Application BASTRI
Fiches Equipes

ZENITH (SR0441RR)

Gestion de données scientifiques
ATLAS (SR0127VR) ZENITH ZENITH (SR0522AR)

Statut: Terminée

Responsable : Patrick Valduriez

Mots-clés de "A - Thèmes de recherche en Sciences du numérique - 2024" : *Aucun mot-clé.*

Mots-clés de "B - Autres sciences et domaines d'application - 2024" : *Aucun mot-clé.*

Domaine : Perception, Cognition, Interaction
Thème : Représentation et traitement des données et des connaissances

Période : 01/01/2011 -> 31/12/2011
Dates d'évaluation : 11/10/2011

Etablissement(s) de rattachement : U. MONTPELLIER 2
Laboratoire(s) partenaire(s) : <sans UMR>

CRI : Centre Inria d'Université Côte d'Azur
Localisation : Montpellier - LIRMM
Code structure Inria : 041127-0

Numéro RNSR : 201121208J
N° de structure Inria: SR0441RR

Présentation

Modern science such as agronomy, bio-informatics, and environmental science must deal with overwhelming amounts of experimental data. Such data must be processed (cleaned, transformed, analyzed) in all kinds of ways in order to draw new conclusions, prove scientific theories and produce knowledge. However, constant progress in scientific observational instruments and simulation tools creates a huge data overload. For example, climate modeling data are growing so fast that they will lead to collections of hundreds of exabytes expected by 2020. Scientific data is also very complex, in particular because of heterogeneous methods used for producing data, the uncertainty of captured data, the inherently multi-scale nature of many sciences and the growing use of imaging, resulting in data with hundreds of attributes, dimensions or descriptors. Processing and analyzing such massive sets of complex data is therefore a major challenge since solutions must combine new data management techniques with large-scale parallelism in cluster, grid or cloud environments. The three main challenges of scientific data management can be summarized by:

- scale (big data, big applications);
- complexity (uncertain, multi-scale data with lots of dimensions),
- heterogeneity (in particular, data semantics heterogeneity).

The overall goal of Zenith is to address these challenges, by proposing innovative solutions with significant advantages in terms of scalability, functionality, ease of use, and performance. We plan to design and validate our solutions by working closely with scientific application partners. To further validate our solutions and extend the scope of our results, we also want to foster industrial collaborations, even in non scientific applications, provided that they exhibit similar challenges.

Axes de recherche

Our approach is to capitalize on the principles of distributed data management. In particular, we plan to exploit: high-level languages as the basis for data independence and automatic optimization; data semantics (taxonomies, folksonomies, ontologies, ...) to improve information retrieval and automate data integration; declarative languages (algebra, calculus) to manipulate data and workflows, with user-defined functions; and exploit user (social) profiles and relationships between participants to help recommendation. Furthermore, we will exploit highly distributed environments in particular, P2P for data sharing between participants and parallel processing to scale up in the cloud. To reflect our approach, we organize our research program in three complementary research themes:

- Data and Metadata Management. This theme addresses the problems of

Contact

- **Responsable :** Patrick Valduriez
- **Tél :** 04.67.14.97.26
- **Secrétariat Tél :** 04.67.41.86.88

En savoir plus

- Site de l'équipe
- Site sur inria.fr
- Site du responsable
- Derniers Rapports d'Activité : 2016 , 2017 , 2018 , 2019 , 2020 , 2021 , 2022 , 2023

Documents sur la structure

- Intranet
- Privés

Décisions

- 7654 (21/12/2010) : création

Localisation

- **Adresse postale :** Plus utilisé
- **Coordonnées GPS :** 43.637, 3.841

managing and integrating data and metadata with uncertainty, in particular, n-way schema matching and distributed probabilistic query processing.
- Data and process sharing. This theme addresses the problems of scientific data and processes in highly distributed and parallel environments, in particular, social-based P2P data sharing and scientific workflow management.
- Scalable data analysis. Given the gap between the growth of computing power and that of data production, our ability to analyze these data is inevitably at stake. This theme addresses the scalability problem by investigating new data mining and content-based retrieval techniques that exploit parallelism in the cloud.

Relations industrielles et internationales

International

- Equipe Associée Sarava (2009-2011) with UFRJ, Rio de Janeiro, on P2P data management for online communities.
- CNPq-INRIA project DatLuge (Data & Task Management in Large Scale, 2010-2012) with UFRJ and LNCC, Rio de Janeiro, and UFPR, Curitiba on large scale scientific workflows.
- EGIDE Picasso project Scaling GraphDB (2010-2011) with UPC, Barcelona on very large graph database support.
- EGIDE Osmoze project SECC (SErvices for Curricula Comparison, 2011-2012), with Riga Technical University on automatic analysis and mapping of conceptual trees and maps acquired from digital documents.

Industry: Amadeus, Data Publica S.A., EADS, France Telecom, Mandriva, Nexedi, Xpertnet