

Progetto e Piano di Formazione

Progettazione di un Sistema di Storage e Processing Scalabile di ontologie per la modellazione di dati relativi a materiali

Le attività dell'assegno di ricerca si inseriscono nel contesto del progetto EU H2020 "Digital Ontology-based Modelling Environment for Simulation of materials (SimDOME)". Questo progetto di ricerca prevede di studiare, progettare e realizzare soluzioni innovative, architetturali e tecnologiche per lo storage e il processing di ontologie per la modellazione di materiali.

In un contesto sempre più integrato, il progetto SimDOME si propone di facilitare la transizione dalle soluzioni di simulazione attualmente impiegate, che sono in diversi casi proprietarie, di difficile modificabilità e cablate su specifici domini applicativi, a una soluzione basata su ontologie per arrivare a un ambiente di modellazione e simulazione comune e altamente integrato. All'interno del progetto SimDOME, le attività dell'assegno di ricerca saranno mirate a studiare, progettare e realizzare soluzioni innovative che consentano di supportare sia lo storage che le necessarie parti di processing delle ontologie impiegate per ottenere un supporto scalabile e completamente controllabile per tutto ciò che riguarda la configurazione del supporto stesso e degli algoritmi e i reasoner che andranno ad eseguire al di sopra di tale supporto.

Il progetto SimDOME si propone pertanto di aumentare la competitività delle imprese EU che lavorano in questo settore migliorandone e innovandone i processi produttivi perseguendo un triplice obiettivo:

- superare l'attuale frammentazione verso l'integrazione di strumenti software e di simulazione proprietari e open-source;
- offrire una soluzione di supporto a storage e processing di ontologie che faciliti la manutenibilità (anche remota) e la massima flessibilità, tenendo conto degli ultimi sviluppi in ambito DevOps, come tecnologie a container (ad esempio Docker/Kubernetes);
- abilitare l'integrazione di reasoner per una gestione efficiente delle ontologie facilitando anche processi di materialization.

Il programma di formazione associato alle attività dell'assegno di ricerca prevede le seguenti attività organizzate nel periodo di diciotto mesi.

I semestre

Analisi e valutazione dello stato dell'arte nel settore e delle realizzazioni attualmente in essere negli attori aziendali coinvolti, sia da un punto di vista metodologico/modellistico, che delle architetture e delle piattaforme open-source più attuali e di potenziale interesse per gli stakeholder di SimDOME. Realizzazione di primi prototipi che validino l'architettura integrata definita e le piattaforme open-source individuate al punto precedente, sia in termini di requisiti funzionali che non funzionali. Valutazioni e sperimentazioni del prototipo realizzato a deployment in scenari di supporto a servizi, con possibilità di estensione dinamica senza necessità di blocco (continuità di servizio), stress test per valutazioni quantitative di scalabilità.

II semestre

Discussione e valutazione complessiva dei risultati di implementazione e di performance ottenuti per il prototipo integrato completo, nei semplici scenari identificati per i casi d'uso di progetto; analisi comparativa delle in ambiente distribuito e virtualizzato. Prototipazione e valutazione sperimentale anche di una piccola demo integrata in un vero scenario di produzione/deployment reale di interesse per le altre aziende partner di progetto.

III semestre

Analisi e valutazione dello stato dell'arte nel settore delle piattaforme di processing per big data.

Di interesse per il progetto è la realizzazione di un sistema di storage e di processing scalabile. Quando si parla di scalabilità per un sistema software generalmente si intende la possibilità per quest'ultimo di poter essere modificato e avere quindi la capacità di adattarsi ad un consistente cambiamento delle capacità prestazionali richieste. Allo stato attuale quasi tutti i principali DataBase Management Systems (DBMS) permettono di realizzare una infrastruttura distribuita, garantendo alta scalabilità nelle operazioni di archiviazione e reperimento dei dati. Tra questi, alcuni utilizzano una struttura di dati schemaless, ovvero senza un preciso schema di archiviazione dei dati, permettendo quindi la creazione "al volo" di nuovi campi di dato non precedentemente previsti. Tra le piattaforme che fanno utilizzo di queste tecnologie, le due principali sono senza dubbio MongoDB e Cassandra. Entrambe fanno uso di un database NoSQL per l'archiviazione dei dati, e propongono soluzioni per infrastrutture distribuite con sistemi di load-balancing and failover. Ai fini della scalabilità si procederà quindi all'analisi dei requisiti per la realizzazione di un sistema di processing scalabile per dati di tipo culturale provenienti da fonti eterogenee

Sono inoltre previste attività di gruppo orientate alla predisposizione di materiali e prototipi da presentare all'interno del progetto. Saranno quindi maturate esperienze significative relativamente a redazione di documenti di carattere divulgativo, presentazione di risultati e prototipi.