ABSTRACT:
I will talk about Google's search cluster architecture that uses more than 15000 commodity class PCs to deliver superior search performance at a fraction of the cost of a system built from fewer, but more expensive, high-end servers. I'll mention some of the challenges we have faced in making such large scale deployment of commodity hardware work well. In particular, I will talk about some of our distributed systems software infrastructure that we have built to manage the underlying unreliable hardware and make rapid development and deployment of new scalable web services like Gmail, Google Earth etc.

PROFILE:
Arvind Jain is a member of Google's systems laboratory and currently the head of Google India R&D center. At Google, he has worked on various infrastructure projects including the crawl and indexing system, distributed file replication system, and compression techniques for large scale storage systems. Before joining Google, he was the founding engineer at Riverbed Corporation, a networking startup that builds caching and compression based appliances to speed up WAN interactions from remote offices in an enterprise. Prior to Riverbed, he was the architect of the streaming and storage division at Akamai Technologies where he helped build Akamai's distributed video on demand and live broadcasting service. He started his career at Microsoft Corporation where he worked on various components in the Windows NT kernel. He holds an M.S. and BTech in Computer Science from University of Washington and Indian Institute of Technology, Delhi, respectively.