

Generic Applications Questionnaire

GRACE Project: <http://www.grace-ist.org/>

Section A: Describe your community

1) Is your community scientific or industrial ?

Both Industrial (2 partners) and Scientific (3 Universities, Cern).

The context is Digital Library and Information Science.

2) Describe concisely the “added value” for your community to run your application(s) at an European-wide scale specifying in particular what could be done on the grid infrastructure and could not be done at a local scale (max 500 words).

GRid seArch & Categorization Engine (GRACE) Project aims at fostering the creation of the next generation of GRID compliant, user-friendly, and cost-effective software for distributed search-and-retrieval through the exploitation of the capabilities offered by the Grid. By running the application at an European-wide scale, it will enable just-in-time, flexible allocation of computational and data storage resources, making terabytes of information distributed on vast amounts of geographically distant locations highly accessible.

3) In how many countries (N) in Europe is your community spread out ?

a. $N < 5$

b. $5 < N < 15$ *

c. $N > 15$

4) How many people (N) in your community will be using a grid infrastructure in the next year ?

• $N < 10$

• $10 < N < 100$ *

• $N > 100$

5) How many people (N) in your community will be using a grid infrastructure in the next two years ?

Do not know – GRACE project started in September 2003 and it’s going to end in February 2005.

• $N < 10$

• $10 < N < 100$

• $N > 100$

6) From how many sites (N) in Europe will people belonging to your community connect to the grid infrastructure in the next year ?

- a. $N < 10$ *
- b. $10 < N < 50$
- c. $N > 50$

7) From how many sites (N) in Europe will people belonging to your community connect to the grid infrastructure in the next two years ?

Do not know – GRACE project started in September 2003 and it's going to end in February 2005.

- a. $N < 10$
- b. $10 < N < 50$
- c. $N > 50$

8) Describe concisely (max 500 words) the security requirements of your community.

No requirements at present.

Section B: Describe your application no. 1

9) Is your application no. 1 scientific or industrial ?

Scientific

10) Describe concisely your application no. 1 (max 500 words) from the point of view of its goals and algorithms.

Expected Achievements:

- A distributed search engine adapted to the Grid-infrastructure and enhanced by a categorization engine which performs a semantic analysis of the retrieved documents and organize them according to the main concepts extracted.
- Handling of structured and unstructured textual information (text files, documents, Web pages, text stored in databases)
- The toolkit is also enhanced with multilingual abilities for the main Western European languages by integrating suitable lexical databases.
- GRACE services are accessible through an API interface and through an advanced Web GUI
- GRACE toolkit is built as a layer of services on top of the grid services (e.g. configuration, security, and data replication).

Technology:

- A set of algorithms handles categorization and classification on the fly and creates a series of semantic maps of large textual corpora.
- Analysis is done in Real Time

- Analysis is performed on the full document
- Ability to be always auto-updated
- No need for prior knowledge

11) Is your application no. 1:

- a. CPU intensive ?
- b. data intensive ?
- c. both ? *

12) Is your application no. 1 mainly intended for:

- interactive use ?
- batch use ?
- both ? * Both.

13) Quantitatively evaluate your application no. 1 in terms of:

- a. CPU power (SpecInt2000/SpecFp2000 per second per job);
- b. Memory consumption per job (Megabytes);
- c. Disk storage needs per job (Terabytes);
- d. Tape storage needs per job (Terabytes);
- e. Number of jobs per user per year;
- f. Number of users per year;
- g. Network bandwidth requirements (Megabit/sec).

Don't know

14) Quantitatively evaluate which percentage of the data of your application no. 1 needs to be replicated in more than one site and the average number of copies per elementary replicated data set (e.g., file).

No need.

15) Does your application no. 1 have a graphic or a command-line user interface ?

Graphic and API

16) Can your application no. 1 be accessed/steered from within a web browser ?

Yes

17) Is your application no. 1 already interfaced to any grid middleware ?

Yes

18) If yes to question 17, to which middleware and in the context of which project?

LCG-2, GRACE project

19) If yes to question 17, can you cite some references (less than 10) to related work ?

20) Did people of your community already attend any grid demos/tutorials ?

Yes

21) If yes to question 20, how many people per event ?

1 or 2

21) If yes to question 20, how many events of each kind (demos, tutorials) ?

3

22) Does your application no. 1 need third party commercial software to run ?

It only uses the search engine components from commercial partner (GL)

23) If yes, which one(s) ?

24) Describe the license under which your application no. 1 can be distributed on the grid infrastructure (max 200 words).

25) Describe concisely (max 500 words) the security requirements of your application no. 1.

26) If yes to questions 17 and 20, describe concisely (max 500 words) what problems you foresee to successfully port your application no. 1 on a distributed computing environment like an European grid infrastructure.

- Grid middleware unsuitability for the support of Real Time information processing and delivery of results
- Difficulties, delays in installation, use, update of European grid infrastructure/mw
- Possible lack of stability Grid mw
- Certification issue for users

Section C: Describe your application no. 2

27) Is your application no. 2 scientific or industrial ?

28) Describe concisely your application no. 2 (max 500 words) from the point of view of its goals and algorithms.

29) Is your application no. 2:

- a. CPU intensive ?
- b. data intensive ?
- c. both ?

30) Is your application no. 2 mainly intended for:

- interactive use ?
- batch use ?
- both ?

31) Quantitatively evaluate your application no. 2 in terms of:

- a. CPU power (SpecInt2000/SpecFp2000 per second per job);
- b. Memory consumption per job (Megabytes);
- c. Disk storage needs per job (Terabytes);
- d. Tape storage needs per job (Terabytes);
- e. Number of jobs per user per year;
- f. Number of users per year;
- g. Network bandwidth requirements (Megabit/sec).

32) Quantitatively evaluate which percentage of the data of your application no. 2 needs to be replicated in more than one site and the average number of copies per elementary replicated data set (e.g., file).

33) Does your application no. 2 have a graphic or a command-line user interface ?

34) Can your application no. 2 be accessed/steered from a web browser ?

35) Is your application no. 2 already interfaced to any grid middleware ?

36) If yes to question 35, to which middleware and in the context of which project ?

37) If yes to question 35, can you cite some references (less than 10) to related work ?

38) Did people of your community already attend any grid demos/tutorials ?

39) If yes to question 38, how many people per event ?

40) If yes to question 38, how many events of each kind (demos, tutorials) ?

41) Does your application no. 2 need third party commercial software to run ?

42) If yes, which one(s) ?

43) Describe the license under which your application no. 2 can be distributed on the grid infrastructure (max 200 words).

44) Describe concisely (max 500 words) the security requirements of your application no. 2.

45) If yes to questions 35 and 38, describe concisely (max 500 words) what problems you foresee to successfully port your application no. 2 on a distributed computing environment like an European grid infrastructure.

Section D: Describe your commitment

46) How many people/FTE's can your community dedicate in the next year to port your application(s) to the grid infrastructure in collaboration with grid experts ?

It has been already done.

47) From how many sites and with what distribution among sites ?

48) How many people/FTE's can your community dedicate in the next two years to port your application(s) to the grid infrastructure in collaboration with grid experts ?

Do not know, GRACE project started in September 2003 and it's going to end in February 2005.

49) From how many sites and with what distribution among sites ?

50) How many people/FTE's can your community dedicate in the next year to support users to use your application(s) on the grid infrastructure ?

4

51) From how many sites and with what distribution among sites ?

4 (1 person per site)

52) How many people/FTE's can your community dedicate in the next two years to support users to use your application(s) on the grid infrastructure ?

Do not know – GRACE project started in September 2003 and it's going to end in February 2005.

53) From how many sites and with what distribution among sites ?

Do not know

54) Which computing/storage resources can your community put as part of the grid infrastructure to run your application(s) in the next year ?

3 sites of about 3WNs, 1 SE, 1CE each

55) From how many sites and with what distribution among sites ?

3 (5/6 machines per site)

56) Which computing/storage resources can your community put as part of the grid infrastructure to run your application(s) in the next two years ?

Do not know – GRACE project started in September 2003 and it's going to end in February 2005.

57) From how many sites and with what distribution among sites ?

Do not know

58) Are those resources targeted for “exclusive use” by your community or can be shared with other applications running on the grid infrastructure ?

Shared

59) If yes to question 58, which is the percentage of use that can be allowed to other communities for their applications ?

Do not know

60) How many people/FTE's can your community dedicate in the next year to install, upgrade and manage grid middleware on the sites of your community ?

2-3

61) From how many sites and with what distribution among sites ?

3 (1 per site)

62) How many people/FTE's can your community dedicate in the next two years to install, upgrade and manage grid middleware on the sites of your community ?

Do not know – GRACE project started in September 2003 and it's going to end in February 2005.

63) From how many sites and with what distribution among sites ?

Do not know