

**FORMULARIO  
PROPOSTA PROGETTUALE  
DATI GENERALI**

Destinatari Titolo di Studio	Laurea specialistica /magistrale / Vecchio ordinamento	Idoneità dottorato di ricerca	Specializzazione	Lingua straniera
	Informatica	SI	No	No
	Ingegneria informatica	SI	NO	NO

**Profilo dei destinatari richiesti**

Il candidato da formare è una persona in possesso di una Laurea Magistrale o Specialistica (o Laurea precedente ordinamento quinquennale) in Informatica o Ingegneria Informatica, che abbiamo nel percorso di studi gli esami di Basi di Dati.

**Descrizione dell'articolazione del percorso nel suo complesso**

L'Advanced Systems dispone di un **Centro Ricerche**, accreditato da dieci anni come laboratorio presso l'Albo Nazionale del MIUR e presso l'Albo Regionale della Campania, ed è, pertanto, tra i soggetti ufficialmente abilitati a svolgere attività di ricerca e realizzazione di prodotti innovativi.

**SADAS** è un DWMS (Data Warehouse management System) sviluppato nei laboratori di ricerca di **Advanced Systems** di Casalnuovo di Napoli (NA), specializzato per ottenere grandi performance in termini di tempi di risposta in ambiente Data Warehouse.

SADAS si inserisce inoltre nel contesto delle nuove tecnologie specializzate nell'analisi di grosse moli di dati in ambiente Data Warehouse; SADAS, rispetto ai DBMS tradizionali, si differenzia per l'approccio innovativo, puntando ad una nuova struttura dei dati ed ad algoritmi sofisticati, per ottenere la sostenibilità della gestione dei grandi volumi grazie alla compressione dei dati e grandi performance in termini di tempi di risposta delle interrogazioni utilizzando configurazioni HW di basso profilo e molto economiche. Si inserisce quindi a pieno titolo nel settore della **Green Economy**, imponendosi sul mercato come una soluzione che richiede minime risorse computazionali, che si traducono in un corrispondente risparmio energetico a parità di performance del sistema.

Il settore di intervento è quello delle attività di ricerca miranti all'innovazione delle tecnologie software per l'analisi dei dati in ambiente Data Warehouse e di Business Intelligence, ed al loro trasferimento tecnologico, con riferimento all'evoluzione di metodi e tecniche relativi all'analisi di dati statici in archivi di dimensioni massive.

Gli obiettivi tecnologici che si intende perseguire sono i seguenti:

- *Clustering*

In particolare l'attenzione è posta sullo sviluppo e l'evoluzione delle tecnologie per l'elaborazione parallela dei DBMS utilizzati nelle funzionalità OLAP dei Sistemi di Supporto alle Decisioni (DSS). I DBMS considerati sono quelli per archivi statici del tipo column-based, perché allo stato dell'arte sono quelli che già mostrano le migliori caratteristiche in termini di performance rispetto all'utilizzo definito.

- *Estensione dell'SQL per l'analisi dei dati*

Sebbene molti tipi di operazioni di analisi dei dati possono essere fatte con strumenti di analisi a partire da dati estratti dal data warehouse, per ragioni di efficienza è opportuno estendere il linguaggio SQL con nuove funzionalità per far eseguire alcune operazioni complesse direttamente dal gestore dei dati. Nascono così le estensioni all'operatore di Group By con gli operatori CUBE e ROLLUP.

- *Strutture di memorizzazione*

Per la quantità di dati gestiti, e per il fatto che i dati sono statici, i data warehouse consentono l'uso di strutture di memorizzazione ridondanti e più complesse di quelle usate nei tradizionali DBMS relazionali (RDBMS) al fine di ridurre i tempi di esecuzione delle interrogazioni. Nei RDBMS i dati di una tabella sono di solito memorizzati in un file con record a lunghezza variabile. Questa soluzione risulta essere la migliore per interrogazioni di carattere generale e in presenza di dati che possono essere modificati. Quando i dati sono statici e le interrogazioni sono di analisi dei dati, si possono prendere in considerazione altre tecniche di memorizzazione dei dati di una tabella basate sull'idea dei file trasposti: invece di memorizzare i dati di una tabella con n campi in un solo file (memorizzazione per righe), si memorizzano i valori degli n campi in file separati (**memorizzazione per colonne**).

- *Viste materializzate legate al "Learn by usage"*

Nei DBMS tradizionali le viste si usano frequentemente per vari motivi e in particolare per semplificare la formulazione di interrogazioni particolarmente complesse oppure per consentire la scrittura di interrogazioni che non sarebbero esprimibili senza ricorrere alle viste. Nei tradizionali sistemi relazionali, l'uso delle viste è esplicito nelle interrogazioni che poi, quando è possibile, prima di essere eseguite vengono automaticamente riscritte sostituendo l'espressione che definisce la vista al posto del suo uso prima di passare all'ottimizzazione dell'interrogazione. Questo modo di procedere aumenta infatti le possibilità di ottimizzazione dell'interrogazione e ne riduce drasticamente i tempi di esecuzione. La tecnica delle viste materializzate prevede invece di memorizzare il risultato dell'espressione che la definisce e quindi di alcune tipiche interrogazioni, in particolare quelle con costose operazioni di giunzione, raggruppamento e calcolo di aggregati. Quando occorre eseguire un'interrogazione complessa, è possibile che esista una vista materializzata per la quale alcune o tutte le operazioni più complesse siano già state eseguite e, quindi, usando i dati precedentemente memorizzati sarà possibile trovare il risultato dell'interrogazione con costi molto bassi e con notevoli miglioramenti dei tempi di risposta. Anche in questo caso l'interrogazione formulata dall'utente viene riscritta in modo da utilizzare le viste materializzate, ma, a differenza del caso precedente, ciò avviene in modo automatico da parte del sistema, senza che l'utente faccia riferimento alle viste materializzate. Avendo supposto i dati statici, la materializzazione di viste richiede però la soluzione di due problemi fondamentali e difficili:

- la scelta delle viste da materializzare e degli eventuali indici su di esse da parte dell'amministratore dei dati, o in modo automatico da parte del sistema in base al tipo di uso che viene fatto dei dati (*learn by usage*);
- l'utilizzo delle viste materializzate per eseguire una data interrogazione. E' una scelta delicata e complessa di cui si deve far carico l'ottimizzatore.

- *Ottimizzazione delle interrogazioni su schemi a stella*

Le particolarità dei data warehouse richiedono nuovi accorgimenti durante la fase di ottimizzazione delle interrogazioni. In particolare, l'ottimizzatore per scegliere il miglior piano d'accesso per eseguire un'interrogazione deve tener conto non solo dei modi disponibili per recuperare i dati di

una tabella, dei vari tipi di indici (tradizionali o a bitmap) e delle viste materializzate presenti, ma anche del fatto che l'interrogazione interessa uno schema a stella o a fiocco di neve.

Per il raggiungimento di ciascun obiettivo è prevista una durata complessiva di 3 anni,

#### **Articolazione annuale delle attività previste dal percorso formativo nel suo complesso**

Per ciascun obiettivo sono previste le seguenti fasi:

- formazione, per permettere al soggetto di acquisire conoscenza in merito alle specifiche tematiche e agli strumenti da utilizzare; studio della letteratura in merito alle specifiche tematiche e analisi dello stato dell'arte della tecnologia SADAS e degli altri DWNS presenti sul mercato, per
- in base al risultato della fase di studio, scelta delle tecnologie da usare, progettazione dell'architettura e del prototipo (1 anno)
- implementazione del prototipo per eseguire test specifici (1 anno)

#### **Descrizione delle metodologie e strumenti per la valutazione in itinere e finale dei partecipanti e per il riconoscimento della formazione in impresa**

Il candidato dovrà acquisire padronanza con gli strumenti per l'analisi dei dati in ambiente data warehouse e dei moduli di gestione e interni (ParserSQL, Ottimizzatore, Gestore delle strutture di memorizzazione, Strutture fisiche di memorizzazione) che compongono SADAS.

Il progetto prevede la frequenza di corsi, seminari, stage di studio presso Università e centri di ricerca italiani ed esteri.

Durante e alla fine del percorso formativo il candidato dovrà mostrare la capacità di progettare basi di dati in ambiente data warehouse e di intervenire nello sviluppo del motore di Sadas.

#### **Descrizione dell'organizzazione del gruppo di lavoro dei profili professionali impiegati con particolare riferimento ai tutor formativo e aziendale**

L'organizzazione del gruppo di lavoro coinvolto nel programma è definito all'interno dell'organigramma aziendale con struttura indipendente, in considerazione del valore strategico delle attività di Ricerca.

Il personale specializzato impiegato è, pertanto, costituito dal seguente organico stabile:

Responsabile del Centro: Laureato con esperienza pluridecennale sia tecnica che gestionale di progetti di ricerca realizzati in equipe con Università italiane ed estere.

Nr. 3 Ricercatori Senior: Laureati con esperienza pluridecennale nello sviluppo di progetti di ricerca industriale e precompetitiva

Nr. 2 Ricercatori junior: Laureati con esperienza quinquennale in progetti di ricerca applicata e sviluppo precompetitivo

Nr. 1 tecnico di laboratorio: diplomato con esperienza decennale in progetti di ricerca applicata e sviluppo precompetitivo.

A tale organico viene di volta in volta aggregato idoneo personale tecnico, di 7/8 dipendenti, per completare la compagine operativa.

Per quanto concerne le attività di tutoraggio formativo ed aziendale, già sono state effettuate varie esperienze qualificate nello specifico settore, sia con Università italiane (Federico II, Unisannio, SUN, Roma 3, Pisa) che su progetti che hanno visto il coinvolgimento di Università straniere (Dresda, Bucarest).



### **Sede di svolgimento del tirocinio formativo**

La Sede di svolgimento del tirocinio formativo è il laboratorio di R&S della Advanced Systems sito in Casalnuovo di Napoli alla Via Napoli, 125 – Centro Meridiana

### **Descrizione sintetica degli obiettivi previsti dall'attività di tutoraggio individuale**

Sono previsti risultati:

*di tipo teorico:*

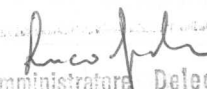
- conoscenza e specializzazione sugli aspetti funzionali e tecnologici legati agli strumenti di Business Intelligence ed alla loro interazione con le moderne tecnologie di analisi delle basi di dati, e con gli operatori specializzati che queste mettono a disposizione;
- capacità di manipolazione dei linguaggi simbolico/formale;
- metodologie di progettazione di un ottimizzatore per un DBMS specializzato per l'analisi dei dati nel data warehouse;
- metodologie di progettazione e programmazione avanzati Object Oriented.
- lo studio e l'analisi delle strutture in memoria che rappresentano un'interrogazione per la l'analisi e lo studio delle viste e dell'estensione del dialetto SQL
- delle dipendenze dei dati statici e degli operatori per queries su di una architettura di tipo cluster e per DBMS per archivi statici del tipo column-based; la definizione di metodologie di distribuzione statica e dinamica dei dati del DBMS column-based sui nodi di un cluster e delle operazioni di calcolo parallelizzabili sugli stessi nodi, con risorse distribuite sia di storage che di calcolo;

*di tipo implementativo relativo all'implementazione di prototipi di strumenti software:*

- per la riscrittura di una query con uso di viste materializzate e test di performance della query riscritta rispetto alla versione originale
- per configurare la distribuzione dei dati sui nodi dell'architettura parallela (configuratore), per effettuare una redistribuzione dinamica dei dati sui nodi del cluster in base all'analisi del carico reale (ristrutturatore) e per distribuire le operazioni computazionali per eseguire le operazioni previste dai piani di accesso (ottimizzatore) su dati distribuiti (per configurazione o ristrutturazione) sullo storage dei diversi nodi del cluster, in modo da avere un bilanciamento del carico elaborativo tra le varie componenti computazionali sui diversi nodi;

*di tipo formativo:*

- con la creazione di specifiche competenze metodologiche e tecnologiche per soluzioni innovative alle problematiche del settore.

  
L'Amministratore Delegato  
( Ing. Lucio Goglia )