Researchers study new ways to handle big data

The amount received from space missions is astronomical

By Adam Poulisse
adam.poulisse@langnews.com
@AdamPoulisse on Twitter

LA CAÑADA FLINTRIDGE » The amount of data that NASA and the Jet Propulsion Laboratory receive from space missions is astronomical: A NASA mission produces hundreds of terabytes every hour.

To put that into context: If researchers transcribed just one terabyte of information to paper, it would kill 50,000 trees.

"A lot of massive data is used for big things," said Chris Mattmann, senior computer scientist at JPL. "(Researchers) will take the information from an earth science satellite and use it to revive scientific assumption. On the astronomy side, it can be used to target potential planets."

These data streams — dubbed "big data" — are getting bigger, faster and more complex thanks to technological advancements. To make it easier for researchers to sift through, mission planners and software engineers at JPL are collaborating on ways to utilize existing data storage technology to better process, archive and access big data.

Mattmann is the principal investigator of JPL's big-data initiative, managing a team of 24 data scientists to develop intelligent approaches to storing big data by studying current data storage "triages," Mattmann said.

His team is sidestepping the traditional relational approaches to storing data and instead focusing on nonrelational data storage, which flattens data and makes it easier find links among related strands of information. Mattmann's case was presented in Nature magazine in January.

"The advantage is, the flattened model is easier to program," Mattmann said. "When you have this data construct, you don't have to worry about how it's stored as much. They tend to scale data, and the technology is a lot easier to understand."

Rather than inventing new hardware, Mattmann says current algorithms, open-source or open-code software (software that is free to build upon) can all be re-tooled to better disseminate big data.

"At the very least, we will have ways of sifting and sorting through the information, even if it exhausts us," Mattmann said. "It doesn't solve the problem 10 years from now, but it makes small, incremental steps."

In 2016, construction will begin on the Square Kilometer Array, a collection of thousands of telescopes in South Africa and Australia. It's expected that 700 terabytes of data will rush into the array every day, according to a news release. JPL officials will be involved with archiving the array's torrents of images.

"The size of the current Internet — it will eclipse that in a couple of days," said Mattmann, who is serving on the project's international governing board.

With big data, trillions of pieces of information is filtered. Websites like Facebook, Twitter, Google, Instagram, Netflix and Pandora are prime examples

DATA » PAGE 5

Data

FROM PAGE 3

of how big data is used every day, according to Steve Groom, manager of NASA's Infrared Science Archive, one of the groups within the Infrared Processing and Analysis Center at Caltech. Collecting big data is what allows advertisers to gauge what you post and where you tag yourself on social media sites, which in turn generates advertisements based on your apparent interests and lifestyle.

"They're looking for patterns and trends from analyzing lots and lots of data," Groom said. "You look at it as the case of what kind of questions people are going to ask." "You take this information you have with astronomy and organize it so if there is a special query, you know how to find an answer."

The big data that registers where you update your status or what musicians you like on Facebook can be employed by researchers who need to find a specific strand of science information out of the hundreds of terabytes of data that is beamed back from space missions.

"You look at it as the case of what people are going to ask, or what programmers are going to ask," Groom said. "We have to anticipate the common types of queries to quickly respond to those kind of queries."

One other way to better sift through the gobs of big data is to turn it into photos or short movies, which are created out of the data that is transmitted in binary code.

"For NASA, the biggest challenge is in returning this data from space," said Eric De Jong, principal investigator for NASA's Solar System Visualization project, which converts NASA mission science into visualization products that researchers can use.

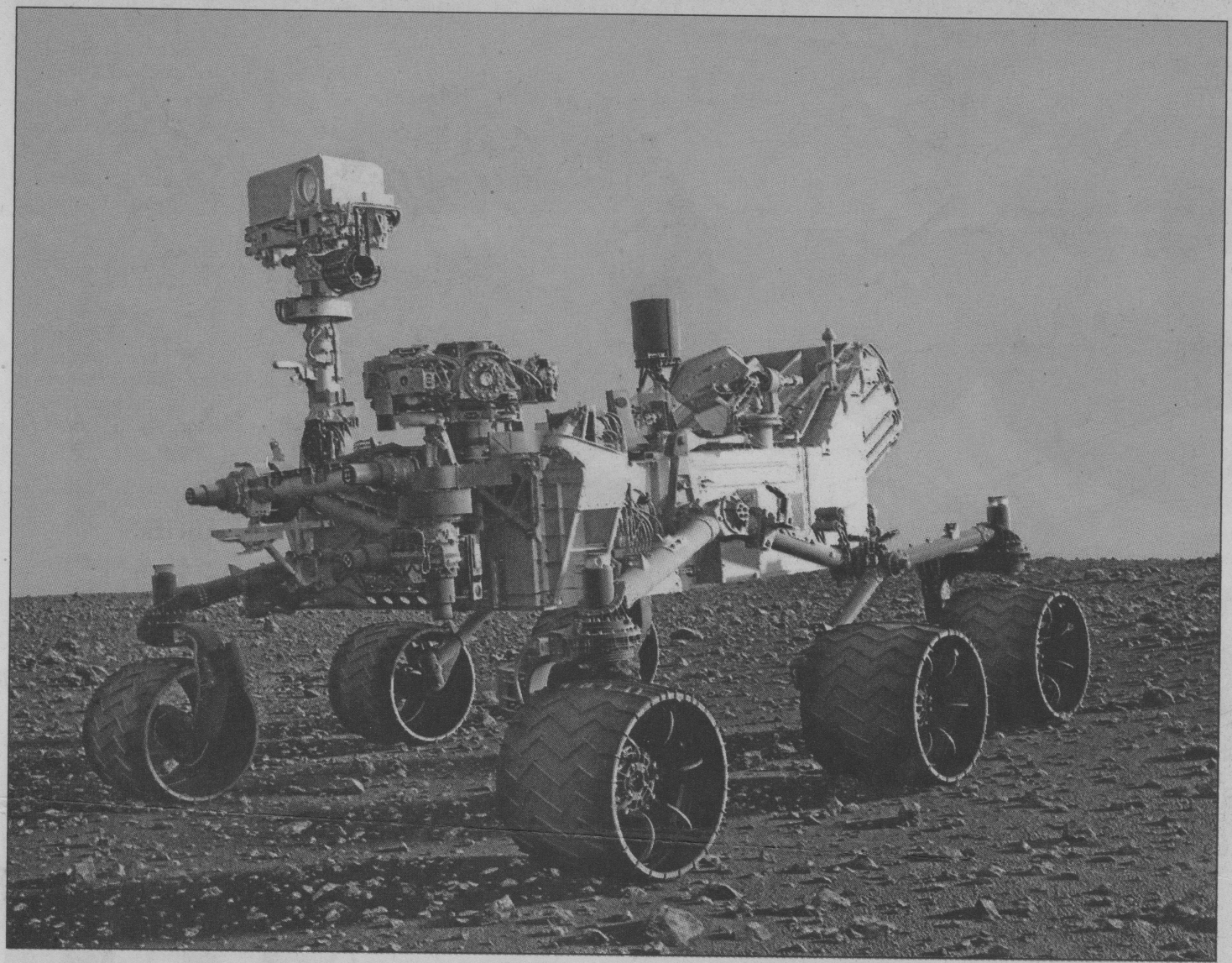
After the data is received, it is sorted and still images are produced, which then are stranded together like a flip book, De Jong said.

"All the inherent information is there, except sometimes you miss a package that contains part of the full image," he said.

Though it's proving tricky to maneuver so much data, the more there is, the more room there it provides for future technological advancement, Mattmann said.

"I think people are consistently inventing things out of new data," he said. "But from a science policy type of perspective, you can never have enough data."

JET PROPULSION LABORATORY



COURTESY NASA/JPL

The Mars Curiosity rover is one example of a mission that uses big data. Scientists are working on the best way to store it.