SQLCAT – Largest SQL Server Projects in the World

Knowing What is Possible

Kevin Cox
SQL Server Customer Advisory Team
kevincox@Microsoft.com

Global Sponsors:
SQL CAT (Customer Advisory Team)

The **SQL Server Customer Advisory Team (SQL CAT)** represents the customer-facing resources from the SQL Server Product Group. SQL CAT is comprised of product and solution experts that regularly engage in the largest, most complex, and most unique customer deployments worldwide.

**Achieving Customer Success**
- Very large, complex projects
- Mission Critical projects

**Azure Projects**
- Are you pushing the limits with Azure? Talk to AppFabric CAT and SQL CAT

**Making a Better Product**
- Drives feedback and product requirements back into SQL Server development teams from deep and strategic customer and ISV engagements

**Sharing with the Community**
- [http://sqlcat.com](http://sqlcat.com)
- Mission Critical – field resources
- App fabric team
Session Objectives and Takeaways

Objectives
- Learn about the largest custom SQL Server projects
- See the techniques they use to scale

Takeaways
- SQL Server will scale to handle any size project
- SQL Server is enterprise ready
# Top statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Largest single database</td>
<td>100 TB</td>
</tr>
<tr>
<td>Largest table</td>
<td>1.5 trillion rows</td>
</tr>
<tr>
<td>Biggest total data 1 application</td>
<td>88 PB</td>
</tr>
<tr>
<td>Highest database transactions per second 1 server (from Perfmon)</td>
<td>130,000</td>
</tr>
<tr>
<td>Fastest I/O subsystem in production (SQLIO 64k buffer serial read test)</td>
<td>18 GB/sec</td>
</tr>
<tr>
<td>Fastest “real time” cube</td>
<td>Millisecond latency</td>
</tr>
<tr>
<td>data load for 1TB</td>
<td>30 minutes*</td>
</tr>
<tr>
<td>Largest MOLAP cube</td>
<td>24 TB</td>
</tr>
<tr>
<td>SQL Server Proof Points</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>VLDB</strong></td>
<td><strong>Mission Critical</strong></td>
</tr>
<tr>
<td><strong>Pan-STARRS</strong></td>
<td><strong>Large U.S. Financial Organization</strong></td>
</tr>
<tr>
<td>&gt;1 peta-byte</td>
<td>Critical operations with &gt; 1,000,000,000 tx/day</td>
</tr>
<tr>
<td><strong>Large Manufacturing Organization</strong></td>
<td><strong>bwin</strong></td>
</tr>
<tr>
<td>80TB</td>
<td>Mission Critical database &gt; 120 TB</td>
</tr>
<tr>
<td>4TB / month growth</td>
<td>&gt;80GB daily growth &gt;50 TB</td>
</tr>
<tr>
<td></td>
<td>&gt;90GB daily growth</td>
</tr>
<tr>
<td><strong>USDA</strong></td>
<td><strong>Hilton</strong></td>
</tr>
<tr>
<td>&gt;90GB daily growth</td>
<td>Critical operations at &gt; 2200 facilities</td>
</tr>
<tr>
<td><strong>Dansk Supermarked</strong></td>
<td><strong>Progressive</strong></td>
</tr>
<tr>
<td>&gt;10B rows in 1 table</td>
<td>Mission Critical table &gt; 1.9 billion rows</td>
</tr>
</tbody>
</table>
Scaling techniques and features

- Compression
- Table Partitioning
- Partitioned Views
- Service Broker
- File Group and File management
- SODA – Services Oriented Database Architecture
- DDR – Data Dependent Routing

- Scale up or Scale out
SQL Server Case Studies

DATA WAREHOUSE / BI

• Building an 80TB data warehouse
• SSAS with write back to ROLAP
• Very fast disk subsystem
• Scale out AS and RS
• Analytics before OLTP
Building an 80TB reporting system

- A large electronics company wanted to build a system to support 80TB of data for their reporting needs.
- The data is coming from sensors in their manufacturing line.
- They hold 36 months of data, one per filegroup, to manage compression and HA.
HP DL980 (Production)
- MS SQL Server 2008 R2
- Windows Server 2008 R2
- 2 CPU * 10 Core = 20 Core
- 512GB memory
- 8Gbps FC * 20port

HP DL980 (Standby)
- MS SQL Server 2008 R2
- Windows Server 2008 R2
- 2 CPU * 10 Core = 20 Core
- 512GB memory
- 8Gbps FC * 20port

SAN Switch
- DS5300B * 2ea
- 8Gbps FC * 96port Total

VMAX SE * 3 ea
- Cache 128GB
- 8Gbps FC * 16 Port
- 300GB 15k * 88ea Raid-5 : U17TB for Data Area
- 300GB 15k * 6ea : Hot Spare
- 600GB 10k * 44ea Raid-5 : U17TB for BCV
- 600GB 10k * 4ea : Hot Spare
- Same Specs. For 3 VMAX SE
- EMC Replication Manager, TimeFinder/Clone, PowerPath
VMAX-SE

88 450GB 15K rpm Disks
- 8 Hot Spares
- 80 “Data disks”

Raid 5 3+1 (3RAID5)
- 20 RAID sets
- Each “set” is a single 3+1 group

2,800 MB/sec target throughput
Analysis Services with High Scale Writeback to ROLAP
LG Electronics

Customer Profile
- LG Electronics is a manufacturing corporation with more than 115 businesses worldwide and more than 80,000 employees

Business Situation
- New demands for analysis based on Demand Forecasting System, the heart of SCM are Increasing.

Solution
- To meet this challenge, LG Electronics built demand forecasting related BI system based on SQL Server 2008 R2, offering much improved performance and convenience than the legacy system.

Benefits
- Improved transaction performance
- Deployed adaptable BI with effective investment
- Increased usage value of BI system
LGE HE (Home Entertainment) - LCD TV Demand Forecast Task

This is a corp-wide Demand Management System for LG Electronics.

- America Area OLAP server should cover a lot of subsidiaries located in America.
- Europe Area OLAP server should cover a lot of subsidiaries located in Europe.
Key Facts

Technologies used:
- Analysis Services 2008 R2 – writeback
- Custom application built by Wise, local ISV

Key Facts:
- Database Size: 5TB
- Avg query response time: 30 seconds.
- 3,500 registered users with average 11 concurrent users.
- LGE used writeback for the bottom-up level changes when they used t-sql updates and MOLAP partitions for the top-down changes to balance performance levels.
  - The cubes were partitioned by divisions so each one can process their partitions frequently without affecting the rest.
- History is 5 years * 80 million * 6 = 2.4 billion records
- Data is increasing daily by minimum of 100 thousand to maximum of 5 million cells.
Hilton Hotels

Room forecasting system
Full suite of SQL products (SQL, AS, IS, RS)

Scale out AS and RS
  ▪ If you need more capacity, just add another server

Load Balanced Analysis Services reader machines
40 to 50 concurrent users per RS server
  ▪ Complex queries
  ▪ Large data sets returned to many clients

Case study:
studyid=49192
OLAP before OLTP

StreamInsight Application at Runtime

**Event sources**
- Devices, Sensors
- Web servers
- Event stores & Databases
- Stock ticker, news feeds

**Event targets**
- Pagers & Monitoring devices
- KPI Dashboards, SharePoint UI
- Trading stations
- Event stores & Databases

**StreamInsight Engine**
- Input Adapters
- Output Adapters

**Standing Queries**
- Query Logic
- Query Logic
- Query Logic

**OLAP** before **OLTP**
SQL Server Case Studies

**OLTP**

Scale out one database
Scale up one database
No downtime allowed – ever
16,000 + instances
Microsoft CTP Data Tier

- The Commerce Transaction Platform supports Billing and subscriptions (eCommerce) for Microsoft products such as adCenter, Xbox Live, Zune, Windows Live Hotmail Plus, and Azure.

- The Commerce Transaction Platform supports payments using 13 payment methods spanning 42 currencies across 65 localized markets.

- 5 DBAs
DB Infrastructure (PROD)

2 Datacenters
5 Webstore Clusters
15 Replicas (Financial Reporting; Fraud etc)

220 SQL 2008 SP2 servers (no VM)
736 databases (excluding Mirrors and Secondary's)
121 TB of datafiles (excluding Mirrors and Secondary's)
420 TB of storage – DAS & SAN (EMC/HDS)

12 TB monthly growth
82 DB Mirror Pairs – in DC most asynch, some with auto-failover
70 Log Shipping pairs – cross DC
400 Replication subscription streams, 6 distributors
Large Health Care Provider

Chose to scale up
17 TB OLTP database
10,000 concurrent user connections
Replication for reporting database
SAN hardware replication used for Disaster Recovery
How do we scale the solution?

- “Performance Engineering” is at the heart of our methodology

- Scale Application layer by adding additional servers into the VIP

- Scale DB by adding CPU’s → MS support for large SMP hardware platforms is fundamental

- Segregation of OLTP and Reporting workloads → this allows specific tuning of the workloads
Securities Trading Organization

Application: “Real-time” high transaction, low latency stock quoting

- Over 500k business transactions/sec (peak), 350k (sustained)
- 6 active SQL Servers
- Over 1 million data manipulation calls per second in single database
- Latency per business transaction under 50 microseconds.
- “Real-time” nature of data flow.

Workload Characteristics:

- Send large batch containing multiple business transactions, parse and end up inserting all records into a large table which is constantly read
- Under 10 tables in the application.

5 DBAs
Securities Trading Organization (2 of 2)

Hardware/Deployment Configuration:
- Distributed on alphabetical split.
- DDR Scale-out.
- Commodity based hardware (4-socket, quad-core) and high performance SAN.

Other Solution Requirements:
- Mission critical to the business in terms of performance and availability.
- Require 99.999% uptime overall and 100% during business day.
- Treat system like their mainframe operations.
- Utilize SQL Server HA features to help support 5 9’s uptime requirement and geographical redundancy.
  - SQL Server Failover Clustering for local (within datacenter) availability.
  - Database Mirroring (High Availability/Async) for geo-availability.
    - Distance 250 miles
    - 30MB/sec log generation with no-redo queue.
• World’s biggest publicly listed online gaming platform

• 15 million page views and up to 980,000 unique users a day

• Environment
  • 5 DBA’s & 1 Database Architect
  • 100+ SQL Server Instances
  • 120+ TB of data,
  • 1,400+ Databases
  • 1,600+ TB storage
  • 5,000+ GB RAM
  • 450,000+ SQL Statements per second on a single server
  • 500+ Billion database transactions per day

• No downtime allowed
High Availability Today

- Principal
- Log Shipping 2nd copy
- Mirror
- Log Shipping 1h delay

All Log Backups and Full Backups Days 1, 3, 5...
All Log Backups and Full Backups Days 2, 4, 6...
• **Scale UP and Scale OUT**
  - Scale UP main financial transactions
  - Scale OUT other application functions

• **High Availability**
  - 2 data centers
  - Synchronous Database Mirroring adds 1 ms per transaction
  - Replication for reporting
  - Log Shipping for DR

• **Backup 2 TB per hour over the network**


• **Case study**

Retail Application

1,200 stores
  - SQL Standard
  - 10 GB average

15,000 cash registers
  - SQL Express
  - 10 MB database average

1 Corporate Server farm
  - SQL Enterprise
  - Windows Cluster + DB Mirroring + Log Ship
  - SQL Server, AS, IS, RS
  - 12+ TB database, 1.5 TB cube.
Retail Application (cont)

Merge replication for products & pricing
Service Broker for all transactions
  - One of the largest Service Broker and replication projects in the world
2 Million SKUs (products)
25 Million Accounts
10 million transactions / day
How do these projects scale?

Scale UP
- Table Partitioning
- NUMA
- Worker Threads

Scale Out
- Table Partitioning
- Partitioned Views
- SODA
- Data Dependent Routing
Scaling Up

Windows Server 2008 R2 the limit is 1024 Cores
- New concept: Kernel Groups
- A bit like NUMA, but an extra layer in the hierarchy

SQL Server 2008 R2 the limit is 256 cores
- Currently, largest x64 machine is 96 Cores
- And largest IA-64 is 128 cores (256 Hyperthread)

2 TB RAM
And it Looks Like This...
Scaling Out SQL Server

How do you:

- Manage an 80 TB database
- Back up a petabyte
- Build an index on a trillion row table
Scaling Out SQL Server

How do you:

- Manage an 80 TB database
- Back up a petabyte
- Build an index on a trillion row table

Answer
- Break it into manageable size pieces
Scalability Feature: Table Partitioning

- Big Table
  - P1
    - I1
    - In
  - P2
    - I1
    - In
  - P3
    - I1
    - In
  - Pn
    - I1
    - In
Scalability Feature: Local Partitioned View

```
VIEW
select * from R1
UNION ALL
select * from R2
```

Big Table: Range 1

Big Table: Range n
Scalability Feature:
Cross Database Partitioned View

**DB1**

**VIEW**
```
select * from db1.sch.R1
UNION ALL
select * from dbn.sch.R2
```

**Big Table: Range 1**

- P1
  - I1
  - I2
- P2
  - I1
  - I2
- P3
  - I1
  - I2
- Pn
  - I1
  - I2

**DBn**

**VIEW**
```
select * from db1.sch.R1
UNION ALL
select * from dbn.sch.R2
```

**Big Table: Range n**

- P1
  - I1
  - I2
- P2
  - I1
  - I2
- P3
  - I1
  - I2
- Pn
  - I1
  - I2
Integration with PDW: “Hub and Spoke”
Scalability Concept: SODA – Services Oriented Database Architecture

Separate your data by business function
  ▪ Example: HR, Payroll, Accounting, etc

Or by user function
  ▪ Example: Login, chat, email, pictures, etc

Each function goes in a different database
A common database for shared tables
Scalability Concept: Data Dependent Routing
Building an 80TB reporting system

- A large electronics company wanted to build a system to support 80TB of data for their reporting needs.
- The data is coming from sensors in their manufacturing line.
- They hold 36 months of data, one per filegroup, to manage compression and HA.
**H/W Configurations**

**Storage**
- **SAN**
  - **VMAX SE * 3 ea**
    - Cache 128GB
    - 8Gbps FC * 16 Port
    - 300GB 15k * 88ea Raid-5: U17TB for Data Area
    - 300GB 15k * 6ea: Hot Spare
    - 600GB 10k * 44ea Raid-5: U17TB for BCV
    - 600GB 10k * 4ea: Hot Spare
    - Same Specs. For 3 VMAX SE
    - EMC Replication Manager, TimeFinder/Clone, PowerPath

- **HP DL980 (Production)**
  - MS SQL Server 2008 R2
  - Windows Server 2008 R2
  - 2 CPU * 10 Core = 20 Core
  - 512GB memory
  - 8Gbps FC * 20port

- **HP DL980 (Standby)**
  - MS SQL Server 2008 R2
  - Windows Server 2008 R2
  - 2 CPU * 10 Core = 20 Core
  - 512GB memory
  - 8Gbps FC * 20port

**SAN Switch**
- DS5300B * 2ea
- 8Gbps FC * 96port Total

**PowerPath**

**Production Server**

**Standby Server**
VMAX-SE

88 450GB 15K rpm Disks
- 8 Hot Spares
- 80 “Data disks”

Raid 5 3+1 (3RAID5)
- 20 RAID sets
- Each “set” is a single 3+1 group

2,800 MB/sec target throughput
Analysis Services with High Scale Writeback to ROLAP
LG Electronics

Customer Profile
- LG Electronics is a manufacturing corporation with more than 115 businesses worldwide and more than 80,000 employees

Business Situation
- New demands for analysis based on Demand Forecasting System, the heart of SCM are Increasing.

Solution
- To meet this challenge, LG Electronics built demand forecasting related BI system based on SQL Server 2008 R2, offering much improved performance and convenience than the legacy system.

Benefits
- Improved transaction performance
- Deployed adaptable BI with effective investment
- Increased usage value of BI system
LGE HE (Home Entertainment) - LCD TV Demand Forecast Task

- America Area OLAP server should cover a lot of subsidiaries located in America.
- Europe Area OLAP server should cover a lot of subsidiaries located in Europe.
Key Facts

Technologies used:
- Analysis Services 2008 R2 – writeback
- Custom application built by Wise, local ISV

Key Facts:
- Database Size: 5TB
- Avg query response time: 30 seconds.
- 3,500 registered users with average 11 concurrent users.
- LGE used writeback for the bottom-up level changes when they used t-sql updates and MOLAP partitions for the top-down changes to balance performance levels.
  - The cubes were partitioned by divisions so each one can process their partitions frequently without affecting the rest.
- History is 5 years * 80 million * 6 = 2.4 billion records
- Data is increasing daily by minimum of 100 thousand to maximum of 5 million cells.
Hilton Hotels

Room forecasting system on SQL 2012
Full suite of SQL products (SQL, AS, IS, RS)

Scale out AS and RS
  - If you need more capacity, just add another server

Load Balanced Analysis Services reader machines
40 to 50 concurrent users per RS server
  - Complex queries
  - Large data sets returned to many clients

Case study:
OLAP before OLTP (near real time)

StreamInsight Application at Runtime

Event sources
- Devices, Sensors
- Web servers
- Event stores & Databases
- Stock ticker, news feeds

Event targets
- Pagers & Monitoring devices
- KPI Dashboards, SharePoint UI
- Trading stations
- Event stores & Databases

StreamInsight Engine
- Input Adapters
- Standing Queries
- Query Logic
- Output Adapters
SQL Server Case Studies

OLTP
Scale out one big database
Scale up one big database
Geo-DR
No downtime allowed – ever
16,000 + instances
Microsoft CTP Data Tier

- The **Commerce Transaction Platform** supports Billing and subscriptions (eCommerce) for Microsoft products such as adCenter, **Xbox Live**, Zune, Windows Live Hotmail Plus, and Azure.

- The Commerce Transaction Platform supports payments using **13 payment methods** spanning **42 currencies** across **65 localized markets**.

- 5 DBAs
DB Infrastructure (PROD)

2 Datacenters
5 Webstore Clusters
15 Replicas (Financial Reporting; Fraud etc)

220 SQL 2008 SP2 servers (no VM)
736 databases (excluding Mirrors and Secondary's)
121 TB of datafiles (excluding Mirrors and Secondary's)
420 TB of storage – DAS & SAN (EMC/HDS)

12 TB monthly growth
82 DB Mirror Pairs – in DC most asynch, some with auto-failover
70 Log Shipping pairs – cross DC
400 Replication subscription streams, 6 distributors
GEO-DR using SQL 2012

High-Volume OLTP application
Data Centers 500 miles apart
Basis for the “Last Man Standing” whitepaper
Large Health Care Provider

Chose to scale up
17 TB OLTP database
10,000 concurrent user connections
Replication for reporting database
SAN hardware replication used for Disaster Recovery
How do we scale the solution?

- “Performance Engineering” is at the heart of our methodology
- Scale Application layer by adding additional servers into the VIP
- Scale DB by adding CPU’s → MS support for large SMP hardware platforms is fundamental
- Segregation of OLTP and Reporting workloads → this allows specific tuning of the workloads
Securities Trading Organization

**Application**: “Real-time” high transaction, low latency stock quoting

- Over 500k business transactions/sec (peak), 350k (sustained)
- 6 active SQL Servers
- Over 1 million data manipulation calls per second in single database
- Latency per business transaction under 50 microseconds.
- “Real-time” nature of data flow.

**Workload Characteristics**:

- Send large batch containing multiple business transactions, parse and end up inserting all records into a large table which is constantly read
- Under 10 tables in the application.

5 DBAs
Security Trading Organization (2 of 2)

Hardware/Deployment Configuration:
- Distributed on alphabetical split.
- DDR Scale-out.
- Commodity based hardware (4-socket, quad-core) and high performance SAN.

Other Solution Requirements:
- Mission critical to the business in terms of performance and availability.
- Require 99.999% uptime overall and 100% during business day.
- Treat system like their mainframe operations
- Utilize SQL Server HA features to help support 5 9’s uptime requirement and geographical redundancy
  - SQL Server Failover Clustering for local (within datacenter) availability
  - Database Mirroring (High Availability/Async) for geo-availability
    - Distance 250 miles
    - 30MB/sec log generation with no-redo queue
• World’s biggest publicly listed online gaming platform

• 15 million page views and up to 980,000 unique users a day

• Environment
  • 5 DBA’s & 1 Database Architect
  • 100+ SQL Server Instances
  • 120+ TB of data,
  • 1,400+ Databases
  • 1,600+ TB storage
  • 5,000+ GB RAM
  • 450,000+ SQL Statements per second on a single server
  • 500+ Billion database transactions per day

• No downtime allowed
High Availability Before 2012

- Principal
- Mirror
- Log Shipping 2nd copy
- Log Shipping 1h delay

All Log Backups and Full Backups Days 1, 3, 5...
All Log Backups and Full Backups Days 2, 4, 6...
High Availability using SQL 2012
• Scale UP and Scale OUT
  • Scale UP main financial transactions
  • Scale OUT other application functions

• High Availability
  • 3 data centers
  • 2 Synchronous Availability Group copies adds 1-4 ms per transaction
  • Replication for reporting
  • Log Shipping for DR replaced by Async Availability Group replica

• Backup 2 TB per hour over the network

• Case study
  [Link to Case Study](http://www.microsoft.com/casestudies/Case_Study_Detail.aspx?casestudyid=400001470)
Retail Application

1,200 stores
- SQL Standard
- 10 GB average

15,000 cash registers
- SQL Express
- 10 MB database average

1 Corporate Server farm
- SQL Enterprise
- Windows Cluster + DB Mirroring + Log Ship
- SQL Server, AS, IS, RS
- 12+ TB database, 1.5 TB cube.
Retail Application (cont)

Merge replication for products & pricing
Service Broker for all transactions
  ▪ One of the largest Service Broker and replication projects in the world

2 Million SKUs (products)
25 Million Accounts
10 million transactions / day
How do these projects scale?

Scale UP
- Table Partitioning
- NUMA
- Worker Threads

Scale Out
- Table Partitioning
- Partitioned Views
- SODA
- Data Dependent Routing
Scaling UP

Windows Server 2008 R2 the limit is 1024 Cores

- New concept: Kernel Groups
- A bit like NUMA, but an extra layer in the hierarchy

SQL Server 2008 R2 the limit is 256 cores

- Currently, largest x64 machine is 96 Cores
- And largest IA-64 is 128 cores (256 Hyperthread)

2 TB RAM

SGI announced Windows and SQL Server running on 256 core machines
And it Looks Like This...
Scaling Out SQL Server

How do you:

- Manage an 80 TB database
- Back up a petabyte
- Build an index on a trillion row table
Scaling Out SQL Server

How do you:

- Manage an 80 TB database
- Back up a petabyte
- Build an index on a trillion row table

Answer

- Break it into manageable size pieces
Scalability Feature: Table Partitioning

Big Table

P1
I1
In

P2
I1
In

P3
I1
In

Pn
I1
In
Scalability Feature: Local Partitioned View

VIEW
select * from R1
UNION ALL
select * from R2

Big Table: Range 1

Big Table: Range n
Scalability Feature:
Cross Database Partitioned View

```
DB1

VIEW
select * from db1.sch.R1
UNION ALL
select * from dbn.sch.R2

Big Table: Range 1

P1  P2  P3  Pn
  I1  I1  I1  I1

DBn

VIEW
select * from db1.sch.R1
UNION ALL
select * from dbn.sch.R2

Big Table: Range n

P1  P2  P3  Pn
  I1  I1  I1  I1
```
Integration with PDW: “Hub and Spoke”
Scalability Concept: SODA – Services Oriented Database Architecture

Separate your data by business function
  - Example: HR, Payroll, Accounting, etc
Or by user function
  - Example: Login, chat, email, pictures, etc
Each function goes in a different database
A common database for shared tables
Scalability Concept: Data Dependent Routing

WEB or App TIER
1. User log in

Data Access Layer
2. Connect to Login Server (or cache)
3. Return Connection String

4. Connect to the correct data server

Hash Table Caching (Velocity)

SQL Server
- Login Server
- Data Server 1
- Data Server 2
- Data Server 3
Session Objectives and Takeaways

SQL Server will scale to handle any size you need

- Database size
- Transactions per second
- Concurrent users

SQL Server is enterprise ready

- Mission Critical
- Flexible HA/DR options
Questions?
Thank You for Attending