DataGrid Project Status

Fabrizio Gagliardi – Project Manager
Fabrizio.Gagliardi@cern.ch

Outline

- Introduction to the review

- Achievements of the project

- Project deliverables

- Project budget

- Summary

Introduction to the Review

- ◆ Review structure

  - Today (AM): 2003 Overview and current status

  - Today (PM): Application assessment, dissemination and quality, Application demos

  - Tomorrow (AM): Middleware development, security, summary and future exploitation

- ◆ Details are provided in the deliverables and periodic reports

- ◆ Need to follow the agenda firmly so please hold questions till the end of presentations

DataGrid Summary

- Leading vehicle for Grid Research and deployment over the last three years
    - **Europe's flagship Grid Project**

- 2001
    - **Established organisational structures**
    - **Gathered/analysed requirements and surveyed software toolkits**
    - **Assembled large scale grid testbed and defined middleware architecture**

- 2002
    - **Improved services provided to application groups (quality issues in software, creation of documentation, tutorial programme)**

- 2003
    - **Built on solid technical base to provide advanced functionality to application groups and Grid related projects around the world.**

    The results will form the basis of EGEE

Themes of 3rd year of the project

- ◆ Complete the work as described in the technical annex

- ◆ Delivery and support the grid software for various national and international production grid programmes

- ◆ Define migration plan for future projects

- ◆ Elaborate exploitation plans

DataGrid Achievements in Brief

◆ **Complete the work as described in the technical annex**

- **Major enhancements in terms of functionality and stability in all middleware areas**
  - VOMS based security model integrated in job-submission chain
  - Replica Location Service (RLS) deployed
  - Re-factorization of WMS, edg-replica-catalog
  - Support for MPI, interactive, and checkpointable jobs
  - Unified access to storage (SE)
  - R-GMA replaced MDS as basis for the information system
  - New automated fabric tools developed
  - Many more (details in the individual WP presentations)

- **Successful deployment of middleware for use by applications**
  - Sept. 03: third major release of EDG middleware made available
  - Important improvements included in final EDG version 2.1 deployed end Nov. 03
  - Testbed focused on geographic distribution (more than 20 sites), not on computing power (many sites moved CPUs to production Grids)

DataGrid Achievements II

◆ **Applications assessed middleware on application testbed**

- Many scientific results achieved exploiting DataGrid testbed in all three application areas

- New applications ported to Grid (WP10)

- EO sites joined the testbed (WP9)

- Grid environments (e.g. LCG) already in production use for HEP (WP8)

◆ **Re-launched Application Working Group (AWG) has delivered important cross-application outputs**

- Re-launched March 2003 - Chaired by Vincent Breton (WP10)

- Interactions with GridLab Project

- Valuable recommendations and requirements

- Excellent forum for discussions

DataGrid Achievements III

◆ **Delivery and support the grid software for various national and international production grid programmes**

- **Successful delivery of middleware to the LCG production infrastructure:**
  - EDG software installed on more than 40 sites throughout the world. Next generation LCG-2 deployed January 2004

- **Datagrid Middleware being exploited by other projects and production facilities:**
  - LCG, EU Crossgrid Project, EU Grace Project, Italian INFN grid/grid.it, Dutch Dutchgrid, UK eScience programme, DataTAG, etc.
  - openlab: Involvement of industrial partners in Grid-related activities
  - EDG middleware being ported to different platforms such as IA64 as part of the CERN openlab project. (Enterasys Networks, HP, IBM, Intel Corporation, Oracle)
  - WP4 Quattor computing management tool now deployed at the CERN computing centre (approx. 2000 nodes). Other sites have expressed interest

DataGrid Achievements IV

- **Further important achievements**
  - **Europe-wide network activity**
    - Upgrade of numerous National research and Education networks
    - National networks inter-connected with GEANT
    - Grid data transport at GBit/s speed anticipated confidently
    - Close collaboration maintained with DataTAG and DANTE
  - **Active participation in international standards bodies**
    - Providing contributions to different GGF Working and Research groups
  - **Coordination and cooperation with related Grid Projects**
    - DataTAG, CrossGrid, GRIDSTART, GRACE, GridLab, etc.
  - **Open-style software license established**
    - Used throughout 2003 and submitted for approval as an OSI compliant license; model for CrossGrid and Globus license.
  - **Tutorials**
    - "Education & Outreach Manager" assigned to coordinate and lead tutorial activities.
    - More than 600 people trained in over 25 events
  - **Two Project conferences organized**
    - May 2003 in Barcelona, September 2003 in Heidelberg

DataGrid Third Year Deliverables

D1.6, D2.5, D3.5, D4.5, D5.5, D6.7: Components and documentation for the third project release

D1.7, D2.6, D3.6, D4.6, D5.6: Final Evaluation report

D6.8: Final evaluation of testbed operation

D7.4: Final report on network infrastructure and services

D7.7: Final security report

D8.4: Report on the results of HEP applications at run #2 and final application report

D9.4: EO application platform interface

D9.5: EP application processing testbed demonstration and final report

D10.4: Final report including report on the 2nd bio-testbed release

D11.6: Second annual conference and grid forum

D11.7: Final conference

D11.9: Contribution to international standards

D12.13-16: Third year quarterly reports

D12.19: Third year annual report

D6.6, D8.3, D9.3 and D10.3 which had been rescheduled for the first quarter of the third year were also delivered

Project Major Issues

- ◆ Build on fast moving technology

  - GT2 -> GT3 -> (GT4); OGSI -> WSRF

  - Decision at the beginning of year 3 to stay with GT2 and only prepare a few components for OGSI by building on plain web-services turned out to be a wise decision

- ◆ Relations with other Grid projects, LCG in particular

  - real production use with LCG required the adoption of more strict procedures

- ◆ Globus and other tools support

  - Now based on VDT; together with LCG an excellent relationship and support structures could be established

- ◆ **Third year funding was marginal for some partners**

  - This affected the work of some Work Packages (WP9, WP11 etc.)

- ◆ Persistency of project results

  - Exploitation of project results (in particular through EGEE)

  - More details will be given at the presentation at the end of the review

Relationship with LCG

- LCG is a grid deployment project
  - Provides the computing infrastructure for the LHC experiments
  - World-wide coverage
  - Production requirements with strict deadlines

- LCG has worked closely with DataGrid over the last two years
  - Has adopted software products from all DataGrid middleware WPs
  - Has adopted many DataGrid procedures and templates
  - Established joint testing-team with software certification and scalability tests
  - Rapid feedback has helped improve the quality of DataGrid software

Project Budget

Partners	Cost basis	Original Plan Y1	Reimbursed	Deviation Y1	Original Plan Y2	Reimbursed	Deviation Y2	Original Plan Y3	Claimed	Deviation Y3	EU ceiling	Total planned	Overall Deviation
CERN		494 597	393 547	-20%	588 221	640 957	9,0%	463 761	541 404	17%	1 546 579	1 575 908	1,90%
CERN coord.		130 666	88 682	-32%	120 055	126 922	5,7%	169 372	157 269	-7%	420 093	372 873	-11,24%
Total CERN	AC	625 263	482 229	-23%	708 276	767 879	8,4%	633 133	698 673	10%	1 966 672	1 948 781	-0,91%
ITC-irst	FC	102 170	78 033	-24%	88 920	66 511	-25,2%	83 120	127 660	54%	274 210	272 204	-0,73%
UH	AC	90 624	13 775	-85%	91 820	71 206	-22,5%	93 394	159 600	71%	275 838	244 581	-11,33%
VR	AC	124 844	0		123 466	132 999	7,7%	123 636	197 890	60%	371 946	330 889	-11,04%
ZIB	AC	90 156	76 487	-15%	93 156	105 785	13,6%	93 156	94 193	1%	276 468	276 465	0,00%
EVG HEI UNI	AC	90 156	59 436	-34%	93 156	79 657	-14,5%	93 156	121 748	31%	276 468	260 841	-5,65%
CNRS	FF	278 547	251 686	-10%	314 111	338 368	7,7%	318 732	343 791	8%	911 390	933 845	2,46%
CSSI	FF	165 100	177 828	8%	165 100	181 371	9,9%	165 200	138 092	-16%	495 400	497 291	0,38%
CEA	FC	60 264	62 281	3%	61 377	62 700	2,2%	62 604	92 963	48%	184 245	217 944	18,29%
IFAE	AC	64 392	70 740	10%	64 392	38 983	-39,5%	0	19 061		128 784	128 784	0,00%
ESA	AC	206 452	206 042	0%	254 873	274 437	7,7%	183 675	164 522	-10%	645 000	645 001	0,00%
INFN	AC	319 189	256 610	-20%	300 592	455 860	51,7%	300 592	219 127	-27%	920 373	931 597	1,22%
DATAMAT	FF	151 629	141 968	-6%	175 974	156 218	-11,2%	142 406	159 360	12%	470 009	457 546	-2,65%
CNR	FF	89 836	101 040	12%	89 836	91 382	1,7%	90 836	78 218	-14%	270 508	270 640	0,05%
CESNET	FF	45 656	51 071	12%	45 656	50 101	9,7%	45 706	47 766	5%	137 018	148 938	8,70%
FOM	AC	77 136	68 349	-11%	69 919	121 221	73,4%	71 125	81 868	15%	218 180	271 438	24,41%
KNMI	FC	59 655	66 580	12%	59 778	49 613	-17,0%	39 504	42 425	7%	158 937	158 618	-0,20%
SARA	FC	71 300	66 346	-7%	55 696	52 775	-5,2%	46 622	48 727	5%	173 618	167 848	-3,32%
PPARC	AC	396 916	97 504	-75%	386 404	386 310	0,0%	414 005	431 160	4%	1 197 325	914 974	-23,58%
MTA SZTAKI	FF	88 850	95 629	8%	94 367	94 417	0,1%	98 543	95 163	-3%	281 760	285 209	1,22%
IBM	FC	77 218	92 911	20%	79 429	85 604	7,8%	81 707	79 170	-3%	238 354	257 685	8,11%
Total		3 275 353	2 516 543	-23%	3 416 298	3 663 397	7,2%	3 180 852	3 441 177	8%	9 872 503	9 621 117	-2,55%

Table shows status at 31st December 2003

Project Effort

		FUNDED				UNFUNDED				F+UF			
		C	E	PM Dev.	% Dev.	C	E	PM Dev.	% Dev.	C	E	PM Dev.	% Dev.
CERN	AC	230	245	-15	-6%	546	424	122	29%	776	669	107	16%
ITC	FC	103	72	31	42%	3		3		106	72	34	47%
UH	AC	84	36	48	133%	60	36	24	67%	144	72	72	100%
VR	AC	48	54	-6	-11%	21	50	-29	-58%	69	104	-35	-33%
ZIB	AC	35	36	-1	-4%	37	36	1	4%	72	72	0	0%
EVG-HEI	AC	37	36	1	1%	45	36	9	24%	81	72	9	13%
CNRS	FF	310	305	5	2%	418	15	403		728	320	408	127%
CS SI	FF	78	81	-3	-4%	0				78	81	-3	-4%
CEA	FC	37	31	6	19%	57		57		94	31	63	203%
IFAE	AC	36	36	0	0%	62	42	20	47%	98	78	20	25%
ESA	AC	86	84	2	2%	82	13	69	532%	168	97	71	73%
INFN	AC	269	228	41	18%	645	733	-88	-12%	914	961	-47	-5%
DATAMAT	FF	116	108	8	7%	0				116	108	8	7%
CNR	FF	65	60	5	9%	64		64		129	60	69	115%
CESnet	FF	132	132	0	0%	12	12			144	144	0	0%
FOM	AC	42	36	6	17%	81	102	-21	-21%	123	138	-15	-11%
KNMI	FC	27	26	1	3%	9		9		36	26	10	37%
SARA	FC	28	30	-2	-7%	5		5		33	30	3	11%
PPARC	AC	167	156	11	7%	820	534	286	53%	986	690	296	43%
SZTAKI	FF	76	75	1	1%	39		39		115	75	40	53%
IBM	FC	38	36	2	5%	0				38	36	2	5%
Others (*)						433	610	-177	-29%	433	610	-177	-29%
Total		2042	1903	139	7%	3438	2643	795	30%	5480	4546	934	21%

Table shows status at 31st December 2003

Summary

◆ Project completed within budget and on schedule

◆ All partners actively participated and made it to the end

◆ DataGrid has gone the extra mile to produce output beyond the original contracted programme of work

◆ Functionality and performance of software and testbed(s) completed according to the original plans

◆ The value of EDG technology has been confirmed through its adoption by many projects including LCG for one of the largest scientific enterprises to-date

◆ Follow-on project successfully proposed and being launched (EGEE)