

# Oracle America, Inc. Response to National Aeronautics and Space Administration – Jet Proposal Laboratory

Open Sourced Science for NASA Earth System  
Observatory Mission Science Data Processing Study  
Request for Information

February 1, 2022

---

# Response Guidelines

---

## Corporate Entity

This Response is being made by Oracle America, Inc., a wholly owned subsidiary of Oracle Corporation. All responses reflect information concerning Oracle Corporation (hereinafter referred to as Oracle) except where otherwise indicated as being information of Oracle America, Inc. (hereinafter Oracle America, Inc.).

---

## Definition

Throughout this Response, the term solution refers to and is interchangeable with approach or system. Solution is not intended to contractually bind Oracle to solve any issues or problems. It is intended to express the concept that an approach to your project has been well thought out and is the result of the use of our products, methods, and experience.

Throughout this Response, the term partner refers to and is interchangeable with ally or collaborator. Partner is not intended to contractually or legally bind Oracle to any third party.

---

## Response Validity

This Response shall remain valid for 180-days, unless otherwise mutually agreed, in writing, by Oracle and NASA JPL.

---

## Terms and Conditions

Oracle's proposal is governed by the terms set forth in the General Terms document embedded below and provided as a separate attachment in this proposal.



# Table of Contents

<b>Response Guidelines.....</b>	<b>i</b>
<b>1. Introduction .....</b>	<b>1</b>
Name and Description.....	1
Points of Contact.....	1
Relevant Past Experience.....	1
<b>2. Recommendations for Approaches and Capabilities .....</b>	<b>2</b>
2.1. Data Processing Systems Architecture.....	2
2.2. Open Science.....	3
2.3. Component Technologies .....	4
2.4. Downstream Interoperability .....	5
2.5. Other Recommendations .....	6
<b>3. Links and References .....</b>	<b>7</b>

# 1. Introduction

---

## Name and Description

Since 1977, Oracle America, Inc. (Oracle) has helped hundreds of thousands of customers of all sizes around the globe simplify their processes by engineering hardware and software to work together. Oracle offers a complete technology stack in the cloud, on premises, and in the data center. Our stack of products gives customers complete deployment flexibility and the unmatched benefits of application integration, powerful performance, high availability, scalability, advanced security, energy efficiency, and low total cost of ownership.

In addition, we not only provide robust products, but Oracle also works with you on every step of the digital journey. NASA will benefit from Oracle's customer support services and can also take advantage of optional services, such as consulting, training, upgrade support, and financing. We will help you get the most out of your Oracle products so that you can meet your business objectives.

Flexibility combined with exceptional price and performance for scientific research mission critical work is the basis of our response.

---

## Points of Contact

Mark Barrett, Sr. Account Executive  
Email: [Mark.C.Barrett@oracle.com](mailto:Mark.C.Barrett@oracle.com)  
Telephone: 925.876.7477

Liz Lydon, Cloud Tech Sales  
Email: [Liz.Lydon@oracle.com](mailto:Liz.Lydon@oracle.com)  
Telephone: 703.862.8213

---

## Relevant Past Experience

Multiple NASA programs and missions have counted on and run on Oracle; NASA's DSN, SGSS, Cassini, MRO, TDRSS, SCA, JWST, HST, and Voyager programs all run on various Oracle infrastructure solutions—Voyager, for over 25 years. Many NASA missions run on Oracle for their entire lifecycle, from the Jet Propulsion Laboratory's (JPL's) ITSD Engineering Product Lifecycle Management CAD drawings and extending beyond the antenna tracking and telemetry data capture of the Deep Space Network. NASA IT leaders are engaging Oracle to better manage database and application workloads—even non-Oracle workloads.

We have leveraged Oracle's Global Solutions architects with specializations in data science, HPC, open science, infrastructure, and cloud to review the open science challenges outlined in your Request for Information (RFI). Oracle provides engineered systems and consulting services for on-premise/hybrid/cloud deployments that directly address the request for recommendations of approaches and capabilities of RFI Topic Areas 1 through 4.

Oracle has many implementations of open big data platforms in different ranges of performance and capacity requirements. Oracle can provide a single vendor solution based on proven, mature products with hundreds of installations around the world. Recent customer examples are also provided in [Section 3. Links and References](#) of this document.

## 2. Recommendations for Approaches and Capabilities

In response to the OSS requirements presented by the ESO Mission Science Data Processing Study team, Oracle offers the capabilities required of the major categories in the areas of data processing system architecture as well as the capabilities required to address the needs of the open science community. This includes achieving efficiencies of component technologies, downstream interoperability, and other areas such as high-performance computing,

The following sections address each of these areas using Oracle's strategic approaches, architecture, and technology. Oracle is committed to open-source projects and is a leader in the development of many of the best-known open source projects in the world, including Java, Linux, MySQL, and VirtualBox. Oracle continues to further demonstrate its commitment by running its own internal IT systems for more than 135,000 employees on open-source infrastructure. At Oracle, more than 45,000 developers use many open-source products as part of the product development process.

### 2.1. Data Processing Systems Architecture

When looking at data processing systems architecture, it is imperative to consider key approaches that this architecture will satisfy in multiple areas such as the need for unstructured and structured data; complex analytics with predictive analysis and modeling scenario-based outcomes; flexibility to run in an on-premises, hybrid, or private/public cloud model; interoperability with common scientific data processing (algorithms) applications available via cloud marketplace; and accomplishing autonomous operational efficiency through artificial intelligence (AI) and machine learning (ML), security, compliance, and high-performance compute capabilities to name a few key criteria.

Oracle Data Lake, a relational database that uses a combination of Hadoop for big data processing of structured data and a combination of hybrid /cloud storage would meet the complex needs of the rapid nature of data ingestion, meet the on-demand analytics needs of the various stakeholders and client agencies, and deliver efficient results and fast turnaround times from the system.

**Figure 2** shows how Oracle Cloud Infrastructure (OCI) and other Oracle on-premises and hybrid offerings could fit into the Earth Science Data System/Open Science program.

- Data is already stored in third-party cloud(s) as well as in their own data centers on premises.
- Sensor data is constantly stored in HDF format (semi-structured data in XML or multi-dimensional arrays plus metadata).
- Data scientists use this data with third-party or self-developed algorithms.

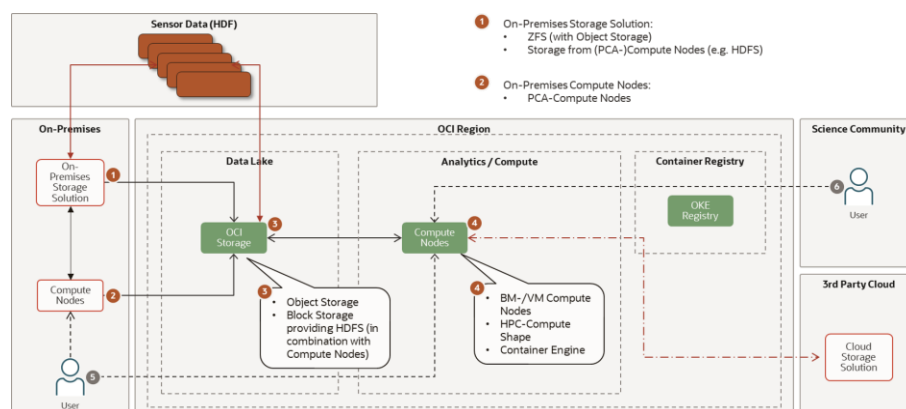


FIGURE 1: CONCEPTUAL SOLUTION OVERVIEW – ESDS



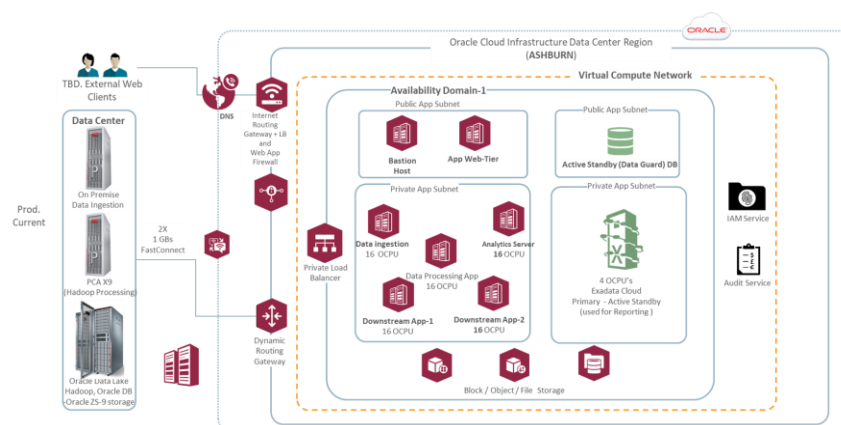


FIGURE 2: SAMPLE ARCHITECTURE – HYBRID CLOUD (ORACLE PRIVATE AND PUBLIC CLOUD)

The following are products/solutions that Oracle can provide for this sample architecture.

#### ON-PREMISES

- PCA as Compute and Storage Solution (Hadoop nodes including HDFS)
- ZFS Storage with Object Storage (and HDFS Connector) - scale up to 10-100 PB

#### ORACLE CLOUD INFRASTRUCTURE

##### Compute

- Bare Metal or VM-Compute nodes with block storage attached (can provide HDFS)
- HPC-Shapes are available if required (based on the analytics performance needs)
- Oracle Kubernetes Engine (OKE) and OKE Registry to provide scalable architectures
- Auto-scaling configurations, metrics or time-based (schedule-based)

##### Storage

- Object Storage with HDFS Connector
- Compute Shapes as Hadoop Cluster providing HDFS)
- OCI HPC File Systems (HFS)

##### Networking

- High-speed/low-latency network within OCI
- FastConnect to access OCI resources from on-premises or third-party clouds

##### Security/Access to Resources

- Internet or dynamic routing gateways to access compute (including attached storage) and policies to limit access
- Routing lists to route traffic between subnets
- Security lists/network security groups to manage traffic

## 2.2. Open Science

Oracle is committed to cultivating, supporting, and promoting popular open-source technologies that customers can confidently deploy in business-critical environments. We are a founding member of the Linux Foundation, Eclipse Foundation, and Java Community Process as well as one of the industry's largest contributors to open source. Oracle invests significant resources in enhancement, testing, optimization, and support for many projects used by millions of people each day. Oracle software, combined with open-source technologies, will enable the open science community to reap the benefits of high performance, reliability, and data security while lowering cost of computing.

**Oracle's Open Source Commitment** – Demonstrated by running its own internal IT systems for more than 135,000 employees on open-source infrastructure, it has realized first-hand the benefit of lower IT costs. Oracle's internal software development organization, with more than 45,000 developers, uses many open-source products as part of the product development process. Oracle Cloud is designed with Oracle open-source technologies at its heart.

**More than just Open Source Code** – Oracle contributes substantial code and bug fixes to various open-source projects. Oracle also offers world-class support and training for its open source

offerings. Oracle Premier Support provides access to product updates, enhancements, and technical support resources to help you maintain and implement new software.

**Oracle Cloud Native Environment** – A fully integrated suite for the development and management of cloud native applications, Oracle Cloud Native Environment (CNE) is a curated set of open-source projects that are based on open standards, specifications, and APIs defined by the Open Container Initiative and Cloud Native Computing Foundation (CNCF) that can be easily deployed, have been tested for interoperability, and for which enterprise-grade support is offered. Oracle CNE delivers a simplified framework for installations, updates, upgrades, and configuration of key features for orchestrating microservices. Oracle even allows developers to move dockerized containers and application packages directly to Oracle Cloud, with built-in support for Kubernetes, CoreOS, PortWorx, and other popular containers.

**Oracle Machine Learning for Spark** – Supported by Oracle R Advanced Analytics for Hadoop, it provides massively scalable machine learning algorithms via an R API for Spark and Hadoop environments for data scientists and application developers to build and deploy ML models. OML4Spark brings custom linear model (LM), generalized linear model (GLM), and MLP neural networks algorithms that execute on Spark infrastructure.

**Oracle Cloud Infrastructure Data Flow** – A fully-managed Apache Spark service to perform processing tasks on extremely large data sets without infrastructure to deploy. This enables rapid application delivery. Developers can focus on app development, not infrastructure management.

**Oracle Data Lake** – Enables open science users to ingest data in the original form without altering it. This can be for speed reasons but also to perform advanced analytics for text mining, data mining, statistical analysis, anything involving clusters, or graph analytics.

**Ingest and Transform** – Move/convert disparate data formats from any open science data source.

**Persist and Access** – Ensure data is secure, can be readily discovered, and has easy scale or access.

**Analyze and Use Data Science** – Uncover insights and trends within data.

---

## 2.3. Component Technologies

Component technologies and techniques that add value to open science efforts are as follows.

### PREDICTIVE ANALYTICS

Big data can help in areas of predictive analytics such as satellite data pertaining to ESO mission science data. Potential issues can be discovered by analyzing both structured data and multi-structured data (log entries, sensor data, error messages, and other factors). With this data, the open science community can be on the cutting edge of research and predict futuristic models and trends while at the same time maximizing operational efficiency and deploying cost-effective maintenance. Oracle Big Data powered by Data Lake addresses challenges related to data integration for data coming from disparate (open science) systems with different data formats at high-velocity volumes.

### OPERATIONAL EFFICIENCY – BIG DATA AND HADOOP

Operational efficiency is one of the areas in which big data can have the most impact on profitability. With big data, you can analyze and assess open science processes, proactively respond to feedback from various internal and external clients stakeholders, and anticipate futuristic trends. Oracle Big Data address challenges that balance data volume due to growth of sources, users, and applications.

Open System Data Ingestion is a diverse data environment, and the data lake can help with that, especially when the data lake is on Hadoop. Hadoop is largely a file-based system because it was originally designed for very large and highly numerous log files that come from web servers. In addition to large data volumes comprised of the HDF format, there may be diverse data formats of file-based data from various open science communities (for example, file-based and document-based data from EDI systems, XML, and recent JSONs data formats) coming in at rapid rates and high volumes. Oracle's hybrid storage solution is capable of fast ingestion and storage of these data formats. These data formats can be retrieved by the open science community from Oracle's hybrid data storage.

### INTERNET OF THINGS

The Internet of Things (IoT) is creating new data sources almost daily in some cases. And of course, as those sources diversify, they create even more data. Increasingly, there are more sensors on

more machinery all the time. As an example, satellites could collect huge set of data elements from potentially huge list of sensors (for example, weather-related information), so it can be tracked through space and time. Enormous amounts of information are coming from these places, and the data lake is very popular because it provides a repository for all of that data.

#### ORACLE HPC

High-performance computing (HPC) on Oracle Cloud Infrastructure brings powerful, cost-effective computing capabilities to solve complex mathematical and scientific problems across industries. OCI's bare metal servers coupled with Oracle's cluster networking provide access to ultra-low latency RDMA (less than two microseconds latency across clusters of tens of thousands of cores) over converged Ethernet (RoCE) v2. HPC on OCI rivals the performance of on-premises solutions with the elasticity and consumption-based costs of the cloud, offering on-demand potential to scale tens of thousands of cores simultaneously. With HPC on OCI, you get access to high-frequency processors, fast and dense local storage, high throughput ultra-low latency RDMA cluster network, and the tools to automate and run jobs seamlessly. With the explosion of business data ranging from customer data to IoT, data scientists need the flexibility to explore and build deep learning models quickly and with more flexibility than traditional on-premises. OCI provides GPU compute instances for deep learning, easy-to-deploy images, and the flexibility to run a single-GPU workstation or cluster of multi-GPU shapes.

---

## 2.4. Downstream Interoperability

Downstream interoperability is ensured for targeted uses of the data lake in certain departments or IT programs with diverse downstream systems and toolsets for data processing, computation, analysis, and so forth. A different approach is for centralized IT to provide a single large data lake that is multitenant. It can be used by lots of different departments, business units, and technology programs with diverse downstream systems and toolsets.

**Interoperability with Third-party Tools** – OCI services natively interoperate with third-party tools.

- **OCI Logging** is a cloud native, fully managed, distributed platform that simplifies ingesting, managing, and analyzing logs from your entire stack. The service brings all your logs (infrastructure, application, audit, and database) into one view while using open standards.
- **Oracle Service Connector Hub** centralizes the management of log data movement and tracking, and securing customer data workflows across Oracle and partner tools.
- **OCI Notifications** supports interoperability with several other third-party messaging platforms, including Slack, PagerDuty, and email.

**Oracle Cloud Native Environment** – Open-source projects based on open standards can be easily deployed to enable interoperability between various downstream systems. Oracle CNE delivers a simplified framework for installations, updates, upgrades, and orchestrating microservices.

Key features of developing microservices-based cloud native applications include the following:

- Build modern, cloud-native applications quickly and efficiently.
- Code in any language and use open source tools such as Hudson, Maven, GitHub, Junit, and so forth.
- Use a built-in API platform for inbound and outbound interactions with microservices.
- Follow DevOps continuous integration/continuous delivery (CI/CD) practices and use Oracle Developer Cloud Service to track issues, review peer code, merge requests, and so forth.
- Run container applications on enterprise-grade Oracle Container Engine for Kubernetes.
- Store and share Docker container images on OCI Registry.

**Oracle Integration** – These platform services drive business agility and flexibility as standout benefits in an environment where interoperability with disparate source systems is vital.

“Oracle is a Leader in this Magic Quadrant. Its EiPaaS offering comprises Oracle Integration (including Process, Insight, File Server, B2B and Visual Builder Studio), Oracle Cloud Infrastructure API Gateway, Oracle Golden Gate, Oracle SOA Suite on Marketplace, OCI Streaming, OCI Data Integration, and Oracle IoT Cloud Service.



Strengths - Offering capabilities and strategy: Oracle Cloud Integration has one of the broadest sets of capabilities in the EiPaaS market. It provides application, data, B2B and API-focused integration capabilities, along with streaming, managed file transfer, RPA, business process management and IoT capabilities. With an increasing focus on digital integration hub enablement, Oracle is enriching its recipes and business accelerators for digital solutions, including human capital management (HCM), ERP, supply chain management (SCM) and customer experience (CX) applications, with prebuilt adapters and flows.”

- Gartner’s Magic Quadrant Report (2021) for Enterprise Integration Platform as a Service

**Oracle Application Adapters** – Application adapters simplify integration by eliminating the complexity associated with web services and other connectivity methods. Connect with virtually any application, service, or data store. Oracle’s comprehensive library of application adapters provides a standardized way to rapidly connect the various protocols required by each application vendor. In addition to application connectors for CX, ERP, and HCM, Oracle offers database, productivity, social, robotic process automation (RPA), and technology adapters. Details about various application adapters can be found at [Application Connectors | Oracle](#).

## 2.5. Other Recommendations

Oracle Open Data is a free service that curates information—spatial images, protein sequences, and annotated text files—from the world’s leading scientific databases. The repository connects researchers, developers, students, and educators with petabytes of open data from trusted resources. Use Oracle Open Data to view important metadata and sample code for each data set, which simplifies technical complexities and makes it easy for researchers to use.

Oracle Open Data from Oracle for Research makes large data sets available to researchers around the world for analysis. We are currently hosting GOES-16 and GOES-17 datasets and hope to add more in the near future. Geostationary Operational Environmental Satellite (GOES) is a series of equatorial satellites operated by NASA and NOAA. The first, GOES-1, was launched in 1975 and decommissioned in 1985.

GOES-16 was launched in 2016 and serves as the GOES East observer centered on the Americas. The satellite carries a primary imaging instrument (Advanced Baseline Imager (ABI)), which is a multi-channel passive imaging radiometer, that provide data across 16 spectral bands. Detecting momentary changes in the optical scenes, there is also a geostationary lightning mapper (GLM) attached.

GOES-17 was launched in 2018 and operates in the GOES West position centered over the Pacific. Much like GOES-16, the satellite also carries an ABI and GLM. However, in 2018 a fault developed in the ABI that impacted data in 13 of the 16 spectral bands. Mitigation was able to make most of these channels available most of the time, but there are plans to replace the satellite as soon as possible in 2022. The ongoing malfunction with the ABI means that there are many more gaps in the GOES-17 dataset than in those from other platforms.

- The [GOES Program](#) falls under the NASA Science Mission Directorate at Goddard. GOES data is collected by the Deep Space Network (DSN) under the Interplanetary Network Directorate (IND) within the Human Exploration and Operations Mission Directorate.
- Most of Deep Space Network operations runs on much of the same Oracle infrastructure mentioned in this proposal; deployed on three continents since 2017.

Further Details can be found at <https://blogs.oracle.com/research/post/oracle-open-data-will-connect-your-research-with-nasa-and-noaa>.

### MISCELLANEOUS ARCHITECTURE CONSIDERATIONS

- Estimates on how much block/object/storage to include for the compute
- Check on storage size to start with and scale given the broad data volume range of 10 to 100 PB
- Is disaster recovery under consideration? If so, what are RTO and RPO objectives?
- Active data guard ship logs live and can be used to offload good deal of reporting
- Oracle’s hybrid (on-premises and cloud) infrastructure can scale up for on-demand processing

NOTE: Cloud (for example, ExaCS) bursting is an option to scale up to unexpected demand surges.

### 3. Links and References

- Oracle for Research <https://www.oracle.com/research/>
- Oracle Open Data <https://opendata.oraclecloud.com/ords/r/opendata/opendata/home>
- Oracle and NASA's Deep Space Network partnership [NASA Cassini Forbes article](#)
- [Cloud Native Private Cloud at Customer](#)
- CERN Tests Data Exploration Using [Big Data, Analytics, And The Cloud](#)
- [Virtual Humans](#) on Oracle Cloud Infrastructure HPC
- [Oracle Private Cloud Appliance for Scalable Big Data Solution](#)
- Data Processing System Architecture <https://www.oracle.com/assets/oracle-wp-big-data-refarch-2019930.pdf>
- Open Science
  - <https://blogs.oracle.com/research/post/oracle-open-data-will-connect-your-research-with-nasa-and-noaa>
  - <https://www.oracle.com/news/announcement/2021-gartner-cloud-database-management-systems-2021-12-20/>
  - <https://www.oracle.com/industries/government/federal/jwcc/?source=:ow:o:h:feb::OHPpn1030>
- [NASA client references](#) available upon request