

# Transparent and integrated access to distributed resources on the Grid

Genoveva Vargas-Solar, CNRS,  
HADAS group, LSR-IMAG, France

[Genoveva.Vargas@imag.fr](mailto:Genoveva.Vargas@imag.fr)

<http://www-lsr.imag.fr/Les.Personnes/Genoveva.Vargas>

## Data on the Grid

---

- Large geographically distributed datasets
- Data intensive queries and computationally intensive analysis
  - Advanced techniques that collectively maximize the use of resources
    - Careful management of terabyte caches
    - Gigabit/s data transfer over WAN's
    - Co-scheduling of data transfers and computation
    - Accurate performance estimations to guide the selection of dataset replicas
  - Complex performance demands not satisfied by any existing data management infrastructure

# Targeted issues

---

- Dynamic and inexplicit registry of data sources
- Discovery mechanisms for the data sources
- Unpredictable availability of data sources
  - High degree of autonomy
  - Intrinsically dynamic environment
  - Adaptable query execution mechanisms
- Data volume and number of sources vs. execution efficiency
- Efficient query evaluation techniques and optimization

## Scheduling

- Grid model: represent resources such as processors, disks, memory, and the network
- Task model: operator tree and the degrees of freedom that the scheduler is allowed
  - Consider space-time tradeoffs

## Optimization

- Execution space model: represents the space of choices
- Cost model definition

## Adaptive and interactive execution

- On-line monitoring services
- Adaptability and fault tolerance

3


# Current actions

---

## Grid: astronomy, geographic, biological and environmental data processing

- Biological data mediation and integration on the Grid
  - MEDIAGRID project, ACI-GRID 2002-2004, LSR, PRISM, LAMI
- Virtual observatory: ANDROMEDA
  - Highly distributed data mediation and integration
  - Intensive query evaluation scheduling on the Grid: V. Cuevas Vicettín
  - Virtual scientific experiments: CASCIOPEA, TAURUS
    - Scientific workflows
    - Semantic oriented experiment engineering
  - U. Cambridge, INAOE
  - Geographical and environmental data monitoring
  - Sensor based databases: National Polytechnic School of Turin, LSR, CENTIA
- Participation on the GRAMA working group of CUDI, Mexico (Internet2 action)

4



---

*“Any tool can be a weapon if it is handled in the correct way”*

*Empire, M. Hardt and T. Negri*

Merci ...  
Thanks ...