Beginning Spatial with SQL Server 2008

ISBN-13 (pbk): 978-1-4302-1829-6

ISBN-13 (electronic): 978-1-4302-1830-2

Printed and bound in the United States of America 9 8 7 6 5 4 3 2 1

Lead Editor: Jonathan Gennick
Technical Reviewer: Evan Terry
Editorial Board: Clay Andres, Steve Anglin, Mark Beckner, Ewan Buckingham, Tony Campbell, Gary Cornell, Jonathan Gennick, Michelle Lowman, Matthew Moodie, Jeffrey Pepper, Frank Pohlmann, Ben Renow-Clarke, Dominic Shakeshaft, Matt Wade, Tom Welsh
Project Manager: Kylie Johnston
Copy Editor: Bill McManus
Associate Production Director: Kari Brooks-Copony
Production Editor: Katie Stence
Compositor: Susan Glinert
Proofreader: April Eddy
Indexer: Julie Grady
Cover Designer: Kurt Krames
Manufacturing Director: Tom Debolski

Distributed to the book trade worldwide by Springer-Verlag New York, Inc., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax 201-348-4505, e-mail orders-ny@springer-sbm.com, or visit http://www.springeronline.com.

For information on translations, please contact Apress directly at 2855 Telegraph Avenue, Suite 600, Berkeley, CA 94705. Phone 510-549-5930, fax 510-549-5939, e-mail info@apress.com, or visit http://www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at http://www.apress.com/info/bulksales.

The source code for this book is available to readers at http://www.apress.com.

CHAPTER 1

Defining Spatial Information

Spatial data analysis is a complex subject area, taking elements from a range of academic disciplines, including geophysics, mathematics, astronomy, and cartography. Although you do not need to understand these subjects in great depth to start using the new spatial features of SQL Server 2008, it is important to have a basic understanding of the theoretical concepts involved so that you use spatial data appropriately and effectively in your applications.

In this chapter you will learn how different spatial reference systems identify positions in space, and how these systems can be used to define spatial objects representing features on the earth. These concepts are fundamental to the creation of consistent, accurate spatial data, and will be used throughout the practical applications discussed in later chapters of this book.

What Is Spatial Data?

Spatial data describes the position, shape, and orientation of objects in space.

In this book, as in most common applications, we are particularly concerned with describing the position and shape of objects on the earth. This is known as *geospatial* data. Geospatial data can describe the properties of many different sorts of "objects" on the earth. These objects might be tangible, physical things, such as an office building or a mountain, or abstract features, such as the imaginary line marking the political boundary between countries.

Uses of Spatial Data

Spatial data provides information that can be used in a wide range of different areas. Some potential applications are as follows:

- Analyzing regional, national, or international sales trends
- Deciding where to place a new store based on proximity to customers and competitors
- Navigating to a destination using a Global Positioning System (GPS) device
- Allowing customers to track the delivery of a parcel

- Monitoring the routes of vehicles in a logistics network
- Optimizing distribution networks to provide the most efficient coverage of an area
- Reporting geographic-based information on a map rather than in a tabular or chart format
- Providing location-based services, such as providing a list of nearby amenities for any given address
- Assessing the impact of environmental changes, such as identifying houses at risk of flooding caused by rising sea levels

All of these examples rely on the ability of spatial data to describe the position and shape of objects on the earth in a structured, consistent way.

Representing Features on the Earth

In real life, objects on the earth often have complex, irregular shapes. It would be very hard, if not impossible, for any item of spatial data to define the exact shape of these features. Instead, spatial data represents these objects by using simple, geometrical shapes that approximate their actual shape and position. These shapes are called *geometries*.

SQL Server 2008 supports three main types of geometry that can be used to represent spatial information: Points, LineStrings, and Polygons. In this section I describe the properties of each of the three types in turn, and then I show how you can use them to represent various features on the earth.

Points

A *Point* is the most fundamental type of geometry, and is used to define a singular position in space. A Point object is zero-dimensional, meaning that it does not have length or area. Figure 1-1 illustrates a representation of a Point geometry.

Figure 1-1. *A Point geometry*

When using geospatial data to define features on the earth, a Point geometry is used to represent an exact location, which could be a street address or the location of a bank, volcano, or city, for instance. Figure 1-2 illustrates several Point geometries used to represent the locations of major cities in Australia.

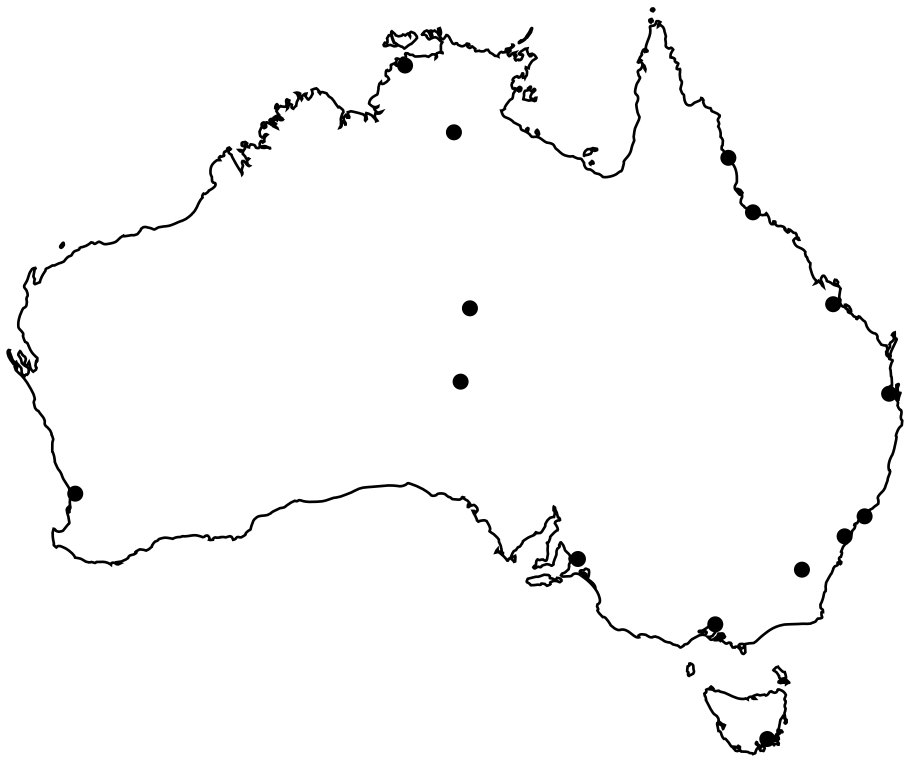

Figure 1-2. *A series of Point geometries representing cities in Australia*

LineStrings

Having defined a series of two or more points in space, we then can draw straight lines connecting each point to the next point in the series, to define a *LineString*. LineStrings comprise a series of two or more distinct points and the line segments that connect them. LineStrings are one-dimensional spatial objects—they have a specified length, but do not contain any area.

LineStrings may be described as having the following additional characteristics:

- A *simple* LineString is one in which the path drawn between the points of the LineString does not cross itself.
- A *closed* LineString is one that starts and ends at the same point.
- A LineString that is both simple and closed is known as a *ring*. Even though a ring appears to represent the perimeter of a closed shape, it does not include the area enclosed within the shape—it only defines the points that lie on the line itself.

Different examples of LineString geometries are illustrated in Figure 1-3.

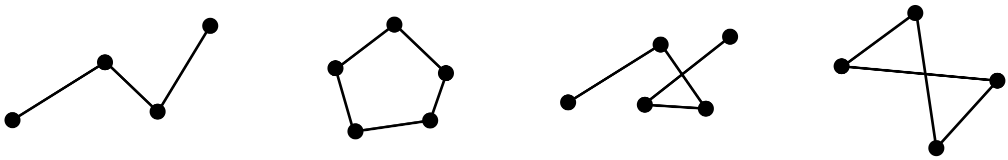

Figure 1-3. *Examples of LineString geometries (from left to right): a simple LineString; a simple, closed LineString (a ring); a nonsimple LineString; a nonsimple, closed LineString*

In geospatial data, LineStrings are commonly used to represent features such as roads, rivers, delivery routes, or contours of the earth. Figure 1-4 shows numerous LineStrings used to represent major rivers in France.

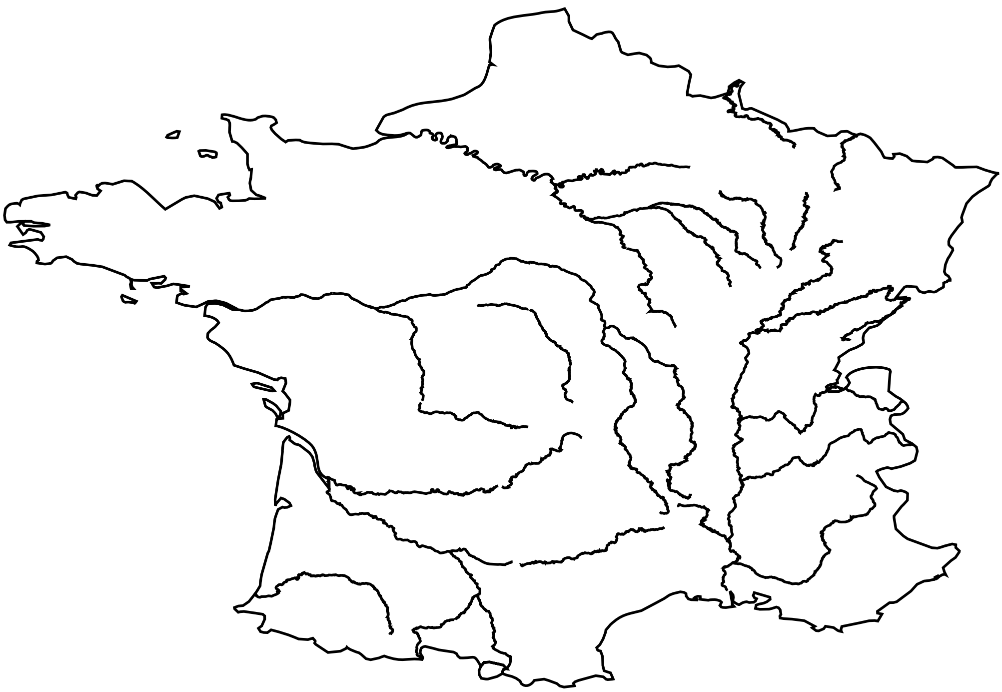

Figure 1-4. *A series of LineString geometries representing major rivers in France*

■ **Note** Some geographic information systems (GISs) make a distinction between a *LineString* and a *Line*. According to the Open Geospatial Consortium (OGC) Simple Features for SQL Specification (a standard on which the spatial features of SQL Server 2008 are largely based), a Line connects exactly two points, whereas a LineString may connect any number of points. Since all Lines can be represented as LineStrings, of these two types, SQL Server 2008 only implements the LineString geometry.

Polygons

A Polygon geometry is defined by a boundary of connected points that forms a closed LineString, called the exterior ring. In contrast to a simple, closed LineString geometry, which only defines those points lying on the ring itself, a Polygon geometry also contains all the points that lie in the interior area enclosed within the exterior ring.

Every Polygon must have exactly one external ring that defines the overall perimeter of the shape, and may also contain one or more internal rings. Internal rings define areas of space that are contained within the external ring but not included in the Polygon definition. They can therefore be thought of as "holes" that have been cut out of the main geometry.

Since Polygons are constructed from a series of one or more rings, which are simple, closed LineStrings, all Polygons themselves are deemed to be simple, closed geometries. Polygons are two-dimensional geometries—they have an associated length and area. The length of a Polygon is measured as the sum of the distances around the perimeter of all the rings of that Polygon (exterior and interior), while the area is calculated as the space contained within the exterior ring, excluding the area contained within any interior rings. Some examples of Polygon geometries are illustrated in Figure 1-5.

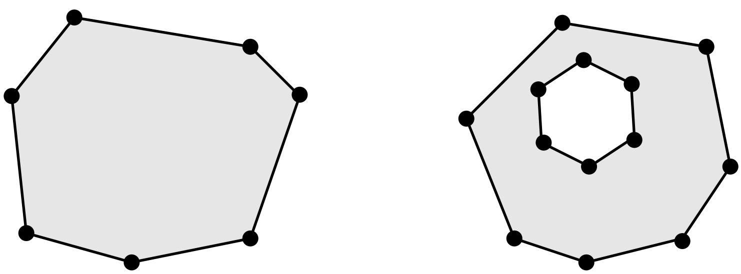

Figure 1-5. *Examples of Polygon geometries (from left to right): a Polygon; a Polygon with an interior ring*

Polygons are frequently used in spatial data to represent geographic areas such as islands or lakes, political jurisdictions, or large structures. Figure 1-6 illustrates Polygon geometries that represent the 48 contiguous states of the mainland United States.

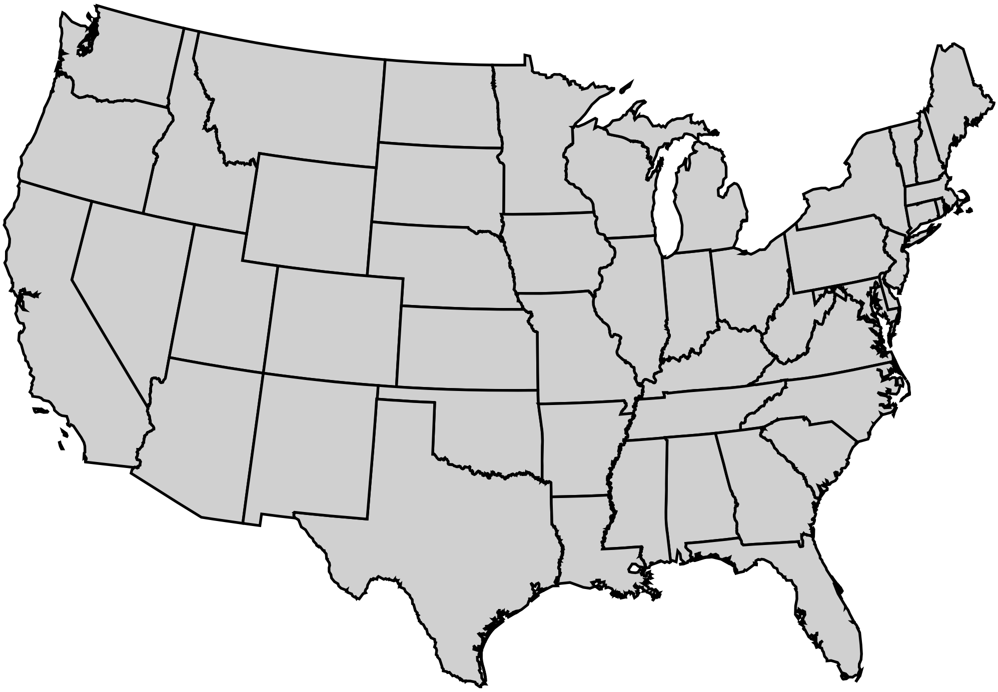

Figure 1-6. *A series of Polygon geometries representing states of the United States*

Choosing the Right Geometry

There is no "correct" type of geometry to use to represent any given object on the earth. The choice of which geometry to use will depend on how you plan to use the data. If you are going to analyze the geographic spread of your customer base, you *could* define Polygon geometries that represent the shape of each of your customers' houses, but it would be a lot easier to consider each customer's address as a single Point. In contrast, when conducting a detailed analysis of a small-scale area for land-planning purposes, you may want to represent all buildings, roads, and even walls as Polygons that have both length and area, to ensure that the spatial data represents their actual shape as closely as possible.

Combining Geometries in a Geometry Collection

Sometimes, what could be considered a single object on the earth may be represented using a combination of several geometry objects. For instance, the Great Wall of China is not a single continuous wall, but rather it is made up of numerous separate sections of wall. As such, the overall shape of the wall may be best represented as a collection of LineStrings. Similarly, a single country spread over several islands, such as Japan, may be represented by a collection of Polygons, each one representing the shape of an individual island. When you define a single object that contains several individual geometries in this way, it is called a *Geometry Collection*. A Geometry Collection may contain any number of any type of geometries. In the specific case in which a Geometry Collection contains only multiple elements of the same type of geometry, it is referred to as a MultiPoint, MultiLineString, or MultiPolygon geometry.

DEFINING AUGUSTA NATIONAL GOLF COURSE

In order to demonstrate the different ways in which spatial data can describe the same object on the earth, let me show you a practical example. Suppose we want to store an item of spatial data describing the course at Augusta National Golf Club (in Augusta, Georgia), home of the annual US Masters Tournament.

If we were storing information for a tourist database of interesting places to visit in Georgia, it would probably suffice to describe the entire golf course using a Point geometry. This Point could describe the approximate location of the course, and would be perfectly sufficient to perform spatial calculations such as finding the distance to the closest airport, or identifying nearby places to stay.

Alternatively, we could choose to represent the course as a geometry collection containing many elements that describe the individual features of the course much more accurately: we could represent the greens and the fairways of each hole as separate Polygons; use Point objects to represent each tee; and use LineString objects to show the optimum drive off the tee. This sort of representation would be more suitable for use by a golfer who is actually playing the course, accessing spatial data via a mobile GPS system to plan their next shot.

Both of these alternative representations are equally valid—the choice simply depends on the application of the data.

Understanding Interiors, Exteriors, and Boundaries

Every geometry shape divides space into three areas relative to that geometry: the *interior*, *exterior*, and *boundary*. In the field of topological mathematics, these terms have very specific definitions, but you can think of them simply as follows:

- The interior of a geometry consists of all the points that lie in the space occupied by the geometry.
- The exterior consists of all the points that lie in the space not occupied by the geometry.
- The boundary of a geometry consists of the points that lie on the "edge" of the geometry. In SQL Server, every geometry is considered to be *topologically closed*; that is, any points that lie on the boundary of a geometry are contained within the interior of the geometry.

Every geometry specifies one or more points in their interior and exterior, although only certain types of geometry contain points in their boundaries. The classification of these different areas of space for each type of geometry follows:

Point and MultiPoint geometries: Represent singular locations, where the interior consists of the individual point(s) defined by that object. However, they do not have a defined boundary.

LineString and MultiLineString geometries: Have an interior consisting of all the points that lie on the straight line segments drawn between the defined series of points. Nonclosed LineStrings and MultiLineStrings have a boundary consisting of the points at the start and end of the LineString. However, closed LineStrings—those that start and end at the same point—do not have a boundary.

Polygon and MultiPolygon geometries: Have an interior consisting of all the points contained within the exterior ring, excluding those contained within any interior ring. The boundary of these types of geometry consists of the closed LineString that forms the exterior ring itself, together with any interior rings defined by that Polygon.

The distinction between these classifications of space becomes very important when expressing the relationship between different spatial objects, since these relationships are generally based on comparing where particular points lie with respect to the interior, exterior, or boundary of the two geometries in question. For instance, two geometries *intersect* each other if they share at least one point in common. However, they are only deemed to *touch* each other if the points that they share lie only on the boundaries of each geometry. This concept is discussed in more detail in Chapter 13.

Positioning a Geometry

After we choose an appropriate geometry (Point, LineString, or Polygon) to represent a given object, we then need to position it in the right place on the earth. We do this by relating each point in the geometry definition to the relevant real-world position it represents. For example, if we want to use a Polygon geometry to represent the US Department of Defense Pentagon building, we need to specify that the five points that define the boundary of the Polygon geometry relate to the location of the five corners of the building. So, how do we do this?

You are probably familiar with the terms *longitude* and *latitude*, and have seen them used to describe positions on the earth. If this is the case, you may be thinking that we can simply express the latitude and longitude coordinates of the relevant position on the earth for each point in the geometry. Unfortunately, it's not quite that simple.

What many people don't realize is that any particular point on the ground does not have a unique latitude or longitude associated with it. There are in fact many systems of latitude and longitude, and the coordinates of a given point on the earth will differ depending on which system is used. Furthermore, latitude and longitude coordinates are not the only way of expressing positions—there are other types of coordinates that define the location of an object without using latitude and longitude at all. In order to understand how to specify the position of your geometry on the earth, you first need to understand how different spatial reference systems work.

COMPARING RASTER TO VECTOR DATA

There are two main ways of modeling spatial information: using a vector model or using a raster model.

Vector data, discussed in this chapter, describes discrete spatial objects by defining the coordinates of geometries that approximate the shape of those features. Vector spatial information is best suited to represent discrete items of spatial data, such as the location of individual customers or warehouses, or the path of roads.

In contrast, raster data represents spatial information using a matrix of cells. These cells are arranged into a grid that is overlaid onto the surface of the earth. The value of each cell in the matrix represents a property of the underlying area covered by that grid cell. One example of raster spatial data is aerial or satellite imagery, in which case the matrix grid is the set of pixels that forms the image, and the value of any individual cell is the color of the associated pixel. However, raster data can also be used to describe any other spatial information. It is particularly suited to data that can take a continuous range of values, such as when depicting the levels of rainfall across an area of land, or the depth of an area of water.

All the spatial features in SQL Server 2008 (and therefore discussed in this book) are based on a vector model of spatial data. There is currently no built-in support for raster data in SQL Server. However, in Chapter 9, I will show you how to overlay vector shape information with raster imagery of the earth, by combining spatial data from SQL Server with the Microsoft Virtual Earth and Google Maps web services.

Describing Positions Using a Coordinate System

The purpose of a spatial reference system is to unambiguously identify and describe any point in space. This ability is essential to enable spatial data to define the positions of points that make up the various kinds of geometry used to represent features on the earth. To describe the positions of points in space, every spatial reference system is based on an underlying coordinate system. A coordinate reference is a conventional and widely accepted way of describing the position of a point from a given origin, in a given dimension. A set of *n* coordinates, such as (1, 2, 3, . . ., *n*), can therefore be used to describe the position of a point from an origin in *n*-dimensional space.

There are many different types of coordinate systems; when you use geospatial data in SQL Server 2008, you are most likely to use a spatial reference system based on either a geographic or projected coordinate system.

Note A set of coordinate values is called a coordinate *tuple*.

Geographic Coordinate System

In a geographic coordinate system, any position on the earth's surface can be defined using two coordinates:

The *latitude* coordinate of a point measures the angle between the plane of the equator and a line drawn perpendicular to the surface of the earth at that point. (This is the definition of *geodetic latitude*. An alternative measure, *geocentric latitude*, is defined as the angle between the plane of the equator and a line drawn from a point on the earth's surface to the center of the earth.)

The *longitude* coordinate measures the angle in the equatorial plane between a line drawn from the center of the earth to the point and a line drawn from the center of the earth to the prime meridian. The prime meridian is an imaginary line drawn on the earth's surface between the North Pole and the South Pole (so technically it is an *arc* rather than a line), chosen to be the line from which angles of longitude are measured.

These concepts are illustrated in Figure 1-7.

Caution Since a point of greater longitude lies further east, and a point of greater latitude lies further north, it is a common mistake for people to think of latitude and longitude as measured on the earth's surface itself, but this is not the case—latitude and longitude are angles measured from the plane of the equator and prime meridian at the center of the earth.

Coordinates of latitude and longitude are both angles, and are usually measured in degrees. In this case, longitude values measured from the prime meridian range from –180° to +180°, and latitude values measured from the equator range from –90° (at the South Pole) to +90° (at the North Pole).

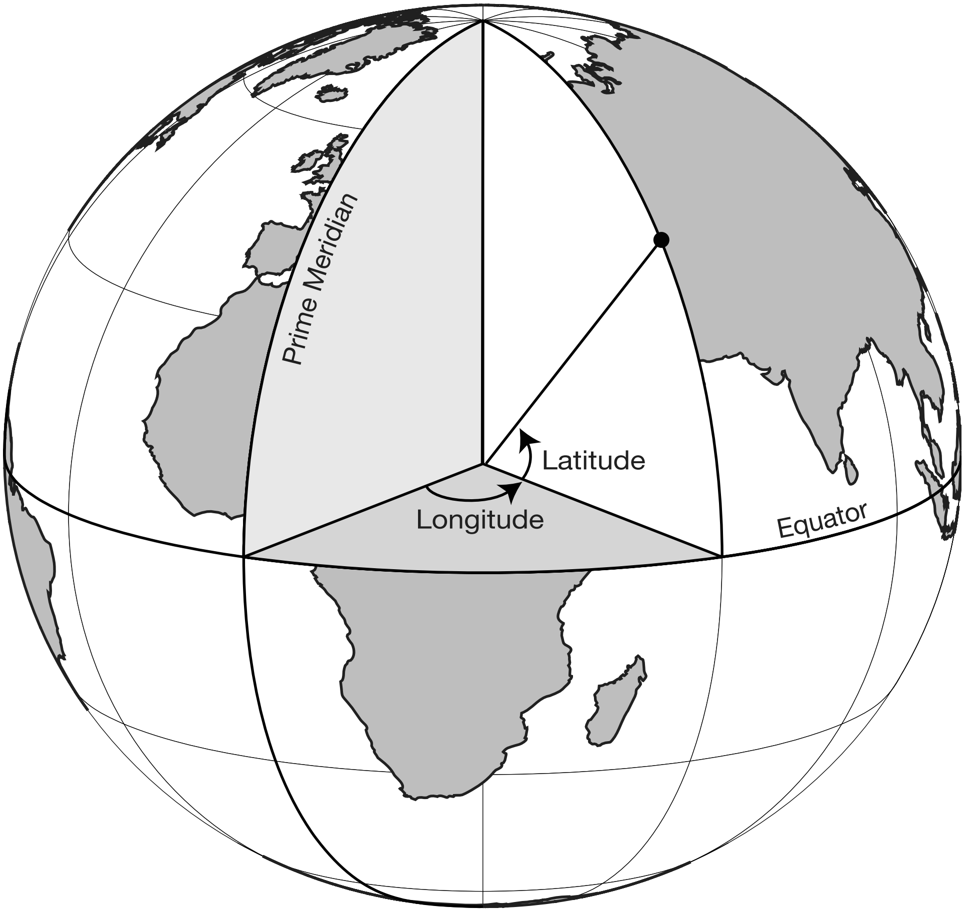

Figure 1-7. *Describing a position on the earth using a geographic coordinate system*

Longitudes to the east of the prime meridian are normally stated as positive values, or suffixed with the letter *E*. Longitudes to the west of the prime meridian are expressed as negative values, or using the suffix *W*. Likewise, latitudes north of the equator are expressed as positive values, or using the suffix *N*, whereas latitudes south of the equator are expressed as negative values, or using the suffix *S*.

There are several accepted notation methods for expressing values of latitude and longitude:

> The most commonly used method is the degrees, minutes, seconds (DMS) system, also known as sexagesimal notation. In this system, each degree is divided into 60 minutes. Each minute is further subdivided into 60 seconds. A value of 51 degrees, 15 minutes, and 32 seconds is normally written as 51°15'32".

The system most commonly used by GPS receivers is to display whole degrees, and then minutes, and decimal fractions of minutes. This same coordinate value would therefore be written as 51:15.53333333.

Decimal degree notation specifies coordinates using degrees and decimal fractions of degrees, so the same coordinate value expressed using this system would be written as 51.25888889.

CONVERTING TO DECIMAL DEGREE NOTATION

When expressing geographic coordinate values of latitude and longitude for use in SQL Server 2008, you should use decimal degree notation. The advantage of this format is that each coordinate can be expressed as a single floating-point number. To convert DMS coordinates into decimal degrees, you can use the following rule:

Degrees + (Minutes / 60) + (Seconds / 3600) = Decimal Degrees

For example, the US Central Intelligence Agency's online edition of *The World Factbook* (`https://www.cia.gov/library/publications/the-world-factbook/geos/uk.html`) gives the geographic coordinates for London as follows:

51 30 N, 0 10 W

When expressed in decimal degree notation, this is

51.5 (Latitude), –0.166667 (Longitude)

When converting a coordinate value from DMS to decimal degree notation, you should state the accuracy of the result with up to 15 significant figures, because this is the precision with which the converted coordinate value will be stored in SQL Server.

Projected Coordinate System

In contrast to the geographic coordinate system, which defines positions on a three-dimensional, round model of the earth, a projected coordinate system describes the position of points on the earth's surface as they lie on a flat, two-dimensional plane. A simple way of thinking about this is to consider a projected coordinate system as describing positions on a map rather than positions on a globe.

If we consider all of the points on the earth's surface to lie on a flat plane, we can define positions on that plane using familiar Cartesian coordinates of x and y, which represent the distance of a point from an origin along the x axis and y axis, respectively. In a projected coordinate system, these coordinate values are sometimes referred to as *easting* (the x coordinate) and *northing* (the y coordinate), as shown in Figure 1-8.

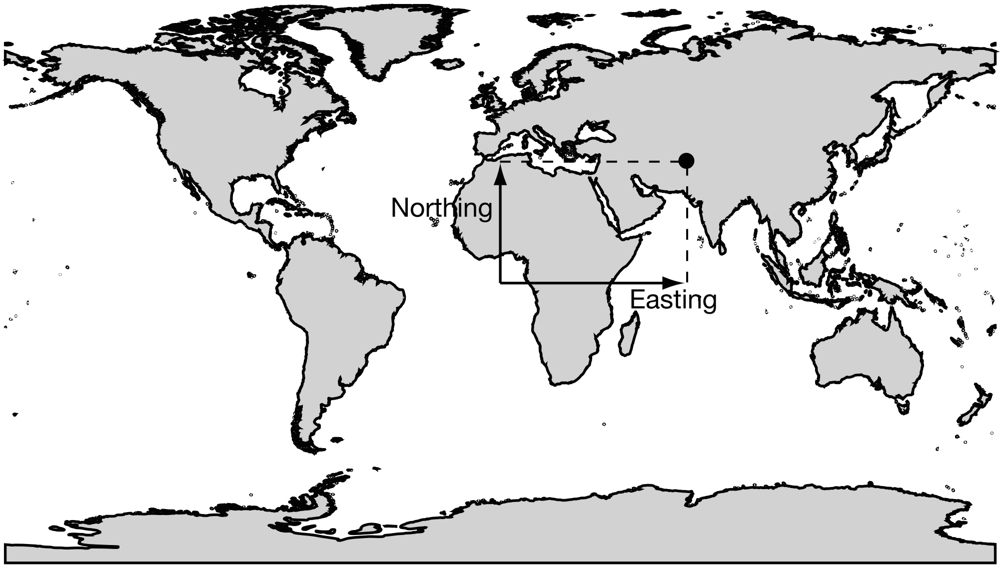

Figure 1-8. *Describing position on the earth using a projected coordinate system*

Since a projected coordinate system describes the position of an object by calculating the distance from an origin along a flat plane representing the earth's surface, northing and easting coordinate values are measured and expressed using a linear unit of measure, such as meters or feet.

Applying Coordinate Systems to the Earth

So far, we have defined two different coordinate systems that can be used to define points in theoretical space: the geographic coordinate system, which uses angular coordinates of latitude and longitude, and the projected coordinate system, which uses x and y Cartesian coordinates. However, a set of coordinates from either of these systems does not, on its own, uniquely identify a position on the earth. We need to know additional information, such as where to measure those coordinates from, in what units, and what shape to use to model the earth. For this, we need to examine the other elements of a spatial reference system—the datum, prime meridian, and unit of measurement.

Datum

A datum contains information about the size and shape of the earth. Specifically, it contains the details of a reference ellipsoid, and a reference frame. We use this information to create a geodetic model of the earth, onto which we can apply our coordinate system.

The actual shape of the earth is very complex. On the surface, we can see that there are irregular topological features such as mountains and valleys. But even if we were to remove these features and consider the mean sea level around the planet, the earth is still not a regular shape. In fact, it is so unique that geophysicists have a specific word solely used to describe the shape of the earth—the *geoid*.

When using spatial data to describe the position of geometries on the earth's surface, ideally, we would like to use coordinates that refer to positions relative to the geoid itself. However, there is no way that we can accurately model the complicated, irregular shape of the geoid, so instead we base our spatial system on an approximation of the geoid. This approximation is called a *reference ellipsoid*.

■ **Note** *Geodesy* is the science of studying and measuring the shape of the earth. A *geodetic* model is therefore a model of the shape of the earth.

Reference Ellipsoid

Despite its name, a reference ellipsoid normally describes an *oblate spheroid*, which is the three-dimensional shape obtained when you rotate an ellipse about its shorter axis. When used in spatial data modeling, spheroid models of the earth are always oblate—they are wider than they are high, and resemble a squashed sphere. This is a fairly good approximation of the shape of the geoid, which bulges around the equator.

The important feature of a spheroid is that, unlike the geoid, it is a regular shape that can be exactly mathematically described by two parameters—the length of the semimajor axis (which represents the radius of the earth at the equator), and the length of the semiminor axis (the radius of the earth at the poles). This is illustrated in Figure 1-9.

■ **Note** A spheroid is a sphere that has been "flattened" in one axis. An ellipsoid is a sphere that has been flattened in two axes—that is, the radius of the shape is different in the x, y, and z axes. Since ellipsoid models of the world are not significantly more accurate than spheroid models at describing the shape of the geoid, reference ellipsoids are rarely based on true ellipsoids, but rather on a simpler spheroid model.

An alternative method of stating the properties of an ellipsoid is to give the length of the semimajor axis and the flattening ratio of the ellipsoid. The flattening ratio, f, is used to describe how much an ellipsoid has been "squashed," and is calculated as

$$f = (a - b) / a$$

where a equals the length of the semimajor axis, and b equals the length of the semiminor axis.

In most ellipsoid models of the earth, the semiminor axis is only marginally smaller than the semimajor axis, which means that the value of the flattening ratio is also small—typically around 0.003. For the sake of convenience, many systems, including SQL Server 2008, use the inverse-flattening ratio of an ellipsoid instead. This is stated as $1/f$, and calculated as follows:

$$1/f = a / (a - b)$$

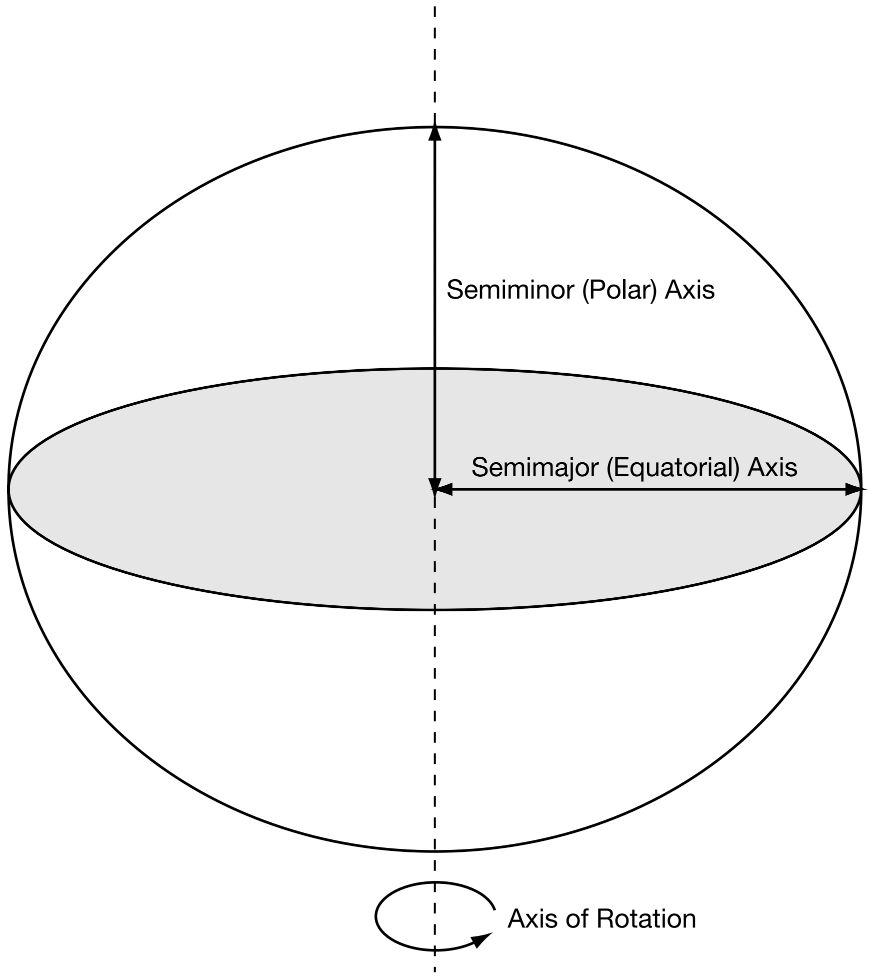

Figure 1-9. *Properties of a reference ellipsoid*

The inverse-flattening ratio of an ellipsoid model typically has a value of approximately 300. There is not a single reference ellipsoid that best represents every part of the whole geoid. Some ellipsoids, such as the WGS 84 ellipsoid used by satellite GPS systems, provide a reasonable approximation of the overall shape of the geoid. Other ellipsoids approximate the shape of the geoid very accurately over certain regions of the world, but are much less accurate in other areas. These ellipsoids are normally only applied for use in specific countries, such as the Airy 1830 ellipsoid commonly used in Britain. Figure 1-10 provides an (exaggerated) illustration of how different ellipsoid models vary in accuracy over different parts of the geoid.

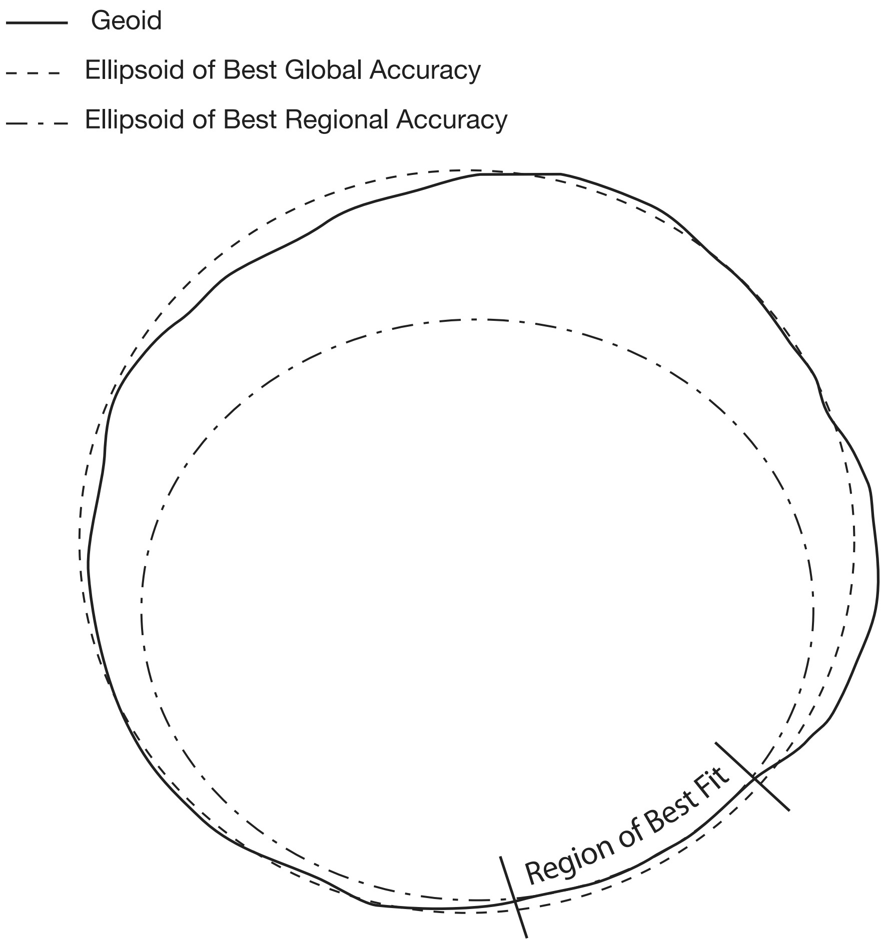

Figure 1-10. *Comparison of cross-sections of different ellipsoid models of the geoid*

It is important to realize that specifying a different reference ellipsoid to approximate the geoid affects the accuracy of how well a set of coordinates that defines a geometry on that ellipsoid reflects the actual position and shape of the feature on the earth that the geometry represents. When choosing an ellipsoid to define spatial data, we must therefore be careful to use one that is suitable for the purpose of the data in question.

SQL Server 2008 recognizes a number of different reference ellipsoids that are designed to best approximate the geoid at different parts of the earth. Table 1-1 lists the properties of some commonly used reference ellipsoids that can be used.

Table 1-1. *Properties of Some Commonly Used Reference Ellipsoids*

Ellipsoid Name	Semimajor Axis (m)	Semiminor Axis (m)	Inverse Flattening	Usage
Airy (1830)	6,377,563.396	6,356,256.909	299.3249646	Great Britain
Bessel (1841)	6,377,397.155	6,356,078.963	299.1528128	Czechoslovakia, Japan, South Korea
Clarke (1880)	6,378,249.145	6,356,514.87	293.465	Africa
NAD 27	6,378,206.4	6,356,583.8	294.9786982	North America
NAD 83	6,378,137	6,356,752.3	298.2570249	North America
WGS 84	6,378,137	6,356,752.314	298.2572236	Global

Reference Frame

Remember that we are going to use our coordinate system to define positions on the reference ellipsoid as a way of approximating positions on the earth itself. Having established our ellipsoid model, we need some way to position that model so that it lines up with the right points on the earth's surface. We do this by creating a frame of reference points.

Reference points are places (normally on the earth's surface) that are assigned known coordinates in the coordinate system relative to the ellipsoid being used. By establishing a set of points of known coordinates, we can use these points to "fix" the reference ellipsoid in the right position. Once the ellipsoid is set in place based on these known points, we can apply our chosen coordinate system to obtain the coordinates of any other points on the earth, based on the ellipsoid model. Reference points are sometimes assigned to places on the earth itself; the North American Datum of 1927 (NAD 27) uses the Clarke (1866) reference ellipsoid, primarily fixed in place at Meades Ranch in Kansas. Reference points may also be assigned to the positions of satellites orbiting the earth, which is how the WGS 84 datum used by GPS systems is realized.

When packaged together, the properties of the reference ellipsoid and the frame of terrestrial reference points form a datum. The most common datum in global use is the World Geodetic System of 1984, commonly referred to as WGS 84. This is the datum used by MapPoint and Google Earth, as well as in handheld GPS systems.

Prime Meridian

As defined earlier in this chapter, the geographic coordinate of longitude is the angle in the equatorial plane between the line drawn from the center of the earth to a point and the line drawn from the center of the earth to the prime meridian. Our spatial reference therefore needs to include a definition of what line we are using for the prime meridian—the axis from which we measure our angle of longitude.

A common misconception is to think that there is a single prime meridian based on some inherent fundamental property of the earth, but this is not the case. The prime meridian of any spatial reference system is arbitrarily chosen simply to provide a line of zero longitude from which all other coordinates of longitude can be calculated. One commonly used prime meridian is the meridian passing through Greenwich, London, but there are many others. If we were to use a different prime meridian, the value of the longitude coordinate of all the points in our system would change.

Unit of Measurement

Geographic coordinates of latitude and longitude are generally measured in degrees, but may also be measured in radians, or other angular units of measure. Every spatial reference system must explicitly state the name of the unit in which geographic coordinates are measured, together with the conversion factor from the specified unit to one radian. For instance, a spatial reference system that uses coordinates measured in degrees would include the value of $\pi/180$ (approximately 0.017453293), since this equals the value of one degree when measured in radians.

When using a projected coordinate system, the individual coordinate values represent a linear distance along the earth's surface to a point. They are measured in a linear unit of measure, such as the meter, foot, mile, or yard. Any spatial reference system based on a projected coordinate system must therefore also state the linear unit of measure in which coordinate values are stated.

Projection

Remember that a projected coordinate system defines positions on the earth as they lie on a flat, two-dimensional plane, such as a map. We see two-dimensional projections of geospatial data on an almost daily basis—in street maps, in road atlases, or on our computer screens. Given their familiarity, and the apparent simplicity of working on a flat surface rather than a curved one, you would be forgiven for thinking that defining a point on the earth using a projected coordinate system is somehow simpler than doing so using a geographic coordinate system. The difficulty associated with a projected coordinate system is that, of course, the world *isn't* a flat, two-dimensional plane. In order to be able to represent it as such, we have to use a map projection.

Projection is the process of creating a two-dimensional representation of a three-dimensional model of the earth. Map projections can be constructed either by using purely geometric methods (such as the techniques used by ancient cartographers) or by using mathematical algorithms. However, whatever method is used, it is not possible to project any three-dimensional object onto a two-dimensional plane without distorting the resulting image in some way. Distortions introduced as a result of the projection process may affect the area, shape, distance, or direction represented by different elements of the map.

By altering the projection method, cartographers can reduce the effect of these distortions for certain features, but in doing so the accuracy of other features must be compromised—there is no single ideal map projection that best represents all features of the earth. Over the course of time, many projections have been developed that balance these distortions in different ways to create maps suitable for different purposes. For instance, when designing a map used by sailors navigating through the Arctic regions, a projection may be used that maximizes the accuracy of the direction and distance of objects at the poles of the earth, but sacrifices accuracy of the shape of countries along the equator.

The full details of how to construct a map projection are outside the scope of this book. However, the following sections introduce some common map projections and examine their key features.

Hammer-Aitoff Projection

The Hammer-Aitoff map projection is an equal-area map projection that displays the world on an ellipse. An *equal-area* map projection is one that maintains the relative area of objects; that is, if you were to measure the area of any particular region on the map, it would accurately represent the area of the corresponding real-world region. However, in order to do this, the shapes of features are distorted. This is illustrated in Figure 1-11.

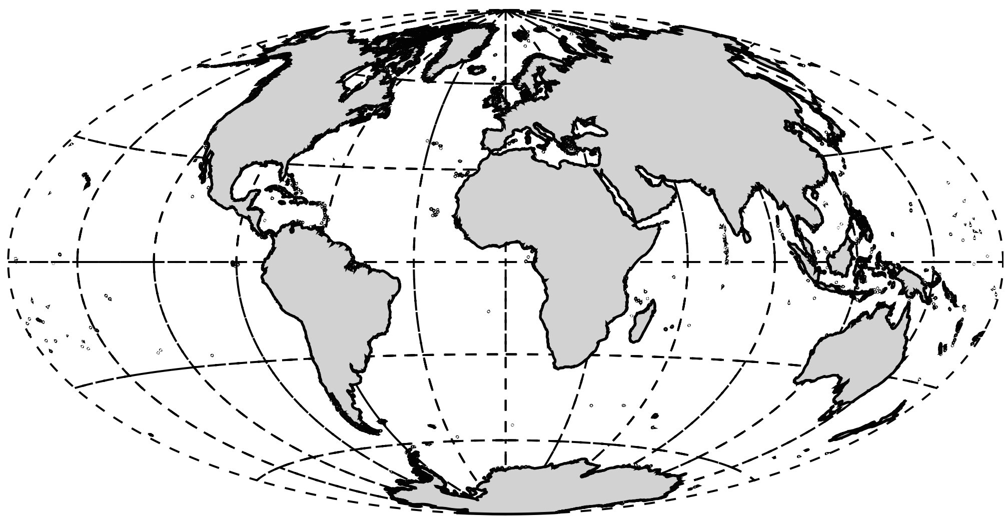

Figure 1-11. *The Hammer-Aitoff map projection*

Mercator Projection

The Mercator map projection is an example of a conformal map projection. A *conformal* map projection is any projection that preserves the local shape of objects on the resulting map.

The Mercator projection was first developed in 1569 by the Flemish cartographer Gerardus Mercator, and has been widely used ever since. It is used particularly in nautical navigation because, when using any map produced using the Mercator projection, the route taken by a ship following a constant bearing will be depicted as a straight line on the map.

The Mercator projection accurately portrays all points that lie on the equator. However, as you move further away from the equator, the distortion of features, particularly the representation of their area, becomes increasingly severe. One criticism of using this projection is that, due to the geographical distribution of countries in the world, many developed countries are depicted with far greater area than equivalent sized developing countries. For instance, examine Figure 1-12 to see how the relative sizes of North America (actual area 19 million sq km) and Africa (actual area 30 million sq km) are depicted at approximately the same size.

Despite this criticism, the Mercator projection is still commonly used by many applications, including the Google Maps web site (`http://maps.google.com/`).

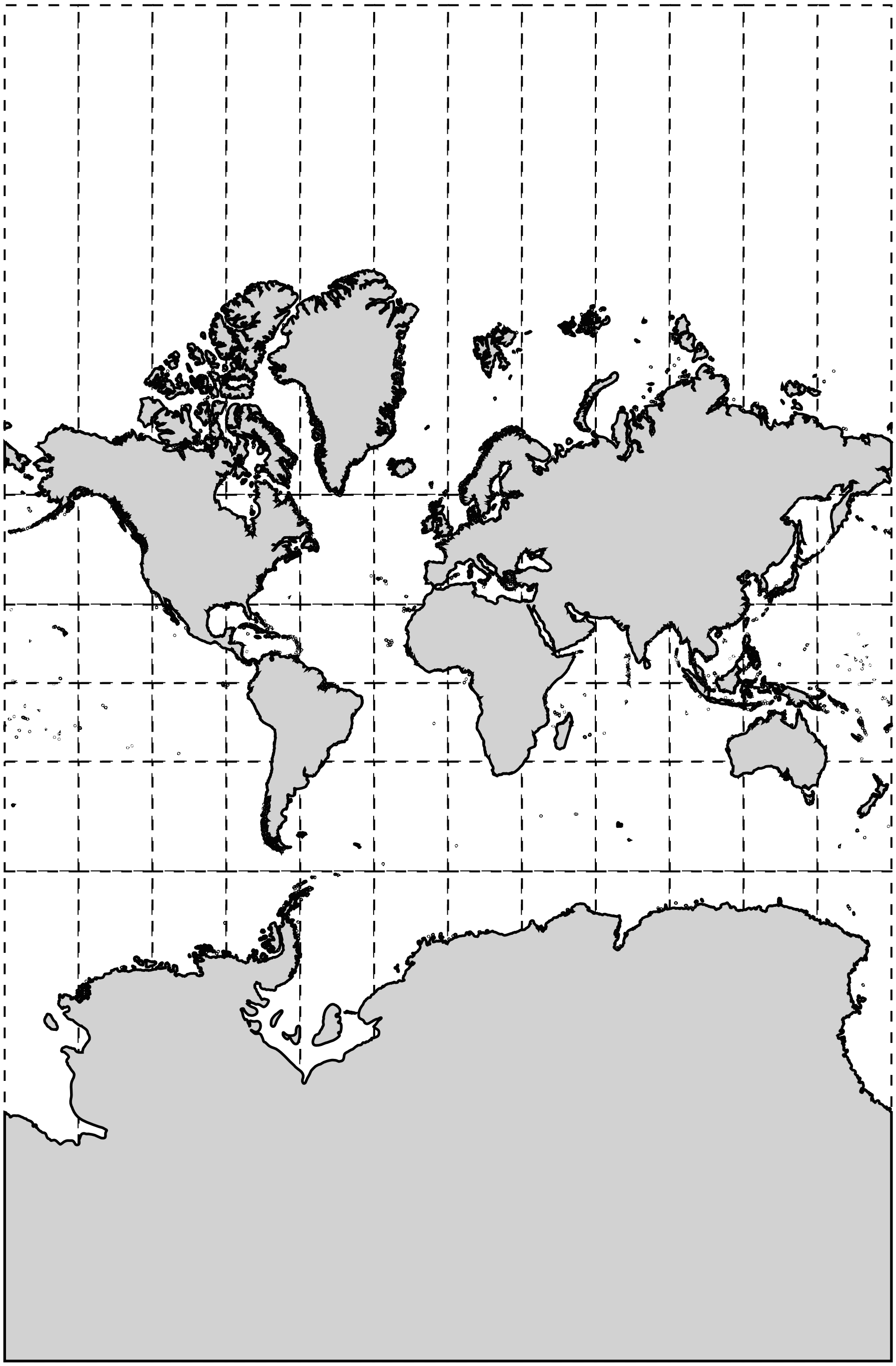

Figure 1-12. *The Mercator map projection*

Equirectangular Projection

The equirectangular projection is one of the first map projections ever to be invented, being credited to Marinus of Tyre in about 100 AD. It is also one of the simplest map projections, because the map projects equally spaced degrees of longitude on the x axis, and equally spaced degrees of latitude on the y axis.

This projection is of limited use in spatial data analysis since it represents neither the accurate shape nor area of features on the map, although it is still widely recognized and used for such purposes as portraying NASA satellite imagery of the world (`http://visibleearth.nasa.gov/`). Figure 1-13 illustrates a map of the world created using the equirectangular projection method.

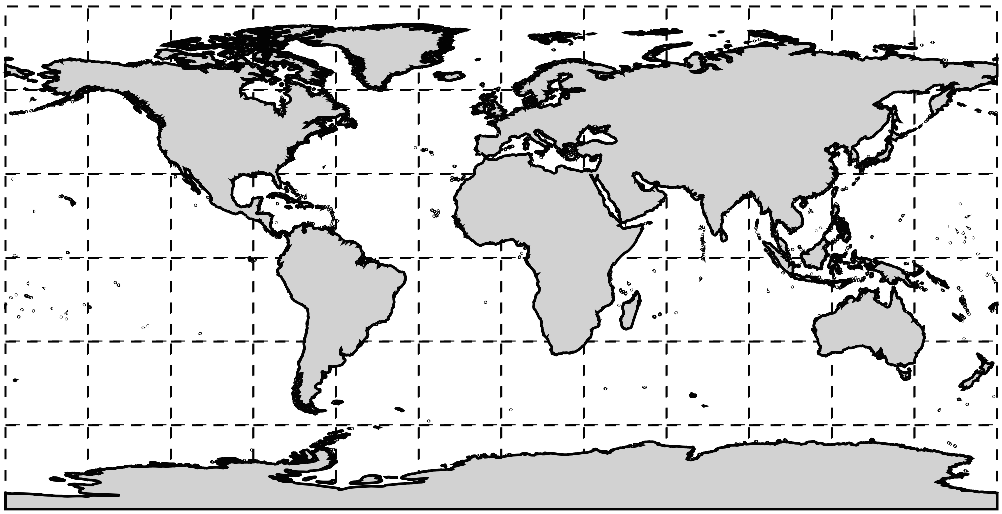

Figure 1-13. *The equirectangular map projection*

Universal Transverse Mercator Projection

The Universal Transverse Mercator (UTM) projection is not a single projection, but rather a grid composed of many projections laid side by side. The UTM grid is created by dividing the globe into 60 slices, called “zones,” with each zone being 6° wide and extending nearly the entire distance between the North Pole and South Pole (the grid does not extend fully to the polar regions, but ranges from a latitude of 80°S to 84°N). Each numbered zone is further subdivided by the equator into north and south zones. Any UTM zone may be referenced using a number from 1 to 60, together with a suffix of *N* or *S* to denote whether it is north or south of the equator. Figure 1-14 illustrates the grid of UTM zones overlaid on a map of the world, highlighting UTM Zone 15N.

Within each UTM zone, features on the earth are projected using a *transverse* Mercator projection. The transverse Mercator projection is produced using the same method as the Mercator projection, but rotated by 90°. This means that, instead of portraying features that lie along the equator with no distortion, the transverse Mercator projection represents features that lie along a central north-south meridian with no distortion. Since each UTM zone is relatively narrow, any feature on the earth lies quite close to the central meridian of the UTM zone in which it is contained, and distortion within each zone is very small.

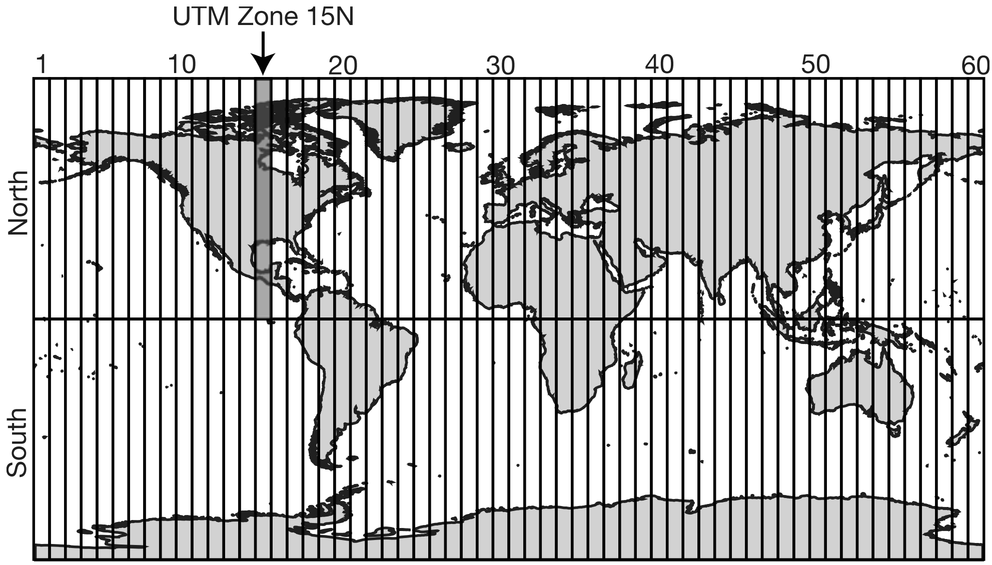

Figure 1-14. *UTM zones of the world*

The UTM projection is *universal* insofar as it defines a system that can be applied consistently across the entire globe. However, since each zone within the UTM grid is based on its own unique projection, the UTM map projection can only be used to accurately represent features that lie within a single specified zone.

Projection Parameters

In addition to the method of projection used, there are a number of additional parameters that affect the appearance of any projected map. These parameters are listed in Table 1-2.

Table 1-2. *Map Projection Parameters*

Parameter	Description
Azimuth	The angle at which the center line of the projection lies, relative to north
Central meridian	The line of longitude used as the origin from which x coordinates are measured
False easting	A value added to x coordinates so that stated coordinate values remain positive over the extent of the map
False northing	A value added to y coordinates so that stated coordinate values remain positive over the extent of the map
Latitude of center	The latitude of the point at the center of the map projection
Latitude of origin	The latitude used as the origin from which y coordinates are measured
Latitude of point	The latitude of a specific point on which the map projection is based
Longitude of center	The longitude of the point at the center of the map projection

Table 1-2. *Map Projection Parameters (Continued)*

Parameter	Description
Longitude of point	The longitude of a specific point on which the map projection is based
Scale factor	A scaling factor used to reduce the effect of distortion in a map projection
Standard parallel	A line of latitude along which features on the map have no distortion

Using Spatial Reference Systems

We have examined several components that make up any spatial reference system—a system that allows us to define positions on the earth's surface, which we can use to construct geometries representing features on the earth. Table 1-3 gives an overview of each component.

Table 1-3. *Components of a Spatial Reference System*

Component	Function
Coordinate system	Specifies a mathematical framework for determining the position of items relative to an origin.
Datum	States a model of the earth onto which we can apply the coordinate system. Consists of a reference ellipsoid (a three-dimensional mathematical shape that approximates the shape of the earth) and a reference frame (a set of points that enables us to position the reference ellipsoid to line up with the right points on the earth).
Prime meridian	Defines the axis from which coordinates of longitude are measured.
Projection[a]	Details the parameters required to create a two-dimensional image of the earth's surface (i.e., a map), so that positions can be defined using projected coordinates.
Unit of measurement	Provides the appropriate unit in which coordinate values are expressed.

[a] *Projection parameters are only defined for spatial reference systems based on projected coordinate systems.*

Through a combination of all these elements, you can use a spatial reference system to uniquely identify any point on the earth.

■ **Note** In order to be able to describe positions on the earth using a projected coordinate system, a spatial reference system must first specify a three-dimensional, geodetic model of the world (as would be used by a geographic coordinate system), and then *additionally* state the parameters detailing how the two-dimensional projected map image should be created from that model. For this reason, spatial reference systems based on projected coordinate systems must contain all the same elements as those based on geographic coordinate systems, together with the additional parameters required for the projection.

Spatial Reference Identifiers

Every time we state the latitude and longitude, or x and y coordinates, that describe the position of a point in a geometry, we must also state the associated spatial reference system in which those coordinates were obtained. Without the extra information contained in the spatial reference system, a coordinate tuple is just an abstract set of numbers in a mathematical system. The spatial reference takes the abstract coordinates from a geographic or projected system and puts them in a context so that they can be used to identify a real position on the earth's surface.

However, it would be quite cumbersome to have to write out the full details of the datum, the prime meridian, and the unit of measurement (and any applicable projection) each time we wrote down a set of coordinates. Fortunately, various authorities allocate easily memorable, unique integer reference numbers that represent all of the necessary parameters of a spatial reference system. These reference numbers are called spatial reference identifiers (SRIDs).

One authority that allocates SRIDs is the European Petroleum Survey Group (EPSG), and its reference identification system is implemented in SQL Server 2008. Whenever you use any of the spatial functions in SQL Server that involve stating the coordinates of a position, you must always supply the relevant EPSG SRID as a parameter.

■ **Tip** You can view the details of all spatial reference systems administered by the EPSG registry at the following web site: `http://www.epsg-registry.org`.

Spatial References in SQL Server 2008

SQL Server 2008 stores the details of all supported geodetic spatial reference systems in a special system table called `sys.spatial_reference_systems`. Every row in this table corresponds to a unique spatial reference system that you can use to define spatial data in SQL Server 2008.

In order to see the list of supported geodetic spatial reference systems, execute the following code in a SQL Server Management Studio query window:

```
SELECT
  *
FROM
  sys.spatial_reference_systems
```

Table 1-4 lists and describes each column of the `sys.spatial_reference_systems` table shown in the results.

Table 1-4. *Columns of the sys.spatial_reference_systems Table*

Column Name	Description
`spatial_reference_id`	The integer identifier used within SQL Server 2008 to refer to this system
`authority_name`	The name of the authority that defines this reference
`authorized_spatial_reference_id`	The identifier allocated by the authority to refer to this system
`well_known_text`	The parameters of the spatial reference system, expressed in well-known text format
`unit_of_measure`	A text description of the unit used to express linear measurements in this system, such as distance and length
`unit_conversion_factor`	A scale factor for converting from meters into the unit of measurement

The `sys.spatial_reference_systems` table only includes those spatial reference systems based on geographic coordinates supported by SQL Server. In addition to the spatial reference systems listed in this table, you can also define data using *any* projected spatial reference system, as you will see in the next chapter.

■ **Note** Currently, the only authority used to define spatial references in SQL Server 2008 is the EPSG, and all internal SRIDs are based on the EPSG numbering system. As a result, the value of the internal `spatial_reference_id` for any system is the same as the `authorized_spatial_reference_id`.

Expressing Spatial References in the Well-Known Text Format

Within the `sys.spatial_reference_systems` table, SQL Server stores the relevant details of each spatial reference using the Well-Known Text (WKT) format, which is an industry-standard format for expressing spatial information defined by the OGC. The WKT description of the spatial reference is stored as a text string in the `well_known_text` column.

To illustrate how spatial references are represented in WKT format, let's examine the properties of the EPSG:4326 spatial reference, by running the following query:

```
SELECT
  well_known_text
FROM
  sys.spatial_reference_systems
WHERE
  authority_name = 'EPSG'
  AND
  authorized_spatial_reference_id = 4326
```

The following is the result (with line breaks and indents added to make the result easier to read):

```
GEOGCS[
  "WGS 84",
  DATUM[
    "World Geodetic System 1984",
    ELLIPSOID[
      "WGS 84",
      6378137,
      298.257223563
    ]
  ],
  PRIMEM["Greenwich", 0],
  UNIT["Degree", 0.0174532925199433]
]
```

Let's examine this result, to identify each of the component elements of a spatial reference system:

Coordinate system: The first line of a WKT spatial reference is a keyword to tell us what sort of coordinate system is used. In this case, `GEOGCS` tells us that EPSG:4326 uses a geographic coordinate reference system. If a spatial reference system is based on projected coordinates, then the WKT representation would instead begin with `PROJCS`. Immediately following the declaration of the type of coordinate system is the name of this spatial reference. In this case, we are describing the `"WGS 84"` spatial reference.

Datum: The values following the `DATUM` keyword provide the parameters of the datum. The first parameter gives us the name of the datum used. In this case, it is the `"World Geodetic System 1984"` datum. Then follow the parameters of the reference ellipsoid. In this spatial reference, we are using the `"WGS 84"` ellipsoid, with a semimajor axis of 6,378,137 m and an inverse-flattening ratio of 298.257223563.

Prime meridian: The `PRIMEM` value tells us that this system defines Greenwich as the prime meridian, where longitude is 0.

Unit of measurement: The spatial reference specifies that the units of angular measurement are expressed as `"Degree"`. The value of `0.0174532925199433` is a conversion factor required to convert to the appropriate units. This represents the value of $\pi/180$, required to convert angular measurements from radians into degrees.

Contrasting a Geographic and a Projected Spatial Reference

Let's compare the result in the preceding section to the WKT representation of a spatial reference system based on a projected coordinate system. The following example shows the WKT representation of the UTM Zone 10N reference, a projected spatial reference system used in North America. The SRID for this system is EPSG:26910.

```
PROJCS[
  "NAD_1983_UTM_Zone_10N",
  GEOGCS[
    "GCS_North_American_1983",
    DATUM[
      "D_North_American_1983",
      SPHEROID[
        "GRS_1980",
        6378137,
        298.257222101
      ]
    ],
    PRIMEM["Greenwich",0],
    UNIT["Degree", 0.0174532925199433]
  ],
  PROJECTION["Transverse_Mercator"],
  PARAMETER["False_Easting", 500000.0],
  PARAMETER["False_Northing", 0.0],
  PARAMETER["Central_Meridian", -123.0],
  PARAMETER["Scale_Factor", 0.9996],
  PARAMETER["Latitude_of_Origin", 0.0],
  UNIT["Meter", 1.0]
]
```

Notice that the spatial reference for a projected coordinate system contains a complete set of parameters for a geographic coordinate system, embedded within brackets following the `GEOGCS` keyword. The reason is that a projected system must first define the three-dimensional, geodetic model of the earth, and then specify several additional parameters that are required to project that model onto a plane.

Comparing Spatial Reference Systems

By now, you should have a good appreciation of the fact that a given point on the earth may be represented using many different sets of coordinates, each one corresponding to a particular spatial reference system. (Conversely, the same set of coordinate values can refer to different places on the earth depending on the spatial reference system from which the coordinates were obtained.) Whenever we define an item of spatial data in SQL Server to represent an object on the earth, three bits of information are required:

- The type of geometry used to represent the object (e.g., Point, LineString, Polygon)
- The coordinates of each of the points that define that geometry, expressed in decimal degree notation (e.g., 37.215, –57.5)
- The unique identifier of the spatial reference system from which those coordinates were obtained (e.g., 4326)

To demonstrate this in practical terms, let's compare how a particular feature on the earth—Loch Ness, in Scotland—can be expressed in different spatial reference systems. For this example, we will use a geographic coordinate system based on the WGS 84 datum, and a projected coordinate system based on the National Grid of Great Britain.

WGS 84

WGS 84 is the most commonly used geodetic spatial reference system and is used by GPS systems. It is based on the WGS 84 ellipsoid, which gives a reasonable approximation over the whole surface of the geoid. It is referenced by the EPSG reference 4326. In the WGS 84 system, Loch Ness can be represented by a Point object positioned at the following coordinates:

- Geometry: Point
- Latitude/longitude coordinates: (57.3, –4.5)
- SRID: 4326

National Grid of Great Britain

Many countries have defined their own grid systems for referencing coordinates of positions that lie exclusively within that country. Great Britain, Ireland, New Zealand, Malaysia, Singapore, the Netherlands, and Sweden all have defined national grid systems that are commonly used to express coordinates for local positioning within those countries.

The National Grid of Great Britain uses a projected coordinate system based on the transverse Mercator projection. The transverse Mercator projection is similar to the Mercator projection, but instead of accurately portraying features lying on the equator of the earth, it has been rotated so that the map accurately portrays features lying along a given meridian—a line running north-south between the poles of the earth. This makes the transverse Mercator projection more suitable for mapping tall, thin countries such as Great Britain. The datum is the Ordnance Survey of Great Britain 1936 (OSGB 36), using the Airy 1830 ellipsoid, which is the ellipsoid that provides the best fit for the geoid over this region. The "true" origin on which the projection is based has a latitude of 49° north and longitude 2° west. However, coordinates in the grid system are actually stated from a "false" origin situated 400 km west and 100 km north of the true origin. Using the false origin as the point from which coordinates are measured ensures that coordinate values for any point in Great Britain are always positive. The SRID of this system is EPSG:27700.

The grid system is constructed by overlaying a series of 100 km by 100 km squares, starting from the false origin, that cover the land surface of Great Britain. Each of these squares is given a two-letter identifier. To refer to any point in the system, you state the identifier of the square that the point lies in, together with the easting and northing coordinates of the point measured in meters from the bottom-left corner of the square. Alternatively, easting and northing coordinates can be expressed in absolute meters east and north from the false origin. As such, Loch Ness could be represented in this system as follows:

- Geometry: LineString
- Easting/northing coordinates: (238172, 808732), (261620, 839938)
- SRID: 27700

The comparison between the coordinate values obtained from the National Grid of Great Britain and WGS 84 is illustrated in Figure 1-15.

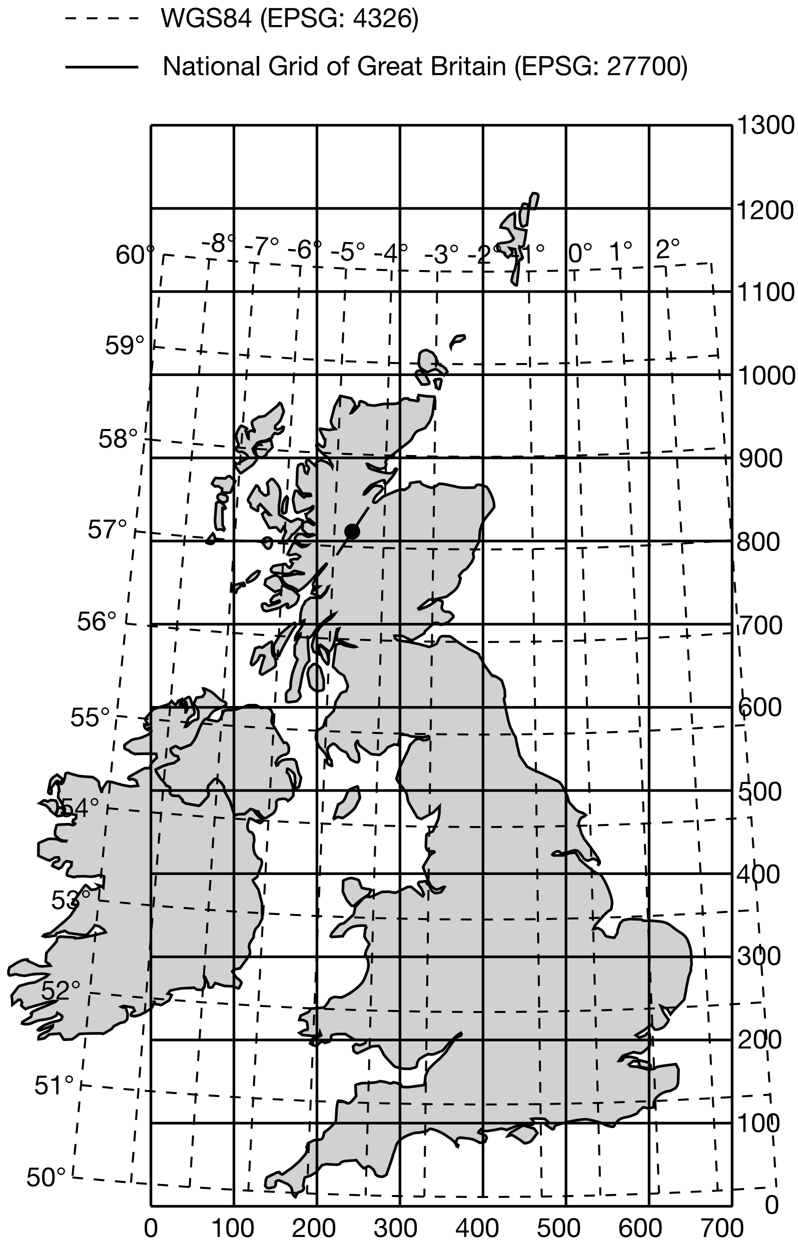

Figure 1-15. *Comparing representations of Loch Ness in two different spatial reference systems*

Note these are only two examples—this same feature could be described by many other sets of coordinates in different spatial reference systems.

Summary

After reading this chapter, you should understand how spatial data can be used to describe the properties of features on the earth:

- Spatial data creates representations of features on the earth by defining regular shapes that approximate the shape and position of those features. The three basic types of shape that can be used in SQL Server 2008 are Points, LineStrings, and Polygons. When used in geospatial data, these shapes are called geometries.
- Each of these geometries can be defined by specifying the coordinate values of a series of points that make up the overall geometry.
- To define the coordinates of the position of a point on the earth, we use a spatial reference system.
- A spatial reference system consists of a coordinate system (which describes a position using either projected or geographic coordinates), a datum (which describes a model representing the shape of the earth), the prime meridian (which defines the origin from which units are measured), and the unit of measurement. When using projected coordinates to describe a point, the spatial reference system also defines the properties of the projection used.
- A geographic coordinate system defines the position of objects using angular coordinates called latitude and longitude, which are measured from the equator and the prime meridian, respectively.
- A projected coordinate system defines the position of objects using Cartesian coordinates, which measure the x and y distance of a point from an origin. These are also referred to as easting and northing coordinates.
- Whenever you state a set of coordinates representing a point, it is essential that you also give details of the associated spatial reference system. The spatial reference system defines the additional information that allows us to apply the coordinate reference to identify a point on the earth.
- For convenience, spatial reference systems may be specified by a single integer identifier—known as a spatial reference identifier (SRID).
- Details of all the geodetic spatial reference systems supported by SQL Server 2008 are contained within a system table called `sys.spatial_reference_systems`.