

Is It Still “Big Data” If It Fits In My Pocket?

Keynote VLDB 2011

David Campbell
Microsoft

<http://www.microsoft.com/presspass/exec/techfellow/Campbell/>

ABSTRACT

“Big Data” is a hot topic but, in many ways, we are still trying to define what the phrase “Big Data” means. For many, there are more questions than answers at this point. Is it about size alone? Complexity? Variability? Data shape? Price/performance? New workloads? New types of users? Are existing data models, data management systems, data languages, and BI/ETL tools relevant in this space? Is MapReduce really a “major step backwards”? I have spent time over the last several years trying to answer many of these questions to my own satisfaction. As part of the journey I have witnessed a number of natural patterns that emerge in big data processing. In this talk I will present a catalog of these patterns and illustrate them across a scale spectrum from megabytes to 100s of petabytes. Finally, I will offer some thoughts around a systems and research agenda for this new world.

1. ABOUT THE PRESENTER



David Campbell is a Microsoft Technical Fellow working in Microsoft Corp.s Server and Tools Business. Campbell joined Microsoft in 1994 from Digital Equipment Corp. as Microsoft began its push to become a credible enterprise software vendor. His early work at Microsoft included creating an OLE DB interface over the existing SQL Server storage engine, which helped to boot-

strap SQL Servers present-generation query processor.

He also worked closely with Mohsen Aghen, another Microsoft Technical Fellow, and the Microsoft Transaction Server team to add distributed transaction support to SQL Server 6.5. Microsoft made a bold move to re-architect SQL Server for the SQL Server 7.0 release. As a key technical member of the storage engine team, Campbell implemented the SQL Server lock manager and other critical concurrency control mechanisms. He also implemented row-level locking in SQL Server 7.0, one of the hallmark features of the release.

Through the SQL Server 2000 and SQL Server 2005 releases, Campbell served in a variety of roles including product-level architect and general manager of product development. After the SQL Server 2005 release, he led a small team in redesigning SQL Server product development methodology. The new process, used to produce SQL Server 2008, resulted in SQL Server 2008 having the highest initial quality levels of any SQL Server release to date. As of August 2010, Campbell is serving as general manager of Microsofts Data and Modeling Group, which oversees Microsofts data modeling and data access strategies.

Campbell holds a number of patents in the data management, schema and software quality realms. He is also a frequent speaker at industry and research conferences on a variety of data management and software development topics. His current product development interests include cloud-scale computing, realizing value from ambient data, and multidimensional, context-rich computing experiences.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Articles from this volume were invited to present their results at The 37th International Conference on Very Large Data Bases, August 29th - September 3rd 2011, Seattle, Washington.

Proceedings of the VLDB Endowment, Vol. 4, No. 11

Copyright 2011 VLDB Endowment 2150-8097/11/08... \$ 10.00.