

# Pilot Data Analysis of PDS Geosciences Data in the Cloud with the Java CRISM Analysis Tool (JCAT)

**Pilot Data Analysis of PDS Geosciences Data in the Cloud with the Java CRISM Analysis Tool (JCAT)**  
 D. V. Politte, R. E. Arvidson, T. C. Stein, L. E. Arvidson  
 Department of Earth and Planetary Sciences, Washington University in Saint Louis, Saint Louis, Missouri, 63130

**Why analyze data on the Cloud?**  
 Doing heavy computing on the Cloud makes sense, especially for the user's own computer, which means less time spent. When you run a cloud VM to do an analysis, you can easily change your "local" environment.  
 Local computer requirements change for those various personal laptops:  
 - Storage space doesn't tend to be enough to fit data out  
 - CPU

**Methodology**  
 In this pilot project, we analyzed data from the Campaign Mars Reconnaissance Orbiter (Mars CRISM) Hyperspectral instrument using the Java CRISM Analysis Tool (JCAT) available from JHU/APL.  
 This technology was used for this project, as well as the necessary for using it in cloud, are applicable to many other PDS datasets and the tools for analyzing them.  
 We used a Windows Server 2019 VM on the Microsoft Azure Cloud IIS web server - single-tenant, free-tier, IIS, v3.  
 - 2 virtual CPUs  
 - 3 GB of RAM  
 - 128 GB of virtual hard-disk space available  
 - Windows Server  
 Our costs: \$0.2118 per hour. Expected time to analyze 10 CRISM datasets with JCAT on Windows 2019 VM is 2.0000 hours, or 7 hours, making the total analysis cost \$1.48. (Note: this cost may vary depending on the size of the data set and the cloud provider.)

**Example CRISM Analysis**  
 (Image showing a CRISM analysis interface with a map and data plots)

**Local vs. Cloud Data Analysis**  
 (Diagram showing a workflow from a local workstation to a cloud VM, which then connects to a data source and returns results to the workstation.)

**Future Plans**  
 The possible future directions include better support for data users who want to use the workflow for their own datasets. Our next steps, by making the process simpler. Possible future include:  
 - "One-click" cloud VM setup on PDS datasets, with tools to be used directly from user's VM (no custom installation script is not necessary)  
 - Cloud data repository in VM as read-only disk  
 - Portable VM image with pre-installed Java enabling software (Python, Java, etc.)  
 - Custom documentation, including examples  
 - Support for other personal web-based VMs, allowing local access of data  
 - Tools for non-interactive software vendors that allow users to

CRIM AUTHOR INFORMATION REFERENCES CONTACT AUTHOR PRISM C&T POSTER

D. V. Politte, R. E. Arvidson, T. C. Stein, L. E. Arvidson

Department of Earth and Planetary Sciences, Washington University in Saint Louis, Saint Louis, Missouri, 63130

PRESENTED AT:



## WHY ANALYZE DATA ON THE CLOUD?

Doing heavy computing on the Cloud reduces requirements for the user's own computer, which lowers hardware costs. Short-term use of a cloud VM to do an analysis is often relatively cheap (see "Methodology").

Local computer requirements change for these common processing bottlenecks:

- Storage space: doesn't need to be enough to fit data set under analysis
- CPU: doesn't need to be as powerful
- RAM: doesn't need to be large enough to run the analysis software

Other benefits:

- Faster to acquire data, if data supplier (e.g., PDS Geosciences Node) stores data in the same cloud data center (see Figure 1 in below section)
- Provides temporary processing and storage resources
- Scalable cloud processing to meeting application requirements
  
- Data supplier also benefits from avoiding data egress fees (i.e., a charge incurred upon download), which are often a significant portion of costs of data storage on the cloud

## LOCAL VS. CLOUD DATA ANALYSIS

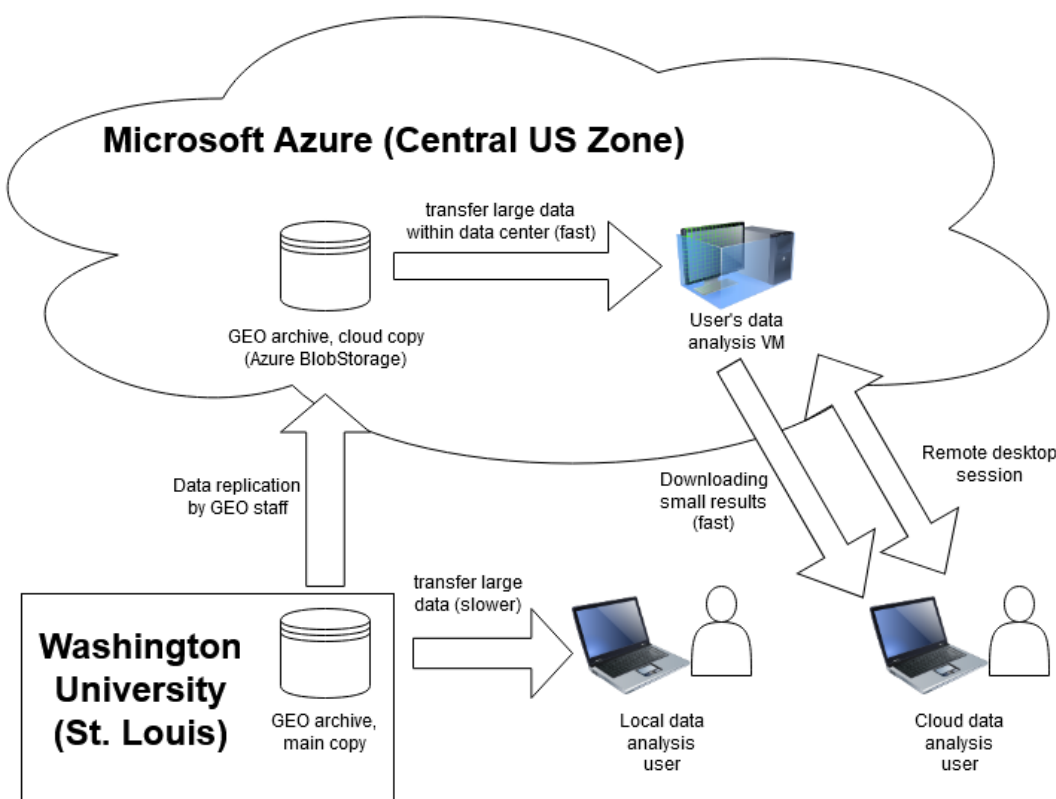


Figure 1: Some differences between local and cloud data analysis of PDS Geosciences Node (GEO) data. In local analysis, large files must be transferred to the user's local machine across the public internet, potentially leading to large delays. In cloud analysis, however, large files are transferred within a cloud data center's local network and so can be ready for analysis sooner. Only the smaller result files need to be transferred back to the user's local machine.

## METHODOLOGY

In this pilot project, we analyzed data from the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) hyperspectral instrument using the Java CRISM Analysis Tool (JCAT) (<http://crism.jhuapl.edu/JCAT/>) available from JHU/APL.

The techniques and tools used in this project, as well as the rationale for using the cloud, are applicable to many other PDS datasets and the tools for analyzing them.

We used a Windows Server 2019 VM in the Microsoft Azure Central US zone with configuration Standard\_D2s\_v3:

- 2 virtual CPUs
- 8 GB of RAM
- ~100 GB of virtual hard-drive space available
- Windows license

Our costs: \$0.2110/hour. Expected time to analyze 10 CRISM scenes with JCAT as described in "Example CRISM Analysis" is 7 hours, making the total analysis cost \$1.48. (Note: this cost may vary between users, depending on relationships with the cloud provider.)

Steps before the analysis:

- From Azure cloud management web portal, selected desired VM configuration and logged into it with remote desktop
- Transferred installers for non-OS software we needed (Python, Java Runtime Environment, JCAT itself) to the VM using our remote desktop client's "local resources" feature
- Installed the above tools on the VM
- Found CRISM products for analysis using Geosciences Node's Mars Orbital Data Explorer (<https://ode.rsl.wustl.edu/mars/>)
- Copied download links for these files to our VM & input them to a Python script that used them to locate & download the files from the archive cloud copy

## EXAMPLE CRISM ANALYSIS

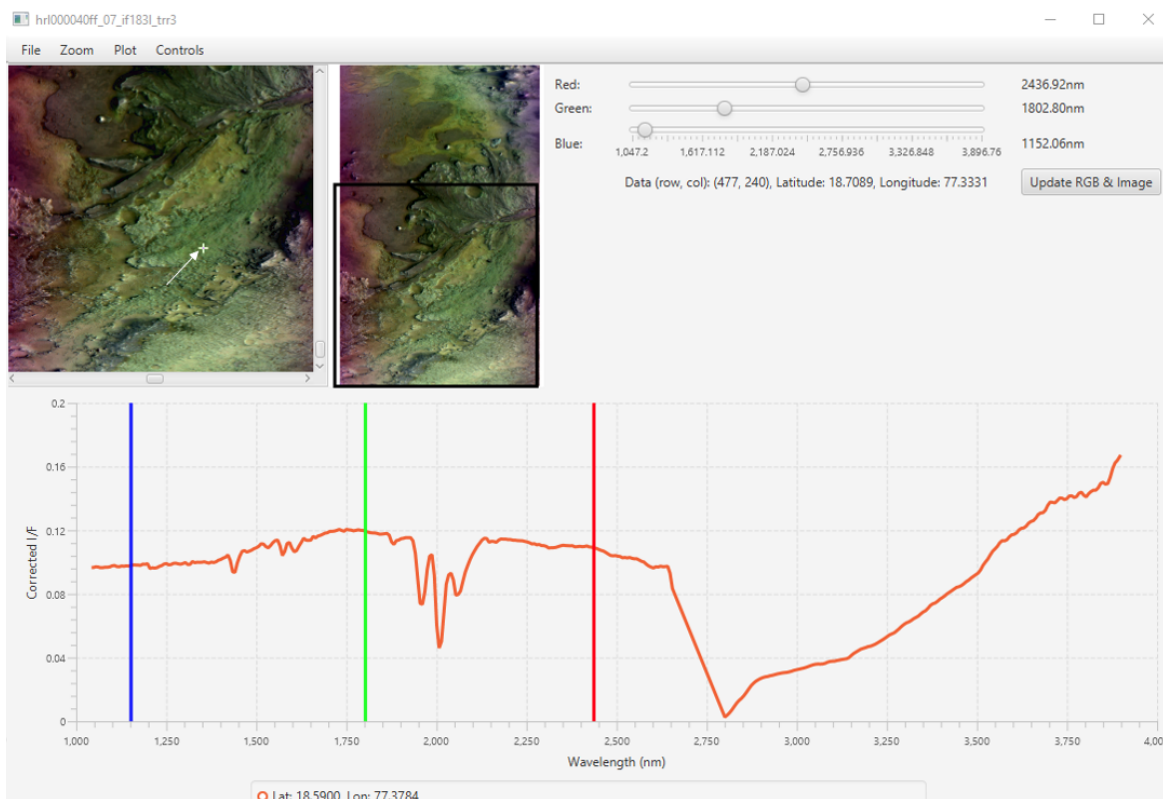


Figure 2: JCAT's initial view of CRISM scene HRL000040FF's data upon loading it. The displayed IOF spectrum was chosen from the candidate carbonate-bearing terrain. Vertical lines show the spectral bands used to display the RGB false color images in the upper left. Carbon dioxide band triplet at ~2 micrometers is evident, along with the deep 3 micrometer band and thermal rise at longer wavelengths. North is toward the bottom of the image views.

To illustrate this proof-of-concept, we analyzed CRISM scene HRL000040FF, which covers the western side of Jezero Crater, including the Mars 2020 rover mission (Perseverance) landing error ellipse, the delta, and the carbonate-bearing terrains up against the western Jezero Crater wall [1]. Our analysis:

- Loaded IOF data for the scene into JCAT (see Figure 2 above)
- Divided the spectrum by a volcano scan observation to remove the gas bands
- Identified diagnostic 2.3 and 2.5 micrometer absorptions in an area of the scene, which demonstrates the presence of carbonate-bearing terrain (see Figure 3 below)
- Exported these results from JCAT as PNG images
- Transferred our results back to our local machine via remote desktop client

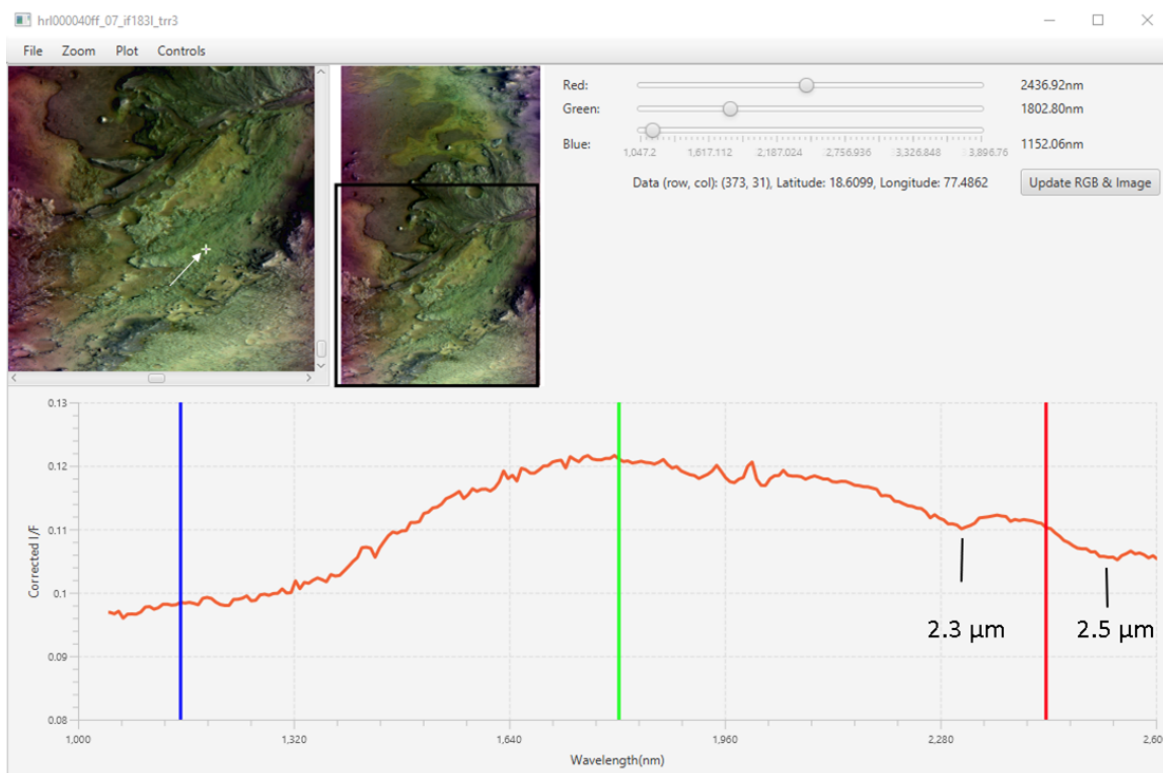


Figure 3: IOF spectrum shown after dividing by a volcano scan to remove gas bands, with the X and Y ranges of the spectral plot configured to better show absorption features at 2.3 and 2.5 micrometers associated with one or more carbonate bearing minerals.

References: [1] Goudge T. A. et al. (2015) JGR Planets, 120, 775-808.

## FUTURE PLANS

Our possible future directions involve better support for data users who want to use this workflow for their own planetary data analysis, by making the process simpler. Possibilities include:

- "Download from Cloud" option in PDS Geosciences search tools, to be used directly from user's VM (so custom download script is not necessary)
- Cloud data mounted in VM as read-only disk
- Pre-made VM image with pre-installed basic enabling software (Python, Java Runtime Environment, etc.)
- Online documentation, including examples
- Support for similar procedures with cloud VMs running Linux instead of Windows
- Tools from proprietary software vendors that allow users to run licenced software, often without licences for certain subsets of use-cases

## Acknowledgments

This work was supported through NASA CAN NNX16AB15A for the Planetary Data System Geosciences Node at Washington University in Saint Louis. The CRISM data used can be found in the PDS datasets MRO-M-CRISM-3-RDR-TARGETED-V1.0 and MRO-M-CRISM-6-DDR-V1.0.

## AUTHOR INFORMATION

Authors: D. V. Politte, R. E. Arvidson, T. C. Stein, L. E. Arvidson

All authors are affiliated with: Department of Earth and Planetary Sciences, Washington University in Saint Louis, Saint Louis, Missouri, 63130

Corresponding Author: Daniel Politte ([dvpolitte@wustl.edu](mailto:dvpolitte@wustl.edu))



## REFERENCES

- [1] Goudge T. A. et al. (2015) JGR Planets, 120, 775-808.