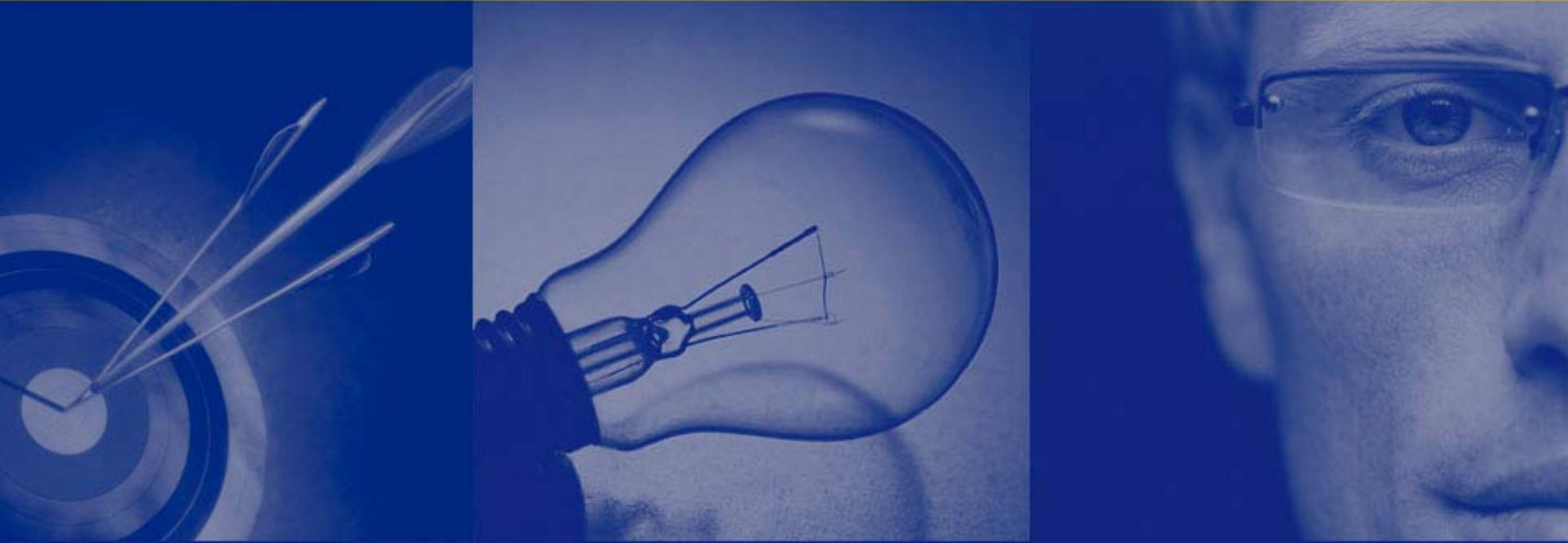


Surviving the Data Avalanche



Presented by Kevin Kline
SQL Server MVP since 2004

Agenda

- Speaker bio
- About Quest Software
- Accolades & Awards
- Surviving the Data Avalanche
- Resources
- Q & A

Kevin Kline



- *Technical Strategy Manager for SQL Server Solutions, Quest Software Inc.*
- Microsoft SQL Server MVP
- President of the International SQL Server User's Group – PASS
- SQL Server expert and author of the O'Reilly titles "SQL in a Nutshell" and "Transact-SQL Programming"

Speaker Bio – Kevin Kline

- Started in IT in 1986. BS in MIS in 1989 from University of Alabama.
- Microsoft SQL Server MVP since 2004
- Author on 7 database books
 - 1 on Oracle, 2 on SQL Server, 2 on SQL, 1 on DB design, 1 on DB Benchmarking
- President of PASS (www.sqlpass.org)
 - Conference is next Nov in Seattle
 - Over 130 sessions on SQL Server, BI & Dev
- Blogs for SQLMag and SQLBlog
- Monthly columns in SQL Server Magazine and Database Trends & Applications
- Worked for NASA, US Army, and Deloitte & Touche before Quest Software.

Accolades & Awards

- TechTarget 2006 Product of the Year with LiteSpeed
- Best of Tech Ed award 2006 with Spotlight on SQL Server
- SQL Server Magazine, Platinum reader's choice award
- SQL Server Magazine Readers Choice Awards, winner in 11 categories, 2005
- No. 1 vendor in Distributed Data Management Facilities, IDC, 2005
- Microsoft ISV Partner of the Year, 2004 and 2007
- Microsoft TechEd Best of Show Winner, 2004
- Microsoft TechEd Europe Best of Show Winner, 2005
- No. 1 vendor in Application Management Software, Gartner Dataquest, 2005
- Jolt Productivity Award Winner
- Network Computing Editor's Choice Winner
- No. 8 in the "Who's Who in Enterprise Software," Investor's Business Daily



What is a VLDB?

- Used to be a set size:
 - >100mb in the 1980's
 - >100gb in the early 1990's
 - >1tb in the early 2000's
 - >10tb in the mid 2000's
- Now, a more flexible definition for VLDB prevails:
 - It's a VLDB when you can no longer use "common and standard" techniques for management and performance

How big are SQL Server databases today?

- Hundreds of multi-terabyte databases on SQL Server are known. At least an equal number are unknown.
- An example Microsoft customer has a 270tb data warehouse
 - Growing 3tb per day, with 1tb of deletions per day
 - Definitely a VLDB
- An example EMC customer in the DOE has a 200tb database (with lots of blob data)
 - Possibly not a VLDB

Why do apps generate more data today?

- **Compliance and Auditing Requirements**
 - Especially for telecom, finance, and health industries
 - Requirements to retain data for 7-10 years
 - SOX
 - Internal processes in industries like insurance
- **Real Time Business Intelligence**
 - Old ETL/Mainframe-derived data is losing ground
 - Live streaming data is producing more accurate, more timely decision-making
 - Data is both more granular and more records (e.g. drive-through fast-food vendors in the restaurant industry)
 - More sophisticated business processes
 - Longer-term data is still kept on-line for better decision making (e.g. the leap-year birthday phenomenon in the retail industry)

Why else is there more data today?

- Backup retention is a HUGE driver
- Recovery from tapes is losing ground to:
 - High-availability alternatives (e.g. hot standby's, clustering, replicated data) are often easier than tapes:
 - Old versions of application aren't around
 - Platform software (OS, database, drivers) aren't the right versions
 - Tape drives or other important hardware may not be around
 - Tiered storage is a growing alternative
 - Replace the "sliding window" of data with a tiered set of storage based on performance and cost
- Recovery is the overlooked component of backup strategy. TEST!!!

What other problems arise with VLDBs?

- Server proliferation becomes a big headache with VLDBs
 - Additional servers are needed for high-availability, replicated, or hot standby servers
 - Often, architecture divides processing onto a several servers
 - ETL servers
- KISS principles dictate that when you have something big and unwieldy, that you break it down into to more manageable components
 - Partitioning in SQL Server 2005
 - Metadata server with partitioned data warehouses in BI VLDBs
- Loading data and, separately, cleaning data is an enormous issue

Database Management Challenges

- Backup and Recovery Operations
- Finding and Fixing SQL Server Performance Issues
- Monitoring and Tracking Changes made to your SQL Server Environment
- Planning and Forecasting Storage and Resource Requirements

Specific Challenges: Backup and Recovery

- How to reduce storage costs in the face of growing datasets
- Improving backup performance to ensure maintenance windows are met
- Ensuring fast and efficient restore in a crisis
- Restoring data at the table or schema level

Specific Challenges: Managing Performance

- Managing for planned and unplanned changes
- Preventing performance issues before they happen
- Diagnosing and resolving issues once identified
- Gathering performance information from multiple instances

Specific Challenges: Capacity and Resource Planning

- Getting a clear picture disk consumption across your production database environment
- Planning for future disk need
- Ensuring that disk space and server resources are being used efficiently
- Effectively report disk and server resource consumption metrics to management
- If table partitioning is implemented, managing this powerful feature

Specific Challenges: Managing Database Changes

- Knowledge of who is accessing or trying to access the database?
- How to quickly tell who made a change to one of my critical databases
- The database is underperforming. Have any significant changes been made since yesterday?
- Producing management reporting relative to database change controls
- Are my junior DBAs making changes they should not be?

Storage Strategies – Tiered Storage

- Tapes don't get upgraded when the system gets upgraded
- Some progressive customers are using tiered storage
 - Active application data is on state-of-the-art disk arrays (e.g. RAID10, high RPM speeds)
 - Near-term older data is kept on less expensive disk arrays (e.g. RAID5, middle RPM speeds)
 - Old, long-term data is kept on very inexpensive SATA drives (e.g. high volume, low RPM speeds) *and never deleted*

Storage Strategies – Disk and IO Design

- Storage admin might give the DBA disk according to the volume s/he needs, but not the IO
 - Schema design is key; not always a knob to turn
 - Schema design (done poorly) often contributes to data bloat through poor normalization and/or poor choice of data types
 - Misaligned block sizes
 - LUNs set up for serial IO transaction load, not large block reads
 - Exchange best practices are carried forward to SQL Server *even when they don't apply!*
 - OLTP and OLAP have conflicting needs (IO/sec versus MB/sec)
- Remedy of first resort is often “throw more hardware at it”

Storage Strategies – Personnel

- Applications need an overall architect!
- Personnel (storage admin, database admin) within a company often don't communicate
- Ensure that there's one version of the "truth"
 - Developers aren't always thinking beyond the deadline
 - Coding to business requirements, not performance requirements
 - Often thinking as in row-based mode
 - Many companies could benefit from a *database programmer* role
- DBAs very often spend a LOT of time fixing the bad code of other people

Storage Strategies – SAN

- Other apps may suck up all the cache.
- Other apps may suck up all of the IO on the SAN
- DBAs often data the volume of storage they ask for, but don't know to also ask for IO as well!
- Test your SAN to ensure it carries the load effectively

Managing VLDBs, part 1

- Backup

- You have to get clever with VLDBs, using something like database snapshots
- Date range backups; e.g. older data doesn't change, so you don't need to back it up
- Partitioning, especially read-only partitions, can amplify your ability to manage a SQL Server system – not only backups, but also indexing and defragmentation
- Serialized may be better than parallel backup
- More thought needs to go into multi-database backup and recovery!
- More databases on a single instance means more IO contention and shorter maintenance windows

- Indexing

- New on-line indexing opens up new capabilities; slower due to trickle effect, but on-line through-out the process

Managing VLDBs, part 2

- Partitioning
 - Multiple IO paths to the SAN
 - Allows parallel indexing, backups, data cleanup operations, and much more
 - Be careful of updating statistics and DBCC commands. They run per tables across all partitions.
 - Sexy!
- Transaction Processing
 - Views can help mitigate transactions that have run amok
 - Views work on all versions of SQL Server
 - Partitioning is a big help in SQL Server 2005
- ETL
 - Smart clustering can facilitate parallel loads, even to a single table
 - Load jobs are long running, single-threaded jobs – thus tied to a processor – meaning that no more load jobs than number of CPUs.

Quest Management Suite for SQL Server



Resources

- Review my recorded webcast on “SQL Server 2005 Disk IO Tuning” at <http://www.quest.com>
- Download my white papers at Quest.com, read my blog at <http://www.sqlmag.com> and <http://sqlblog.com>
- See the full panel discussion from the PASS 2006 Community Summit at <http://www.quest.com/events/listdetails.aspx?ContentID=4688&site=&prod=&technology=&prodfamily=&loc=>
- Project REAL at <http://www.microsoft.com/sql/solutions/bi/projectreal.mspx>
- SQL Server Customer Advisory Blog at <http://blogs.msdn.com/sqlcat/>
- Storage Top 10 Best Practices at <http://www.microsoft.com/technet/prodtechnol/sql/bestpractice/storage-top-10.mspx>

Call to Action – Next Steps

- Learn more about Quest Management Suite for SQL Server: http://www.quest.com/sql_server/
 - Download trials
 - Read white papers
 - Review case studies
- Ask for a proof-of-concept or personalized demo!

Questions & Answers

Thank you for attending this session and the

More questions? Email me at kevin.kline@quest.com