

Інтеграція даних в семантичному Грід

Лунченко О.О., НТУУ «КПІ», ІПСА, кафедра СП

У різних областях науки спостерігається експоненціальний ріст одержуваних експериментальних даних. Велика кількість організацій, які виконують спостереження, та їх незалежність, розмаїття об'єктів спостереження, неперервне і швидке вдосконаленням техніки спостережень призводить до необхідності використання неоднорідної, розподіленої інформації, накопиченої протягом значного періоду спостережень технологічно різними інструментами[1].

Завданням інтеграції даних є об'єднання інформації з різних джерел[2]. Вона має справлятися з різноманітними даними, породженими розбіжностями моделей даних, часу отримання, якості тощо.

Існує ряд проблем, таких як відсутність інтерфейсів і відкрито доступної документації користувача, залежність від конкретних фреймворків і приватних програмних пакетів або бібліотек, прив'язка процесів до конкретних типів наборів даних. Для подолання цих проблем було запропоновано розбиття на модулі всього процесу інтеграції даних в багаторазові, придатні до зміни, багатоцільові веб-сервіси - сервіс-орієнтована архітектура.

Під час сумісної праці з ресурсами учасникам необхідно обмінюватись інформацією у взаємо-зрозумілому форматі[3]. Тому необхідна додаткова обробка для інтеграції компонентів, що використовують різні синтаксичні структури, так зване синтаксичне посередництво. Для автоматичного здійснення узгодження даних у випадку виникнення синтаксичної невідповідності розроблені мова відображення (mapping language), яка може використовуватись для коментування XML структур даних за допомогою OWL понять і властивостей, ядро мови відображення (Mapping Language Engine) для реалізації цієї мови, і компонент динамічного виклику веб-сервісу (Dynamic Web Service Invocation) для виконання веб-сервісів[3].

Розглядаються такі компоненти пакету продуктів SAS® як SAS® Data Integration Studio і SAS® Grid Manager, що надають можливості масштабувати типові навантаження інтеграції даних на типових апаратних засобах Грід і завантаження збалансованих чисельних задач інтеграції даних по Грід-ресурсам[4].

В даній статті досліджується проблема узгодження гетерогенних даних. Також розглядаються розроблені на даний час підходи, методи та інструменти інтеграції даних в Семантичному Грід, такі як мова mapping language та інструменти SAS® Data Integration Studio і SAS® Grid Manager, описано підходи до інтеграції даних на прикладах геологічних даних, даних про дорожні мережі, соціологічні дані (проект SDI-Grid).

1. Брюхов Д.О., Вовченко А.Е., Захаров В.Н., Желенкова О.П., Калиниченко Л.А., Мартынов Д.О., Скворцов Н.А., Ступников С.А. Интеграция неоднородных информационных ресурсов в грид-инфраструктуре виртуальных обсерваторий на основе предметных посредников **[Електронний ресурс] : (Московская Секция АСМ SIGMOD)** — Режим доступу до електронного журн.: http://synthesis.ipi.ac.ru/sigmod/seminar/s20090429-назва_з_екрану_.
2. Werder, S. (2010). Data Integration in a modular and parallel Grid-Computing workflow. Leibniz Universität Hannover, Institute of Cartography and Geoinformatics.
3. Szomszor, M., Payne, T. R., Moreau, L. Using Semantic Web Technology to Automate Data Integration in Grid and Web Service Architectures. School of Electronics and Computer Science University of Southampton Southampton.
4. Doninger, C., Mehler, G., Rausch, N. (2008). Data Integration in a Grid-Enabled Environment. SAS Institute Inc.