Analysis of Data Science in Rocket Science

Mrs. R.Pradheepa (Head of the Department)[1], Ms. R.Menaka (Research Scholar)[2]

[1]Department of Information Technology *Sankara college of Science & Commerce Saravanampatty, Coimbatore-*
[2]*Sankara college of Science & Commerce Saravanampatty, Coimbatore*

Abstract: *When sending humans to the place, where no one has gone before, there are a multitude of variables to consider, and NASA (National Aeronautics and Space Administration) is hard at work researching the health and safety risks of a future Mission to Mars. The stakes are high, but NASA realized from the get-go that it needed to focus less on developing the perfect analytic model and more on building a data science process that empowers decision-makers to use analytics to answer a multitude of continually changing questions. This paper investigates, "How data science coupled with rocket science to get humans to Mars?"*
Keywords: *Analytic – mission - data science – rocket science*

I. Introduction

For the Mars project, there are so many levels of data to look at, ranging from health data collected from astronauts like Scott Kelly who have completed previous space missions, to non-astronaut test studies, to studies done in simulated space environments like the Human Exploration Research Analog (HERA) at Johnson Space Center in Houston. Getting all the data in one place is the critical first step. For this reason, NASA is using the Collaborative Advanced Analytics and Data Sharing platform developed by Lockheed Martin and several analytic partners, such as Alpine Data, to analyze data at its source. Because there's no waiting to download data into a separate analytic environment to work with it, researchers can focus their time and energy on asking questions and getting the answers that will help them plan a mission to Mars.

This paper addresses the following issues:
Data Science and Big Data
What will be NASA's Biggest Data Challenge?
Instruments that are used to gather the data
An Overview of Advanced Analytics
Premier Example of a Collaborative Data
Collaborative Advanced Analytics and Data Sharing

II. Data Science and Big Data

A. An Overview of Data Science
Data Science is a discipline that incorporates varying degrees of Data Engineering, Scientific Method, Mathematics, Statistics, Advanced Computing, Visualization, Hacker Mindset and Domain Expertise.

B. About Big Data
Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, manage and process the data within a tolerable elapsed time.
For Example, Google's self-learning voice recognition through probability calculations Nestle saves $1B annually by analyzing big data. Also, Visa sped up crunching 36 TB of data from one month to 13 minutes with Hadoop in the cloud.

• Traditionally, data from the Mars Science Laboratory (MSL) is posted to specific Sol journal entries, in PowerPoint and PDF file formats. While this process is reasonable for record-keeping, on-the-fly analysis is limited. Because MSL data could be difficult to interact with, methods to integrate images and thermal data from the rover were developed using time series data format. The resulting tools can query over many days, months, or years of data, and represent an early win for NASA in the Big Data realm.

C. NASA's big data challenge
NASA's big data challenge is not just a terrestrial one and it goes beyond the stereotypical challenge. Many of their "big data" sets are described by significant metadata, but on a scale that challenges current and future data

management practice. They regularly engage in missions where data is continually streaming from spacecraft on Earth and in space, faster than we can store, manage, and interpret it. NASA has two very different types of spacecraft.

- Deep space spacecraft that sends back data in the order of MB/s.
- Then earth orbiters that can send back data in GB/s per second.

In their current missions, data is transferred with radio frequency, which is relatively slow. In the future, NASA will employ technology such as optical (laser) communication to increase the download and mean a 1000x increase in the volume of data. This is much more than they can handle today and this is what we are starting to prepare for now. IT at NASA is planning missions, which will easily stream more than 24TB's a day. That's roughly 2.4 times the entire Library of Congress – every day. (For one mission)

- It's still very expensive to transfer just one bit down from a spacecraft so they want to make sure to collect what is most important. Once the data makes its way to their data centers, storing, managing, visualizing and analyzing it becomes an issue. For example, the size of the Climate Change data repositories alone is projected to grow to nearly 350 Petabytes by 2030. 5 PB's is equivalent to the total number of letters delivered by the US Postal Service in one year!

III. Instruments that are used to gather the DATA

A. *About Insight*

Information gathered by the *"Insight"* mission will boost understanding of how all rocky planets formed, including Earth, NASA said. "InSight", which stands for *Interior Exploration using Seismic Investigations Geodesy and Heat Transport*, is an international mission designed to help us understand how rocky planets, like Mars and the Earth, formed and evolved. The mission will place a stationary lander near Mars' equator. With two solar panels that unfold like paper fans, the lander spans about six meters. Within weeks after the landing, InSight will use a robotic arm to place its two main instruments directly and permanently onto the martian ground, an unprecedented set of activities on Mars, according to NASA.

- One of the two instruments is a *seismometer*, which is shielded from wind, and has sensitivity fine enough to detect ground movements half the diameter of a hydrogen atom. It will record seismic waves from "mars-quakes" or meteor impacts that reveal information about the planet's interior layers.
- The other instrument is a heat probe, designed to hammer itself to a depth of three meters or more and measure the amount of energy coming from the planet's deep interior.
- A third experiment will use radio transmissions between Mars and Earth to assess perturbations in how Mars rotates on its axis, which are clues about the size of the planet's core.

IV. WHAT WILL BE NASA'S BIGGEST DATA CHALLENGE?

A. *An Overview*

For NASA, the biggest data challenge is that the astronaut sample size is small — only 300 individuals have been accepted to NASA's Astronaut Corps. Researchers have to mine the heck out of the data collected from this small sample and extrapolate. For example, based on how a 35-year-old female with a starting weight of 120 pounds responded to a five-month trip in space, how would a 32-year-old weighing 123 respond to two years? What about a 30-year-old weighing 118? Furthermore, since an astronaut has yet to step foot on the Red Planet, there's no data about the health impacts of actually living on Mars.

- What can NASA learn from astronauts who have gone to the Moon, or spent a year in the International Space Station?
- What happens when data from test subjects living in simulated space environments is plugged into a predictive model?

With analytic tools that support rapid model deployment and refinement, organizations can keep trying different ways to extract insight from the data they do have to make better predictions, even when key information is missing.

B. *Break the metaphorical black box*

Like any consumer of analytics, NASA needs to be able to trust in the recommendations that are being generated, but this is hard to do if predictions are computed in a "black box" that only data science experts can manipulate or understand. For a project like this, empowering analytic consumers who aren't necessarily data science PhDs (such as the health researchers, equipment engineers and others planning the mission) to actually build and launch queries and use the data on their own is the key. This requires tight collaboration between business and IT stakeholders, modeling tools that are simple to use and modify and the ability to push insights to the people who need them.

This is why NASA has chosen a collaborative analytic platform that includes tools that extend outputs directly into the systems and applications that are used by the scientists and decision-makers working on the Mars Mission.

V. An Overview of Advanced Analytics

A. Short Note

Advanced analytics is a broad category of inquiry that can be used to help drive changes and improvements in business practices. Because analytics is the "combustion engine of business," organizations invest in business intelligence even when times are tough.

Some of the analytical categories that fall under advanced analytics are as follows,

- Predictive Analytics
- Data-Mining
- Big-data Analytics
- Location intelligence

These technologies are widely used in industries including marketing, healthcare, risk management, and economics.

IV. Premier Example of a Collaborative Data

A. NASA Earth Exchange

NASA Earth Exchange is the premier example of a collaborative data capability's impact on Earth science research. The NASA Earth Exchange (NEX), a key capability offered jointly by the NASA Advanced Supercomputing (NAS) and the Earth Sciences Divisions at the Ames Research Center, represents a collaboration platform for the Earth science community that provides a unique mechanism for scientific cooperation and knowledge sharing. Combining the NAS facility's powerful supercomputing resources with NASA ground- and space-based remote sensing data feeds, Earth system modeling capabilities, tools, and a social networking component, the NEX platform offers scientists a place to explore and analyze large datasets, run modeling codes, collaborate on new or existing projects, and quickly share results.

Collaborative Advanced Analytics And Data Sharing

A. Introduction to CAADS

The Collaborative Advanced Analytics & Data Sharing Platform (CAADS™) is an end to end scalable healthcare data analytics solution from Leidos. CAADS™ seeks to address some of healthcare's major challenges in data analytics by using existing technologies

B. About Leidos and its role

Leidos is a global science and technology solutions and services leader working to solve the world's toughest challenges in the defense, intelligence, homeland security, civil, and health markets.

Its role is to expertise in analytics and big data which continues to transform healthcare technologies by translating "*disparate data into meaningful and actionable information*." With stakeholders in healthcare, public health, biomedical research and health policy, they deliver actionable information that aids decision-making to optimize care and identify health threats. Their quality health solutions in predictive analytics and big data, lifecycle management, business intelligence, investigation tracking, enterprise search and data management help improve performance and expand capabilities, thus maximizing value for its users.

Acknowledgement

NASA employs data science within business operations to transform data into useful information and inspire questions that they did not think to ask before. Their definition of data has expanded to include "dark data"- hidden data we have not traditionally considered to be data. For example, we can analyze our calendars to answer the question "how can we make better use of our time?" Continued prototyping across additional business areas will help enhance analytical thinking across the organization and promote a data-driven approach to decision making. In the future, they anticipate expanding beyond descriptive analytics into the most predictive and prescriptive phases.

Large and complex data sets pose challenges for any organization about to embark on an analytics deployment. But NASA's example of harnessing data to plan the most complicated of journeys — an expedition to Mars — proves that the challenges are not insurmountable. With the right tools and, most importantly, a consistent and well-planned approach, data science doesn't have to be as daunting as rocket science.

References

[1] Instruments that are used to gather data: http://economictimes.indiatimes.com/articleshow/60273566. cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst.

[2] What will be NASA's biggest data challenge?: https://techcrunch.com/2016/11/19/how-data-science-and-rocket-science-will-get-humans-to-mars/

[3] An Overview of Advanced Analytics: http://searchbusinessanalytics.techtarget.com/definition/advanced-analytics

[4] About Leidos and its role:

[5] https://www.cloudera.com/solutions/gallery/lockheed-leidos-collaborative-advanced-analytics-and-data-sharing-platform-caads.html