Needle in a haystack:
How to find one galaxy among half a billion others

Niharika Sravan and GionMatthias Schelbert

Purpose

As technology becomes ever more tightly integrated in our lives, our ability to harness the power of computation determines our potential to make breakthroughs in scientific knowledge. Advances in petascale computing will herald the next big boom in discoveries, in practically every area of science, engineering, and technology. It will be guided by exploiting vast quantities of data, also known as 'data driven discovery'. In astronomy, the advent of rapid-cadence big-data telescopes marks the beginning of Time-Domain Astronomy, the study of the time-variable Universe.

The Sloan Digital Sky Survey or SDSS contains photometric observations of around 500 million objects and spectra for more than 3 million objects. The SDSS database, with a total data volume of about 116 TB, is publically available over the Internet. Its design and management is a model for upcoming big-data telescopes hoping to make its data publically available in real-time.

SkyServer is the data access site for the Sloan Digital Sky Survey. It provides a range of interfaces to an underlying Microsoft SQL Server, which is a relational database management system developed by Microsoft. SkyServer is an incredibly powerful tool for conducting astronomy education and research. The techniques described in this lesson are many of the same tools professional astronomers use to conduct cutting edge research.

Overview

In this lesson students will learn basics of SQL and use it to find objects of their choice from the SDSS database. They will examine one or more of these objects in detail using images and other scientific data.

Student Outcomes

After completing this lesson, students will be able to

- utilize SQL to construct a query for extracting objects satisfying criteria specified by them from the SDSS database,
- evaluate scientific and computational search parameters needed to accomplish this, and
- interpret results returned by the SDSS database, select one for further examination and analyze it.

Reach for the Stars is a GK-12 program supported by the National Science Foundation under grant DGE-0948017. However, any opinions, findings, conclusions, and/or recommendations are those of the investigators and do not necessarily reflect the views of the Foundation.

Standards Addressed

Performance expectation: DCI code HS-ESS1-2. - *Construct an explanation of the Big Bang theory based on astronomical evidence of light spectra, motion of distant galaxies, and composition of matter in the universe.*
Disciplinary Core Ideas: The Big Bang theory is supported by observations of distant galaxies receding from our own.

Time

1 class period or approx. 1 hour

Level

11th - 12th grade

Materials and Tools

- ✓ Students will need ready access (should not need to type out web addresses) to:
  - ➢ **SQL Search Form**: http://skyserver.sdss.org/dr12/en/tools/search/sql.aspx
  - ➢ **specObj table**: http://skyserver.sdss.org/dr12/en/help/browser/browser.aspx#&&history=description+SpecObj+V
  - ➢ **Object Explorer**: http://skyserver.sdss.org/dr12/en/tools/explore/Summary.aspx?
- ✓ A printed copy of the SQL tutorial worksheet (appended at the end of this lesson outline)

Preparation

Ensure that students have ready access to the websites mentioned in the earlier section and a printed copy of the SQL tutorial worksheet (4 pages). Ensure that they know what each webpage is called (in boldface above) as they will be referred to by these names in the tutorial.

Prerequisites

None

Background

To complete this lesson you should know how astronomers locate objects in the sky and about one of the ways they specify how far away they are. The former is much easier to do: it is analogous to pointing to star in the night sky to your friend. However, much like how you cannot tell how far away an object is by just looking at it without knowing how intrinsically big or bright it is, astronomers rely on our understanding of how the Universe works to infer how far away objects in the Universe are without ever leaving the vicinity of earth. A widely used convention for the former is based on **the equatorial coordinate system.**

The equatorial coordinate system is a widely used celestial coordinate system that is used to specify the positions of celestial objects. It is implemented in spherical coordinates. The origin of this coordinate system is at the center of the Earth. To ensure that the coordinate system, while aligned with the Earth's equator and pole, does not rotate with the Earth, but remains relatively fixed against the background stars, a fundamental plane and a primary direction are defined. The fundamental plane consists of the projection of the Earth's equator onto the celestial sphere (forming the celestial equator), while the primary direction points towards the vernal equinox. The system uses the right-handed convention, i.e. coordinates are positive toward the north and toward the east in the fundamental plane.

The location of any object in the sky is specified as a pair of coordinates known as **right ascension and declination**. The **declination** (symbol δ, abbreviated *dec*) measures the angular distance of an object perpendicular to the celestial equator, positive to the north, negative to the south. For example, the north celestial pole has a declination of +90°. The origin for declination is the celestial equator, which is the projection of the Earth's equator onto the celestial sphere. Declination is analogous to terrestrial latitude. The **right ascension** (symbol α, abbreviated *RA*) measures the angular distance of an object eastward along the celestial equator from the vernal equinox to the hour circle passing through the object. The vernal equinox point is one of the two where the ecliptic intersects the celestial equator. Analogous to terrestrial longitude, right ascension is usually measured in sidereal hours, minutes and seconds instead of degrees, a result of the method of measuring right ascensions by timing the passage of objects across the meridian as the Earth rotates. There are (360° / 24h) = 15° in one hour of right ascension, 24h of right ascension around the entire celestial equator. When used together, right ascension and declination are usually abbreviated **RA** and **Dec**, respectively.
The gif here is helpful in visualizing this coordinate system.

To infer the distance of objects is more complicated. There are various ways astronomers can do this, but the most useful metric for the most distant objects in the Universe is the **cosmological redshift**. Light from distant galaxies is not featureless, but has distinct spectral features characteristic of the atoms in the gases around them. Because these objects are moving away from us due to the expansion of the Universe[1], these features are Doppler shifted toward the red end of the electromagnetic spectrum. The higher the redshift (denoted *z*), the faster the objects are moving away from us, the further they are.

Teaching Notes

Teachers should practice working out exercises in the handout before teaching it. This will allow them to anticipate the problems students might have, saving class time that will otherwise be spent troubleshooting.

Assessment

Included in the SQL tutorial worksheet (appended at the end of the lesson outline).
The correct query for the last SQL exercise is:

```
select
     bestObjID, ra, dec, z
from
     specObj
where
     class='qso'
     AND z > 5.0
```

[1] The expansion of the universe is a result of the original explosion that created the universe, also known as the Big Bang.

Additional Information

The material for this lesson was derived from http://cas.sdss.org/dr5/en/help/howto/search/ and http://cas.sdss.org/dr5/en/help/docs/sql_help.asp. Teachers should consult these webpages for more help or ideas on how to expand this lesson.

The SQL tutorial worksheet starts on the next page.

SQL Tutorial

SQL is a standard language for accessing databases. A database most often contains one or more tables. Each table is identified by a name. Tables contain rows of each with a specific type of data. An SQL query is a request to the database to return specific data. It consists of three pieces, or blocks: the `select` block, the `from` block, and the `where` block.

The `select` block tells the database which type of data you want it to return. Each identifier must be separated with a comma. For example, if you wanted to find the celestial coordinates right ascension (`ra`) and declination (`dec`) of an object, the `select` block will read:

```
select
     ra, dec
```

The `from` block specifies which table (or tables) you want to search. If you want to retrieve information from a table called `galaxies`, the `from` block will read:

```
from
     galaxies
```

The `where` block allows you to search for records matching certain criteria. Your list of criteria must be separated by the word "AND". Suppose you only want galaxies in a patch of the sky with RA between 194 and 195 degrees and declination between 2 and 3 degrees, your `where` block would read:

```
where
     ra BETWEEN 194 AND 195 AND
     dec BETWEEN 2 AND 3
```

Now, look at the following query:

```
select
     z, ra, dec
from
     specObj
where
     specClass = 2 AND
     z > 0.3
```

The `select` and `from` blocks tell the database to look in the SpecObj table (which contains spectroscopic information) and send you the redshift, right ascension, and declination. The `where` block tells the database to look only at galaxies (which have a specClass value of 2) with redshifts above 0.3.

In this query, the word AND appears between the characteristics in the `where` block. Both these characteristics must be met by a given record for the search to return that record.

AND is just one of three logical operators used by SQL; the others are OR and NOT. The meanings of the three logical operators are given in the table below:

Logical Operator	Meaning
AND	all characteristics met
OR	at least one characteristic met
NOT	characteristic not met

You can combine the logical operators using parentheses. For example, for the characteristics A, B, and C, A AND (B OR C) means that either characteristics B or C, as well as characteristic A, must be met for records to match.

Question: What does A AND (NOT B) mean?
Answer:

SQL includes a variety of mathematical operators that perform math functions in a query. You can use multiplication, division, addition and subtraction. SQL uses the same symbols for these operators that most other computer languages use: + for addition, - for subtraction, * for multiplication, and / for division.

You can also use the conditional operators listed below.

Conditional	Meaning
=	equal to
>	greater than
<	less than
>=	greater than or equal to
<=	less than or equal to
<>	not equal to

Now, complete the following query to search for quasars with redshifts greater than 5.0 and retrieve information on their right ascension, declination, redshift, and the (final, or ``best'') object ID.

```
select

        ____________________________________________________

from

      specObj

where

      class = ____________

      AND ________________
```

Hint: Look at the **specObj table** to find the relevant search parameters.

Type out the completed query in the **SQL Search Form** and click Submit. A new tab will open showing the list of returned results. Select a couple of objects with non-zero Obj IDs and write their Obj IDs here:

1. ____

2. ____

Now go to the **Object Explorer** and click **Search**. A **Search by** form will appear at the top of the page. Type out one of the Obj IDs you write down in the **ObjID** field and click **Go**. Examine the result. Do the object properties satisfy the criteria you specified in the `where` block?

____

____

What kinds of astronomy questions do you think you could answer by studying these objects?

____

____

____

____

Now, click on the object image to open it in the **Navigation Tool**. Pan across/zoom into the image. Click on nearby objects to get a quick summary about them on the right. Click Explore on the bottom right side to open it in **Object Explorer**. Examine it.

What type of objects do you generally find? How do their characteristics differ from the ones returned by your SQL search? Are they closer to us or further away?

Based on the objects you find, what can you say about the general distribution of objects in our Universe?