

WFP SPATIAL DATA INFRASTRUCTURE: **SVILUPPO DI UN AMBIENTE DI ANALISI E PUBBLICAZIONE** **SU PIATTAFORMA *OPEN SOURCE***

Andrea AJMAR (*), Marco ARCIDIACONO (**), Daniele BUSIN (***), Olivier TERZO (****)

(*) ITHACA, via P.C. Boggio 61, 10138 Torino, Italy, tel. +39-011-19751852, fax +39-011-19751122,
e-mail: andrea.ajmar@ithaca.polito.it

(**) tel. +39-335-5400195, e-mail: marco.arcidiacono@studenti.polito.it

(***) tel. +39-333-8351066, e-mail: daniele.busin@studenti.polito.it

(****) Istituto Superiore Mario Boella, via P.C. Boggio 61, 10138 Torino, Italy, tel. +39-011-2276855,
e-mail: terzo@ismb.it

Riassunto

La capacità di risposta in tempi brevi durante le emergenze, la riduzione del rischi a lungo termine e lo sviluppo di attività per la protezione dell'ambiente sono settori in cui la disponibilità di un'Infrastruttura di Dati Spaziali (SDI) permette di migliorare notevolmente le capacità analitiche, consentendo l'accesso a informazioni geografiche sulla base di un insieme di procedure, protocolli e specifiche standard.

L'associazione ITHACA (*Information Technology for Humanitarian Assistance Cooperation and Action*) opera a supporto del *World Food Programme* delle Nazioni Unite nello sviluppo di una propria SDI. Il progetto prevede l'implementazione di un'architettura a 2 livelli: un ambiente di gestione e produzione del dato, in cui sono disponibili procedure analitiche complesse per l'amministrazione dei dati, e un ambiente di analisi e pubblicazione, destinato a un'utenza allargata. L'ambiente di analisi e pubblicazione deve garantire i più elevati livelli di interoperabilità e, su richiesta del WFP, essere basato per quanto possibile su prodotti *Open Source*.

Il presente contributo ha come obiettivo l'analisi delle problematiche di *porting* da un *geodatabase* Oracle a uno PostgreSQL e la descrizione delle caratteristiche di un ambiente middleware per la gestione della conversione tra i diversi *datatype* supportati.

Abstract

Response capacities during emergencies, long-term risk reduction and the development of environmental protection measures are all fields where the availability of a Spatial Data Infrastructure (SDI) allows to enhance analytical capacities through an optimized and controlled geographic data access managed by standard procedures and protocols.

ITHACA (*Information Technology for Humanitarian Assistance Cooperation and Action*) association supports the UN World Food Programme in developing their own SDI solution. The proposed solution is basically conceived on two levels: a management and production environment, providing complex analytical procedures for data administrators, and an analysis and publication environment, dedicated to the distribution of the information to the broader humanitarian community. This second environment, to be really effective, must grant high interoperability levels and, on specific request by WFP, be based on Open Source software.

Present contribution wants to analyze some porting problems from an Oracle database to a PostgreSQL database, including the description of a middleware environment for the datatype conversion.

1. Introduzione

Il *World Food Programme* (WFP) è la maggiore agenzia delle Nazioni Unite; conta più di 10.000 dipendenti, il 90% dei quali dislocati negli uffici decentrati presenti in più di 40 nazioni dell'Africa sub-sahariana, del Medio-Oriente, dell'America Latina, in Asia e nel Pacifico. Il suo compito è quello di utilizzare la distribuzione di aiuti alimentari per gestire eventi calamitosi (naturali o prodotti dall'uomo) e per supportare lo sviluppo economico e sociale. Il WFP inoltre supporta lo *UN Joint Logistic Centre* (UNJLC), un'unità inter-agenzia responsabile della fornitura di supporto e servizi nel campo del *Logistic Information Management* per tutte le agenzie delle UN.

Ormai da più di un anno l'associazione ITHACA (*Information Technology for Humanitarian Assistance, Cooperation and Action*) opera come centro di ricerca applicata per la fornitura di servizi di *Information Technology* a supporto di azioni umanitarie. Sulla base di un *Memorandum of Understanding* (MoU) siglato con il WFP, ITHACA sta sviluppando e implementando una soluzione di SDI con l'obiettivo di:

- Riorganizzare le informazioni a carattere geografico attualmente disponibili e acquisire in maniera organizzata nuovi dati;
- Creare regole e procedure per la gestione, l'accesso e la documentazione dei dati;
- Fornire strumenti per la diffusione e l'utilizzo delle informazioni, principalmente per la gestione delle emergenze (fasi di *early-warning* e *early-impact*).

2. L'architettura di SDI per il WFP

Le principali problematiche relative alla progettazione di una soluzione di SDI per il WFP sono legate all'esigenza di coniugare da un lato funzionalità complesse di gestione (grossi volumi di dati, complesse procedure di aggiornamento con dati provenienti da fonti molto diversificate) e di analisi, e dall'altro elevate prestazioni (soprattutto durante le emergenze) e facilità di utilizzo per un'utenza allargata e potenzialmente non esperta nell'utilizzo di dati geografici. Queste condizioni hanno portato all'idea di sviluppare due ambienti paralleli, ma strettamente interconnessi: uno per le attività di gestione e mantenimento del dato (*Production Layer*) e uno per la sua pubblicazione e diffusione (*Publication Layer*), come evidenziato in *Figura 1*.

Questa soluzione presenta una serie di vantaggi, tra cui:

- La possibilità di sfruttare le procedure complesse e consolidate che derivano dall'utilizzo di un ambiente commerciale, e per cui le soluzioni *Open Source* attualmente disponibili non sono considerate ancora mature;
- L'incremento delle prestazioni complessive del sistema, diminuendo il carico di lavoro di ogni ambiente e consentendo un *tuning* specifico, dovendo i due ambienti rispondere a richieste sostanzialmente differenti;
- La capacità di mantenere la struttura aggiornata alle evoluzioni della comunità *Open Source* e di sviluppare nuove *feature* per un progressivo *porting* completo verso una soluzione completamente *Open Source*.

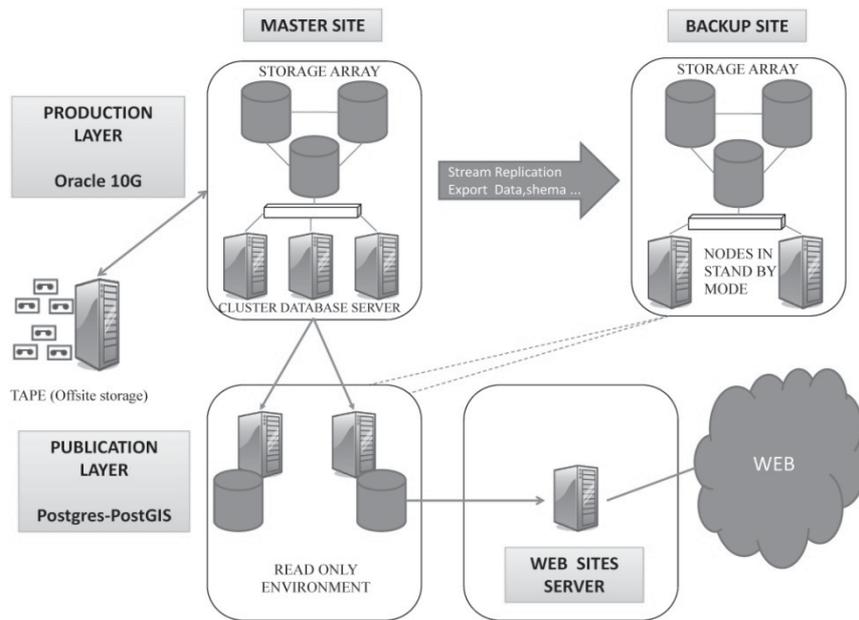


Figura 1 - WFP SDI, architettura di sistema

Sulla base dei requisiti indicati dal WFP, l'ambiente di gestione e produzione del dato è stato realizzato utilizzando ESRI come piattaforma di gestione del dato geografico; è stato costruito un *geodatabase* utilizzando Oracle 10G come DBMS e ArcSDE come *gateway*. Per garantire adeguati livelli di *backup* e *recovery*, a tale ambiente sono associati un sistema di cassette per un *backup* fisico e un nodo parallelo che costituisca al contempo un *backup* logico e un nodo in *stand-by*, da attivare rapidamente in caso di improvvisa indisponibilità del nodo principale. Il nodo parallelo viene mantenuto allineato a quello *master* tramite procedure automatiche.

L'accessibilità al sistema per la comunità allargata di operatori umanitari viene garantita dall'ambiente di pubblicazione, costituito da un *database* PostgreSQL, allineato al nodo master. I dati in questo nodo saranno utilizzati in modalità *read-only*, per la loro pubblicazione mediante WMS (*Web Map Service*). Le operazioni su WMS sono chiamate da client che supportano il protocollo http.

2.1. L'architettura di un database Oracle 10g

Un *database* Oracle è una collezione di dati immagazzinati in uno o più *files*, sulla base di una struttura relazionale basata su tabelle. Un'istanza comprende una serie di processi a livello di sistema operativo e strutture di memoria che interagiscono con lo *storage*; un tipico esempio di processi include il PMON (*process monitor*) e lo SMON (*system monitor*).

La struttura di memoria di un *server* Oracle è chiamata SGA (System Global Area), e tipicamente contiene informazioni *cache* come i *data-buffer*, i comandi SQL e le informazioni sugli utenti. In aggiunta alla componente di *storage*, il database consiste di *online redo logs*, che mantengono la *transactional history*; processi sono in grado di trasferire il contenuto degli *online redo logs* in *archive logs* (o *offline redo logs*), che costituiscono la base per i processi di *data recovery* e, in parte, di *data replication*.

2.2. L'architettura di un database PostgreSQL 8.2

PostgreSQL è un *object relational DBMS* (ORDBMS, *Object Relational Database Management System*). Ha un'architettura basata su un modello client/server. Il processo *server*, denominato

postgres, gestisce i file del database, accetta nuove connessioni dalle applicazioni *client* ed esegue le operazioni richieste sul database. Per ogni nuova connessione il *server* genera un processo *child* che si prende carico del *client*, mentre il processo *parent* comunica con il *socket* per gestire connessioni multiple.

3. Il porting da Oracle a PostgreSQL

L'obiettivo è quello di generare procedure automatiche per l'estrazione di schema e/o dati da un database Oracle e per la loro conversione in una sintassi interpretabile da un database PostgreSQL. I tipi di dati esportabili sono essenzialmente di diversi tipi: vettoriali, numerici, blob e stringhe.

L'attività si basa essenzialmente sull'utilizzo di due *tool*: *Ora2Pg* e *ogr2ogr*.

Ora2Pg, attualmente in versione 4.7, sviluppato in Perl. Il suo modulo principale, *Ora2Pg.pm*, contiene le funzioni che operano l'*export* dei dati. La distribuzione comprende uno *script* di default, *ora2pg.pl*, e un file di configurazione, *ora2pg.conf*, che contiene parametri da impostare secondo necessità per l'esecuzione in un determinato ambiente. *Ora2Pg* richiede le librerie Perl DBI e DBD::Oracle. Lo *script* consente di eseguire sia un *export* diretto, tramite connessione tra i database, sia indiretto, tramite la generazione di un file contenenti le istruzioni SQL necessarie per la creazione in PostgreSQL della struttura e dei dati copiati da Oracle (*Figura 2*).

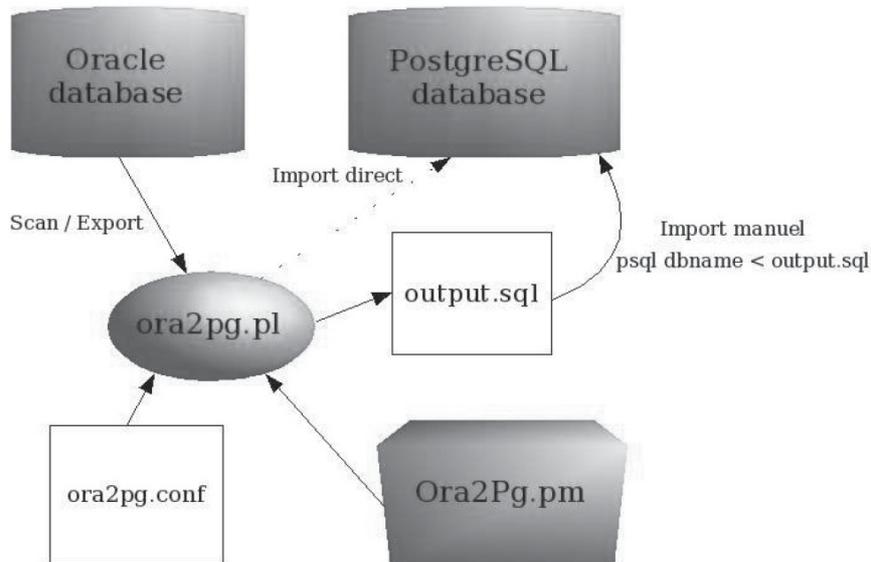


Figura 2 - Schema di funzionamento del tool Ora2Pg

Lo *script* consente di esportare tabelle (struttura, indici, *primary keys*, *foreign keys*), viste, diritti, funzioni e dati; è possibile anche l'*export* dello *schema*, anche se non in maniera esplicita.

Il *package* di base è stato modificato ed integrato allo scopo di:

- esportare singole tabelle esistenti nel database Oracle e non in quello PostgreSQL. Per questo scopo sono stati creati due *script* in Perl (
- *Figura 3*), Il primo consente di esportare da Oracle il codice sql di definizione delle strutture delle tabelle per poi essere create in ambito postgres, Il secondo consente di esportare i dati.
- eseguire l'*export* completo di tutte le tabelle incluse in un determinato *schema* e di tutti i dati in esse inclusi. In questo caso 3 *script* in Perl sono stati sviluppati. Il primo consente di esportare l'intero user schema sul database Oracle e il secondo consente la creazione dello schema sul database PostgreSQL e infine il terzo script sulla base dell'elenco degli oggetti esportati provvede all'esportazione dei dati.

- eseguire l'aggiornamento di tabelle esistenti, mediante lo sviluppo di 5 script in Perl (Figura 4). Il primo passo consente di esportare il contenuto delle tabelle da aggiornare dai rispettivi database, il secondo passo provvede a determinare se ci sono delle differenze e in tale caso si procede con l'aggiornamento dei dati sul database postgresQL.

Inoltre, il modulo Ora2Pg.pm è stato modificato per utilizzare uno *schema* diverso da quello di default.

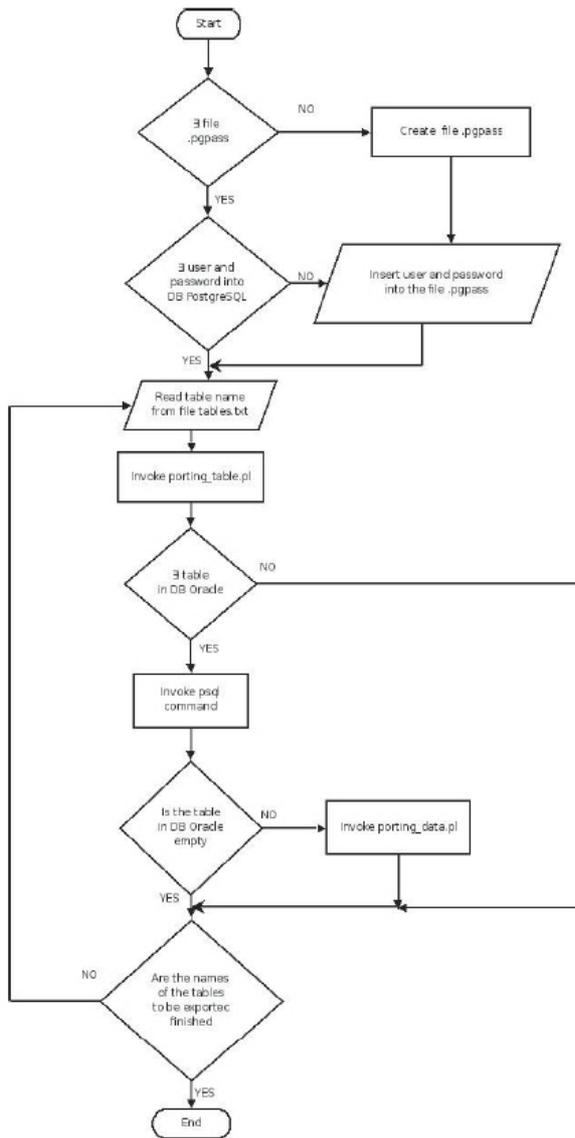


Figura 3 - Workflow per l'export di tabelle

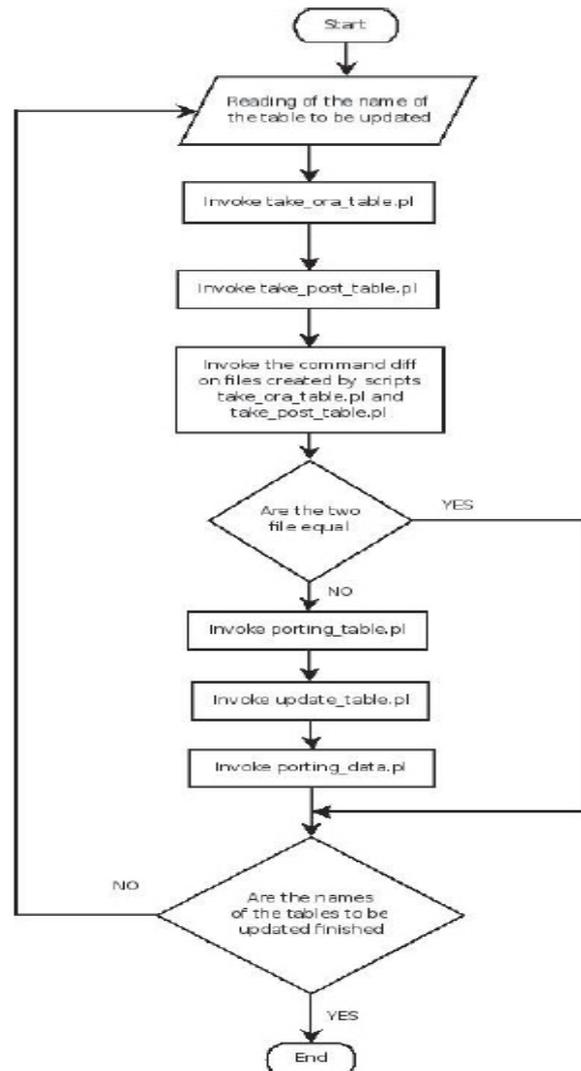


Figura 4 - Workflow per l'aggiornamento di tabelle

Ogr2ogr è un tool disponibile con le librerie gdal (versione 1.5.1) e consente di esportare i dati di tipo vettoriali. L'impiego di ogr2ogr passa attraverso uno script bash che è stato creato appositamente. Nel nostro caso specifico abbiamo la necessità di avere due tipi di export dei dati geometrici. Il primo a partire di un elenco statico di oggetti da esportare e il secondo su richiesta specifica.

4. Conclusioni

Volendo costruire un ambiente per l'analisi e la pubblicazione in ambiente *Open Source*, la prima problematica affrontata è stata quella di generare una serie di procedure in grado di replicare in *near real-time* le tabelle di sistema e i dati su PostgreSQL/PostGIS, il DBMS scelto come contenitore per tale ambiente. L'analisi include la comparazione tra i *data type* supportati e le modalità di gestione degli indici nei 2 DBMS. Procedure di automazione basate sui *tool* Ora2Pg e Ogr2Ogr sono utilizzate per la migrazione dei dati tra le due piattaforme.

Un'infrastruttura a due livelli come quella proposta presenta il duplice vantaggio di incrementare le performance, separando l'ambiente di gestione e quello di produzione, e di creare le condizioni per una migrazione progressiva da un DBMS commerciale a uno *Open Source*.

Riferimenti bibliografici

Executive Office Of The President - Office of Management and Budget (2002), "Circular No. A-16", http://www.whitehouse.gov/omb/circulars/a016/print/a016_rev.html

Groot R. and McLaughlin J.(2000), "Geospatial Data Infrastructure: Concepts, Cases and Good Practice", New York: Oxford University Press.

<http://www.samse.fr/GPL/squidclamav/>

Loney K.(2005), "ORACLE Database 10g: la Guida Completa", McGraw-Hill, Milano.

Masser I. (2005), "GIS Worlds - Creating Spatial Data Infrastructures", ESRI Press.

Nebert D. (2005), "Developing Spatial Data Infrastructures: The SDI Cookbook", Version 2.0, Global Spatial Data Infrastructure Association, <http://www.gsdi.org/gsdicookbook/index.asp>

The PostgreSQL Global Development Group (2005), "PostgreSQL Reference Manual", Volume 1-3 Network Theory LTD.

UNGIWG (2005), "UNSDI COMPENDIUM. A UNSDI Vision, Implementation Strategy and Reference Architecture",

http://www.ungiwg.org/docs/unsdi/UNSDI_Compndium1020.pdf

UNGIWG (2005), "STRATEGY for developing and implementing a United Nations Spatial Data Infrastructure in support of Humanitarian Response, Economic Development, Environmental Protection, Peace and Safety",

http://www.ungiwg.org/docs/unsdi/UNSDI_Strategy_ImplementationPaper.pdf