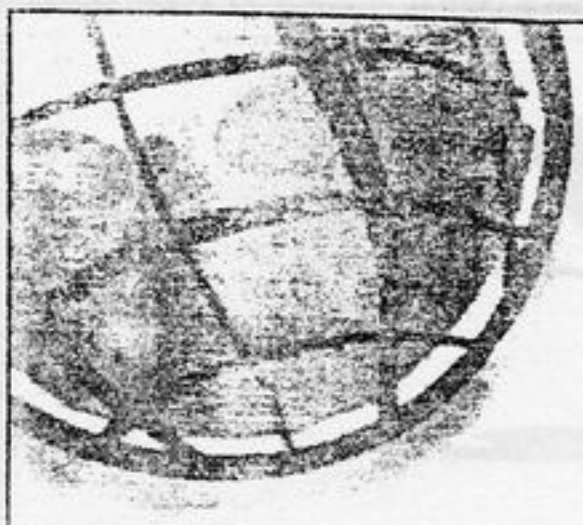


Data loading

5

Copyright © 19XX-19XX Environmental Systems Research Institute, Inc. All Rights Reserved.

第五章　加载数据

章节概述
数据源属性
***ArcSDE** 整数坐标系统*
整数坐标的特征
将坐标将转换为正整数
章节概述
地理坐标系
投影坐标系
选择空间索引单元大小
建立空间索引需要考虑的问题
练习 5a
图层的存贮结构
加载图层数据
改变图层参数
删除图层
使用脚本
练习 5b

Section overview

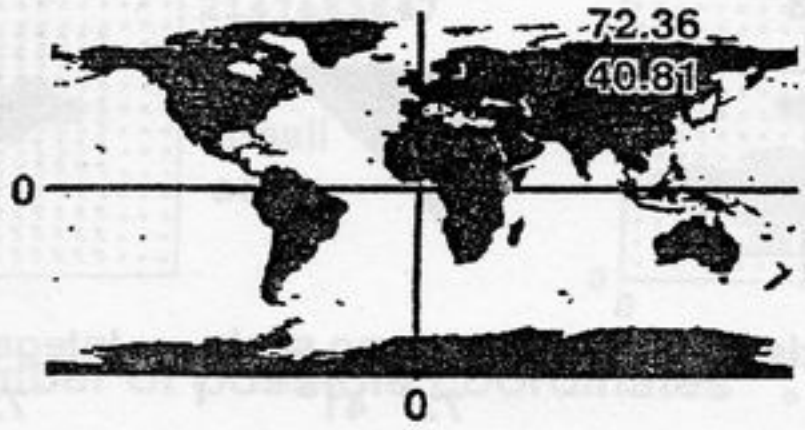
- Determine layer properties (5a)
  - Integer coordinates
  - Coordinate system and projections
  - Spatial index resolution
- Load new layer (5b)
  - Create RDBMS data structures
  - Insert data

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

章节概述

本章节第一部分描述把图层加载 ArcSDE 前对图层的评估和对属性的确定。图层的属性将增加到 ArcSDE 中元数据表中。第二部分着重讲述把数据加载 ArcSDE。

Source data properties



- Global dataset—*admin98.shp*
  - ESRI shapefile format
- Measured in decimal-degrees
- Positive and negative coordinates

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

数据源属性

有几个数据属性影响到数据如何加载 ArcSDE 并决定图层加载的参数。

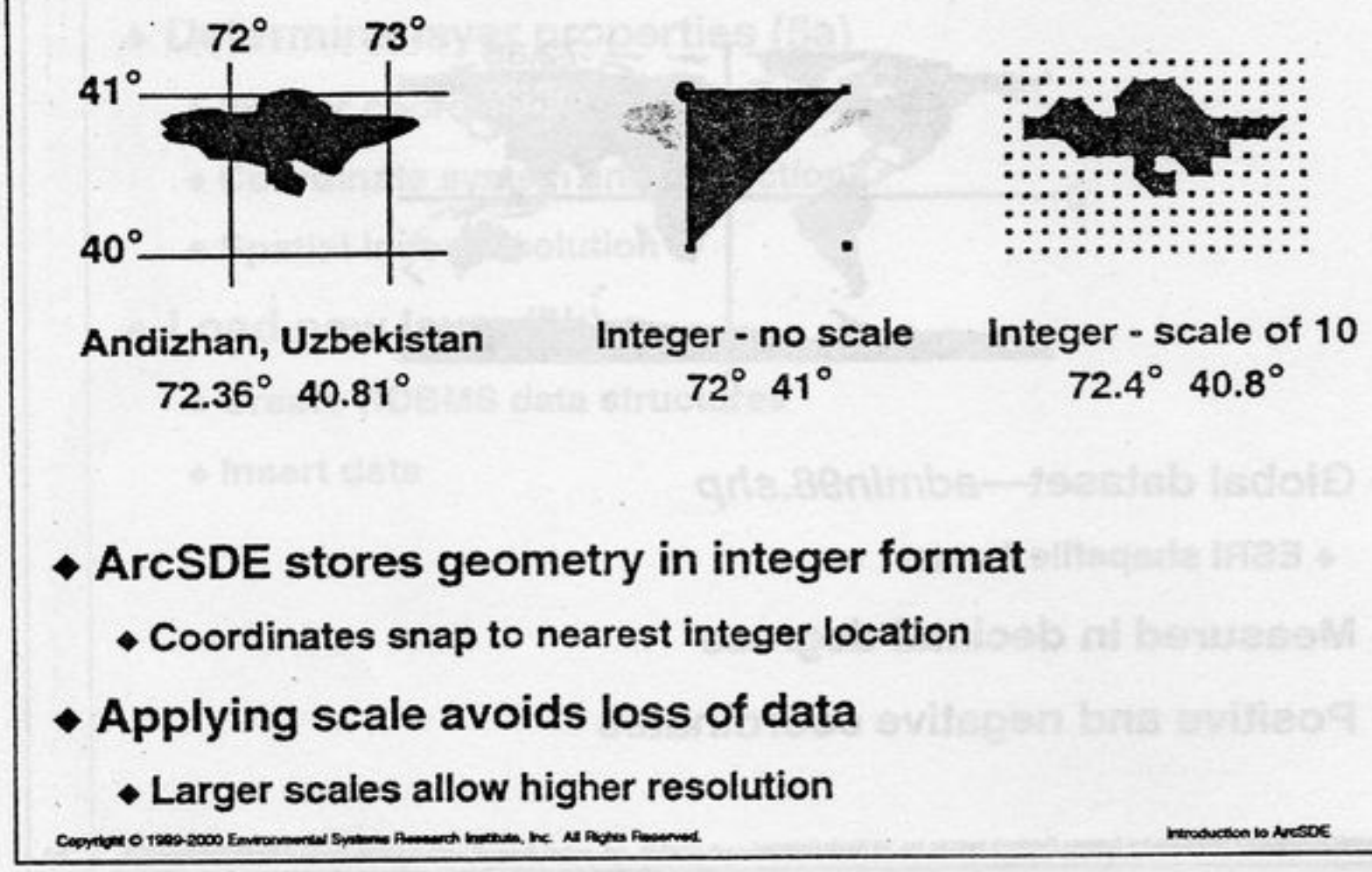
数据属性

ArcSDE 提供将 ArcView Shapefile、ArcInfo Coverage、INFO 表和 DBF 表直接加载 ArcSDE 的功能。如果数据文件不是这几种形式，那么在加载之前须转换格式。如果安装了 CAD 客户端，CAD 文件也可以加载。

在数据加载前，数据的空间参考系必须已知。如果文件没有正确的坐标系，在加载前，必须进行正确投影。在 ArcInfo 中有把坐标转换到不同参考系的功能。

ArcSDE 以正整数存贮坐标值。在加载数据前，要确定一个偏移量以便把所有的坐标值转化成正值。这个偏移量在转入时指定，一旦数据加载后不能改变。

练习 5 中的被加载到图层的数据是 ArcView 的 shapefile，它是一个用十进制度表示的全球数据，其中四分之三的数值有负值。



ArcSDE 整数坐标系

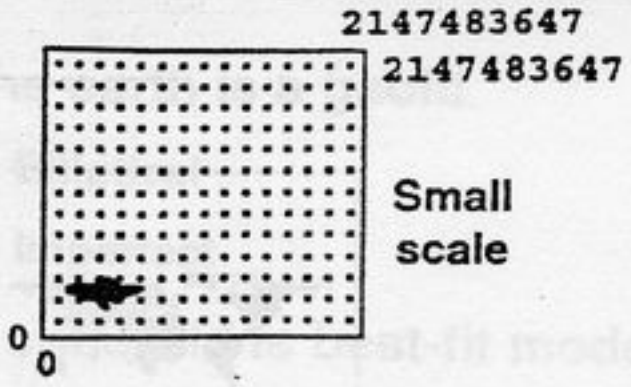
ArcSDE 以整数存贮空间数据。因为使用整数比浮点数有更高的压缩率，并且对整数进行二进制搜索比较快。如果数据源是浮点数，坐标将被取为最近似的整数。如果不使用一个比例系数，这将导致坐标值的综合。

比例系数

当数据加载时定义比例系数。通过定义系数可以“放大”数据，从而允许 ArcSDE 以更高分辨率存贮整数数据。大的整数值将消耗大量的计算机物理内存，所以确定的系数不要大于必要的系数。

一旦数据存入 ArcSDE，比例系数不能改变。比例系数由 ArcSDE 实施，客户端看不到它。当数据从数据库提取显示时，ArcSDE 通过比例系数把坐标转换回浮点数。注意，这个比例只能被 ArcSDE 使用，并且对图层有明显的影响。如果比例太小或没有使用比例，坐标被取整后，要素的 shape 可能变形。

Characteristics of integer coordinates



- Finite number of possible coordinates
  - Positive, 32-bit integers
- Larger scales consume more coordinate / storage space
  - Integer system must contain all data
- Choose smallest scale that supports data's resolution
  - Cannot alter after layer creation
  - Allow for update with higher resolution data

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

整数坐标的特征

X、Y 的比例属性

ArcSDE 使用正整数坐标系存贮空间坐标，其最大值不能超过 3 2 位正整数所能表示的范围，即 21.4 亿。任何小于 0 或大于 21.4 亿的坐标值将不能存贮于 ArcSDE，图层认为其超出了范围。

大的比例把数据放大到一个大的等级。坐标值的放大会消耗更多的计算机内存，这是因为大整数比小整数的压缩比小。但大整数坐标可以存贮 shape 更详细的细节。

选择合适的比例系数很重要，因为数据入库后比例就不能改变了。要改变比例，必须以不同的比例重新把数据加载到 ArcSDE。因此既要保证图层有最好的分辨率，同时又要使比例系数最小，以供今后数据更新和增长的需要。



将坐标转换为正整数

正整数坐标

ArcSDE 用正整数存贮空间数据。比例系数防止数据被过度综合。所有的数值必须小于 2^{32} 的正整数。如果任何坐标是负的或大于 2^{32} 将不能被加载。为了确保数据是正整数，当数据加载时须设置偏移。在数据库内该偏移量一旦设定就不能改变。要改变它，必须用正确的偏移量重新加载数据。

X 和 Y 的偏移量

偏移的目的是重新定位数据的坐标空间。通常数据存贮在坐标空间的中心以满足数据在空间上增长的需要。如果数据太靠近坐标空间的边界，今后加载的数据可能超出范围。当这种情况发生时，须确定新的偏移并需要重新加载数据。

通过计算数据的中间值和坐标系中间值的差异来确定偏移量。数据的坐标将加上这一偏移量后加载到 ArcSDE 的正整数坐标系中。

xy 的偏移量在 ArcSDE 内部使用，仅用来存贮数据。当数据被一个 ArcSDE 客户端提取显示时，它又被转换回原始位置。偏移完全是用于存贮目的并强制把坐标转换到 0 —2^{32} 间的正整数。

课堂用的练习数据以十进制的度存贮并且中心点在 0， 0 。坐标空间的中心点在 2^{16}, 2^{16}。坐标系的中心点需被转换成与数据相同的系统单位以使它们有可比性。根据数据集中心与整数系统中心的差异计算偏移。

Geographic coordinate systems

- The earth is a geoid
  - Elliptical
  - Imperfect
- A spheroid is best-fit model
  - Simplifies spatial calculations
- A datum is a spheroid aligned with a known point
- Coordinates are measured on a longitude, latitude grid
  - Origin is equator and prime meridian



Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

地理坐标系

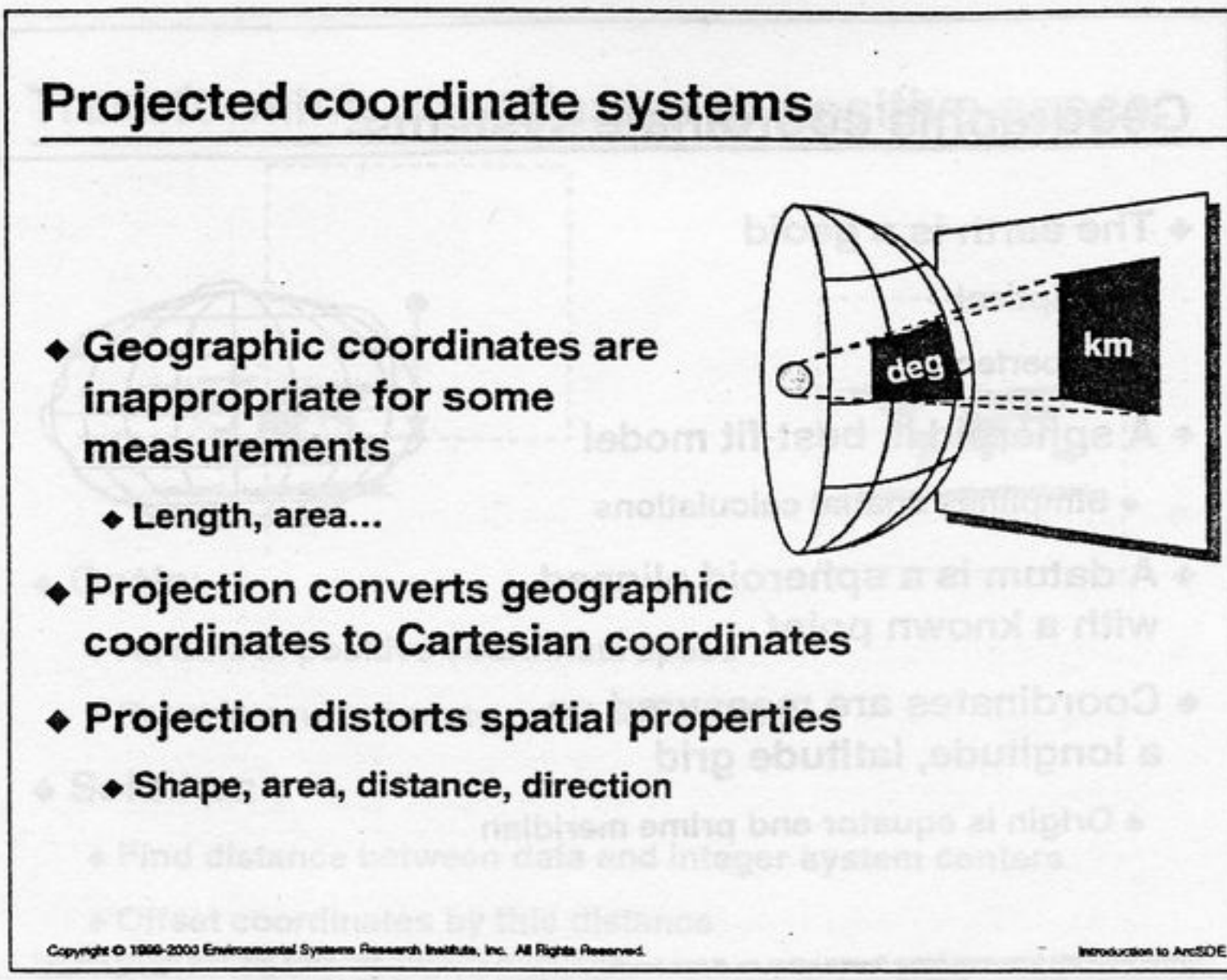
坐标系属性

坐标系用于确定地球表面的地理位置。地球是一个三维的不规则的椭球体。这种形状的椭球体有一个比 y 轴（短轴）长的 x 轴（长轴）。

因为地球形状不规则，地理学家用一个椭球体表示它的形状。这个椭球体是地球形状的一个规则近似体，在上面可以确定位置。在实际应用中产生了很多种椭球体，为了更精确地模拟地球，每个都使用了复杂的测量方法。

椭球体被定位到地球上一个特定的位置产生一个基准面。有两种类型的基准面：地心基准面和本地基准面。地心基准面表示椭球体的中心被调整到地球的已知质心。本地基准面则被设计成与地球某一特定的位置相对应。本地基准面用于精确地模拟特定的区域，但它不能象地心基准面那样精确地模拟整个地球。

在三维球体中，现实世界用经度和纬度来量测。经度和纬度的原点是赤道和本初子午线的交点。全球的经度和纬度的度是不均一的，在赤道上 1 个经度的长度为 111 公里，而在极地（ 9 0 度纬度）的长度为 0 公里。



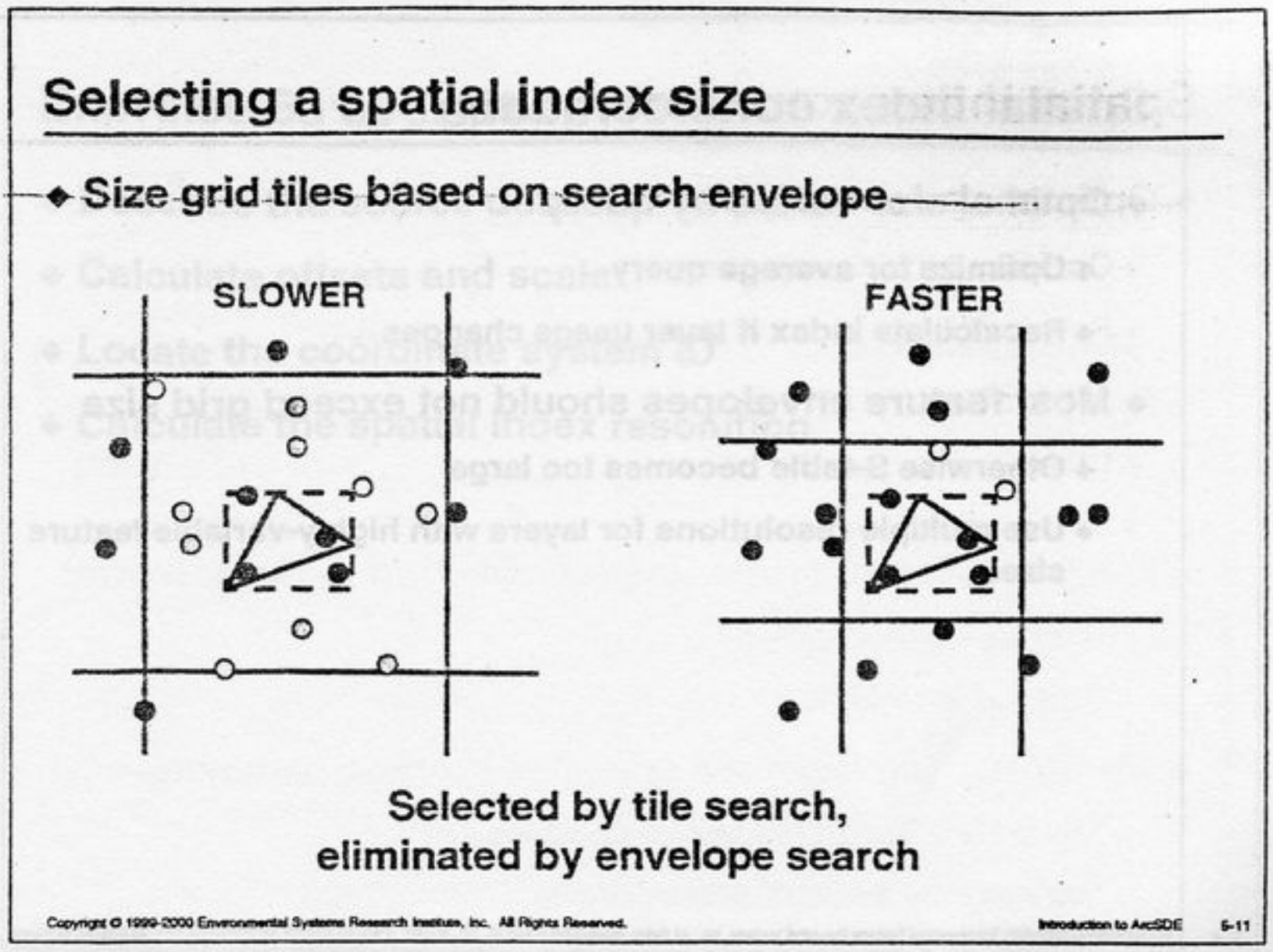
投影坐标系

平面或笛卡尔坐标系

地球表面的地理坐标（经纬度）是不均一的。因为地球上的度在值上有差异，不适合用它量度长度、面积、方向或形状。通过投影把球面位置转换到笛卡尔坐标系。

笛卡尔坐标系是建立在一个平面上并且用统一的 x,y 坐标量测。坐标值通常以米或英尺度量且多为正值。从球面坐标到笛卡尔坐标的投影会导致部分空间属性扭曲（形状、面积、距离和方向）。

ArcSDE 不对数据进行投影变换。可以为图层定义一个投影系统，但这并不会导致物理地改变任何坐标值。数据投影操作必须在加载 ArcSDE 前进行。在 ArcSDE 中改变投影信息不会改变坐标值，它只是改变元数据表中的参考信息。如果要改动投影，先从 ArcSDE 中提取出数据，用一个客户端进行投影，然后重新加载数据库。也可以用 API 改变投影。



确定空间索引格网单元大小

空间查询

当 ArcSDE 客户端在空间查询中使用一个空间过滤器时，它将减少被 RDBMS 服务器返回的要素的数目。空间查询利用空间索引表即 S 表来加速从 F 表中检索数据的速度。RDBMS 用 S 表选择一个子记录集。ArcSDE 从 F 表选择这些记录并执行最后的 shape 比较以获取满足检索条件的记录，最终被返回到 ArcSDE 客户端。

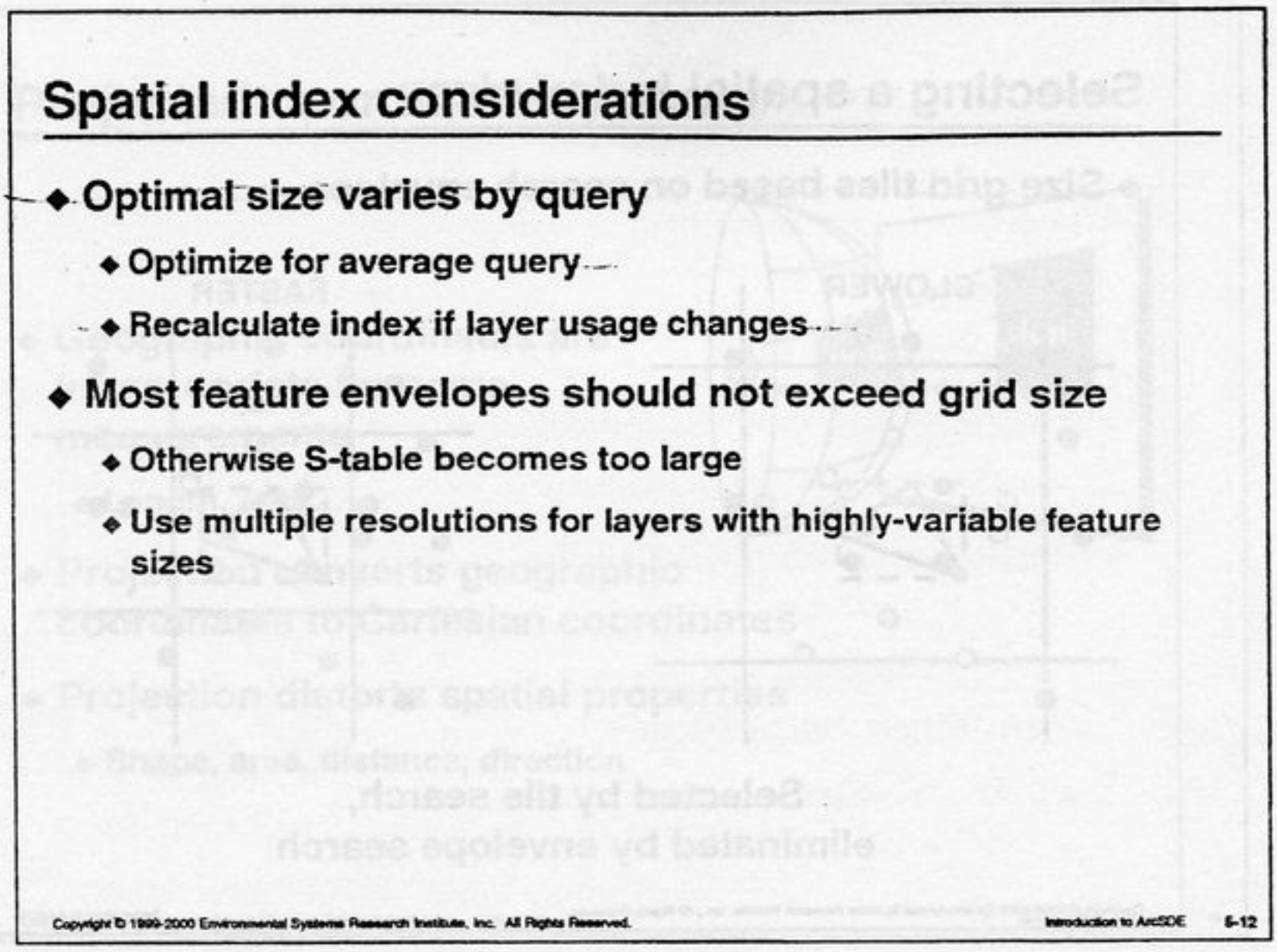
空间索引表

空间索引表含有格网单元、要素封装边界和要素标识码。第一阶段使用来自 S 表的格网单元，ArcSDE 确定哪些格网单元与空间过滤器相覆盖，RDBMS 搜索 S 表中匹配的格网单元。第二阶段，一旦格网单元被找到，通过格网单元访问到的要素封装边界将与空间过滤器的封装边界比较，这将段排除初始格网单元中索引出来的一些不相关的要素。从第一和第二阶段获取的记录再到 F 表检索出要素返回给 ArcSDE 以用于最后的 shape 比较阶段。

格网单元大小

对于许多给定的空间查询，被 ArcSDE 处理的要素的数目是不变的。封装边界查询总是给 ArcSDE 返回相同数目的要素用于下一步空间查询，而不管格网单元的大小。然而，封装边界比较过程（第二阶段）将受格网单元大小的影响。

与要素封装边界大小相比，格网单元很大时，将导致每个格网单元内有很多要素封装边界。第一阶段将选择格网，第二阶段将不得不处理大量格网内的要素封装边界，潜在地增加了查询的时间。相对于要素封装边界大小，一个很小的格网单元将导致产生大量格网都含有同一要素封装边界的 S 表。这将使得格网单元搜索缓慢，因为 RDBMS 有较多的记录要搜索和选择。



建立空间索引需要考虑的问题

选择格网单元的大小

格网单元的大小不是一个确定性的问题，需要多次尝试和努力才会得到好的结果。有一些确定格网初始值的原则，用它们可以进一步确定最佳的格网大小。可在任何时候重新计算格网的大小，使 RDBMS 重建 S 表。

最佳格网单元的大小可能受图层平均查询的影响。如果用户经常对图层执行相同的查询，格网的大小应是查询空间范围的 1.5 倍。

格网的大小不能小于要素封装边界的平均大小，否则会导致 S 表很多格网单元都索引出相同的要素封装边界。当要素封装边界被格网单元的边框切割，S 表变大因而查询格网单元时间增加长。为了减少每个格网单元有多个要素封装边界的可能性，格网单元的大小应取平均格网单元的 3 倍。

如果要素封装边界大小变化比较大，可以选择多种格网大小。ArcSDE 最多提供三级格网单元。每一要素封装边界在一合适的级内，这减少了每一封装边界有多个格网的可能性。但在空间索引搜索过程中，RDBMS 则必须搜索所有 3 个格网单元级，这将消耗大量的时间。

显示范围

当显示图层时（例如：放大和缩小），须执行空间搜索。当全范围显示图层时，ArcSDE 不执行空间搜索；系统知道显示整个图层并从 F 表获取所有的记录。ArcSDE 绕过空间索引表所以加快了显示大图层的速度。也可使用数据加载时定义的显示范围参数来指定空间搜索时的缩放分辨率，而不必搜索整个 F 表。在特定的缩放分辨率下，显示整个 F 表比执行一个空间搜索快得多。

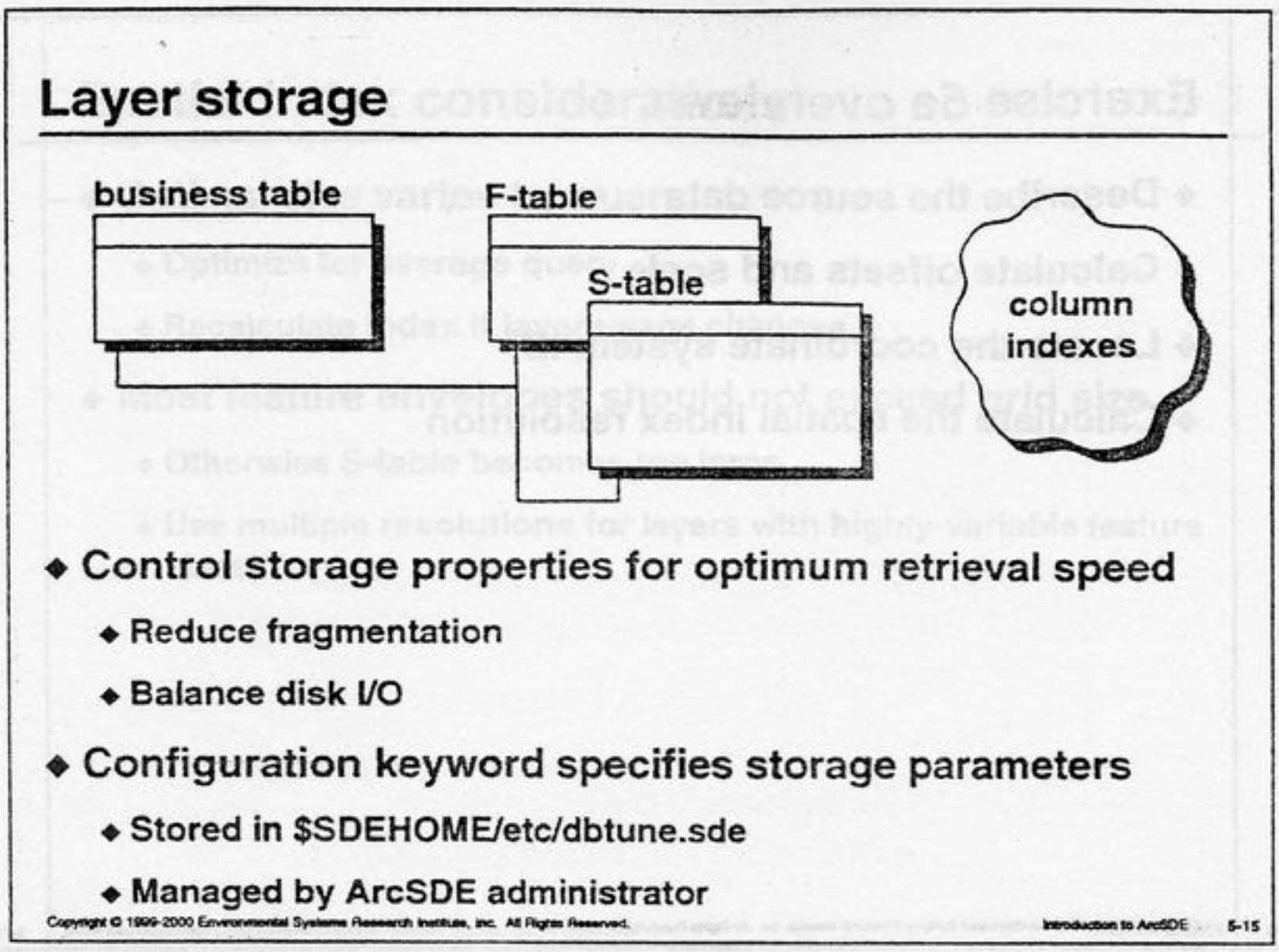
Exercise 5a overview

- Describe the source data
- Calculate offsets and scale
- Locate the coordinate system ID
- Calculate the spatial index resolution

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

练习 5a

在本练习中，你将计算练习 5b 中加载数据需要的参数。



图层的存贮结构

ArcSDE 的图层是通过 RDBMS 中的组合表来存贮的。业务表存贮图层属性和唯一的要素标识码。F 表存贮几何 shape，S 表包含空间索引信息。各表通过要素标识码连接并由 ArcSDE 维护。

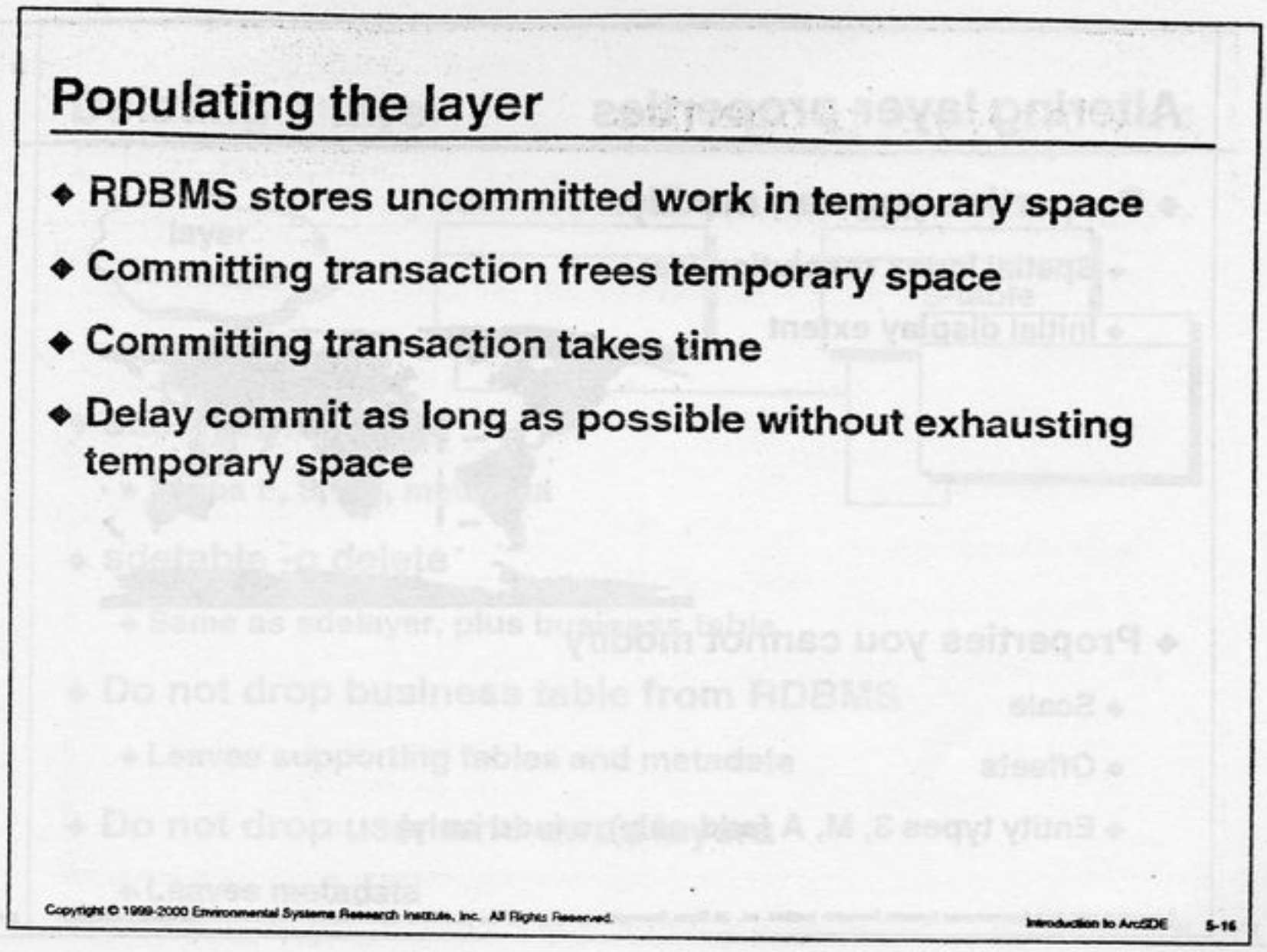
ArcSDE 通过网络向客户端传输数据，因此控制图层各表的存贮结构非常重要，这样才能获得最佳查询响应。每个 RDBMS 系统都有自己特殊的调整参数，它们由数据库管理员指定。数据库管理员应该平衡网络中硬盘的输入/输出情况以减少资源争夺。减少硬盘碎片也是很重要的，因为它们会使服务器减慢。

ArcSDE 索引

通常，索引定义在频繁访问的数据库列上。ArcSDE 为图层建立特定的索引。这些索引由 ArcSDE 生成和维护且不能改动。数据库管理员可以在业务表存贮的属性信息上产生索引，但不可以在 ArcSDE 中用来进行空间查询的数据列上产生索引。ArcSDE 自身已经经过优化，去使用特定的索引和程序访问数据。业务表空间列存贮着要素标识码，它用于连接 S 表和 F 表，这一列已被建立索引。在 F 表中，要素标识码、面积列和长度列也建立有索引。在 S 表上，要素标识码上被定义了一个索引，格网单元和要素封装边界上定义了一个复合索引。

Dbtune.sde

ArcSDE 用 dbtune.sde 文件（在%sdehome%\etc 目录下）指定图层的存贮参数。ArcSDE 管理员可以利用该文件得知图层在数据库内是如何存贮的。文件使用关键字细化了图层的存贮参数。##DEFAULT 关键字用作所有没有指定关键字的图层。被加载的图层的存贮参数可以被增加到该文件。存贮参数依赖于 RDBMS。



加载图层数据

ArcSDE 使用 INSERT 语句把数据加载到 RDBMS。在数据加载到 RDBMS 时先被存在一个临时位置直到执行 COMMIT。执行后，临时数据被物理地写入硬盘。把数据存在临时位置可以在必要时进行撤消操作（undo）。

在多个处理已经执行完毕并且已把数据存入临时位置后才可执行 COMMIT，因为写盘需花费时间。适时执行 COMMIT 是重要的，因为临时空间满了，数据库将自动进行撤消操作。如果 COMMIT 间隔频繁，写入数据的时间又会很长。

ArcSDE 以 load_only_io 模式加载图层，这意味着对图层的访问仅限于加载。在 load_only_io 模式内，将不产生空间索引表和列索引表。一旦要素被加载到数据库，图层转入 normal_io 模式。该模式将建立 S 表和所有的索引并允许对图层的正常存取。

Altering layer properties

- Properties you can modify
  - Spatial index resolution (5a)
  - Initial display extent



- Properties you cannot modify
  - Scale
  - Offsets
  - Entity types 3, M, A (add only), c (add only)

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

改变图层参数

可修改的参数

在加载图层时定义的一些属性，可以在加载后可以改动。通过“sdelayer –o grant”命令授予或取消用户的权限来改变图层存取/权限。

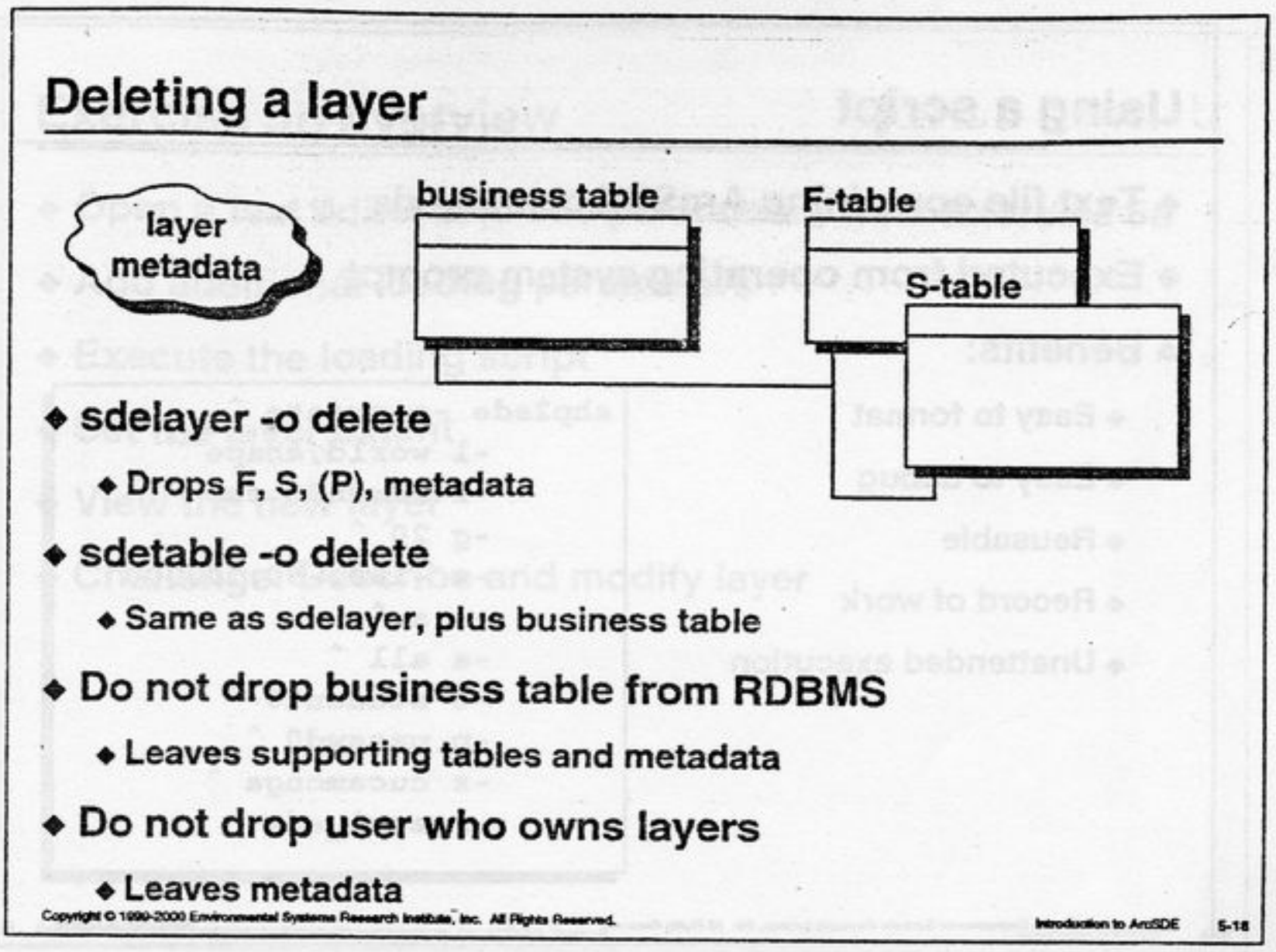
第一次显示时，图层的初始显示范围决定了它今后将如何被显示。在每次把图层传输给 ArcSDE 客户端时，它可用于把图层缩放到这个特定的空间位置。这个初始值也决定了绕过空间索引表和整个 F 表存取的缩放分辨率。“sdelayer –o alter”命令用于改变这个属性。

在任何时候，空间索引分辨率（格网单元的大小）可通过“sdelayer –o alter”命令改变。新的大小值将用于重新生成 S 表。在改变格网大小前，ArcSDE 自动把图层变成 load-only-io 模式，防止重新生成 S 表时重建索引。一旦格网大小被改变且 S 表被建立，ArcSDE 把图层变回 normal-io 模式以生成索引。

不能修改的参数

一旦图层存入 RDBMS，图层的很多加载时定义的参数不能修改。比例系数和偏移量决定了 F 表中的坐标值。要改变这些参数须把数据重新加载到数据库。

ArcSDE 内的实体类型可被有限地修改。高程和度量值类型不能加到图层上。因为高程或度量值类型将影响数据的存贮并迫使 RDBMS 给 vertex 点分配更多的空间。如果加入了高程、度量值类型，F 表内的记录会出现链接错误。注记和 CAD 实体可以加入图层，一旦加入后就不能删除。要删除高程、度量值、注记或 CAD 实体，需要重新加载图层且加载时不指定这些实体类型作为参数。



删除图层

ArcSDE 图层的完整性

ArcSDE 控制着图层表和 RDBMS 中的元数据。使用 ArcSDE 工具、API 或 ArcSDE 客户端去维护数据库的完整性是很重要的.

使用"sdelayer -o delete"或"sdetable -o delete"命令可以完成删除图层操作。这些命令删除图层和 ArcSDE 内所有元数据的参考信息。

"sdelayer -o delete"命令将删除空间信息（F、S、P 表）和元数据中相关联的记录。业务表仍然在数据库中存在且可被其它应用程序使用，但空间数据不存在了。

"sdetable -o delete"命令将删除 F、S（P）表、元数据和业务表。该命令删除图层的所有参考信息。

一定要使用 ArcSDE 去删除图层。如果直接在 RDBMS 中删除表，那么数据库表和元数据的完整性被破坏，在数据库中会留下一些孤立的对象。这些对象需要手工删除，或在元数据表中删除某些记录。

Using a script

- Text file containing ArcSDE commands
- Executed from operating system prompt
- Benefits:
  - Easy to format
  - Easy to debug
  - Reusable
  - Record of work
  - Unattended execution

```
shp2sde -o create ^
        -l world,shape ^
        -f admin98.shp ^
        -g 20 ^
        -x -180,-90,100000 ^
        -e a ^
        -a all ^
        -u student0 ^
        -p passwd0 ^
        -s cucamonga ^
        -i esri_sde
```

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

使用脚本

加载脚本

通过命令行加载数据灵活快捷。编写脚本或批处理文件加载非常方便，因为它容易编辑，可用于重复工作，还提供说明。脚本包含 ArcSDE 命令，从命令行运行。

脚本必须包含跟有参数的 ArcSDE 命令。格式化的脚本可使用户很快地理解命令和指定参数。如果使用续行符（在 NT 中是“^”，在 UNIX 中是“\”），确保在回车前没有额外的空格。

Exercise 5b overview

- Open a text editor and add parameters from Exercise 5a
- Add additional loading parameters
- Execute the loading script
- Set the layer extent
- View the new layer
- Challenge: Describe and modify layer

Copyright © 1999-2000 Environmental Systems Research Institute, Inc. All Rights Reserved.

练习 5d

在本练习中，你将使用练习 5a 中计算的参数来向 ArcSDE 加载图层数据。