Kent State University

MIS-64082: Database Management and Analytics

Fall 2019

Non-Relational Database Practicum:

Exoplanet NoSQL Database

Author:

Steven Spence
(Student ID: sspenc12)

Table of Contents

The Scenario

The business scenario for this paper will focus on the search for exoplanets in the universe. Specifically, how all of the information related to current and newly discovered exoplanets are stored and updated. Only within the past few decades have humans been able to detect and study planets around stars outside our solar system. The main driver for this has been increases in telescope and imaging technologies. It began with the Hubble space telescope surveying the sky, and now it has led to what will soon be dozens of new telescopes scanning distance stars. Some examples of these telescopes soon to launch into order are TESS, CHEOPS, James Webb, and PLATO [1].

Several space agencies are behind this push for surveying of exoplanets; however, NASA and ESA are the largest contributors. Private industries, such as SpaceX, are making space even more accessible and cheaper for companies, so there is bound to be an even larger increase in satellites in the sky scanning for potential habitable planets and signs of life. The current method for scanning for exoplanets is to view the brightness of a particular star for an extended period of time. If the measured brightness of a star tends to dip on a consistent basis, then that could be the sign of a planet passing in front of the star and blocking some light. These surveys collect brightness data over a long period of time and are called light curves.

Light curves are only one pieces of data and information that is gathered while surveying the stars throughout the universe. Other pieces of information include planet size, mass, host star composition, potential habitable zones, atmospheric conditions of the planet, etc. For a single solar system there may be hundreds of different measurements, notes and observations made. Conservative estimates of the Milky Way galaxy put the number of stars at 250 billion, and the Milky Way is just one of billions of galaxies in the universe. Therefore, it is difficult to comprehend the amount of data that will need to be collected as we further expand our knowledge of the universe. Research companies and space administrations will need to find flexible, scalable system to store and share all this information.

Reason for NoSQL

As stated in the previous section, there is a wide range of data that is collected for stars and exoplanets. This can include calculations such as distance and mass; however, it also includes a lot of notes that take unstructured forms and lengths. The notes can be about the potential conditions of the atmosphere, new studies that will be completed, or general researcher notes. Additionally, this additional information can include full data sets from light curves or other measurements, images from satellites, or general image data from satellites. Another benefit to utilizing a NoSQL database for this application would be the fact that it can be scaled much easier than a SQL database. Our understanding of the universe is constantly changing, so scientists are always finding new and creative ways to collect data. Sometimes this data may not be useful at the time, so it may not be cleaned and properly formatted at the time of collection. However, the data may be re-visited in the future when a new discovery is made. Additionally, a lot of exoplanets are initially marked as potential candidate planets until further studies are completed. Therefore, the system must be able to handle the addition and/or subtraction of data on large scales.

Exoplanets and stars may not have the same type/amount of information stored about them as well. There are some stars and planets that are just too far away to study. These observations may only have distance and mass calculation, since the equipment available at the time cannot collect any more useful information. However, stars that are closer to Earth are more easily studied, so these will have significantly more data collected. Therefore, the database must be able to able large amounts of entries with different amounts of data. This is where NoSQL databases and JSON type documents become more beneficial than a standard SQL database. The flexibility to add new documents and change current documents to add new information will be extremely beneficial to those utilizing the data. This will be especially useful when full datasets can be stored/attached to a document for a specific planet. It will be easily accessible for researchers and companies to utilize as needed and add findings and discoveries to the database.

NoSQL Form

The preferred NoSQL database for this type of application would be a document storage database. For this paper, the CouchDB system will be utilized. This database system allows a team to upload documents with as much or as little information gathered at the time. Additionally, users of this database can add attachments to the document at any time. Such attachments could include images, datasets, research papers, etc. Another benefit of using CouchDB is the fact that is has a user-friendly API that can easily be used by someone whom has little to no database experience, as well as linkage with the command line interface for more experienced users. This is an added benefit because citizen scientists that often interface with the public datasets may not have the technical background as someone regularly interacting with the database.

For the purposes of this paper, the documents in the database will be scaled down for simplicity. The documents will include the following information: planet name, host star, distance to planet, number of other planets in system, orbital period, estimated planet mass, discovery method, additional notes, light curve, and additional attachments. These are high-level items that would be collected and calculated by various teams; however, they will be used to test the functionality of the proposed database system. The following section will show the implementation of this system, as well as screen shots of the examples. Additional examples were constructed and can be included by request but were excluded from the paper to consolidate the information.

Implementation and Examples

As stated in the previous section, the CouchDB system was utilized to create an example document database system for storing exoplanet related information. The example documents will include the following information: planet name, host star, distance to planet, number of other planets in system, orbital period, estimated planet mass, discovery method, additional notes, light curve, and additional attachments. This is only a small subset of the potential information collected by satellites but will be used to prove the functionality of the system. See the following screenshot and image descriptions for more details.

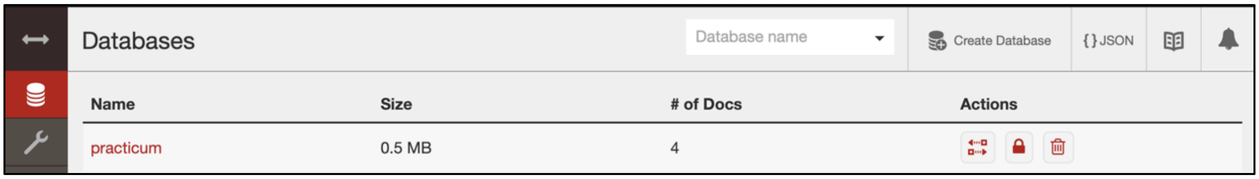

Image 1: Initial Creation of Document Database on CouchDB for the Practicum.

As shown in Image 1, the document database for this project was created in CouchDB and named "practicum" for simplicity. This is where all of the documents would be added to as exoplanets are discovered and added to as more information become available. At the time of the screenshot, there were only four documents; however, this number would climb to thousands and hundreds of thousands in the actual research field.


```
{
  "_id": "c9ab41de1b106b063ec90f8e5600660d",
  "planet_name" : "Proxima Centauri b",
  "host_star" : "Proxima Centauri",
  "distance_to_system" : 1.3,
  "planets_in_system" : 1,
  "orbital period" : 11.19,
  "planet_mass" : 0.004,
  "disovery method" : "Transit",
  "additional notes" : "Planet appears to be in this system's habital zone"
}
```


Image 2: Document for "Proxima Centauri b" that includes basic information about the planet and a small additional note.


```
{
  "_id": "c9ab41de1b106b063ec90f8e56007cc1",
  "_rev": "6-3c7cfd70ca10bad330b0611d52fc3aca",
  "planet_name": "WASP-33 b",
  "host_star": "WASP-33",
  "distance_to_system": 122.36,
  "planets_in_system": 1,
  "orbital period": 1.22,
  "planet_mass": 2.093,
  "disovery method": "Transit",
  "additional notes": "Planet atmosphere conditions to be studied in 2020.",
  "_attachments": {
    "WASP-33b.jpg": {
      "content_type": "image/jpeg",
      "revpos": 3,
      "digest": "md5-6P7PIHEj8FL6Gtdd81Eb7g==",
      "length": 193181,
      "stub": true
    }
  }
}
```


Image 3: Document for "WASP-33 b" with similar information as before, but also include an attachment for the observed light curve.

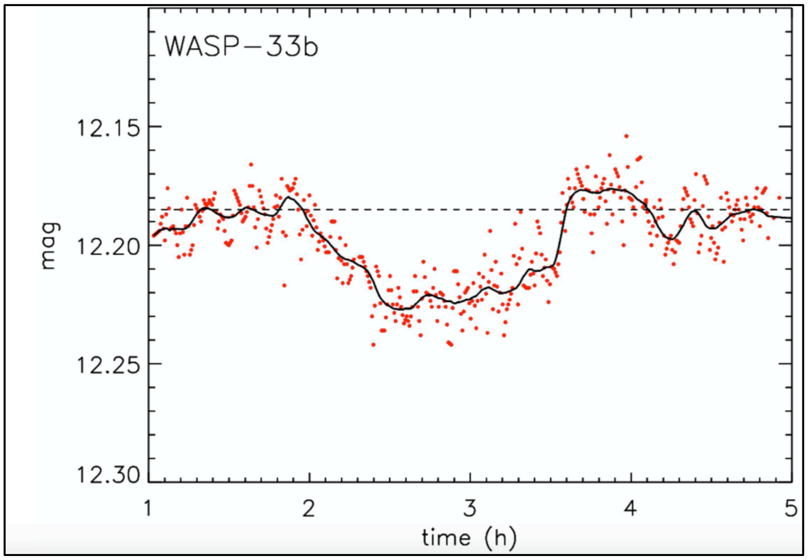

Image 4: View of the light curve for "WASP-33 b" upon opening the attachment [2].

Images 2, 3 and 4 show two different examples of documents in the database. The first entry for Proxima Centauri b only has the basic information about the planet and solar system along with a small note about being the habitable zone of the star. However, the second example for WASP-33 b has the same information with the addition of a light curve created from satellite data collection. In addition to the light curve chart, there would also be all of the dataset this chart was created from. This is what companies and researchers would utilize when trying to confirm a star as having exoplanets orbiting it. The next example will show a similar case, but with the addition of an image.


```
{
  "_id": "c9ab41de1b106b063ec90f8e5600c18a",
  "_rev": "3-c29cab89b30ec3424510cfd35340b24c",
  "planet_name": "Jupiter",
  "host_star": "Sol",
  "distance_to_system": 0.00003,
  "planets_in_system": 8,
  "orbital period": 4380,
  "planet_mass": 1,
  "disovery method": "Telescope",
  "additional notes": "Largest planet in system. Atmosphere contains mostly hydrogen.",
  "_attachments": {
    "Jupiter.jpg": {
      "content_type": "image/jpeg",
      "revpos": 3,
      "digest": "md5-9qPT3wIv1MFxyr1ImTVfXQ==",
      "length": 5117,
      "stub": true
    }
  }
}
```

Image 5: Example document for Jupiter and the data collected about it along with an image attachment.

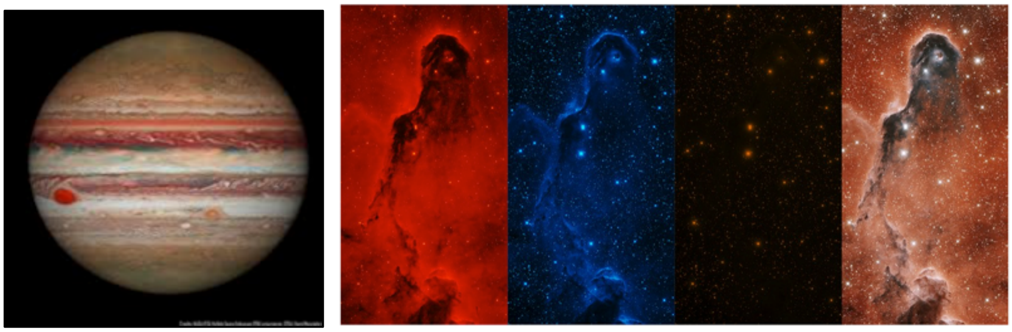

Image 6: (Left) Satellite image of Jupiter included in the document for further review [3]. (Right) Example image of galaxy photos taken in different wavelengths: infrared, ultraviolet, visible, and x-ray [4].

This example shows another document with basic information on the planet; however, it also includes an image of the planet. Technically, Jupiter is not an exoplanet, but this is to simulate the addition of exoplanet images once technology is capable of collecting these types of images. However, the satellite image dataset can also be added into the document, such as the image on the right that shows the collected image data in several different wavelengths. This will then be merged together by photographers and scientists to get the spectacular space images you typically see on the internet.

```
{
  "_id": "c9ab41de1b106b063ec90f8e5600d98c",
  "_rev": "1-bcd51d1c8192acfbe9f4f724e5beec6d",
  "planet_name": "16 Cyg B b",
  "host_star": "16 Cyg B",
  "distance_to_system": 21.5,
  "planets_in_system": 1,
  "orbital period": 798.5,
  "planet_mass": 1.78,
  "disovery method": "Radial Velocity",
  "additional notes": "Potential planet in habital zone. James Webb telescope to further analyze in 2024.",
  "secondary_study_start" : 12-2-2024
}
```

Image 7: Example of document with additional information on future studies for a planet.

The example in Image 7 shows the format for a document with slightly different information regarding a potential exoplanet. This is a current exoplanet candidate, so it has an additional timeline set for observing the light curves in 2024. This example would also be useful to have the light curves attached so future studies can compare the values.

Table | Metadata | {} JSON | Create Document

id	key	value
c9ab41de1b106b063ec90f8e56...	c9ab41de1b106b063ec90f8e56...	{ "rev": "3-f4b2564f7afc9a87741...
c9ab41de1b106b063ec90f8e56...	c9ab41de1b106b063ec90f8e56...	{ "rev": "7-a29fac113e716988f2...
c9ab41de1b106b063ec90f8e56...	c9ab41de1b106b063ec90f8e56...	{ "rev": "5-811cb6e4722c08c3f6...
c9ab41de1b106b063ec90f8e56...	c9ab41de1b106b063ec90f8e56...	{ "rev": "2-df732fd0e9b8d750c8...

Image 8: Overview of the database metadata in CouchDB's API.

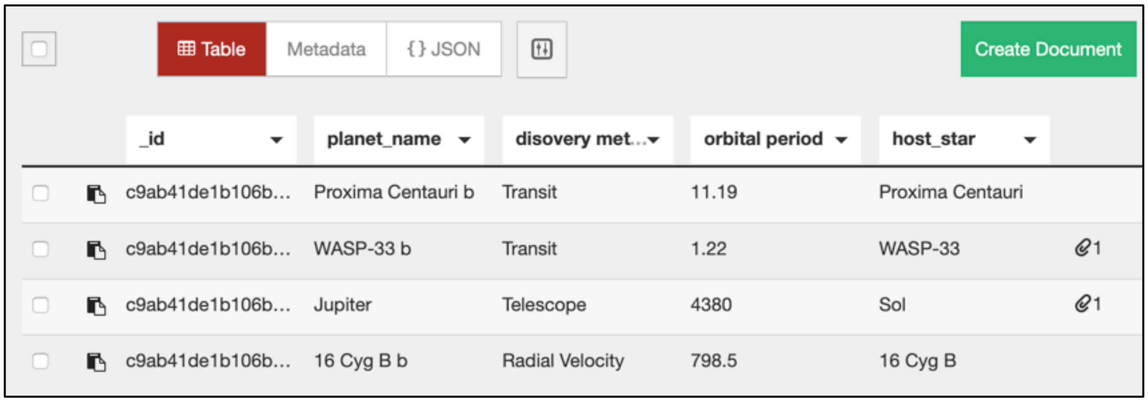

Image 9: Overview of the database table format in CouchDB's API.

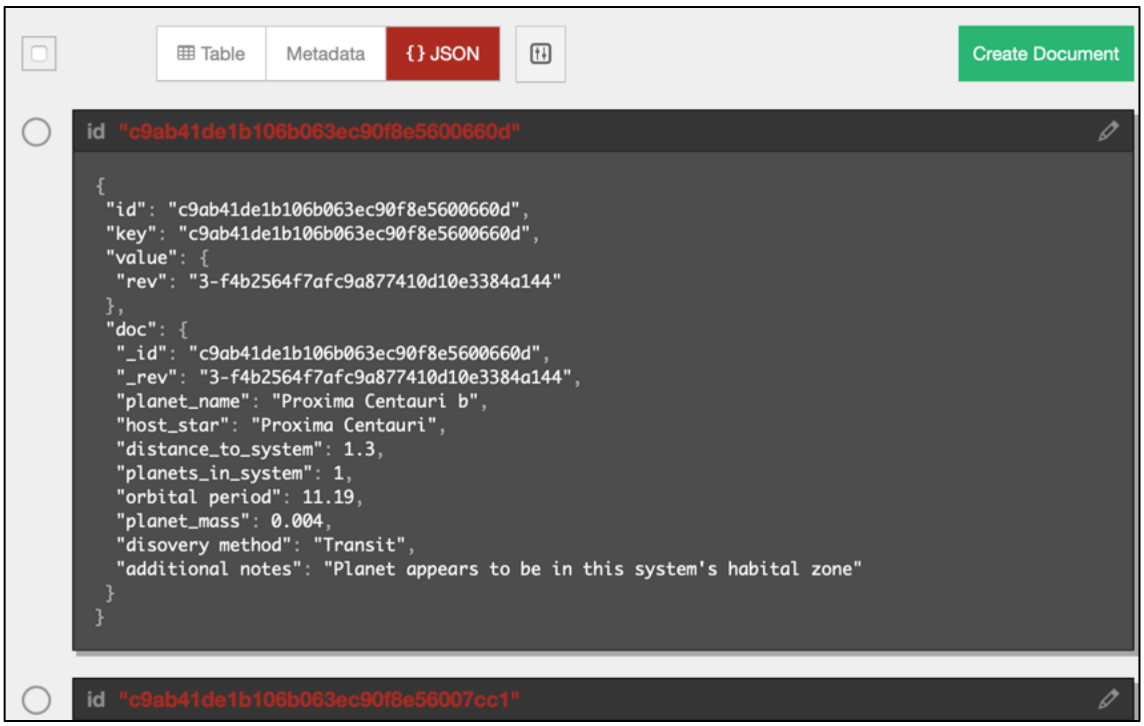

Image 10: Overview of the database in JSON document format in CouchDB's API.

Images 8, 9 and 10 above show the different views that are available in the CouchDB API. It also shows a high-level overview of the database and what it would look like as the number of documented exoplanets are entered into the system. As previously stated, this databased would expand very quickly as more and more data is collected. Additionally, it will have to be very dynamic to accommodate the large number of different data types that will be entered into the system.

References

[1] “Exoplanets”. European Space Agency. ESA. 1 September 2019.
< https://sci.esa.int/web/exoplanets/-/60657-the-future-of-exoplanet-research>

[2] “Studying Exoplanets from Transit Light Curves”. Durham University, Department of Physics. Astrolab. <https://community.dur.ac.uk/physics.astrolab/exoplanets.html>

[3] Masztalerz, Karol. “Hubble’s Jupiter and the Shrinking Great Red Spot”. NASA. 25 April 2018. <https://apod.nasa.gov/apod/ap180425.html>

[4] Mallonee, Laura. “The Secret Behind All Those Gorgeous Photos of Space”. Wired. 24 November 2015. < https://www.wired.com/2015/11/travis-rector-coloring-the-universe/>