

Microsoft | Business Intelligence

MICROSOFT BI/DW DAY 2011 ROAD SHOW



SQL Server Analysis Services Best Practices

한국마이크로소프트(유)

Senior Premier Field Engineer

이준규

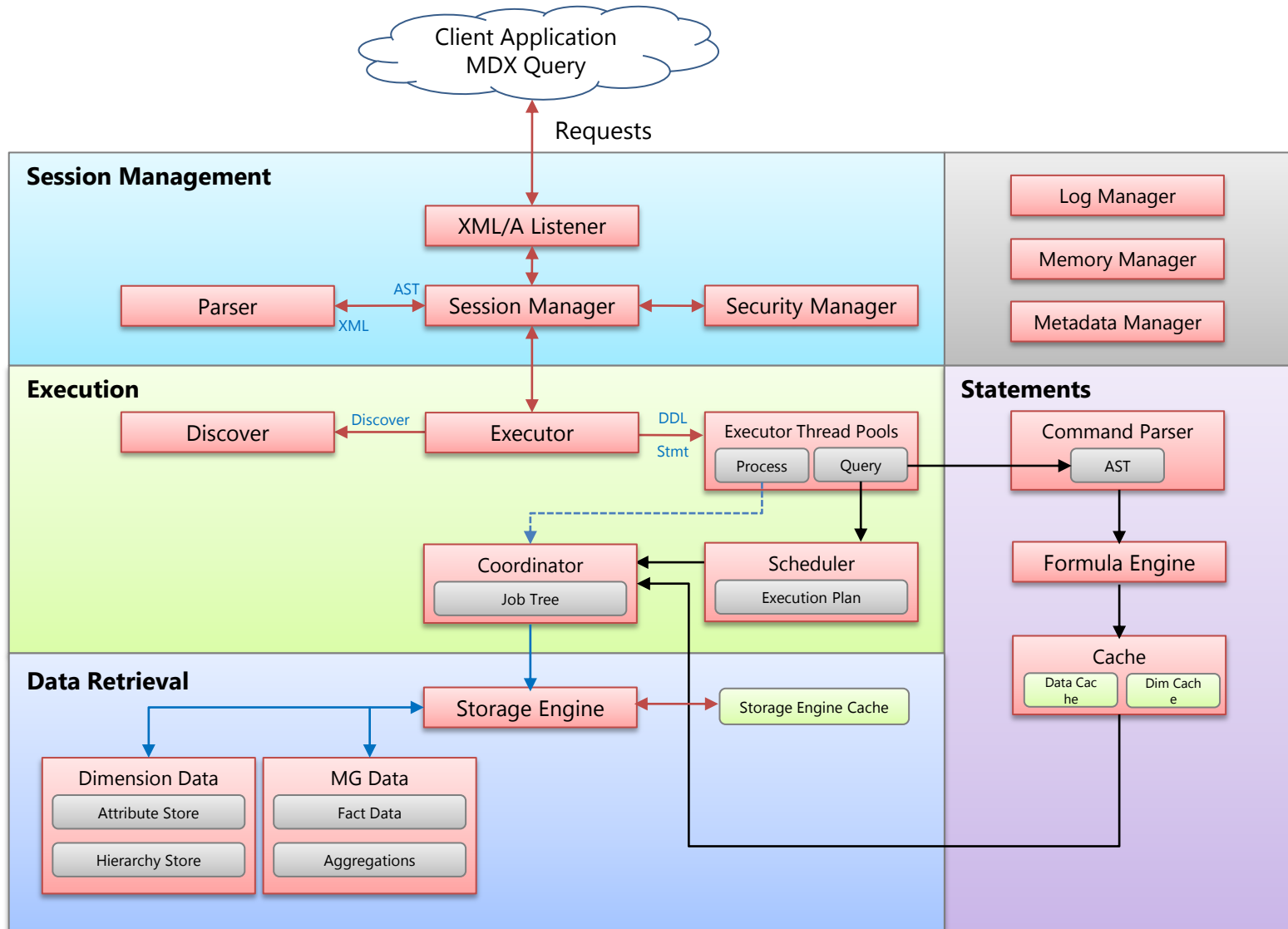
목차

- 아키텍처
- 유용한 도구들
- 가이드라인
 - Processing (Dimension, Partition, Memory, Thread)
 - Query (Aggregation, UBO, Storage/Formula engine)
 - Other Best practices
- 고려사항
 - Read-only database
 - Distinct count
 - Near Real-time
- 참조자료

SQL Server Analysis Services
Best Practices

아키텍처

아키텍처



SQL Server Analysis Services
Best Practices

유용한 도구들

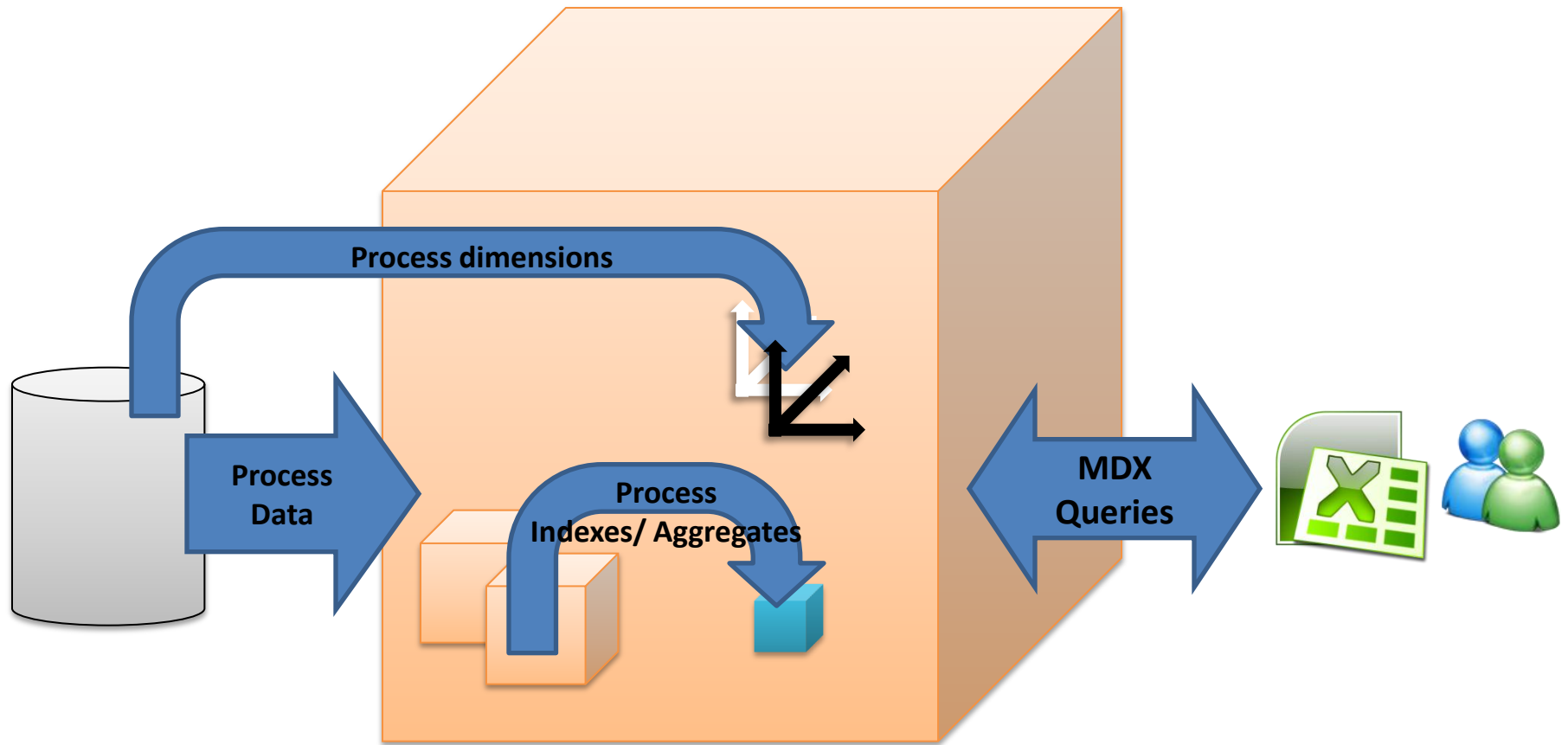
유용한 도구들

- Profiler
- Performance Monitor
- BIDS Helper
- SQLBPA
- Dynamic Management Views
- Ascmd
- MDX Studio
- Best Practices Warnings
- Aggregation Designer
- Resource Monitor

SQL Server Analysis Services
Best Practices

가이드라인

트러블슈팅



Processing

- Dimension Processing : 차원의 멤버를 추가, 변경, 제거하는 작업
 - ProcessUpdate – 현 구조에서 데이터를 변경
 - ProcessAdd – 현 구조에서 데이터를 추가
- Partition Processing : 전체 혹은 일부의 measure group에 대해서 프로세싱
 - ProcessData – RDBMS에서 데이터를 읽기
 - ProcessIndex – Index와 Aggregation 만들기

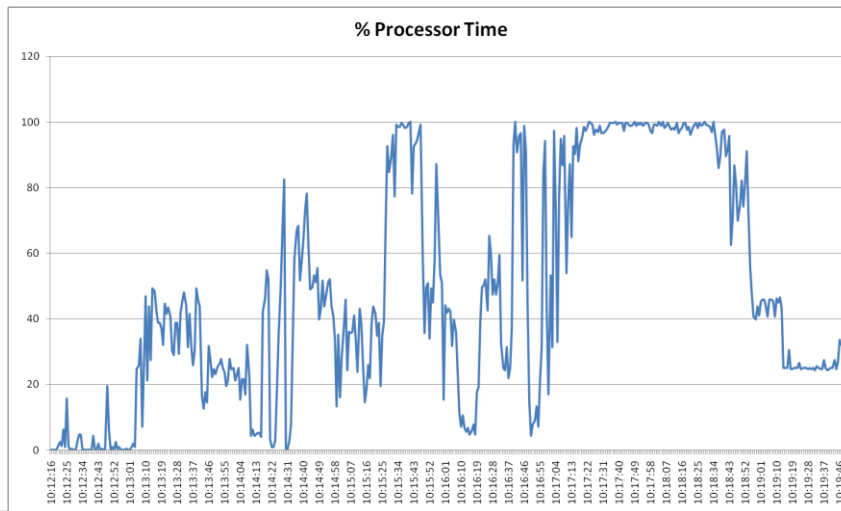
Dimension Processing

- 적절한 Process 모드 사용
 - 데이터추가만 있을 경우 ProcessAdd 사용
- Attribute Relationship 정의 필요
- Attribute의 Overhead를 줄여야 함
 - 필요한 Attribute만 추가
 - Aggregation 생성이 필요하지 않는 Attribute에 대해서 계층 생성을 Disable, Member property를 통해서 사용
 - AttributeHierarchyEnabled = false
 - 차원의 키와 비교하여 상호 유일성이 높은 Attribute는 Bitmap Index를 생성하지 말 것
 - 예) 고객번호와 주민번호
 - AttributeHierarchyOptimizedState = Not Optimized
- Distinct 쿼리 개선
 - Key, Name 컬럼에 Index 추가

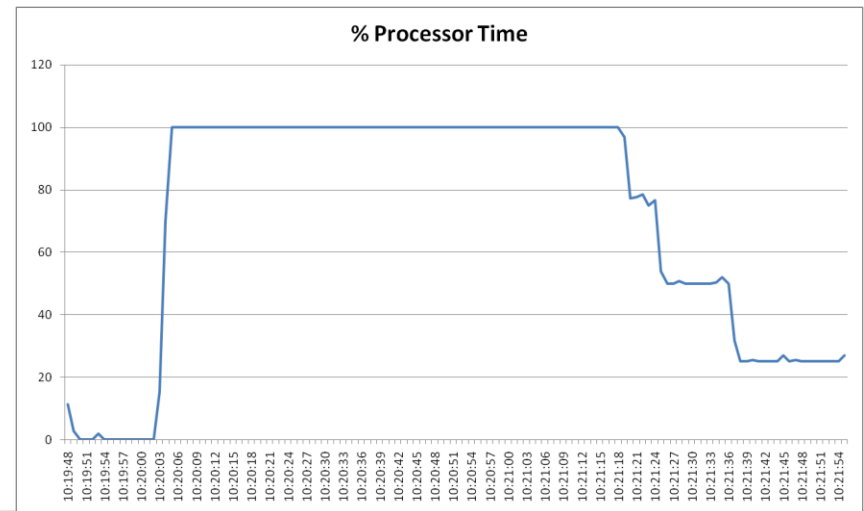
Partition Processing

- ProcessFull을 ProcessData와 ProcessIndex로 나누어 수행
 - ProcessData
 - RDBMS 소스에서 Process Buffer로 데이터 읽는 단계
 - ProcessIndex
 - AS 저장소에서 데이터를 읽어서 Index와 Aggregation을 만드는 단계

ProcessData



ProcessIndex



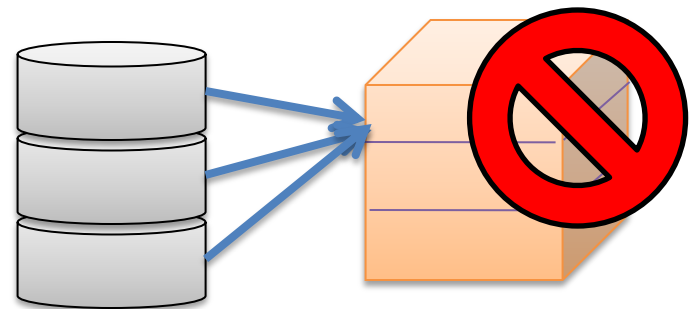
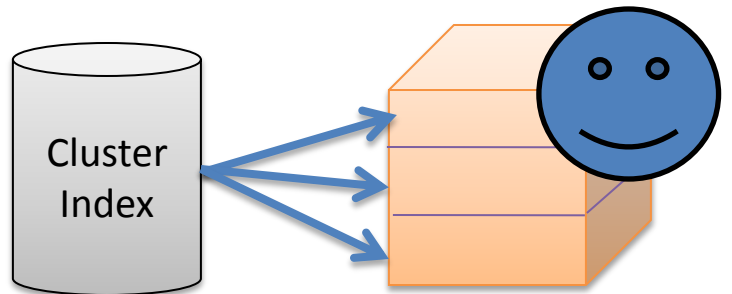
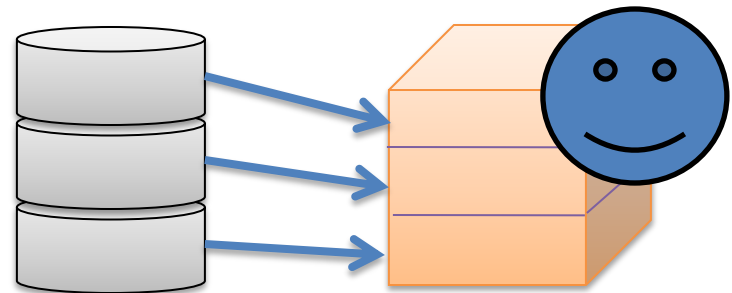
Partition Processing

병렬 처리 성능

- ProcessData 수행시에는 하나의 Cube Partition은 하나의 CPU Core와 매핑시 최적의 성능
- ProcessIndex 수행시에는 하나의 Cube Partition은 두개의 CPU Core와 매핑시 최적의 성능

디자인

- 하나의 Cube Partition은 하나의 Table Partition과 매핑
- 하나의 Table Partition과 여러개의 Cube Partition으로 구성되었을 경우 테이블에 Clustered Index 필요
- Distinct Count에 대해서는 다른 Partition 전략이 필요



Partition Processing

- ProcessData

- CPU Core당 하나의 Cube Partition이 최적의 성능

- 최대한 병렬로 처리 필요
 - AS에서 SQL 연결수 조정 필요
 - AS와 SQL이 같은 장비를 공유할 경우 Memory 설정 조정 필요. AS와 SQL 분리 권고

- SQL 최적화

- 사이즈가 작은 Integer 등의 가공키 사용
 - SQL 조인 최소화 필요
 - SQL Maxdop 값 조정 필요

Partition Processing

- ProcessIndex
 - 두개의 CPU Core당 하나의 Cube Partition이 최적
 - 병렬 처리 전략
 - 적절한 Partition 개수
 - Partition 내에서 적절한 Thread 개수
 - 구성 값
 - CoordinatorBuildMaxThreads : Partition Processing 작업 당 할당되는 최대의 Thread수
 - AggregationMemoryMin : Partition Processing 작업을 위해 전체 Memory중에서 할당되는 Memory 양
 - 큐브 최적화
 - 적절한 Aggregation 필요
 - 적절한 Attribute 개수 : Index 생성 시간 증가

리소스 사용

- ProcessData, ProcessIndex 수행시 Memory, Thread 필요
- Sorting/Aggregating은 Memory에서 수행 권고
 - Proc Aggregations#Temp file bytes written/sec : Temp file에 write 발생시 memory를 추가 혹은 병렬처리되는 개수를 줄어야 함
- Memory
 - ProcessData : Merge buffer (OLAP#Process#BufferMemoryLimit)
 - ProcessIndex : Aggregation buffer(OLAP#Process#AggregationMemoryMin)
- Threads
 - ProcessData : RDBMS에서 데이터 읽기
 - ProcessIndex : Sorting/Aggregating/Indexing
- 구성 파일
 - C:#Program Files#Microsoft SQL Server#MSAS10_50.MSSQLSERVER#OLAP#Config#msmdsrv.ini

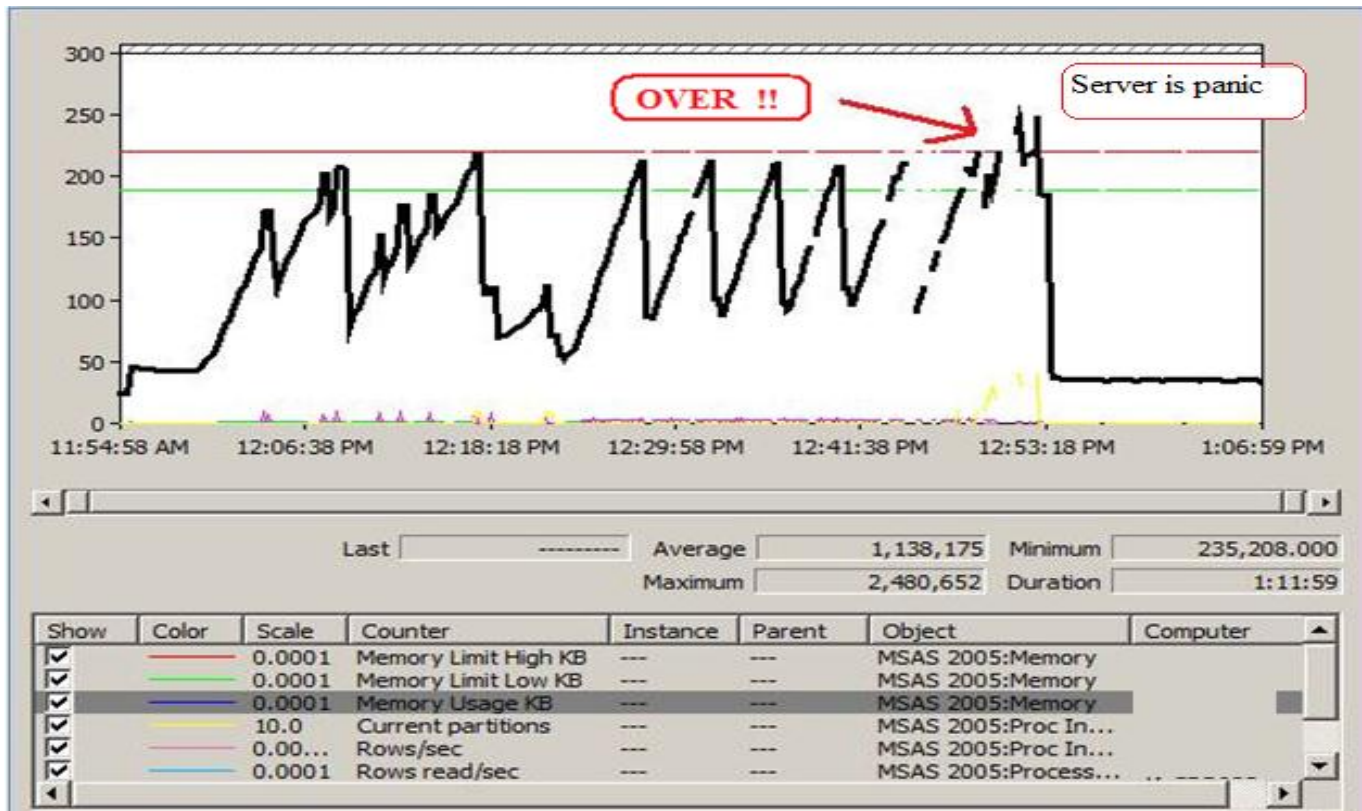
Memory 구성값 & 성능 카운터

- 구성값
 - Memory\TotalMemoryLimit : 최대 Memory(기본값 80%)
 - Memory\LowMemoryLimit : 최소 Memory(기본값 65%)
- 성능 카운터

성능 카운터	정의
MSAS 2008:Memory\Memory Limit Low KB	Memory\LowMemoryLimit 값
MSAS 2008:Memory\Memory Limit High KB	Memory\TotalMemoryLimit 값
MSAS 2008:Memory\Memory Usage KB	Analysis Services 프로세스가 현재 사용하는 Memory 양 다음의 두 값과 비교 필요 : Memory\LowMemoryLimit, Memory\TotalMemoryLimit.
MSAS 2008:Memory\Cleaner Memory nonshrinkable KB	Memory 부족시 background cleaner에 의해서 shrink되지 않는 Memory
MSAS 2008:Memory\Cleaner Memory shrinkable KB	Memory 부족시 background cleaner에 의해서 shrink되는 Memory

Memory Pressure

- Low memory pressure
 - Memory Limit Low KB < Memory Usage KB < Memory Limit High KB
- High memory pressure
 - Memory Limit High KB < Memory Usage KB

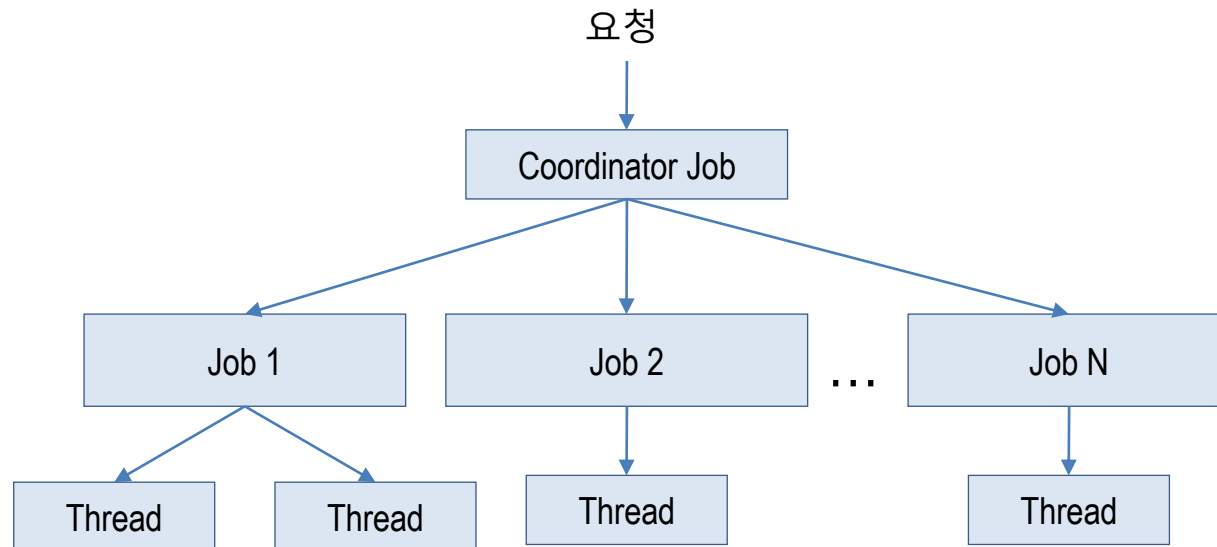


Job & Thread

- 구성값

- CoordinatorExecutionMode (기본값 = -4)

- 4 core 장비에서는 16개의 Job이 동시에 수행될 수 있음
- 양수일 경우 core 수와 관계없이 동시에 수행될 수 있는 Job의 개수를 의미



Thread

- 구성값
 - ThreadPool\Process\MinThreads, MaxThreads
 - ThreadPool\Query\MinThreads, MaxThreads
- 성능카운터

성능 카운터	정의
MSAS 2008:Threads\Processing\Query pool job queue length	Processing/Query thread pool의 대기열에 있는 Job 수
MSAS 2008:Threads\Processing\Query pool busy threads	Processing/Query thread pool내에 busy Thread 수
MSAS 2008:Threads\Processing\Query pool idle threads	Processing/Query thread pool내에 Idle Thread 수

- 가이드라인

Situation	Action
Processing pool job queue length > 0 and Processing pool idle threads = 0 for longer periods during processing.	Increase Threadpool\Process\MaxThreads and retest.
Both Processing pool job queue length > 0 and Processing pool idle threads > 0 at same time during processing.	Decrease CoordinatorExecutionMode and retest.

Aggregation 사용

SQL Server Profiler - [Untitled - 2 (bi-real-es64)]

File Edit View Replay Tools Window Help

EventClass	EventSubcl...	TextData	Connectio...	NTUser...	ApplicationN...
Existing Session		*	13	lenwy	Microsoft SQ...
Notification	6 - Object ...		59	dwickert	
Query Begin	0 - MDXQu...	SELECT { [Measures].[Sales Qty] } ON COLUMNS , NON EMPTY { [Item].[...	63	dwickert	Microsoft SQ...
Query Cube Begin			63		
Calculate Non Empty Begin			63		
Get Data From Aggregation		Query 5a 19b Store Sales	63		
Get Data From Aggregation		Query 5a 19b Store Sales	63		
Get Data From Aggregation		Query 5a 19b Store Sales	63		
Get Data From Aggregation		Query 5a 19b Store Sales	63		
Get Data From Aggregation		Query 5a 19b Store Sales	63		
Progress Report Begin	14 - Query	Started reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report Begin	14 - Query	Started reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report Begin	14 - Query	Started reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report Begin	14 - Query	Started reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report Begin	14 - Query	Started reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report End	14 - Query	Finished reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report End	14 - Query	Finished reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report End	14 - Query	Finished reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report End	14 - Query	Finished reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Progress Report End	14 - Query	Finished reading data from the 'Query 5a 19b Store Sales' aggregation.	63	dwickert	
Query Subcube	2 - Non-ca...	000,000000001000,000,0,01000110...	63		
Query Subcube Verbose	22 - Non-c...	Dimension 0 [Buyer] (0 0 0) [Buyer Name]:0 [Buyer Index]:0 [Buyer ID]:0 Dim...	63		
Get Data From Aggregation		Query 5b 19c Store Sales	63		
Get Data From Aggregation		Query 5b 19c Store Sales	63		
Get Data From Aggregation		Query 5b 19c Store Sales	63		

Aggregation 미사용

SQL Server Profiler - [Untitled - 1 (bi-real-as)]

File Edit View Replay Tools Window Help

EventClass	EventSubcl...	TextData	Connectio...	NTUser...	ApplicationN
Query Begin	0 - MDXQu...	SELECT { [Measures].[cAvg Retail] , [Measures].[Sales Qty] } ON COLUMNS , N...	200	dwickert	Microsoft SC
Query Cube Begin			200		
Calculate Non Empty Begin			200		
Progress Report Begin	14 - Query	Started reading data from the '_v Tbl Fact Store Sales WE 2004 02 14' partition.	200	dwickert	
Progress Report Begin	14 - Query	Started reading data from the '_v Tbl Fact Store Sales WE 2004 02 28' partition.	200	dwickert	
Progress Report Begin	14 - Query	Started reading data from the '_v Tbl Fact Store Sales WE 2004 01 03' partition.	200	dwickert	
Progress Report Begin	14 - Query	Started reading data from the '_v Tbl Fact Store Sales WE 2004 02 07' partition.	200	dwickert	
Progress Report End	14 - Query	Finished reading data from the '_v Tbl Fact Store Sales WE 2004 01 03' partition.	200	dwickert	
Progress Report Begin	14 - Query	Started reading data from the '_v Tbl Fact Store Sales WE 2004 02 21' partition.	200	dwickert	
Progress Report End	14 - Query	Finished reading data from the '_v Tbl Fact Store Sales WE 2004 02 14' partition.	200	dwickert	
Progress Report End	14 - Query	Finished reading data from the '_v Tbl Fact Store Sales WE 2004 02 07' partition.	200	dwickert	
Progress Report End	14 - Query	Finished reading data from the '_v Tbl Fact Store Sales WE 2004 02 28' partition.	200	dwickert	
Progress Report End	14 - Query	Finished reading data from the '_v Tbl Fact Store Sales WE 2004 02 21' partition.	200	dwickert	
Query Subcube	2 - Non-ca...	000,111,000,0,010001100000...	200		
Query Subcube Verbose	22 - Non-c...	Dimension 0 [Buyer] (0 0 0) [Buyer Name]:0 [Buyer Index]:0 [Buyer ID]:0 Dimensio...	200		
Calculate Non Empty Current	1 - Get Data		200		
Calculate Non Empty Current	1 - Get Data		200		
Get Data From Cache	1 - Get dat...	Dimension 0 [Buyer] (0 0 0) [Buyer Name]:0 [Buyer Index]:0 [Buyer ID]:0 Dimensio...	200		
Query Subcube	1 - Cache ...	000,111,000,0,010001100000...	200		
Query Subcube Verbose	21 - Cache...	Dimension 0 [Buyer] (0 0 0) [Buyer Name]:0 [Buyer Index]:0 [Buyer ID]:0 Dimensio...	200		
Get Data From Cache	1 - Get dat...	Dimension 0 [Buyer] (0 0 0) [Buyer Name]:0 [Buyer Index]:0 [Buyer ID]:0 Dimensio...	200		
Query Subcube	1 - Cache ...	000,111,000,0,010001100000...	200		
Query Subcube Verbose	21 - Cache...	Dimension 0 [Buyer] (0 0 0) [Buyer Name]:0 [Buyer Index]:0 [Buyer ID]:0 Dimensio...	200		
Calculate Non Empty Current	2 - Process...		200		

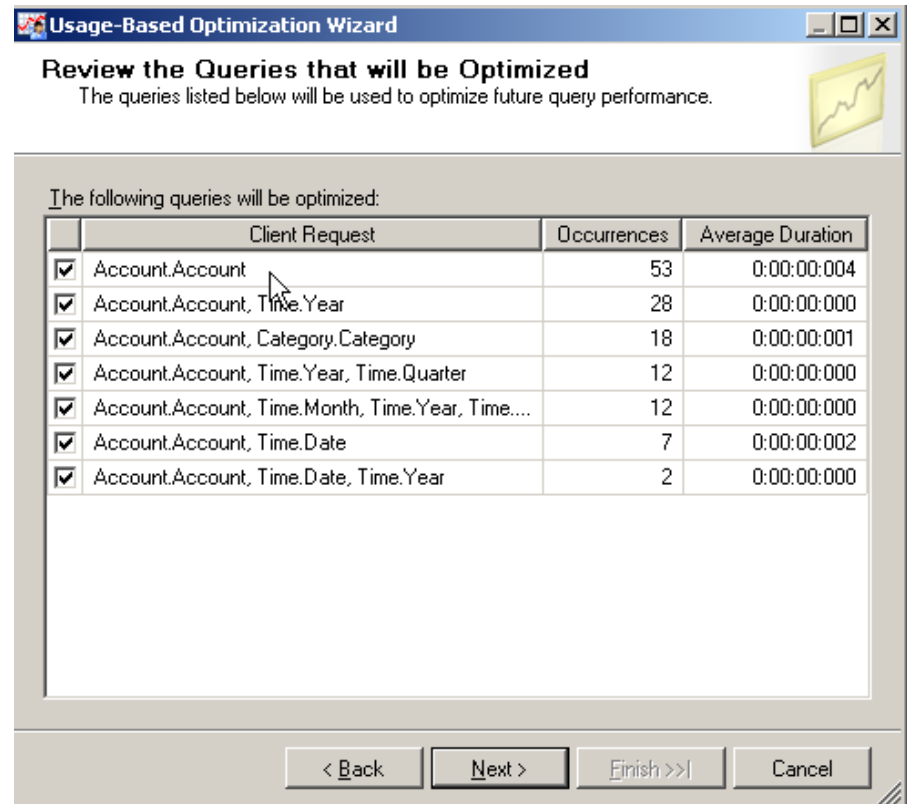
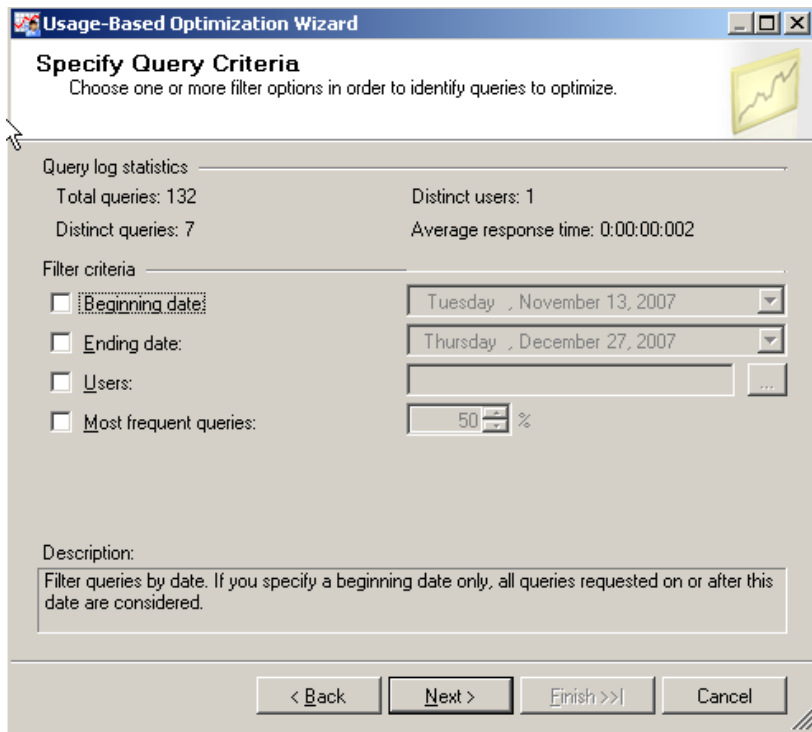
MISS

1

2

Usage Based Optimization

- 가이드라인
 - Processing 시간과 Size를 고려하여 적절한 Aggregation 추가 필요
 - 사용자 쿼리 기반으로 Aggregation을 주기적으로 반영
- 도구 및 접근 방향
 - UBO 사용하여 Aggregation 최적화



쿼리 최적화

- Storage Engine 병목
 - Partition 조정
 - Aggregation 조정
- Formula Engine 병목
 - Cell by cell 계산을 Subspace로 변경
 - Cache warmer 고려

Cache Warmer

- 자주 실행되는 쿼리를 주기적으로 수행하여 Caching
- Profiler에서 쿼리를 도출
- 방법
 - Create cache statement 혹은 with cache 구문 사용
 - Asccmd를 이용하여 MDX 쿼리를 수행
 - Table에 쿼리를 저장하고 SSIS Package에서 MDX 쿼리를 수행
 - Reporting Service report 실행

Other Best practices

- Data source
 - OLE DB provider 사용, .Net SqlClient data provider 사용하지 말 것
- Dimension
 - Numeric key 사용
 - 중복된 Attribute relationship 정의하지 말 것
 - ErrorConfiguration에서 KeyDuplicate를 IgnoreError로 설정하지 말 것
 - UnknownMember를 Hidden으로 설정하지 말 것
 - Member의 수가 클 경우 64bit 서버를 사용할 것
 - Unnatural hierarchy를 피할 것
- Cube
 - Parent-child dimension 사용시 유의
 - Many-to-many dimension 구조에서 너무 큰 intermediate measure group이나 dimension을 사용하지 말 것
 - 동일한 dimension이나 granularity를 가진 measure group은 통합을 고려할 것
 - Distinct count measure는 별도의 measure group으로 정의할 것

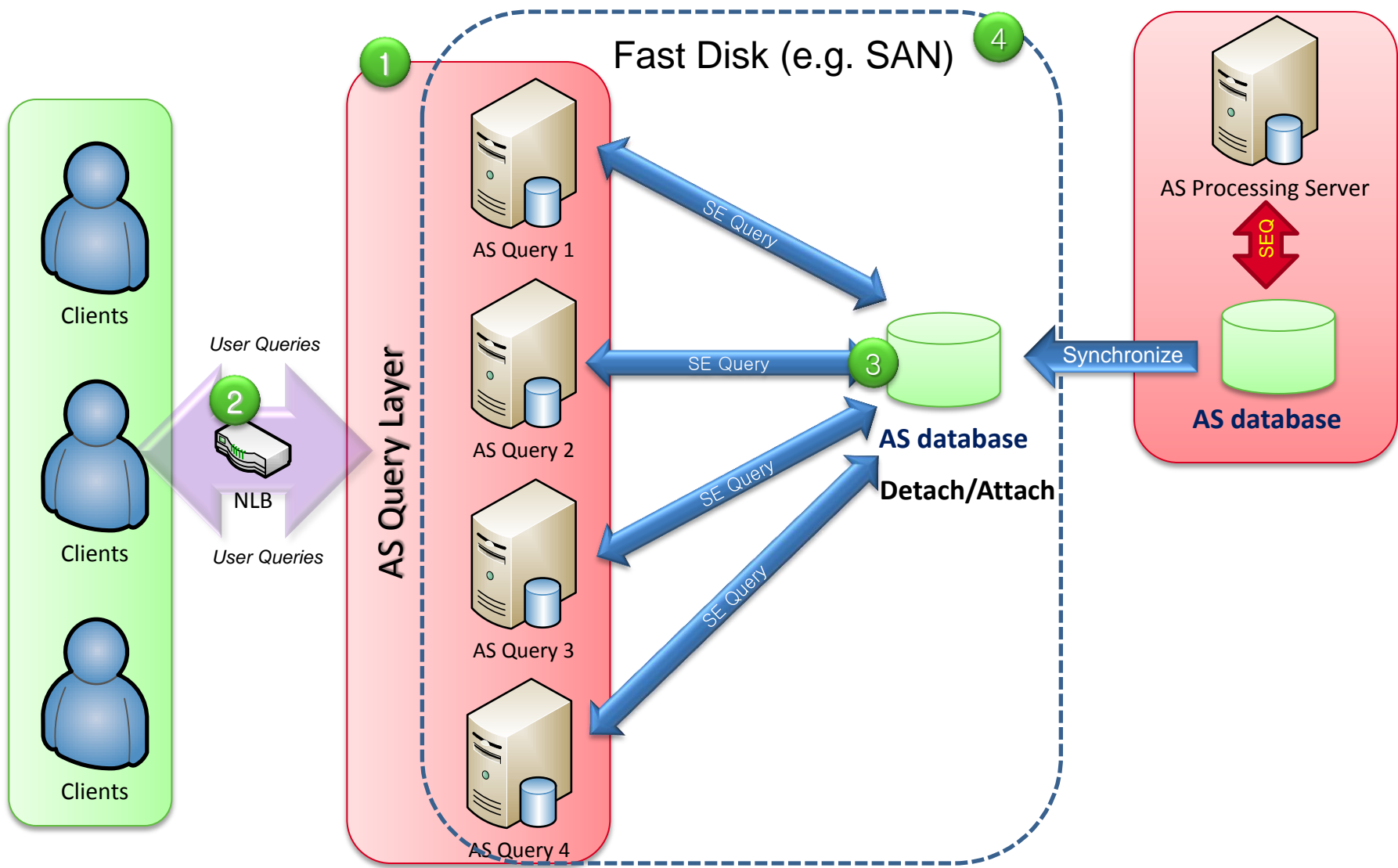
Other Best practices

- Partition
 - 2천만 이상의 row를 가진 partition은 분할을 고려
 - 사이즈가 작은 너무 많은 partition은 통합을 고려
 - ROLAP이나 proactive caching을 사용하는 partition은 Slice 속성 설정 필요
- Aggregation
 - 적절한 AggregationUsage 속성 사용
 - Partition별로 다른 Aggregation 전략을 사용 가능
- Security
 - 너무 많은 member를 가진 attribute에 role security 사용을 자제할 것
- Configuration
 - 동시성을 높이기 위한 설정
 - CoordinatorQueryBalancingFactor = 1
 - CoordinatorQueryBoostPriorityLevel = 0
 - 다중사용자 throughput을 높이기 위한 설정
 - MemoryHeapType = 2
 - HeapTypeForObjects = 0

SQL Server Analysis Services
Best Practices

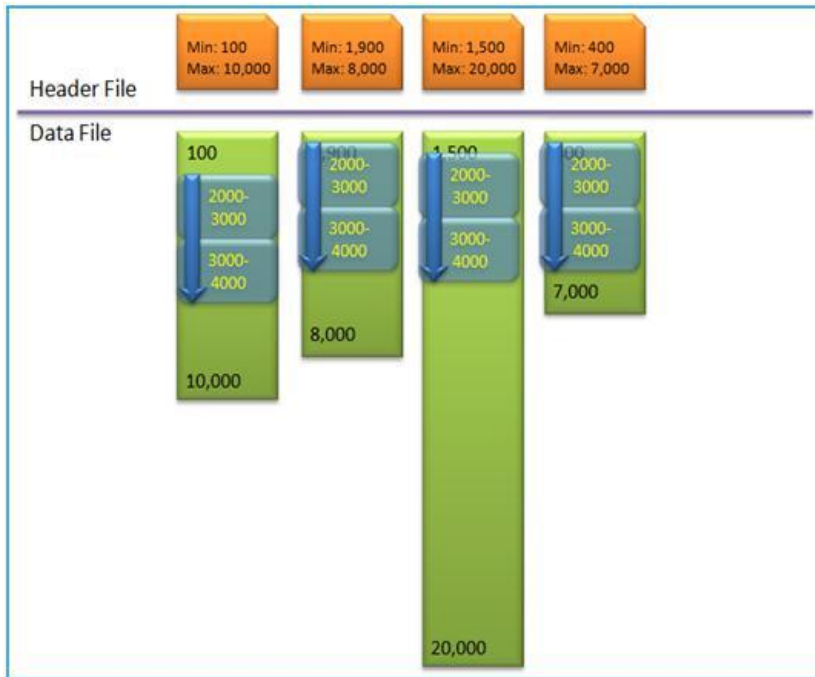
고려사항

Read-only database



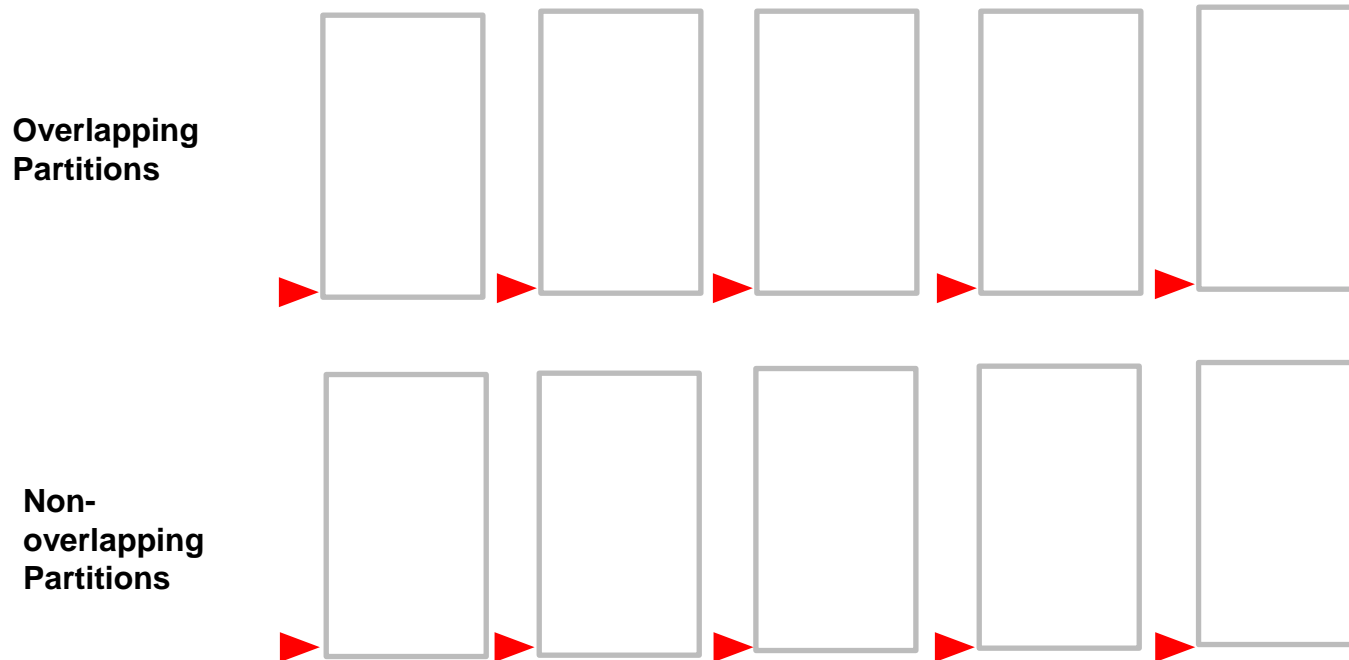
Distinct count

- 파티션키에 distinct count 키가 포함되지 않는 경우(왼쪽)
 - Thread들은 다른 Partition의 Thread에서의 읽기가 끝날 때 까지 대기 발생
- 파티션키에 distinct count 키가 포함되어 있는 경우(오른쪽)
 - Thread는 다른 Partition의 Thread에 대한 대기 없이 읽기 지속



Distinct count

- 5개의 Thread가 데이터를 스캔할 때 대기가 필요한 경우와 그렇지 않는 경우

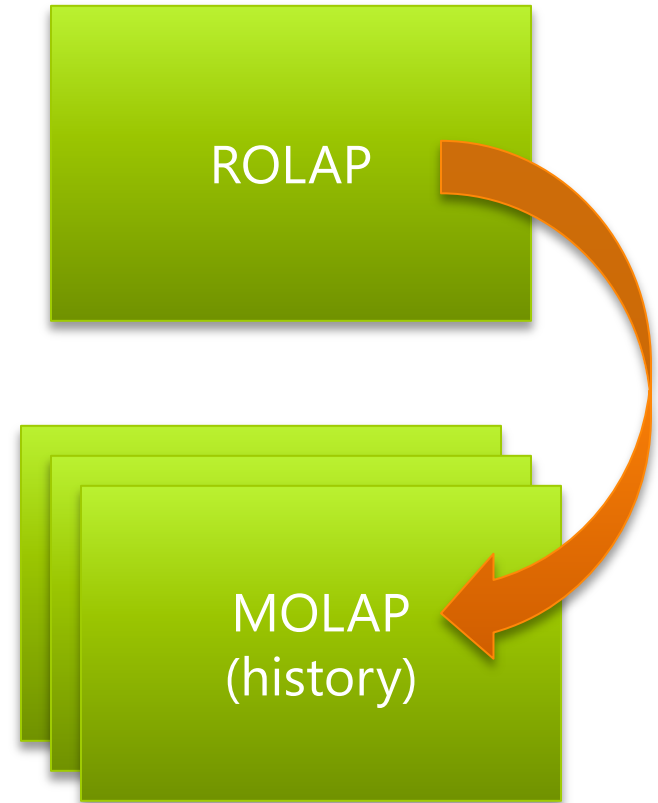


- 결과 예)

대상 데이터	Partition 변경전 속도(ms)	Partition 변경후 속도(ms)
테스트 데이터 1	4096	1563
테스트 데이터 2	42532	11406

Near Real time

- Proactive caching 고려
- 실시간 데이터가 필요한 파티션은 ROLAP으로 구현
 - ROLAP 파티션은 Memory에 캐쉬되도록 최소한 작게 유지
 - RDBMS 튜닝
- 대량의 이력 데이터는 점진적으로 MOLAP 파티션으로 전환



SQL Server Analysis Services
Best Practices

참조자료

참조자료

- Articles

- SQL Server 2008 White Paper: Analysis Services Performance Guide
 - <http://www.microsoft.com/downloads/en/details.aspx?FamilyID=3be0488d-e7aa-4078-a050-ae39912d2e43&displaylang=en>
- SQL Server Best Practices Article: Identifying and Resolving MDX Query Performance Bottlenecks in SQL Server 2005 Analysis Services
 - <http://www.microsoft.com/downloads/en/details.aspx?FamilyId=975C5BB2-8207-4B4E-BE7C-06AC86E24C13&displaylang=en>
- SQL Server Best Practices Article
 - <http://technet.microsoft.com/en-us/library/cc966525.aspx>

- Tools

- BIDS Helper
 - <http://bidshelper.codeplex.com/>
- Microsoft® SQL Server® 2008 R2 Best Practices Analyzer
 - <http://www.microsoft.com/downloads/en/details.aspx?FamilyID=0fd439d7-4bff-4df7-a52f-9a1be8725591>
- MDX Studio
 - <http://www.mosha.com/msolap/mdxstudio.htm>