PPASCA Department Meeting
October 8, 2010

XLDB4
http://www-conf.slac.stanford.edu/xldb10/

Anders W. Borgland

XLDB

Attendance – Rough Breakdown

Invitational workshop				conf	
xldb1	xldb2	xldb3	xldb4		
43%	41%	45%	37%	24%	Data-intensive scientific users
21%	19%	15%	26%	42%	Data-intensive industrial users
30%	19%	23%	18%	9%	Vendors, incl. startups
6%	21%	15%	15%	24%	Academia, db research & programmers
0%	0%	2%	4%	1%	DOE, NSF, EC
53	62	52	57	150	←Total count

XLDB4 attendees:
North & South Americas, Europe, Asia, Australia

11

Workshop Program

Workshop Program

Printer-friendly version [PDF]

TUESDAY, OCTOBER 5, 2010		
08:00 AM	Continental Breakfast	
9:00 AM	**Welcome**	Jacek Becla
9:10 AM	**Oil/Gas**	moderator: Oliver Ratzesberger
10:10 AM	Coffee Break	
10:30 AM	**Finance**	moderator: Raghu Ramakrishnan
11:30 AM	**Medical / Bioinformatics** Including medical bioinformatics, genomics, perspectives from science and industry.	moderator: Kian-Tat Lim
12:30 PM	Lunch, discussion in small groups about expanding and strengthening XLDB community	
1:30 PM	**Science Benchmark** Current status, next steps	Michael Stonebraker
2:00 PM	**Approaches to Big Data Statistical Analytics**	moderator: Curt Monash
3:00 PM	Coffee Break	
3:20 PM	**Emerging Challenges** Solid state disks, GPUs, very-many-cores, fast networks	moderator: Alex Szalay
4:20 PM	**Writable Extreme Scale Databases** Extremely large non-read-only databases (beyond bulk load)	moderator: Magdalena Balazinska
5:00 PM	**Plan Collaborative Projects and Next Steps** Planning collaborative projects (within USA and international), future funding, next steps	moderator: Jacek Becla
5:30 PM	Adjourn	

Conference Program

Conference Program

WEDNESDAY, OCTOBER 6, 2010		
08:00 AM	Continental Breakfast	
9:00 AM	**Welcome -- Science and Computing at SLAC** Official "welcome" and introduction to new exciting science and computing challenges at SLAC.	Donald Lemma (SLAC CIO and Computing Division Director)
9:15 AM	**Conference Introduction and Logistics** Main objectives, logistics and agenda.	Jacek Becla (SLAC, XLDB4 chair)
9:30 AM	**Complex Scientific Analytics at Extreme Scale** A comprehensive overview of how big science approaches complex analytics at extreme scale.	Gregory Dubois-Felsmann (HEP), Andrew Connolly (astronomy), John Caron (atmospheric research), Bill Howe (ocean sciences, bio), Jacek Becla (summary)
10:50 AM	Coffee Break	
11:10 AM	**Complex Industrial and Government Analytics at Extreme Scale** A comprehensive overview of how data-intensive industries approach complex analytics at extreme scale, it will highlight similarities and differences comparing to approaches taken by big science.	Irina Vayndiner (MITRE), Steve Hirsch (NYSE Euronext), Mike McIntire (Yahoo!), Damian Reeves (Quantcast) moderator: Kian-Tat Lim (SLAC)
12:30 PM	Lunch	
1:30 PM	**Operational Issues with Managing Large Database Clusters** Practical, operational issues with managing large database clusters, based on experiences from at least two large-scale industrial setups.	Oliver Ratzesberger (eBay), Jeffrey Rothschild (Facebook)
2:10 PM	**Behind the Scenes of Big Science Projects** A talk explaining how big and long-term scientific projects get started, what the decision processes are, how requirements and data volumes are decided, how vendors are evaluated.	Amber Boehnlein (DOE)
2:55 PM	**Existing Scientific Tools/Formats - netCDF, HDF5, fits, xts** A set of short talks to make non-scientific communities aware of most commonly used scientific formats and related tools custom-built by scientists.	Daniel L. Wang (SLAC)
3:15 PM	Ice Cream Social - Poster Session for Gold Sponsors	
3:55 PM	**Existing Scientific Tools** A continuation of short talks to make non-scientific communities aware of most commonly used scientific tools custom-built by scientists. Considering covering: root, castor, xrootd.	Richard Dubois (SLAC)
4:15 PM	**Lightning Talks (8 x 5 min)** 1. *A (Hypothetical) Data to Discovery Engine* 2. *Big data analytics - use of a highly scalable MPP database on a cloud-based multi-instance platform to reliably deliver near real-time analytics* 3. *Exabyte plus initiative* 4. *Damasc* 5. *Horizontal virtualization on commodity hardware without requirements for database optimization techniques* 6. *Array Versioning System* 7. *Smorgasbord of Real World Extreme Scale Database Analytics* 8. *Introduction to MongoDB*	1. Mark Stalzer, Caltech 2. DJ Das, Third Eye Cloud 3. Leon Guzenda, Objectivity 4. Neoklis Polyzotis, UCSC 5. Stefan Groschupf, Datameer 6. Philippe Cudre-Mauroux, MIT 7. Andrew Lamb, Vertica 8. Shulman, William, MongoDB
5:00 PM	Adjourn	
6:30 PM	**Reception and dinner**	

THURSDAY, OCTOBER 7, 2010		
08:00 AM	Continental Breakfast	
8:40 AM	**Announcements and Logistics**	Jacek Becla
8:45 AM	**Emerging Technologies for Complex Extreme Scale Analytics** Unifying emerging technologies such as map/reduce, streaming databases, and workflow management into a coherent tool set for extreme scale data analytics.	Jeff Hammerbacher (Cloudera)
9:25 AM	**Emerging Scientific Tools - SciDB** Lessons learned from studying scientific use cases and interacting with scientific communities by the SciDB team. SciDB response to these needs.	Mike Stonebraker (MIT)
9:45 AM	**Science Benchmark** Introducing the idea behind science benchmark, current status and plans.	Mike Stonebraker (MIT)
10:00 AM	Coffee Break	
10:40 AM	**Extreme Scale Architectures and New Hardware Trends** Impact of new hardware trends such as solid state disks, GPUs, servers with very-many-cores.	Alex Szalay (JHU)
11:25 AM	**Data Preservation and Integration** Challenges related integrating data from multiple sources and preserving petabytes of data.	Jane Mandelbaum (Library of Congress)
12:05 PM	**Automated Information Extraction, Content Curation and Machine Learning** Needs in the area of automating data processing and analytics. Perspectives from industry and science.	Raghu Ramakrishnan (Yahoo!), Kirk Borne (GMU)
12:45 PM	**Closeout** Next conference planning, final conclusions and closeout.	Jacek Becla
1:00 PM	Lunch	
2:00 PM	Adjourn	

10/1/2010