AREA DI RICERCA SCIENTIFICA E TECNOLOGICA DI TRIESTE- AREA SCIENCE PARK

SELEZIONE PUBBLICA, PER TITOLI ED ESAME - COLLOQUIO, PER L'ASSUNZIONE CON CONTRATTO DI LAVORO A TEMPO PIENO E DETERMINATO DI N. 1 TECNOLOGO DI III LIVELLO PROFESSIONALE DEL C.C.N.L. ISTRUZIONE E RICERCA. PROFILO: SPECIALISTA IN DATA MANAGEMENT - SVILUPPO MODELLI DATI IMPIANTI ENERGETICI.

SETTORE TECNOLOGICO: "SUPPORTO ALLA RICERCA ED INNOVAZIONE".

RIFERIMENTO BANDO TD5/2025.

Prova orale – 19 dicembre 2025

allegato n. 1 al verbale di data 19 dicembre 2025

DOMANDE

Domanda 1 – Digital Twin e modellazione e simulazione

1A – What are the main advantages of a Digital Twin compared to a traditional offline numerical simulation?
Consider aspects such as real-time monitoring, closed-loop feedback, operational optimization, and predictive maintenance.

1B – How is sensor (IoT) data integrated into a Digital Twin model?
Describe the acquisition flow, temporal synchronization, pre-processing, and data-fusion techniques.

1C – Which numerical simulation methods (e.g., CFD, FEM, DEM) are most suitable for reproducing the mechanical phenomena of an industrial plant, and why?
Justify the choice based on scale, non-linearity, material types, and real-time requirements.

1D – How do calibration and validation of a numerical model affect the fidelity of a Digital Twin?
Explain calibration procedures (parameters, boundary conditions), experimental benchmarks, and error metrics.

1E – What challenges are associated with the computational scalability of a Digital Twin in an edge-cloud hybrid environment?
Consider latency, load balancing, data compression, and distributed-computing architectures.

1F – How can machine-learning techniques be used within a Digital Twin to improve system-behavior predictions?
Provide examples such as surrogate models, reduced-order neural networks, and online learning for continuous model updates.

1G – What security and privacy criteria must be considered when implementing a Digital Twin of a critical infrastructure (e.g., power grid, chemical plant)?
Discuss access management, data-flow encryption, resilience to cyber-attacks, and compliance with regulations (GDPR, IEC 62443, etc.).

1H – How to integrate signals and data from multiple sources?

1I – How can Digital Twins be leveraged for lifecycle management of industrial assets?
Discuss integration with PLM systems, version control of models, and strategies for long-term data retention.

Domanda 2 – AI, machine learning e data management

2A – How do you design a robust data-ingestion strategy for an AI pipeline that must handle both batch and streaming sources?
Consider schema evolution, data validation, fault tolerance, and the role of tools such as Apache Kafka, Spark Structured Streaming, or cloud-native event hubs.

2B – What are the key steps of feature engineering in a production-grade ML workflow, and how can you automate them while preserving interpretability?
Discuss techniques such as automated feature synthesis (e.g., Featuretools), statistical profiling, handling of categorical embeddings, and documentation of feature provenance.

2C – Explain the concept of "continuous training" (or "online learning") in an AI pipeline. What architectural patterns enable models to be updated safely without disrupting downstream services?
Address model versioning, canary releases, shadow-mode inference, and strategies for drift detection.

2D – In the context of MLOps, how do you implement automated model validation that goes beyond traditional accuracy metrics?
Include checks for fairness/bias, robustness to adversarial perturbations, calibration quality, and compliance with domain-specific regulatory constraints.

2E – What is the role of model registries (e.g., MLflow Model Registry, AWS SageMaker Model Registry) in an end-to-end AI pipeline, and how do they interact with CI/CD tooling?
Describe lifecycle stages (Staging → Production → Archived), governance policies, and integration with pipelines such as GitHub Actions, Jenkins, or Azure Pipelines.

2F – How would you architect a scalable inference serving layer that can serve both low-latency online requests and high-throughput batch predictions?
Compare options like model-as-a-service (REST/gRPC), serverless functions, GPU-accelerated serving (TensorRT, Triton Inference Server), and batch-or-queued processing pipelines.

2G – What ethical considerations should be embedded into the design of an AI pipeline from data collection to model deployment, and how can you operationalize them?
Discuss consent management, privacy-preserving techniques (differential privacy, federated learning), bias audits, transparency (model cards, datasheets), and mechanisms for human-in-the-loop oversight.

2H – How do you design a feature store for enterprise-scale ML pipelines?
Cover aspects like real-time feature serving, consistency guarantees, and integration with data lakes.

2I – What are best practices for implementing explainability in deployed ML models?
Discuss SHAP, LIME, counterfactual explanations, and their integration into monitoring dashboards.