Welcome to CIDR 2009 at the Asilomar Conference Grounds! This year's conference contains two distinct "tracks":

> Proceedings of CIDR (Academic Peer Review): This is a traditional conference proceedings for peer-reviewed scholarly papers.

> CIDR Perspectives (Editorial Selections): This track is a venue for technical presentations to inform and provoke discussion in the research community.

Talks for the two tracks are interleaved. CIDR Perspectives papers are marked in the program with an asterisk (✳)

Sunday January 4, 2009

12-6 PM	Check-In at Asilomar Conference Grounds
6:00 PM	Dinner
7:00 PM	Evening Get-Together

Monday January 5, 2009

7:00 AM	Breakfast
8:30 AM	Invited Talk 1

Jeff Heer, Stanford: Voyagers and Voyeurs: Supporting Social Data Analysis

In recent years, researchers and entrepreneurs have introduced new online services for data collection and analysis, often relying on interactive visualizations to enable mass interaction with data. These sites represent the first step in what looks to be a growing online phenomenon: social data analysis, that is, collective analysis of data supported by social interaction. Engaging crowds of both experts and non-experts in the process of data exploration has applications ranging from political transparency to business intelligence to citizen science. Achieving this vision, however, will require further innovation in the design of systems and interfaces for collaborative data management.

In this talk I will highlight recent efforts to use the web as a platform for collectively creating, managing, and analyzing data. I will share how web citizens are currently collaborating with data, and discuss how we might design systems that better catalyze social forms of data management and exploration.

Biography:
Jeffrey Heer is an Assistant Professor of Computer Science at Stanford University, where his research focuses on human-computer interaction, interactive visualization, and social computing. His work has produced novel visualization techniques for exploring data, software tools that simplify visualization creation and customization, and collaborative analysis systems that leverage the insights of multiple analysts. He is the author of the prefuse and flare open-source visualization toolkits, currently in use by the visualization research community and numerous corporations. Over the years, he has also worked at Xerox PARC, IBM Research, Microsoft Research, and Tableau Software. He holds B.S., M.S., and Ph.D. degrees in Computer Science from the University of California, Berkeley.

9:30 AM	People, Games and Trust

Social Systems: Can We Do More Than Just Poke Friends?
Georgia Koutrika, Benjamin Bercovitz, Robert Ikeda, Filip Kaliszan, Henry Liou, Zahra Mohammadi Zadeh, Hector Garcia-Molina (Stanford University)

✻ From Declarative Languages to Declarative Processing in Computer Games
Benjamin Sowell, Alan Demers, Johannes Gehrke, Nitin Gupta, Haoyuan Li, Walker White (Cornell University)

Declarative Reconfigurable Trust Management
William Marczak (University of Pennsylvania), David Zook (LogicBlox), Wenchao Zhou (University of Pennsylvania), Molham Aref (LogicBlox), Boon Thau Loo (University of Pennsylvania)

10:30 AM	Coffee Break
11:00 AM	Text, XML and Integration

Interactive Data Integration through Smart Copy & Paste
Zachary Ives (University of Pennsylvania), Craig Knoblock (University of Southern California - Information Sciences Institute), Steve Minton (Fetch Technologies), Marie Jacob (University of of Pennsylvania), Partha Talukdar (University of Pennsylvania), Rattapoom Tuchinda (University of Southern California - Information Sciences Institute), Jose Luis Ambite (University of Southern California - Information Sciences Institute), Maria Muslea (University of Southern California - Information Sciences Institute), Cenk Gazen (Fetch Technologies)

Search Driven Analysis of Heterogenous XML Data
Andrey Balmin (IBM Almaden Research Center), Latha Colby (IBM Almaden Research Center), Emiran Curtmola (UC, San Diego), Quanzhong Li (IBM Almaden Research Center), Fatma Ozcan (IBM Almaden Research Center)

Qunits: queried units in database search
Arnab Nandi (University of Michigan), H V Jagadish (University of Michigan)

12:00 PM	Lunch

1:30 PM	Streaming and Remembering

[Boosting XML filtering through a scalable FPGA-based architecture](#)
 Abhishek Mitra (UC Riverside), Marcos Vieira (UC Riverside), Petko Bakalov (ESRI), Vassilis Tsotras (UC Riverside), Walid Najjar (UC Riverside)

[Capturing Data Uncertainty in High-Volume Stream Processing](#)
Yanlei Diao (UMass Amherst), Boduo Li (UMass Amherst), Anna Liu (UMass Amherst), Liping Peng (UMass Amherst), Charles Sutton (UC Berkeley), Thanh Tran (UMass Amherst), and Michael Zink (UMass Amherst)

[Inter-Operator Feedback in Data Stream Management Systems via Punctuation](#)
Rafael Fernandez-Moctezuma (Portland State University), Kristin Tufte (Portland State University), Jin Li (Portland State University)

✳ [Remembrance: The Unbearable Sentience of Being Digital](#)
Ragib Hasan (University of Illinois), Marianne Winslett (University of Illinois), Radu Sion (Stony Brook University)

3:00 PM	Coffee Break
3:30 AM	Storage Redux

[uFLIP: Understanding Flash IO Patterns](#)
Luc Bouganim (INRIA), Björn Þór Jónsson (Reykjavík University), Philippe Bonnet (University of Copenhagen)

✳ [Unbundling Transaction Services in the Cloud](#)
David Lomet (Microsoft Research), Alan Fekete (University of Sydney), Gerhard Weikum (Max Plank Institute), Mike Zwilling (Microsoft SQL Server)

✳ [The Case for RodentStore: An Adaptive, Declarative Storage System](#)
Philippe Cudre-Mauroux (MIT), Eugene Wu (MIT), Sam Madden (MIT)

 [Teaching an Old Elephant New Tricks](#)
 Nicolas Bruno (Microsoft)

6:00 PM	Dinner
7:15 PM	Gong Show

Tuesday January 6, 2009

7:00 PM	Breakfast
8:30 PM	Giant Steps

[Extracting and Querying a Comprehensive Web Database](#)
Michael Cafarella (University of Washington)

✳ [Interactive Analysis of Web-Scale Data](#)
Christopher Olston, Edward Bortnikov, Khaled Elmeleegy, Flavio Junqueira, Benjamin Reed (Yahoo! Research)

✳ [Continuous Analytics: Rethinking Query Processing in a Network-Effect World](#)
Michael Franklin (Truviso), Sailesh Krishnamurthy (Truviso), Neil Conway (UC Berkeley), Alan Li (Truviso), Alex Russakovsky (Truviso), Neil Thombre (Truviso)

[A Scalable Data Platform for a Large Number of Small Applications](#)
Fan Yang (Cornell University), Jayavel Shanmugasundaram (Yahoo! Research), Ramana Yerneni (Yahoo! Research)

10:30 AM	Coffee Break
11:00 AM	Gong Show Awards
11:15 AM	Schemas and Processes

✳ [The Role of Schema Matching in Large Enterprises](#)
Ken Smith, Michael Morse, Peter Mork, Maya Li, Arnon Rosenthal, David Allen, Len Seligman (MITRE)

✳ [A Case for A Collaborative Query Management System](#)
Nodira Khoussainova, Magda Balazinska, Wolfgang Gatterbauer, YongChul Kwon, Dan Suciu (University of Washington)

[Do-It-Yourself custom forms-driven workflow applications](#)
Keith Kowalczykowski (app2you.com), Kian Win Ong (UCSD), Kevin Keliang Zhao (UCSD), Alin Deutsch (UCSD), Yannis Papakonstantinou (UCSD), Michalis Petropoulos (SUNY Buffalo)

12:30 PM	Lunch

1:30 PM	Feedback on Queries

Why Did My Query Slow Down?
Nedyalko Borisov (Duke University), Sandeep Uttamchandani (IBM), Ramani Routray (IBM), Aameek Singh (IBM)

Visualizing the robustness of query execution
Goetz Graefe, Harumi Kuno, Janet Wiener (Hewlett-Packard Co.)

❋ DBMSs Should Talk Back Too
Alkis Simitsis (HP Labs), Yannis Ioannidis (University of Athens)

3:00 PM	Snack and Free Time
6:00 PM	Dinner
7:15 PM	Perspective Panel 1: Gathering Clouds

❋ Building on Quicksand
Pat Helland, David Campbell (Microsoft)

❋ Principles for Inconsistency
Shel Finkelstein, Dean Jacobs, Rainer Brendle (SAP)

❋ SCADS: Scale-Independent Storage for Social Computing Applications
Michael Armbrust, Armando Fox, David Patterson, Nick Lanham, Beth Trushkowsky, Jesse Trutna, Haruki Oh (UC Berkeley)

8:15 PM	Perspective Panel 2: Pouring Data

❋ Harnessing the Deep Web: Present and Future
Jayant Madhavan (Google), Loredana Afanasiev (Universiteit van Amsterdam), Lyublena Antova (Cornell University), Alon Halevy (Google)

❋ The Case for a Structured Approach to Managing Unstructured Data
AnHai Doan, Jeffrey F. Naughton, Akanksha Baid, Xiaoyong Chai, Fei Chen, Ting Chen, Eric Chu, Pedro DeRose, Byron Gao, Chaitanya Gokhale, Jiansheng Huang, Warren Shen, Ba-Quy Vuong (Wisconsin)

❋ SocialScope: Enabling Information Discovery on Social Content Sites
Sihem Amer-Yahia (Yahoo! Research), Laks Lakshmanan (UBC), Cong Yu (Yahoo! Research)

❋ Sailing the Information Ocean with Awareness of Currents: Discovery and Application of Source Dependence
Laure Berti-Equille (Universite de Rennes 1), Anish Das Sarma (Stanford University), Xin (Luna) Dong (AT&T Labs-Research), Amelie Marian (Rutgers University), Divesh Srivastava (AT&T Labs-Research)

Wednesday January 7, 2009

7:00 AM	Breakfast
8:30 AM	Invited Talk 2

James Hamilton, Amazon: Internet-Scale Data Center Power Efficiency

The cost advantages of high-scale services make them the dominant means of delivering consumer software solutions, a growing part of enterprise IT infrastructure, and an inevitable part of our future. The application of high-scale services to an ever expanding set of problems is bounded only by the value of the problem and the cost of the solution. As solution costs fall, the number of problems that can be cost effectively addressed continues to increase. Power is the limiting factor. This talk inventories 1) where power is spent in a high-scale data center, focusing first on power distribution, from the property line, through PDUs, switchgear, transformers and the UPS, to CPU & Memory, 2) how servers are designed, 3) the software problems around raising utilization and efficiency, and 4) mechanical cooling systems design. The focus is to understand exactly where the power goes in a high-scale data center, some of what can be done to improve efficiency, including results from the Cooperative, Expendable, Micro-Slice Servers (CEMS) project, and some of the interesting software issues.

Biography:
James Hamilton recently joined Amazon.com after twelve years at Microsoft, where he was Data Center Futures architect. Prior to that, James was architect on the Microsoft Live Platform Services team, and earlier led the Exchange Hosted Services team that provided email-related services to over two million users. He spent much of his first eight years at Microsoft as a member of the SQL Server team, where he led many of the core engine development teams.

Before joining Microsoft, James was lead architect for IBM's DB2 UDB relational database system, and earlier led the delivery of IBM's first C++ compiler. In the late 70's and early 80's he worked as a licensed auto mechanic servicing and racing exotic Italian cars.

James' web site is http://mvdirona.com/jrh/work, his blog is http://perspectives.mvdirona.com and his email is james@amazon.com.

9:30 AM	Green Data

✳ Energy Efficiency: The New Holy Grail of Data Management Systems Research
Stavros Harizopoulos (HP Labs), Mehul Shah (HP Labs), Parthasarathy Ranganathan (HP Labs)

Towards Eco-friendly Database Management Systems
Willis Lang (University of Wisconsin-Madison), Jignesh Patel (Wisconsin)

10:15 AM	Coffee Break
11:00 AM	Scientific and Numeric Data

Data Management for High-Throughput Genomics
Uwe Roehm (University of Sydney), Jose Blakeley (Microsoft)

LifeRaft: Data-Driven,Batch Processing for the Exploration of Scientific Databases
Xiaodan Wang (Johns Hopkins University), Randal Burns (Johns Hopkins University), Tanu Malik (Purdue University)

✳ Requirements for Science Data Bases and SciDB

Michael Stonebraker (MIT), Jacek Becla (SLAC), David Dewitt (Microsoft), Kian-Tat Lim (SLAC), David Maier (Portland State University), Oliver Ratzesberger (eBay, Inc.), Stan Zdonik (Brown University)

RIOT: I/O-Efficient Numerical Computing without SQL
Yi Zhang (Duke University), Herodotos Herodotou (), Jun Yang (Duke)

12:20 PM	Best Paper Award and Closing Remarks
12:30 PM	Lunch

9