

Arc42 Documentation

Navid Gharapanjeh Delvin Bacho

2025-02-07

Table of contents

arc42 Data Science Architecture Template	1
1. Introduction and Goals	1
2. Constraints	1
3. System Scope and Context	2
4. Solution Strategy	2
5. Building Block View	2
6. Runtime View	2
7. Deployment View	2
8. Cross-cutting Concepts	3
9. Architecture Decisions	3
10. Quality Requirements	3
11. Risks and Technical Debt	3
12. Glossary	3

arc42 Data Science Architecture Template

1. Introduction and Goals

- **Objective:** Describe the purpose of the project.
- **Stakeholders:** Identify key roles (Data Scientists, Engineers, Business Users, etc.).
- **Key Performance Indicators (KPIs):** Define success metrics.

2. Constraints

- **Technical constraints:** Hardware, software, cloud providers.
- **Regulatory & Compliance:** GDPR, AI ethics, data security.
- **Operational constraints:** Budget, team skills, deadlines.

3. System Scope and Context

- **Scope:** Define what is included/excluded in the project.
- **Context Diagram:** Illustrate data sources, models, and outputs.

4. Solution Strategy

- **Data pipeline strategy:** Batch vs. real-time.
- **ML model strategy:** Pre-trained vs. custom models.
- **Deployment strategy:** On-premises, cloud, or edge.
- **Scaling strategy:** Horizontal vs. vertical scaling.

5. Building Block View

- **Components Overview:**
 - Data ingestion
 - Data preprocessing
 - Feature engineering
 - Model training
 - Model evaluation
 - Model deployment
- **Component Diagram:** Visualize these blocks and their interactions.

6. Runtime View

- **Data flow:** How data moves through the system.
- **Model serving process:** How predictions are generated and served.
- **Monitoring & logging:** Performance tracking and debugging.

7. Deployment View

- **Infrastructure choices:** Cloud services, local servers.
- **CI/CD pipeline for ML models.**
- **Containerization strategy:** Docker, Kubernetes.
- **Versioning & rollback mechanisms.**

8. Cross-cutting Concepts

- Security & Access Control.
- Data Governance & Lineage.
- MLOps Best Practices.
- Bias & Fairness Considerations.

9. Architecture Decisions

- ADR (Architecture Decision Records)
 - Choice of model framework
 - Database selection
 - API design choices
 - Model versioning policy

10. Quality Requirements

- Model Performance Metrics.
- Scalability & Maintainability.
- Explainability & Interpretability.
- Fault tolerance & disaster recovery.

11. Risks and Technical Debt

- Potential biases in data and models.
- Computational cost and resource constraints.
- Integration challenges with existing systems.

12. Glossary

- Define key terms related to ML, AI, and MLOps.

This template provides a structured way to document your data science and ML project architecture while integrating MLOps and software engineering best practices. Let me know if you'd like adjustments!