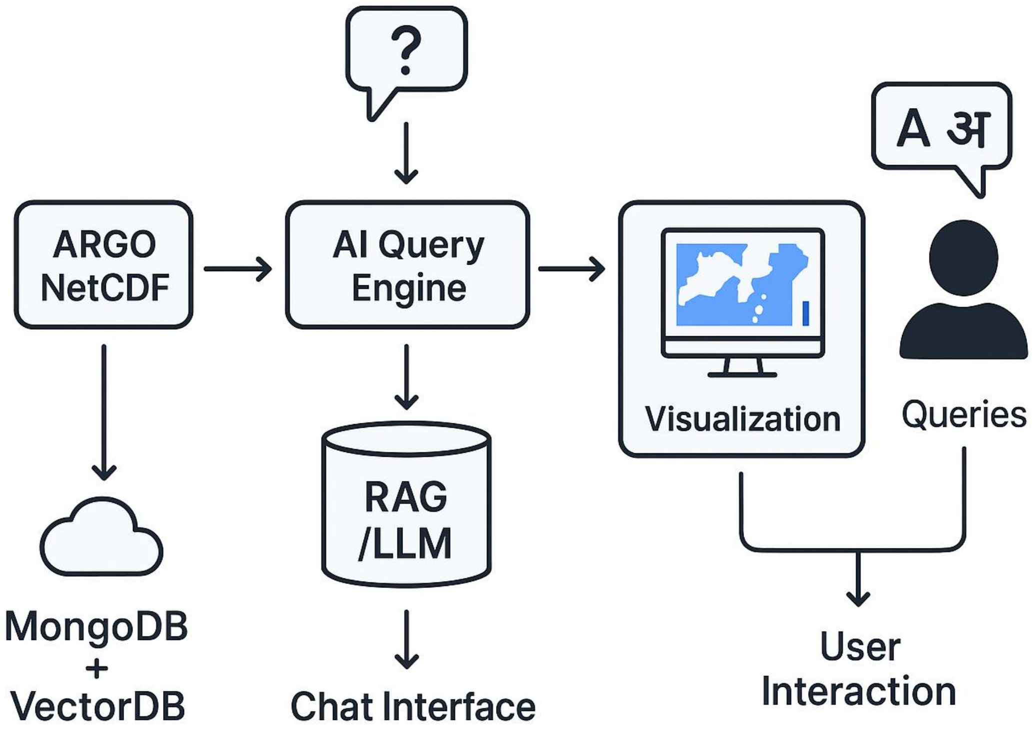
- Fully feasible using **Python, SQL, LLMs, interactive dashboards, and cloud infrastructure**.
- **Hybrid scalable databases** (MongoDB + VectorDB) enable efficient handling of large ARGO datasets.

Challenges & Risks:

- Managing **massive data volumes**, ensuring **real-time updates**, and driving **user adoption**.
- Risks include **misinterpretation of scientific queries** and **high infrastructure costs**.

Strategies & User Support:

- Optimize storage, deploy **domain-tuned LLMs**, and leverage **cloud GPU scaling**.
- Provide a **simple, multilingual UI with guided workflows** for effortless use by non-experts.



EPIC INNOVATORS

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2025

Easy Insights: Data-driven tools for Researchers & Policymakers.

Economic Growth: Boosts Fisheries, Shipping & Renewable Energy.

Environmental Health: Enhances Climate & Marine Ecosystem Monitoring.

Social Equity: Democratizes Access, Boosts Awareness & Education.

Scientific Breakthroughs: Accelerates Oceanographic Research & Innovation.

Global Resilience: Strengthens Climate Resilience & Sustainable Development.

EPIC INNOVATORS

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2025

Argo Data Anomaly Detection with Transformer Models – https://www.sciencedirect.com/science/article/pii/S1385110124000169
Machine Learning Quality Control for Argo Float Profiles – https://www.sciencedirect.com/science/article/pii/S1674283422001751
OceanGPT: Large Language Model for Ocean Science Tasks – https://arxiv.org/html/2310.02031v6
AI Language Models for Marine Environmental Policy – https://www.nature.com/articles/s44183-025-00132-7
Machine Learning with BGC-Argo for Biogeochemical Model Assessment – https://bg.copernicus.org/articles/20/1405/2023/
NetCDF-Based Marine Environment Data Visualization using Virtual Earth – https://www.scientific.net/AMR.518-523.5719
USGS Oceanographic Time-Series NetCDF Documentation – https://pubs.usgs.gov/of/2007/1194/netcdf.html
World Ocean Circulation Experiment (WOCE) NetCDF Format – https://www.bodc.ac.uk/data/hosted_data_systems/sea_level/international/woce_netcdf.html
IDEA: AI Assistant for Geoscience Data Exploration – https://www.eurekalert.org/news-releases/1094334
Klarety: Geospatial Data Analysis with Conversational AI – https://klarety.ai/features/geospatial-ai-analysis
Esri Geospatial AI and Machine Learning – https://www.esri.com/en-us/geospatial-artificial-intelligence/overview
Anthropic's Official MCP Introduction – https://www.anthropic.com/news/model-context-protocol
Model Context Protocol Official Documentation – https://modelcontextprotocol.io
Moveworks Guide to Model Context Protocol – https://www.moveworks.com/us/en/resources/blog/model-context-protocol-mcp-explained
World Ocean Database - NOAA's Largest Ocean Profile Collection – https://www.ncei.noaa.gov/products/world-ocean-database
argoFloats R Package for Analyzing Argo Data – https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2021.635922/full
Argo Float Data Visualizations and Access Tools – https://argo.ucsd.edu/data/data-visualizations/
GO-BGC Data Access and Visualization – https://www.go-bgc.org/data/access-and-visualization
AI for Ocean Monitoring Special Issue – https://www.sciencedirect.com/special-issue/10FCZL672Q7