

Spatial Index 개념과 사용

Spatial Index 와 사용 목적

Index 는 사용자가 데이터베이스에 저장되어 있는 원하는 데이터를 검색할 경우, 데이터베이스의 처음부터 끝까지 하나하나 비교하며 찾는 것이 아니라, 전체 데이터의 일부분만을 검색하여 더 빠르고 간단하게 찾을 수 있는 방법을 제공한다.

Spatial Index 란 Oracle 의 컬럼 색인과 비슷하며, '공간'과 '색인'의 합성어로 공간 정보를 검색하는데 있어서 좀 더 빠르게 사용자가 데이터를 이용할 수 있도록 미리 생성해 놓은 정보이다. 이 spatial Index 를 사용하면 테이블을 전체 탐색(table full scan)하지 않기 때문에 디스크의 I/O 를 줄일 수 있다.

Spatial Index 는 일반적인 컬럼 색인과 비교해서 2 가지의 특징이 있다. 첫째, 이 데이터는 바이너리 형태로 Long RAW 나 BLOB 컬럼에 저장되어 오라클에서 검색에 사용된다. Spatial index 는 오라클이 검색할 수 있는 표준 데이터 타입을 이용하여 공간 데이터나 데이터의 접근을 표현한다. 둘째, spatial index 는 일반적인 컬럼에 저장된 숫자를 찾는 방법과 같이 2 차원의 선형을 이룬다. GX, GY 로 표현되는 이러한 정보는 공간 데이터를 검색하는데 탁월한 성능을 지녔다.

R-Tree Index

Spatial Index 의 검색은 일반적인 데이터의 1 차원 B-tree 검색과는 다르다. 공간 데이터는 2 차원의 R-tree 검색 방법을 사용하고, B-tree 방식으로는 데이터를 검색할 수 없다. 마찬가지로, 비 공간 데이터는 R-tree 방식으로 데이터를 검색할 수 없다. R-tree 검색은 실제 데이터의 위치를 이용하는 것이 아닌, 각 노드가 속한 경계를 사용하는 tree-shaped 구조를 이용한다.

검색은 R-tree 의 색인을 사용하여 찾으려고 하는 일반 영역에서 개체를 발견하고 트리를 내려 다음 개체 자체에 대한 테스트를 수행한다. 따라서 R-tree 를 이용하게 되면 자신이 원하는 영역 이외의 공간에서 개체를 찾으려는 노력은 하지 않는다. 하지만 R-tree 를 이용하지 않는다면 쿼리 기준에 대해 모든 개체를 테스트 하는 과정이 수반되어 성능에 치명적인 결과를 가져오게 된다.

Spatial Index 특이 사항

ArcSDE 는 일반 인덱스와 비교하여 추가적인 사항이 존재한다. 그 중 첫 번째는 데이터는 Long Raw 나 Blob 과 같은 바이너리 형태로 인덱스화 된다. 이는 오라클에서 인지할 수 있는 형태이며, 공간데이터를 표현한다. 두 번째, 공간 인덱스는 일반적인 컬럼 인덱스와 같이 선형 검색이 가능한 2 차원의 데이터로부터 생성 된다.

Multi-Level Spatial Index

ArcSDE는 총 3 Level 의 spatial index level 을 지원한다. 그 형태는 S_IX1, S_IX2, S_IX3 형식으로 생성이 된다. 여기서 값은 9.3 버전 이상에서 ST_Geometry 를 사용할 경우 sde.st_geometry_index 테이블에 저장되어 있다. 9.2 버전 이하의 경우는 default 로 S 형태로 저장되며 이 값은 sde.layers 테이블에 저장되어 있다.

Spatial Index 테이블 Naming 규칙

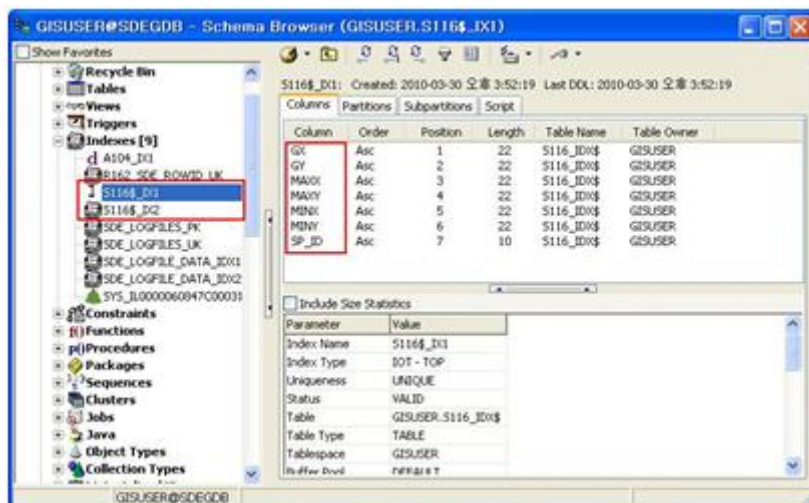
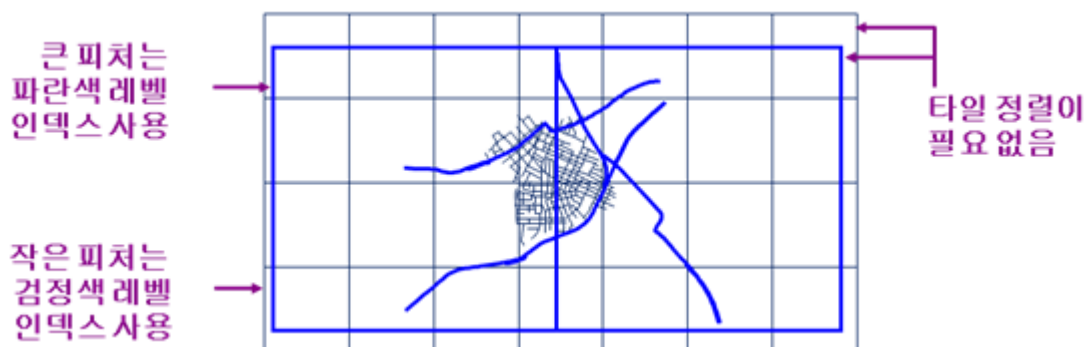
ArcSDE 9.2 이하 버전 : select

```
select Table_name, Layer_id from sde.layers where Table_name='<테이블 이름>'
```

ArcSDE 9.3 이상 버전 :

```
select Table_name, Index_id from sde.st_geometry_index where Table_name='<테이블 이름>'
```

이처럼 3 개의 그리드 레벨을 이용하는 이유는, 한 레벨로 이용했을 때 발생할 수 있는 데이터베이스 내의 spatial index segment 를 줄이기 위해서 이다. 아래의 그림처럼 피처의 envelop 의 크기가 다양한 경우 작은 피처는 첫 번째 그리드 레벨에서 인덱스를 생성하고, 큰 피처는 두 번째 그리드 레벨에서 인덱스를 생성하게 되어 데이터베이스내의 segment 발생을 줄이게 된다.



<실제 Spatial-Index 테이블 모습>

Spatial Index 구성요소

ArcSDE spatial index 인 S-table 에는 3 가지의 요소가 포함되어 있다: grid tile, geometry envelope, feature ID.



Grid tile

Spatial index 그리드 타일은 공간을 나누는 논리적인 격자이다. 각각의 타일은 S-table 내에서 두 가지의 숫자에 의해 유일한 ID 값을 갖는다. GX 는 격자의 X 축을 담당하고, GY 는 격자의 Y 축을 담당한다. 타일의 ID 는 0 부터 시작하고, X 와 Y 값이 증가할수록 1 씩 증가하여 모든 공간 도메인에 설정된다.

이렇게 각 타일에 유일한 번호가 부여되어 있으므로 공간 질의에 대해 매우 쉽게 검색이 가능하다. 하지만 이 정보는 특정 영역에 대한 상세한 정보가 아닌 대략적인 정보만을 제공한다.

Geometry envelope

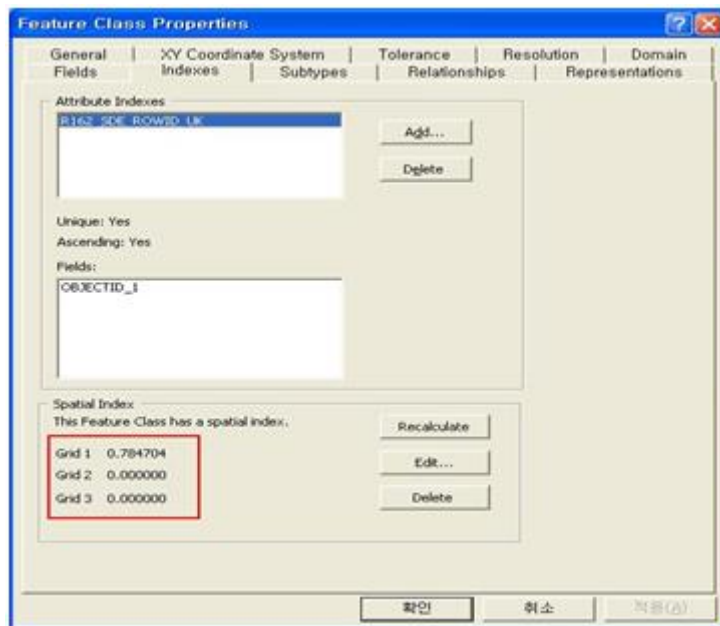
Spatial index 에 의해 참조되는 각각의 지오메트리 영역은 상수 값으로 변경되고 NUMBER 컬럼에 저장된다: EMINX, EMINY, EMAXX, EMAXY. 이 값은 표준 데이터 타입을 사용하기 때문에 오라클에서 해석이 쉽고 검색이 빠르다. 이 또한 그리드 타일과 같이 대략적인 정보를 표출하지만 타일보다는 나은 지역과 피처의 사이즈를 지정하며, 두 배 정도의 정보를 더 요구한다.

Geometry 와 Spatial ID

S-table 의 각 레코드는 지오메트리의 피처 ID 값을 포함한다. 이 피처 ID 는 SP_ID 컬럼에 저장되며, 오라클 컬럼의 인덱스인 ROW ID 처럼 인덱스의 포인터 역할을 한다. 지오메트리 값은 비즈니스 테이블에 바이너리 형식으로 저장되기 때문에 공간 인덱스를 포함하지 않는다.

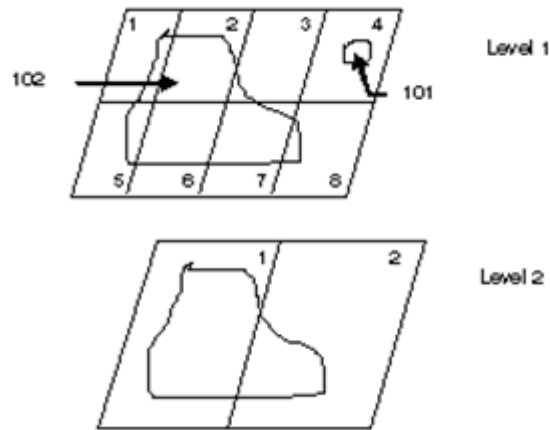
Spatial Grid Index

ArcSDE는 Oracle와 DB2를 이용할 경우 grid index를 사용한다. 이 spatial index는 spatial 컬럼의 데이터에 대한 격자이다. Spatial grid index는 2차원으로 표현되며, 마치 지도의 모습과 같이 피쳐들을 연결하고 있다. 사용자는 이 grid index를 1, 2, 3레벨로 값을 할당할 수 있으며, 각각은 별개의 grid cell의 크기를 결정짓는다. Spatial grid index의 필수 요소인 첫 번째 레벨은 가장 작은 cell 크기를 나타내며, 2, 3레벨의 값은 '0'으로 설정하면 사용하지 않을 수 있다. 만약 2번째 레벨을 사용하고 싶다면, 이 크기는 첫 번째 레벨보다 적어도 3배 이상으로 설정하는 것이 좋다. 마찬가지로 3번째 레벨의 사이즈도 2번째 레벨의 크기보다 3배 이상 설정해야 한다.



<3 레벨의 그리드 타일 사이즈>

다음의 예제에서 피쳐 클래스는 총 2개의 그리드 레벨을 갖는다. 피쳐 101은 영역 4에 존재하고 있으며, 4개의 타일 내에(이 경우 단 한 개에 포함) 존재하고 있으므로 레벨 1에 기록된다. 또한 102 피쳐의 Envelope은 1부터 8영역에 모두 포함되어 있다. 102의 경우 4개 이상의 타일을 사용하기 때문에 두 개의 타일로도 그 영역을 모두 포함할 수 있는 레벨 2에 정보를 기록한다. 그리고 이 두 정보가 spatial index에 기록된다.



<101 피처는 레벨 1 에, 102 피처는 레벨 2 에 정보가 저장된다>

피처의 생성, 변경, 삭제작업이 이뤄지면 spatial index 도 변경된다. 이 경우 그리드 셀의 개수를 얻기 위해 각 피처의 extent 들이 가장 낮은 그리드 레벨로 입혀진다. 만약 피처의 extent 가 4 개의 셀보다 넓게 자리 잡고 있다면, 그 지오메트리는 다음 그리드 레벨로 이동한다. 만약 하나를 정의한 경우, 이 피처는 4 개의 그리드 타일에 포함되거나 그 이하의 개수에 포함될 때까지 레벨을 증가 시키며 확인한다. 최 상위 그리드 레벨에서는 피처가 4 개의 셀 이상에 존재 하더라도 그 레벨에 정의 된다. 오라클 데이터베이스의 경우, spatial index 의 저장 위치를 dbtune.sde 파일의 s_storage 파라미터에 설정해서 피처가 생성될 때 키워드로 지정하여 자동으로 설정할 수 있다.

Single-part point 피처의 Grid Size

일반적인 single-part point 피처의 경우, 각 포인트의 envelop 값은 '0'이다. 이 때문에 포인트는 최대 / 최소 envelop 가 존재 하지 않으며 라인 / 폴리곤 피처의 envelop 계산식을 이용해도 그 값이 '0'으로 나타난다. 이 경우 포인트 피처의 grid size 는 그 피처가 일반적으로 표시되는 영역에서의 수에 관련되어 있다. 이 말은 ArcMap 을 실행 할 경우 한 화면에 표시 되는 포인트의 수를 envelop 의 값으로 설정하면 된다. 예를 들어 ArcMap 에서 포인트 피처를 불러와서 작업하는 경우 평균적으로 화면에 표시되는 값이 1000 개 정도라면 첫 번째 그리드 사이즈는 1000 으로 설정하면 된다. 포인트의 경우 그리드 레벨은 1 개만 존재하며 2, 3 레벨은 의미가 없게 된다. (모든 점들이 각각 하나의 그리드 타일에 포함되기 때문에) 일반적인 공간 데이터를 이용하는 경우 포인트는 '1000'의 값을 많이 이용한다.

Spatial Index 크기 제한

그리드 타일의 제약 사항은 다음과 같다.

그리드 타일 사이즈

최대 : 2,147,483,847 integer unit

최소 : 256 integer unit

최소 : 피처당 1,000 그리드 타일 이하로 유지 (S-segment 가 증가하기 때문에)

각 레벨간 크기 제약 : 전 레벨 보다 3 배 이상

만약 한 개의 피처의 크기가 매우 클 경우 위의 제약 사항이 문제를 발생할 수 있으며 이 경우 다음과 같은 해결책이 제시된다.

1. 피처 클래스를 분리하여 작은 조각으로 만든다.
2. 두 번째 Spatial Index Grid Level 을 추가한다.
3. 첫 번째 Spatial Index Grid 크기를 늘린다.

Spatial Index Grid Size 선택

그리드 셀 사이즈는 spatial index 테이블에 영향을 준다. Spatial index 를 조정 한다는 뜻은 셀 사이즈를 적당하게 변경하는 작업이다 - 작은 셀 사이즈는 shape 당 많은 셀이 존재 한다는 뜻이며, spatial index 테이블에 많은 정보가 기록 된다는 뜻이다.

만약 ArcGIS Desktop 클라이언트를 이용해서 데이터를 생성한다면, Spatial index 그리드는 자동으로 생성되고 변경할 필요는 없을 것이다. 하지만 다른 프로그램을 이용해서 데이터를 생성하는 경우 ArcGIS 프로그램과는 다르게 Spatial index 그리드가 생성될 수 있다. 이 경우 사용자는 spatial index 그리드를 테스트 해보고 재 조정 작업이 필요할 수 있다.

다음의 내용처럼 spatial index 의 그리드 크기를 조절하여 검색의 성능을 향상 시킬 수 있다.

- 얼마나 많은 그리드 레벨이 필요한지 고려하고, 데이터베이스의 옵티마이저가 그리드 레벨 당 spatial index 테이블을 검색함을 기억하라. 가끔, spatial index 의 내용이 가장 적게 저장 되어 있는, 레벨이 한 개인 경우가 최상의 성능을 발휘하는 경우도 있다.
- 순수하게 포인트 타입의 피처 클래스의 경우는 1 레벨만 정의하고 셀 크기는 크게 잡는 것이 좋다. 또한 포인트의 경우 다른 지오메트리를 갖는 피처클래스보다 더 빠르게 검색 된다.
- Spatial index 를 모니터링 하라. 만약 데이터가 자주 변경되는 피처 클래스라면 spatial index 는 튜닝 하기가 쉽지 않다. 튜닝은 공간 데이터의 구조와 관련이 있으며, 주기적으로 변경이 생긴 피처클래스의 spatial index 를 재 정의 하는 것이 필요하다.

- 피처 클래스를 이용하는 애플리케이션에 기반하라. 애플리케이션 윈도우의 extent 를 그리드 셀 사이즈와 동일하게 정의한다. 이렇게 함으로써 애플리케이션은 spatial index 테이블의 정보와 정확하게 일치하는 화면을 볼 수 있다. 이는 spatial index 테이블의 크기를 조절하고, 프로세싱 작업을 줄이는 효과가 있다. 왜냐하면 피처 테이블에 대해 더 적은 수의 피처 ID 값이 비교 되기 때문이다.
- 알지 못하거나 다양한 윈도우 애플리케이션에서 사용되는 경우라면, 해당 피처 클래스의 평균 피처 extent 를 가지고 첫 번째 그리드 사이즈를 정하는 작업을 수행하면 된다. 다음과 같은 SQL 문장을 실행하여 적정 값을 확인한다.

<9.3 이상의 버전>

```
SQL> SELECT ROUND((AVG(sde.ST_MAXX(shape) - sde.ST_MINX(shape)) +
AVG(sde.ST_MAXY(shape) - sde.ST_MINY(shape)))/ 2 * 3,0)
<컬럼 별칭> from <소유자>.<피처 이름>
```

<9.2 이하의 버전>

```
SQL > SELECT AVG(((emaxx - eminx) + (emaxy - eminy)) / 2 * 3) <컬럼 별칭>
from
```

(* 3 의 값은 피처의 평균 extent 에 2 ~ 4 배 중 선택한 값. 각 피처가 한 개나 두 개의 그리드 셀에 포함되게 만들기 위해서 평균 extent 값에 2 ~ 4 배를 함.)

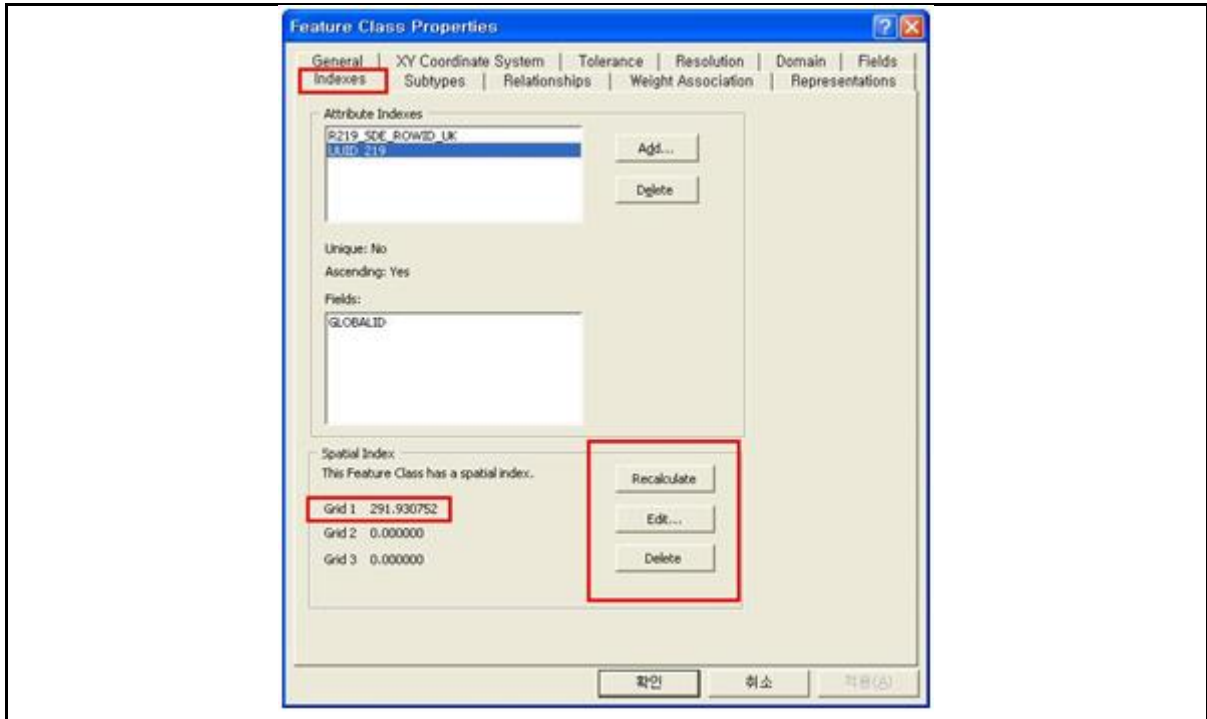
Spatial Index Size 변경

1. SDE Command 사용

```
SDElayer -o alter -l GISUSER.CITY,SHAPE -g 1000,10000,100000 -u sde -p sde
```

2. ArcCatalog 사용

```
해당 피처 클래스의 등록정보 > Indexes > Recalculate 버튼 or Edit 버튼
```



Spatial Index Statistics

Sdlayer 명령어를 이용하면 현재 사용중인 spatial index 의 통계 정보를 생성할 수 있다. 이 통계 정보는 spatial index 레벨당 하나씩 생성되며, -l 옵션을 통해 피쳐 클래스를 지정할 수 있다.

Layer ID 는 이 피쳐 클래스의 Layer ID 값으로 SDE.Layers 에서 확인할 수도 있다. Level 은 통계정보를 생성하는 Grid 레벨을 뜻하며, Grid Size 는 이 레벨의 Grid Size 를 나타낸다. Grid Records 와 Feature Records 는 각각 Spatial Index 레코드와 Business 테이블(F Table)의 레코드를 나타낸다. Grids / Feature Ratio 값은 Grid 레코드 값을 Feature 레코드 값으로 나눈 비율이며, 항상 4 가 넘지 않도록 해야 하며, 최적은 2 이하로 유지 하는 것이다. AVG, Max 값은 선택 정도를 나타내며, 하단의 그룹 편중 정도를 살펴 보면 모든 피쳐가 Grid 4 개 안에 포함되어 있고, Grids / Feature Ratio 가 2 보다 작으므로 Grid 크기를 5000 으로 준 것은 최적의 선택을 한 것이다.

S-segment 크기

- ◆ 4 이하로 유지
- ◆ 최적은 2 이하

Index 선택 정도

- ◆ 작을수록 좋음
- ◆ 평균 Grid당 포함된 피처의 수

Layer 1 Spatial Index Statistics:
Level 1, Grid Size 5000

Grid Records:	39828
Feature Records:	32426
Grids/Feature Ratio:	1.23
Avg. Features per Grid:	54.86
Max. Features per Grid:	1541
% of Features Wholly Inside 1 Grid:	84.35

Spatial Index Record Count By Group								
Grids:	<=4	>4	>10	>25	>50	>100	>250	>500
Features:	32390	36	4	1	1	1	1	1
% Total:	100%	0%	0%	0%	0%	0%	0%	0%

대부분 이곳에 배치됨

Grid 4개 이상 차지하는 피처가 0개 (상태 좋음)

통계 정보를 이용한 Spatial Index 조절

현재 Grid Level 을 1 개로 운영 했을 때 통계 정보를 살펴보면 다음과 같다.

```

C:\WINDOWS\system32\cmd.exe
C:\Documents and Settings\Administrator>sdelay -o si_stats -l gisuser.AAA,shape -i esri_sde -u sde -p sde

ArcSDE 9.3.1 for Oracle10g Build 1632 Thu Feb 26 12:05:37 2009
Layer Administration Utility

-----
Layer 174 Spatial Index Statistics:
Level 1, Grid Size 1000
-----
| Grid Records: 146
| Feature Records: 34
| Grids/Feature Ratio: 4.29
| Avg. Features per Grid: 2.86
| Max. Features per Grid: 11
| % of Features Wholly Inside 1 Grid: 11.76
-----
| Spatial Index Record Count By Group
| Grids: <=4 >4 >10 >25 >50 >100 >250 >500
|-----
| Features: 23 11 1 0 0 0 0 0
| % Total: 68% 32% 3% 0% 0% 0% 0% 0%
|-----

C:\Documents and Settings\Administrator>
    
```

Grids / Feature Ratio 값이 4.29 로 4 를 넘은 상태이며, 4 개 이상의 Grid 에 배치된 피처가 11 개 로 32%나 된다. 이 경우 spatial index 의 조정이 필요한데, 방법은 두 가지가 있다 : 1. spatial index level 을 2 단계로 증가, 2. Grid 의 크기 확장. 이 두 가지 방법으로 spatial index 를 조절해 본다.

```

C:\WINDOWS\system32\cmd.exe
C:\Documents and Settings\Administrator>sdelayer -o si_stats -l gisuser.AAA,shape -i esri_sde -u sde -p sde

ArcSDE 9.3.1 for Oracle10g Build 1632 Thu Feb 26 12:05:37 2009
Layer Administration Utility

-----
Layer 174 Spatial Index Statistics:
Level 1, Grid Size 1000
-----
| Grid Records: 62
| Feature Records: 23
| Grids/Feature Ratio: 2.70
| Avg. Features per Grid: 2.21
| Max. Features per Grid: 7
| % of Features Wholly Inside 1 Grid: 17.39
-----
| Spatial Index Record Count By Group
| Grids: <=4 >4 >10 >25 >50 >100 >250 >500
|-----|
| Features: 23 0 0 0 0 0 0 0
| % Total: 100% 0% 0% 0% 0% 0% 0% 0%
|-----|

Level 2, Grid Size 4000
-----
| Grid Records: 24
| Feature Records: 11
| Grids/Feature Ratio: 2.18
| Avg. Features per Grid: 3.00
| Max. Features per Grid: 8
| % of Features Wholly Inside 1 Grid: 18.18
-----
| Spatial Index Record Count By Group
| Grids: <=4 >4 >10 >25 >50 >100 >250 >500
|-----|
| Features: 11 0 0 0 0 0 0 0
| % Total: 100% 0% 0% 0% 0% 0% 0% 0%
|-----|

```

먼저 spatial index level 을 한 단계 증가 시키고 Grid size 값을 4000 으로 설정하면 Grids / Features ratio 가 2.79 과 2.18 로 낮아졌으며, 각 레벨에서 모든 피처가 4 개의 Grid 안에 들어 오게 되었다. 만약 Grid level 을 증가시키지 않고 Grid 크기 조절만으로 이와 같은 효과를 얻을 수 있다면 인덱스를 생성하는 것보다는 다음의 방법이 더 효과적 일 수 있다.

```

C:\WINDOWS\system32\cmd.exe
C:\Documents and Settings\Administrator>sdelay -o si_stats -l gisuser.AAA,shape -i esri_sde -u sde -p sde

ArcSDE 9.3.1 for Oracle10g Build 1632 Thu Feb 26 12:05:37 2009
Layer Administration Utility
-----
Layer 174 Spatial Index Statistics:
Level 1, Grid Size 3000
-----
| Grid Records: 70 |
| Feature Records: 34 |
| Grids/Feature Ratio: 2.06 |
| Avg. Features per Grid: 5.83 |
| Max. Features per Grid: 18 |
| % of Features Wholly Inside 1 Grid: 35.29 |
-----
| Spatial Index Record Count By Group |
| Grids: <=4 >4 >10 >25 >50 >100 >250 >500 |
| Features: 34 0 0 0 0 0 0 0 |
| % Total: 100% 0% 0% 0% 0% 0% 0% 0% |
-----
C:\Documents and Settings\Administrator>

```

이처럼 추가의 인덱스를 생성하지 않고 Grid 크기를 조절하여 Grids / Features의 값을 2.06으로 낮추고, Grid 4개 안에 모든 피쳐들을 수용할 수 있게 설정할 수 있다.