



Intel® Parallel Studio

Un caso di studio

Ciro Fiorillo
Lead Software Architect

Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, *Ciro Fiorillo*. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.



ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS





Agenda

09:15 Saluto di benvenuto e apertura lavori (Assintel, Adalta)

09:30 Architettura Parallela: lo sviluppo dell'hardware (Intel)

10:00 Parallel Programming, Today and Tomorrow (Intel)

10:45 *Coffee Break*

11:05 Dal seriale al parallelo – analizzare, creare, implementare, correggere, ottimizzare

Il ciclo di sviluppo utilizzando i software HPC Intel; anteprima della nuove versioni autunno 2010. (Intel)

Come ottimizzare le proprie applicazioni Visual Studio per macchine Multi-Core: Intel Parallel Studio. (C. Fiorillo)

12:30 - 13:30 *Buffet Lunch*

13:30 **Un caso di studio (C. Fiorillo)**

14:15 Parallel Programming Methods and Tools (Intel)

15:00 Ottimizzazione di applicazioni (C. Fiorillo)

16:00 Q&A



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS


assintel
Associazione Nazionale
Imprese ICT

Il caso di studio



- Sviluppato da Advanced Systems Srl
 - Centro Ricerche MURST (2000)
- Precedenti esperienze
 - Infomanager (1982)
 - OverMillion (1990)
 - GAS – Gestione Archivi Storici (1992)
 - Sherlock – antiriciclaggio bancario
- Progetto di ricerca MIUR
 - Università degli Studi di Pisa (Prof. Antonio Albano)
 - Università degli Studi del Sannio (Prof. G. Canfora)
 - Presentato ad EDBT (2006), SEBD (2006), IADIS(2009)
- Industrializzazione: banche, enti, telecom.



Caratteristiche principali



- DBMS Engine per archivi statici
- Non solo column-based
- Strutture aggiuntive
 - Indici B-Tree
 - Indici Bitmap
 - Aggregazioni (cubi)
- “Learn By Usage”
- SQL-99 standard
- Estensioni proprietarie al linguaggio SQL
 - DISTRIBUTE
 - NOT STAR JOIN



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS



Algoritmo di sort

- Utilizzato in
 - Order by
 - Group by
 - Distinct
 - Join

- Ostacoli alle performance
 - Cardinalità dei dati
 - Accesso ai dischi
 - Complessità computazionale degli algoritmi
 - Memoria disponibile e complessità spaziale
 - Tipi di dati



Implementazione sort in Sadas

- Strategie diverse
 - Memoria
 - Numero processori
 - Architettura 32 o 64 bit
- Quicksort + Merge
 - Funzioni di copia e di confronto personalizzate per tipo di dato
 - Accesso ai dischi ridotto al minimo
 - Utilizzo efficiente della memoria disponibile
 - Utilizzo del multithreading solo in presenza di più core
- Utilizzo di OpenMP
- Implementazione multi-thread a basso livello



Il caso di prova

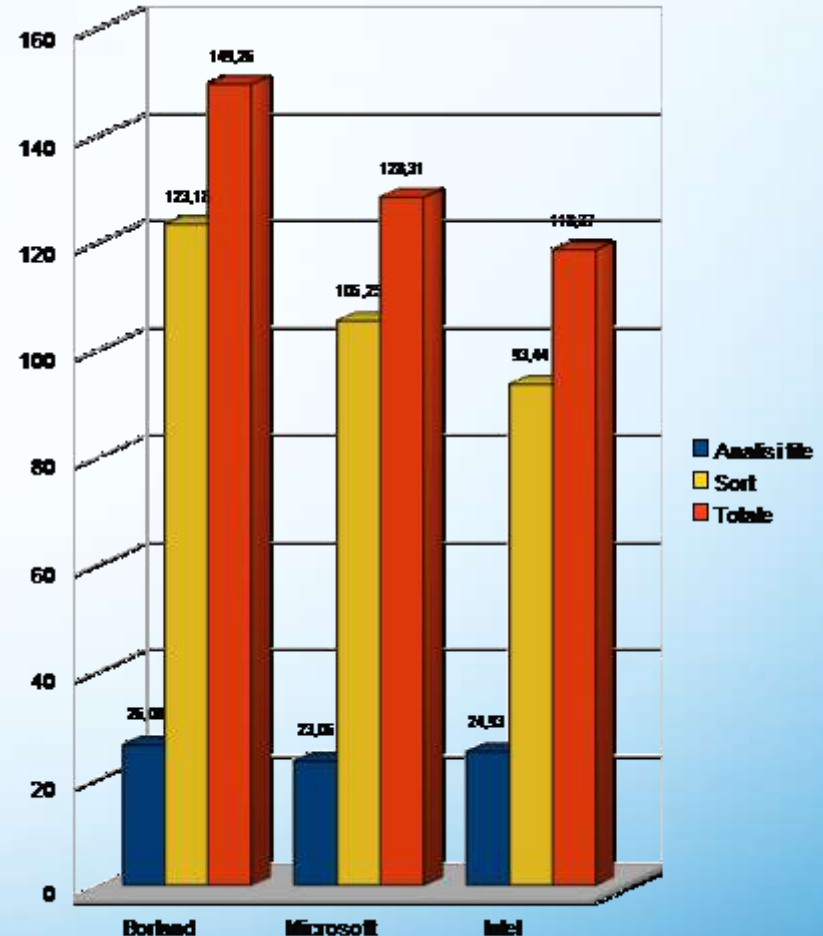
- Progetto Bsort
 - Command line tool
 - Stesso algoritmo di sort del prodotto SADAS
 - Stesse strutture dati
- Dati utilizzati
 - Più di 8 milioni di record (131 byte/record)
 - Dimensioni file input > 1 GigaByte
 - Estratto di dati bancari (piani di ammortamento)
 - Campi numerici, alfanumerici, data, duplicati...
- Ambiente di prova
 - Intel Core2 Duo T7500 2.20 Ghz – 4 MB Cache L2
 - 4 GB RAM Dual Channel / 800 Mhz Non-ECC
 - Windows Vista Business 32 bit

Strumenti utilizzati

- Borland Developer Studio 2006
 - IDE utilizzato per lo sviluppo del prodotto
 - Utilizzo del compilatore Borland C++
 - Versione 5.82
- Visual Studio 2008
 - Utilizzo del compilatore Microsoft C++
 - Versione 15.00.30729.01
- Intel Parallel Studio
 - Utilizzo del compilatore Intel
 - Versione 11.1 – Build 20090421
 - Utilizzo degli altri strumenti della suite...

Primo benchmark

- Risultati ottenuti con i diversi compilatori sulla versione iniziale del codice sorgente
- Valori ottenuti dalla media di cinque elaborazioni
- Utilizzate le opzioni di build di default (release) con tutte le ottimizzazioni attive



Utilizzo di Intel Parallel Studio nel caso di Studio



- Intel Parallel Composer
 - OpenMP (#PRAGMA OMP PARALLEL FOR)
 - IPP (ippCopy, ippCompare)
- Intel Parallel Inspector
 - Threading Errors
 - Memory Errors
- Intel Parallel Amplifier
 - Find HotSpots
 - Lock & Waits
 - Concurrency



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

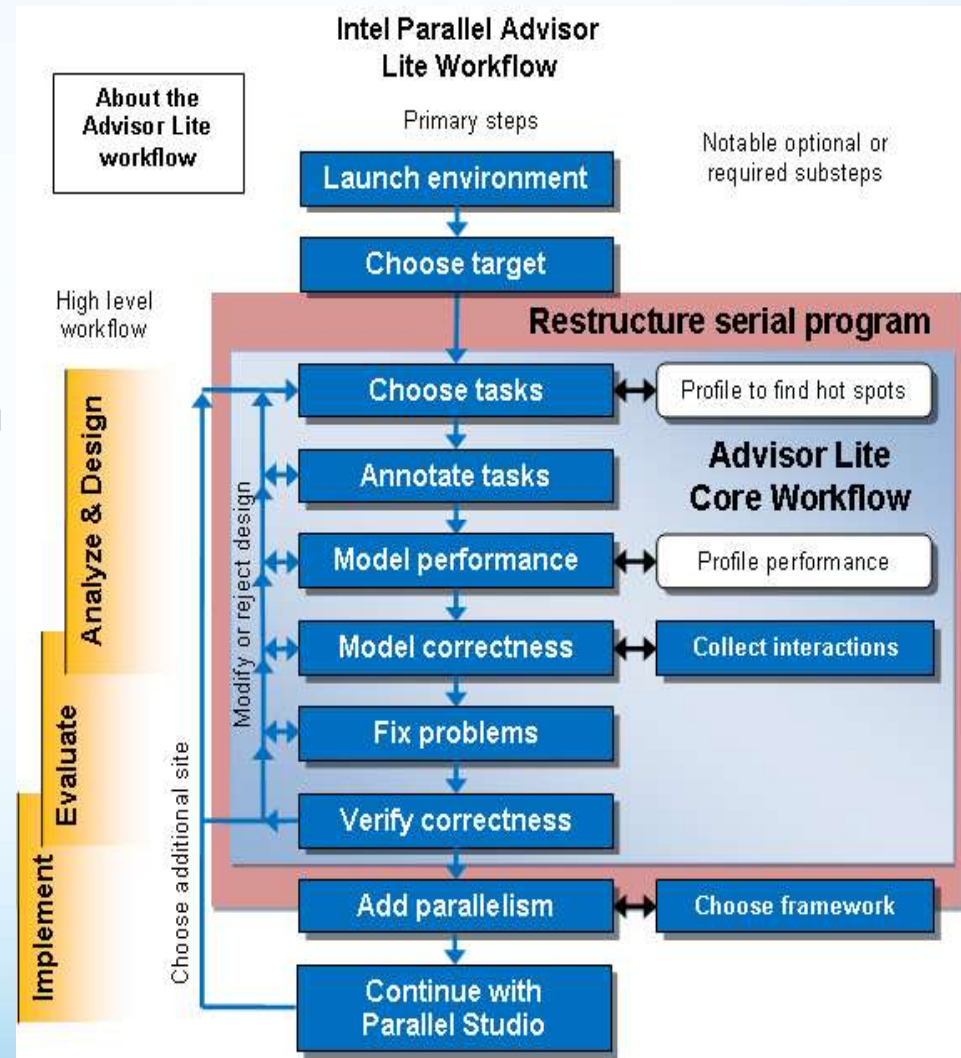
ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS



Intel Parallel Advisor - Funzionalità



- Dal codice seriale al codice parallelo:
 - Definizione di un workflow
 - Modellazione di una soluzione
 - Annotazioni
- Strategie di parallelizzazione
- Best Practices



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS

assintel
Associazione Nazionale
Imprese ICT

Intel Parallel Advisor

Caso di Studio



- Find Hotspots
 - Dove il nostro programma consuma il maggior tempo di elaborazione? (Intel Parallel Amplifier)
 - Stessi parametri e dati di test utilizzati in precedenza
 - Configurazione Release per analisi reale
 - Visualizzazione Sorgente + Call Stack
 - Informazioni sul timing nel sorgente
 - Possibilità di confrontare diverse misurazioni, per raffrontare i risultati ottenuti con l'adozione di soluzioni tecniche diverse (ad es. OpenMP vs IPP)



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS





Intel Parallel Composer

- OpenMP
 - Semplicità di utilizzo mediante direttive di precompilazione
 - Poche modifiche al codice sorgente
- Intel Integrated Performance Primitives (IPP)
 - Modifica del codice sorgente per utilizzare le funzioni di libreria
 - Ottimizzazione degli algoritmi con riferimento alle primitive disponibili
- Intel Threading Building Block (TBB)
 - Non utilizzato nel caso di studio: la gestione dei thread era già implementata a basso livello



Risultati Find Hotspots



r000hs Start Page

Hotspots

Bottom-up Top-down Tree

Function - Caller Function Tree	CPU Time:Self	Module
TSadasExternalSort::mergeNew	7.738s	BSort.exe
sortLin	6.317s	BSort.exe
TSdsQSort::SDSCopiaStrRec4	6.182s	BSort.exe
ricavaInfo	5.023s	BSort.exe
TSdsQSort::sdsCompareStrO4	3.368s	BSort.exe
TSadasExternalSort::inserisciCorsa	2.278s	BSort.exe
LeggiLargeFile	1.861s	BSort.exe
TSdsQSort::sdsSortTHStr	0.679s	BSort.exe
__intel_new_memcpy	0.655s	BSort.exe
diffOra	0.204s	BSort.exe
TSdsQMultiSort::sdsCompareStrO4	0.198s	BSort.exe
DateToPrg	0.080s	BSort.exe
TSdsQMultiSort::funcMultiCompare	0.077s	BSort.exe
TSdsQSort::sdsTestCorse	0.061s	BSort.exe
_allmul	0.031s	BSort.exe
lanciaThreadStr	0.000s	BSort.exe
__intel_fast_memcpy	0.000s	BSort.exe
TSdsQSort::sdsSortStrOtt	0.000s	BSort.exe
TSdsQSort::defSortAlf	0.000s	BSort.exe
new[]	0.000s	BSort.exe
TSdsQSort::defMetodi	0.000s	BSort.exe



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS



Analisi Hotspots



Hotspots Intel® Parallel Amplifier

Bottom-up Top-down Tree SadasSort.cpp

Line	Source	CPU Time:Self
396	/*-----*/	
397	void __fastcall TSdsQSort::SDSCopiaStrRec4(const char *origine , char *destinazione)	
398	{	0.153s
399	SDS_UI ind_Copia;	
400	for (ind_Copia = 0; ind_Copia<qsLenRecDiv4;ind_Copia++)	0.110s
401	*(((SDS_UI*)destinazione) +ind_Copia) = *(((SDS_UI*)origine)+ind_Copia);	3.604s
402	ind_Copia = qsLenRecRes4;	1.862s
403	for (;ind_Copia<qsLenRec;ind_Copia++)	0.233s
404	*(c /*-----*/	
405	void __fastcall TSdsQSort::SDSCopiaStrRec4(const char *origine , char *destinazione)	
406	{	
	/* FC IPP */	
	ippsCopy_8u((Ipp8u*)origine, (Ipp8u*)destinazione, qsLenRec);	
	/*	
	SDS_UI ind_Copia;	
	for (ind_Copia = 0; ind_Copia<qsLenRecDiv4;ind_Copia++)	
	(((SDS_UI)destinazione) +ind_Copia) = *(((SDS_UI*)origine)+ind_Copia);	
	ind_Copia = qsLenRecRes4;	
	for (;ind_Copia<qsLenRec;ind_Copia++)	
	*(destinazione+ind_Copia) = *(origine+ind_Copia);	
	*/	
	}	



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS



Confronto soluzioni (nativo – IPP)



Hotspots Intel® Parallel Amplifier

Bottom-up

Function	CPU Time:Self: r000hs	CPU Time:Self: r001hs	CPU Time:Self:Difference	Module
TSadasExternalSort::mergeNew	7.757s	7.588s	0.169s	BSort.exe
sortLin	6.771s	6.677s	0.094s	BSort.exe
TSdsQSort::SDSCopiaStrRec4	6.078s	0.248s	5.830s	BSort.exe
ricavaInfo	5.039s	4.897s	0.142s	BSort.exe
TSdsQSort::sdsCompareStrO4	3.651s	3.490s	0.161s	BSort.exe
TSadasExternalSort::inserisciCorsa	2.372s	2.574s	-0.203s	BSort.exe
LeggiLargeFile	2.239s	1.997s	0.242s	BSort.exe
TSdsQSort::sdsSortTHStr	0.488s	0.810s	-0.322s	BSort.exe
_intel_new_memcpy	0.405s	0.530s	-0.125s	BSort.exe
diffOra				
TSdsQMultiSort::sdsCompareStrO4				
DateToPrg				
TSdsQSort::sdsTestCorse				
_allmul				
CreaFile				

Hotspots

Bottom-up Top-down Tree

Module	CPU Time:Self
- Function	
- Caller Function Tree	
BSort.exe	27.636s
ippsv8-6.1.dll	5.235s
ipps-6.1.dll	0.060s



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS





Demo Intel Parallel Studio



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Cisco Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS





Parallel Studio: Pro & Contro

- Vantaggi riscontrati nel corso dei test
 - Integrazione fra i diversi strumenti della suite
 - Integrazione con l'IDE di Visual Studio 2008
 - Reporting
 - Confronto fra diverse misurazioni
 - Identificazione del codice e call-stack
 - Evidenza del codice in relazione ad un problema
 - Semplicità di utilizzo
 - Documentazione in linea e sul web
 - Stabilità dell'ambiente
- Svantaggi riscontrati nel corso dei test
 - N.C.



Cosa vedremo nella prossima sessione



- L'utilizzo di Intel Parallel Inspector
 - Verifica degli accessi alla memoria
 - Verifica dei problemi di data race
- L'utilizzo di Intel Parallel Amplifier
 - Localizzazione di lock & wait
 - Analisi dei colli di bottiglia (Hotspots)
 - Analisi dell'utilizzo dei thread



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Cisco Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS



assintel
Associazione Nazionale
Imprese ICT



Domande & Risposte

Intel® Parallel Studio

www.intel.com/go/parallel

www.adalta.it

cfiorillo@gmail.com



Intel Parallel Studio Seminar – Milano 22 Giugno 2010

Copyright © 2010, Ciro Fiorillo. Tutti i diritti riservati.

*Tutti i marchi registrati citati sono di proprietà esclusiva degli aventi diritto.

ADALTA
SOFTWARE PER LA SCIENZA E IL BUSINESS

