

Applicazioni GIS operanti su infrastrutture elaborative collocate in Internet Data Centers

Vincenzo CONSORTI

Università degli Studi di Teramo, Viale Crucoli 122, 64100, Teramo, 0861-248780, vconsorti@unite.it

Riassunto

I modelli architetture tipici dell'Information e Communication Technology (ICT) allo stato attuale evolvono sempre più verso schemi logici che prevedono centri elaborativi di rilevanti dimensioni (Data Centers) con caratteristiche di scalabilità, affidabilità, sicurezza, flessibilità, e che sono in grado di erogare via rete le prestazioni elaborative richieste dalle diverse tipologie di utenti.

In un ambito di tale complessità risulta quindi necessario progettare e definire le applicazioni GIS con schemi che siano in grado di adattarsi pienamente ai nuovi modelli e di utilizzare in modo efficace ed efficiente le possibilità offerte da queste nuove infrastrutture elaborative.

L'articolo esamina in dettaglio i nuovi modelli di infrastrutture elaborative collocate presso Internet Data Centers e gli schemi, i modelli e le tecniche da adottare per lo sviluppo di applicazioni GIS operanti in tali ambienti.

Sono definite delle "guidelines" progettuali di livello generale per la classificazione dei modelli elaborativi e per l'integrazione degli stessi con i principali prodotti GIS attualmente sul mercato.

I modelli architetture e le tecniche descritte consentono ad una vasta platea di utenti operanti direttamente nel settore dei GIS di poter ottenere una conoscenza su queste problematiche che molto spesso risultano poco note e di difficile comprensione.

Sono riportati esempi concreti che facilitano la comprensione dei modelli e degli schemi proposti.

Abstract

The development, deployment, and operation of GIS enterprise applications require a rich application infrastructure. Such an infrastructure must meet the requirements of Information and Communication Technology (ICT) organization for availability, security, scalability, and manageability and provide facilities for applications to meet these requirements.

The Internet Data Centers (IDC) are fundamental elements designed to support an extensive use of Geographical Information System (GIS) and Web-enabled applications that connect organizations to their customers.

This paper examines in detail the new enterprise-level architecture design options in an IDC infrastructure and outlines, the models and the techniques to adopt for the development of operating GIS applications in such specific environment.

Concrete examples are brought back that facilitate the understanding of the models and the outlines proposals.

1. Introduzione

La crescente complessità dell'Information e Communication Technology (ICT) che necessita una centralizzazione delle risorse elaborative per assicurare costi di gestione accettabili e livelli di servizio di assoluta eccellenza ha giustificato l'individuazione di nuovi modelli architetture centralizzati per le infrastrutture elaborative in grado di erogare servizi in rete ed in modalità distribuita.

La diffusione di questi nuovi modelli architetturali è facilitata anche dalle direttive emanate dal Ministero dell'Innovazione e delle Tecnologie, che ha in corso la realizzazione del Sistema Pubblico di Connettività (SPC) che sarà in grado di connettere tutte le Pubbliche Amministrazioni e che prevede espressamente la realizzazione di infrastrutture elaborative centralizzate per l'erogazione di servizi ai cittadini, Enti ed aziende.

L'adozione di tali modelli implica però necessariamente una diversa strutturazione delle applicazioni GIS ed in particolare di quelle collegate alla produzione delle componenti applicative GIS Web oriented necessarie per realizzare una infrastruttura applicativa GIS che sia pienamente integrata all'interno di un Internet Data Center e che consenta di soddisfare in modo completo i requisiti di sicurezza, privacy ed integrità nelle transazioni.

Le implicazioni connesse a queste nuove strutturazioni determinano di conseguenza l'aggiunta di nuovi fattori competitivi quali il piano di licenze, gli elementi gestionali e la flessibilità operativa connessa alle metodologie di sviluppo agili che devono essere pienamente considerati per l'ottimizzazione in termini di costo/benefici nella realizzazione di applicativi GIS operanti in ambito IDC.

2. Le infrastrutture elaborative per l'erogazione di servizi

L'insieme dei servizi in ambito GIS che gli Enti e le aziende possono erogare attraverso una infrastruttura elaborativa centralizzata richiede la realizzazione di un'architettura che soddisfi i requisiti di scalabilità, affidabilità, prestazioni e flessibilità nonché semplicità di gestione anche remota.

Le infrastrutture elaborative centralizzate, utilizzabili anche per l'erogazione di servizi in ambito GIS in modalità distribuita, in generale costituiscono l'elemento centrale dei sistemi informatici nazionali e regionali e consentono la fruizione e l'erogazione di servizi tra le amministrazioni, le imprese ed i cittadini.

In generale i servizi e sistemi messi a disposizione da tali infrastrutture sono quelli di seguito riportati:

- **Infrastruttura di rete** basate su backbone ad alta velocità (Gigabit) che garantiscono il collegamento fra i principali server applicativi assicurando le appropriate prestazioni richieste dai servizi avanzati da supportare. L'architettura di rete degli IDC soddisfa inoltre i più recenti criteri di disponibilità e modularità grazie ad un impianto fisico di cablaggio specifico per data center a supporto di una architettura a più livelli, funzionalmente distinti, che ottimizza le prestazioni dell'infrastruttura IT e della logica applicativa, e fornisce le caratteristiche di scalabilità ed alta affidabilità richieste.
- **Servizi web ed applicativi** posizionati in aree di sicurezza distinte ma interoperanti tra loro. Una serie di web servers sono infatti posizionati all'interno della rete direttamente accessibili dagli utenti esterni e forniscono il livello di presentazione per tutti i servizi messi a disposizione dall'IDC. A monte dei server web sono posizionati una serie di Application Servers, nella rete sicura, che rendono disponibili gli strumenti idonei all'implementazione della logica applicativa che è al cuore dei servizi offerti consentendo inoltre una facile integrazione con eventuali sistemi pre-esistenti.
- **Alta disponibilità dei servizi** dovuta non solo ad una infrastruttura di rete ridondata ma al ricorso a due classi di ridondanza per evitare i "single point of failure", una interna alla singola macchina e una applicativa. L'insieme di queste due componenti garantisce la continuità del servizio offerto dai singoli blocchi che costituiscono l'infrastruttura. Così ogni sistema critico prevede la doppia alimentazione, la doppia connessione di rete ed il mirroring dei dischi di sistema tramite sistemi RAID. Inoltre tutti i componenti essenziali per il funzionamento dell'infrastruttura sono ridondata e configurati in clustering in modo garantire la continuità del servizio a fronte di guasti di sistema.
- **Scalabilità e prestazioni dei servizi** in virtù dei meccanismi di bilanciamento previsti ai vari livelli, dalla rete ai servizi. Difatti in ambito IDC l'infrastruttura elaborativa consente

di effettuare un bilanciamento di carico fra più macchine indirizzando contestualmente sia le problematiche di disponibilità che di scalabilità; infatti l'aggiunta di serveri addizionali (scalabilità orizzontale) è estremamente facilitata dall'architettura stessa sia per la configurazione delle macchine che per la loro connettività. I sistemi serveri, garantiscono inoltre una scalabilità verticale, cioè un incremento delle risorse di sistema per l'ottenimento di un livello prestazionale superiore.

- **Sicurezza completa dell'infrastruttura** grazie a meccanismi di sicurezza perimetrale contro minacce esterne ma anche interne, alla divisione in zone a privilegi distinti dei servizi, alla presenza di sistemi di rilevamento intrusioni e di controllo degli accessi alle macchine ed ai servizi
- **Sicurezza completa dell'accesso ai servizi** grazie alla implementazione di connessioni sicure via VPN-IP dalla sede del percettore dei servizi (Ente, azienda, cittadino, ecc.) all'IDCma anche alla definizione di politiche di autenticazione e privilegi granulari a livello di specifico utente e servizio
- **Sicurezza completa dei dati** grazie al ricorso sia ad architetture con unità di memorizzazione su nastri o dischi condivisi dai servizi tramite infrastrutture su fibra (SAN) o su rete IP (iSCSI) sia a meccanismi di autenticazione e controllo degli accessi ai dati.
- **Gestione centralizzata delle problematiche relative alle infrastrutture locali** con vantaggi conseguenti in termini di efficacia e di efficienza dei servizi erogati.

E' del tutto evidente quindi, che per assicurare l'erogazione di applicazioni GIS ad una diversa, variegata ed ampia classe di utenti, un IDC per poter essere realizzato pienamente in linea con gli obiettivi prefissati necessita in fase progettuale di un'astrazione concettuale particolarmente rilevante.

3. Le architetture e le tecnologie di riferimento in ambito IDC

La realizzazione di infrastrutture elaborative in ambito IDC prevede schemi architetture particolarmente complessi ed in grado di soddisfare i diversi requisiti richiesti dalle varie applicazioni software per funzionare in maniera ottimale. In questo ambito le applicazioni GIS ed i servizi collegati alla posizione geografica (Bolstad 2002), (De Mers, 2002) presentano ulteriori specificità in quanto oltre alla notevole mole di dati coinvolti, esiste una ampia eterogeneità degli stessi ed una modalità di fruizione in modalità multicanale (PC, palmari, smartphone, cellulari, ecc.) orientata all'ambiente Internet/Web su infrastrutture IP oriented fisse e wireless.

Un modello architetture che presenta caratteristiche di elevata complessità ma anche di estrema generalità viene proposto nella Figura 1, e presenta un insieme di aree logicamente distinte ed in particolare:

Area Perimetrale - zona che separa l'area interna da quella esterna, è collocata tra due insiemi di firewall e contiene un insieme di server che offrono principalmente servizi di rete e di front-office in ambito Web.

Area Interna - zona posta all'interno dei firewall interni e contiene un insieme di server in grado di erogare l'insieme dei servizi di back-office per le Web Applications, collocati in aree funzionalmente distinte e separate da VLAN.

Area Locale - zona, fisicamente collocata nello stesso luogo dell'infrastruttura elaborativa, posta all'interno del perimetro ma esternamente all'area dei server interni individua l'insieme degli apparati con i relativi software che consentono agli utenti operanti a questo livello sia di realizzare le Web Application che di fruire di tali applicativi con un livello di performance particolarmente elevato.

Area di Supporto - zona, in generale fisicamente collocata nello stesso luogo dell'infrastruttura elaborativa, posta all'interno del perimetro ma esternamente all'area dei server interni individua l'insieme degli apparati con i relativi software che consentono la gestione ed il monitoraggio dell'infrastruttura. Il posizionamento intermedio tra i firewall è originato dalla necessità di supportare sia l'area perimetrale che quella interna ma in realizzazioni particolarmente sofisticate che prevedono il più alto livello di sicurezza tale area con i relativi apparati viene duplicata sia a livello di perimetro che a livello interno.

Aree Esterne - zona, posta all'esterno dei firewall perimetrali individua ed include tutti gli apparati con i relativi software che costituiscono in generale gli originatori delle richieste e gli utilizzatori delle informazioni. Il modello architetturale delinea l'estrema variabilità di tali apparati che vanno dal PC al cellulare e questo implica che il sistema deve prevedere in modo automatico la possibilità a partire da richieste standard di differenziarsi in funzione del tipo di apparato ricevente. La Geographic Network va intesa come l'insieme delle risorse di tipo geografico disponibili in "rete" (inclusi i Web Services) per evidenziare esempi classici di applicazioni che richiedono elevati volumi di dati, e quindi attiva con il sistema un collegamento bidirezionale (il nostro Web può ad esempio reindirizzare le richieste direttamente in rete), mentre le connessioni remote sono tipiche per collegamenti di telelavoro realizzati in generale tramite i protocolli RDP o ICA che permettono il controllo centralizzato di terminali remoti.

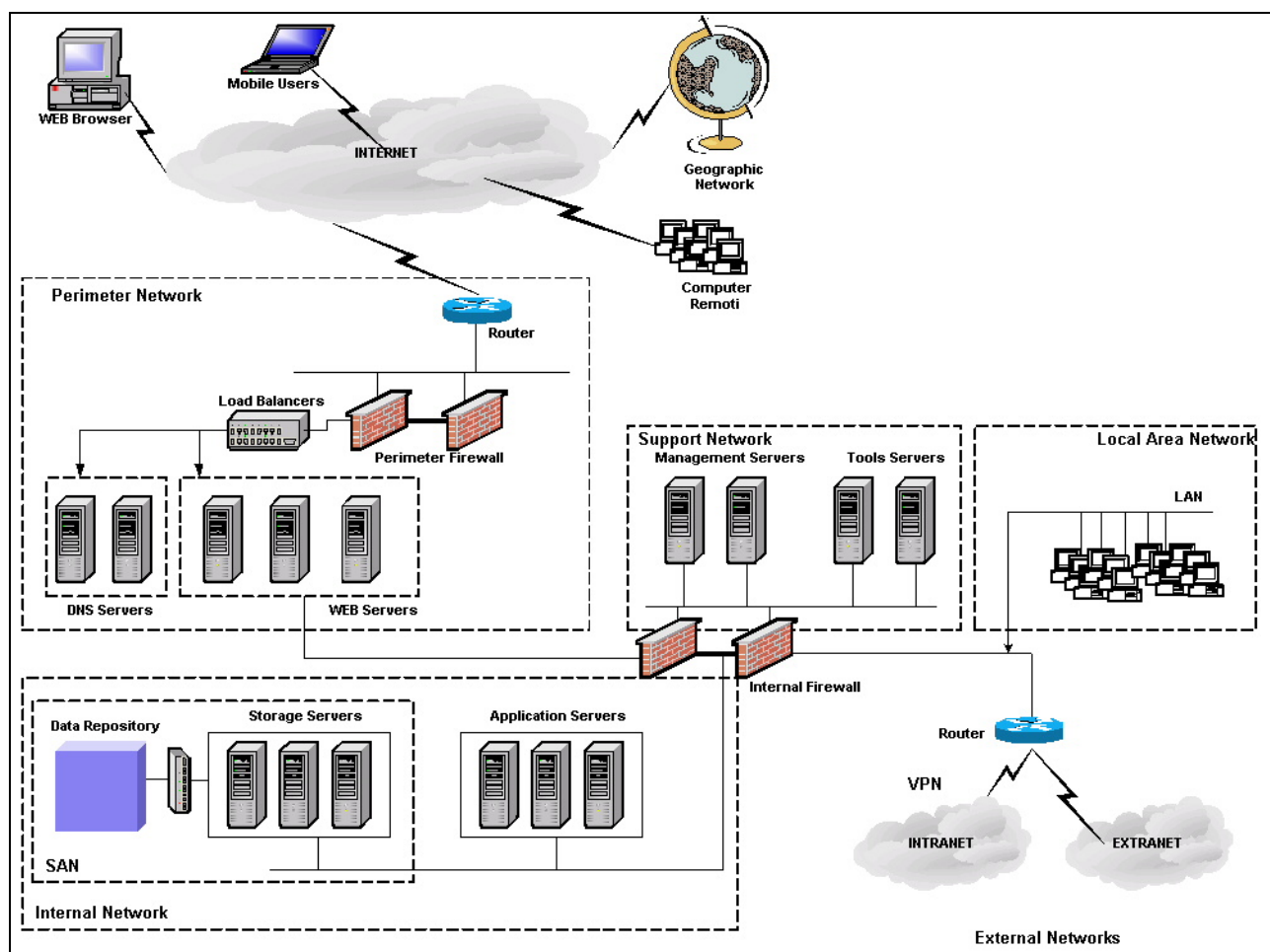


Figura 1 - Architettura di un'infrastruttura elaborativa operante in ambiente IDC

Ad ogni area sono associati un insieme di elementi che possono essere più o meno approfonditi e bilanciati in funzione dei progetti da realizzare. Tutti gli elementi sono collegati da infrastrutture di comunicazione ad alta velocità.

Inoltre la necessità di integrare più componenti (GIS, Internet; Wireless), di fornire i servizi in modalità multicanale (l'accessibilità ai servizi è attuata attraverso più canali di comunicazione per rendere più agevole la fruizione) e di seguire l'evoluzione generale del sistema ed assecondarne le richieste che nel tempo con costanza si amplieranno e si differenzieranno dinamicamente impone l'adozione di paradigmi architetturali e tecnologici particolarmente sofisticati. (Consorti, 2003) (Pothukuch, 2002), (Ren Pheng, 2003).

4. I modelli architetturali per applicazioni GIS in ambito enterprise

I modelli architetturali da utilizzare nell'ambito di architetture enterprise come quella evidenziata nel paragrafo precedente fanno riferimento allo schema SOA (Service Oriented Architecture) che consente applicativamente l'integrazione di dati dei differenti sistemi informativi mediante protocolli standard in ambito Web e permette quindi lo sviluppo di servizi applicativi efficaci ed efficienti in modalità geograficamente distribuita. (Kropla, 2005), (Krygier, 2005)

I servizi applicativi GIS oriented sono in generale pubblicati mediante protocolli Web oriented e di conseguenza possono essere utilizzati da un'ampia classe di applicazioni Web attraverso il comune strato applicativo utilizzato per la comunicazione (enterprise service bus).

Le applicazioni GIS operanti in ambiente enterprise, prevedono come altre applicazioni di tale classe tre livelli di componenti (Orfali, 1998) e precisamente:

- **Componenti di presentazione** - elementi direttamente relazionati agli utenti che hanno il compito di fornire una serie molto importante di funzionalità di base quali il display di informazioni geospaziali, il rendering con le operazioni di zoom, pan, ecc., le interrogazioni e la selezione delle features e l'editing che permette di gestire, aggiornare o aggiungere informazioni alla base dati.
- **Componenti di logica applicativa** - elementi che hanno il compito e la funzione di ricevere le richieste dai client degli utenti, eseguire le richieste e restituire i risultati alle componenti di presentazione. Le principali componenti di questo livello sono incluse all'interno dei web server, degli application server e dei map server.
- **Componenti di accesso ai dati** - elementi che sono in grado di contenere e fornire dati mediante interfacce comuni ai diversi client. In generale tali componenti sono rappresentate all'interno dei modelli architetturali da data server o DBMS server.

In un tale contesto i modelli architetturali per applicazioni GIS enterprise devono essere particolarmente curati, oltre che per la parte applicativa, per gli aspetti collegati ai **cataloghi** ed al **geodatabase**. (Sonnen, 2005)

I cataloghi permettono agli utenti di selezionare i servizi disponibili e di conoscere i dati disponibili, le loro caratteristiche e la loro qualità (metadati). La gestione dei cataloghi deve essere integrata all'interno di specifici segmenti applicativi per consentire nelle diverse fasi elaborative di poter aggiornare sia i metadati, in diretta congruenza con il contenuto dei geodatabase, sia i servizi di gestione, elaborazione e consultazione dei dati.

Il geodatabase è l'archivio che fornisce i dati all'intero sistema. I dati non necessariamente risiedono in un'unica banca dati, ma possono essere distribuiti ed in questo caso l'integrità dei dati per applicazioni multiutente come quelle che risiedono in ambito IDC è un aspetto di particolare criticità che richiede un'attenta ed approfondita valutazione sui prodotti da utilizzare.

Il geodatabase deve basarsi ed essere completamente integrato con un DataBase Management System (DBMS) che deve garantire la gestione sia della componente geometrica che alfanumerica del dato.

I principali vantaggi derivanti dall'utilizzo di una struttura dati logicamente centralizzata riguardano:

- repository uniforme e centralizzato per i dati geografici che sono memorizzati e gestiti in un unico DBMS;
- possibilità di editing multiutente grazie a particolari strutture temporanee che consentono di preservare l'integrità dei dati e di effettuare la riconciliazione degli stessi in caso di eventuali conflitti;
- insieme di features continuo in quanto anche per rilevanti insiemi non si ha necessità di mosaicature o partizionamenti;
- editing ed immissione dati particolarmente efficiente mediante utilizzo di sottotipi, domini e regole di validazione;
- collegamento tra le diverse caratteristiche degli oggetti geografici e questo consente ad esempio in caso di cancellazione di un oggetto di rimuovere in modo automatico e trasparente agli utilizzatori tutti i collegamenti ad esso associati;
- utilizzo di oggetti intuitivi e famigliari agli utenti che invece di punti, linee ed aree manipolano e gestiscono particelle, strade, edifici, ecc.

Gli standard di riferimento per i Geodatabase sono: gli standard ISO serie 19100, "l'OpenGIS Simple Features Specification for SQL" dell'Open Geospatial Consortium, le specifiche del linguaggio GML, le specifiche nazionali IntesaGIS. (Stefanakis, 2005)

I geodatabase possono essere sviluppati sia con tecnologia commerciale (es. ESRI, IBM, Oracle) che con soluzioni Open Source (es. PostGIS, MySQL) o miste in tecnologia ibrida. (Stones, 2005)

Occorre comunque considerare che il percorso di transizione verso modelli di applicazioni GIS in ambiente IDC risulta particolarmente complesso in quanto tradizionalmente i livelli informativi all'interno degli Enti/Aziende risultano memorizzati in varie localizzazioni e questo implica che di solito nei dati ci sono notevoli duplicazioni ed inconsistenze di vario genere che per essere recuperate richiedono un effort lavorativo particolarmente rilevante e di conseguenza costoso.

5. Conclusioni

Allo stato attuale le tecnologie GIS necessitano di una stretta integrazione con quelle di Information Technology. Questo implica che le funzionalità GIS devono essere realmente collegate ad un crescente numero e varietà di applicazioni software.

I nuovi modelli architetturali per applicazioni GIS in ambito IDC offrono la soluzione a queste problematiche ma per essere realmente efficaci necessitano di una progettazione accurata e di una notevole esperienza nelle diverse tematiche presenti in ambiente enterprise.

Bibliografia

- Bolstad P.V. (2002), "GIS Fundamentals: A first textbook on GIS", Bookmaster Dist
- Consorti V. (2003), "Architetture Hardware, Software e di Rete per servizi ed applicazioni Gis oriented", MondoGis N.34 – MondoGis editore
- DeMers M. (2002), "Fundamentals of Geographic Information Systems", John Wiley & Sons
- Kropka B. (2005), "Beginning Map Server: Open Source GIS Development", Apress
- Krygier J., Wood D. (2005), "Making Maps: A visual guide to map design for GIS", The Guilford Press
- Matthew N., Stones R. (2005), "Beginning Databases with PostgreSQL", Apress
- Morris H.D., Sonnen D. (2005), "Extending GIS to Enterprise Application", ESRI Press
- Orfali R., Harkey D. (1998), "Client/Server Programming with Java and CORBA", John Wiley & Sons
- Pothukuch S. (2002), "Enterprise Modeling of Business Objects in GIS Development", Proceedings of Map Asia
- Ren Peng Z., Hsiang Tsou Z. (2003), "Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Network", John Wiley & Sons
- Stefanakis E. (2005), "Geographic Databases and GIS", Springer