

**POR FESR SARDEGNA 2007/2013: ASSE VI - LINEA DI ATTIVITÀ 6.1.1.A "PROMOZIONE E SOSTEGNO ALL'ATTIVITÀ DI RSI DEI POLI DI INNOVAZIONE E DEI PROGETTI STRATEGICI": AVVISO PUBBLICO PROGETTI CLUSTER TOP DOWN**

**Titolo Progetto:** DASSIA Data scalable solutions for industrial applications

**Soggetto Attuatore:** CRS4

**Presentazione progetto:**

La produzione di grosse quantità di dati è diventata pervasiva nell'industria e nelle istituzioni pubbliche, si prevede che l'uso intelligente (analisi ed estrapolazione di conoscenza) di tali risorse diventerà nel prossimo futuro una delle principali basi di competizione e crescita.

Gran parte delle informazioni utili per le analisi di mercato è oggi fornita da sistemi di social networking e microblogging. Facebook, ad esempio, già nel 2008 ospitava ben dieci miliardi di immagini, per un totale di circa 1 Petabyte (un milione di miliardi di byte); il volume di traffico generato da Twitter ha recentemente raggiunto i 400 milioni di *tweet* al giorno. Si stima che la dimensione totale dei dati memorizzati a livello mondiale sia dell'ordine di alcuni Zettabyte (mille miliardi di miliardi di byte), un numero confrontabile con quello delle stelle nell'universo osservabile. A questi ci si aspetta che si aggiungeranno miriadi di altre sorgenti di dati legati, ad esempio, a sistemi distribuiti per l'acquisizione di dati per la gestione e controllo di infrastrutture cittadine.

Collezioni di dati di tali dimensioni rendono particolarmente difficoltosa, se non addirittura impossibile, la loro memorizzazione ed elaborazione tramite sistemi tradizionali quali, ad esempio, i database relazionali. Nell'ambito della tecnologia dell'informazione è recentemente diventato d'uso comune riferirsi a tale classe di problemi con la locuzione "big data", che per esteso viene applicata all'intero insieme di tecnologie hardware e software di nuova generazione progettate per risolverli efficacemente. Elemento comune a gran parte delle suddette tecnologie è la capacità di adattarsi dinamicamente al crescere delle dimensioni del problema, tramite l'impiego di modelli di calcolo parallelo progettati per garantire un funzionamento robusto e scalabile (da poche unità a migliaia di macchine) anche su hardware di basso costo. Tra le varie infrastrutture di calcolo con tali caratteristiche, una delle più importanti è senza dubbio Hadoop, una infrastruttura open source ampiamente utilizzata ed il cui sviluppo è direttamente supportato da aziende quali Facebook, Twitter, EBay ed Amazon.

In questo progetto cluster, l'obiettivo è quello di trasferire parte della conoscenza ed esperienza del CRS4 nel campo del Big Data ad un cluster di imprese a forte spinta innovativa.

Nello specifico, il CRS4 intende portare avanti la progettazione e lo sviluppo di un sistema scalabile di back-office per la gestione ed analisi integrata di dati digitali testuali (documenti, posta elettronica,).

Il sistema utilizzerà tecnologie scalabili ETL come hadoop per il processamento dei documenti e la costruzione dei grafi di interrelazioni, tecnologie di database no-sql per la gestione e l'accesso ai dati e tecnologie di scalable analytics come impala per l'analisi. La realizzazione di questo

**Obiettivi e risultati attesi**

La principale ricaduta attesa per le imprese coinvolte è di crescita di professionale. Al termine delle attività le imprese avranno un panorama chiaro dell'offerta tecnologica disponibile e delle potenzialità correlate. La attività di trasferimento tecnologico, partendo dalla descrizione generale delle tematiche legate al Cloud Computing e al trattamento dei Big Data, porterà alla presentazione e conoscenza degli strumenti attualmente disponibili. L'interazione con le aziende sarà strutturata come un processo dinamico volto a raffinare le specifiche del sistema in modo da poter esplorare gli aspetti tecnologici ritenuti da esse di maggior interesse.

**Conclusione:** Giugno 2015

**Contatti:**

CRS4 - Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna - S.r.l. Uninominale  
soggetta all'attività di direzione e coordinamento di Sardegna Ricerche,  
sede legale in 09123 Cagliari, Via Palabanda, 9,  
sede operativa in 09010 Pula, Località Piscina Manna, Edificio 1,  
tel. 0709250-1, fax 0709250216

**Responsabile Scientifico:** Zanetti Gianluigi [gianluigi.zanetti@crs4.it](mailto:gianluigi.zanetti@crs4.it)