Apache OODT

Grid Middleware Framework for Science Data Processing, Information Integration and Retrieval

Apache Object Oriented Data Technology (OODT) is the smart way to integrate and archive your processes, your data, and its metadata. OODT allows you to:

- Generate Data
- Process Data
- Manage Your Data
- Distribute Your Data
- Analyze Your Data

Allowing for the integration of data, computation, visualization and other components.

OODT also allows for remote execution of jobs on scalable computational infrastructures so that computational and data-intensive processing can be integrated into OODT’s data processing pipelines using cloud computing and high-performance computing environments.

Components of OODT

### OODT is Middleware for Metadata

- Transparent access to distributed resources
- Data discovery and query optimization
- Distributed processing and virtual archives
OODT is a Software Architecture Models for information representation Solutions to knowledge capture problems Unification of technology, data, and metadata

Why OODT?

Traditional processing pipelines are commonly made up of custom UNIX shell scripts and fragile custom written glue code. Apache OODT uses structured XML-based capturing of the processing pipeline that can be understood and modified by non-programmers to create, edit, manage and provision workflow and task execution.

OODT is used on a number of successful projects at NASA's Jet Propulsion Laboratory/California Institute of Technology, and many other research institutions and universities, specifically those part of the:

- National Cancer Institute's (NCI's) Early Detection Research Network (EDRN) project - over 40+ institutions all performing research into discovering biomarkers which are early indicators of disease.
- NASA's Planetary Data System (PDS) - NASA's planetary data archive, a repository and registry for all planetary data collected over the past 30+ years.
- various Earth Science data processing missions, including Seawinds/QuickSCAT, the Orbiting Carbon Observatory, the NPP Sounder PEATE project, and the Soil Moisture Active Passive (SMAP) mission.
- Apache DRAT - A distributed release audit tool written on top of OODT's capabilities.

OODT is a Top Level project of the Apache Software Foundation http://www.apache.org/.

Getting Started

OODT is a middleware, hence it is not an out-of-the-box solution. However, we created the RADiX build system to make it easy to get started with OODT. RADiX manages all the hard work behind-the-scenes, and compiles a fully-operational OODT platform that is ready for development and deployment. We use RADiX to provide pre-configured build environments using Vagrant and Docker.

Option 1 - Using Vagrant

This user-guide will help you understand, and set up a Vagrant powered Virtual Machine containing a pre-configured OODT installation.

Link - Vagrant Powered OODT

Option 2 - Using Docker

Link - Docker powered OODT

Option 3 - Using RADiX

You can also ignore Vagrant and Docker altogether and build OODT using RADiX on your local machine. Check out this video to see the RADiX build system in action.

Link - RADiX Powered By OODT

Option 4 - Build Straight From Source
Key Components

<table>
<thead>
<tr>
<th>OODT Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
</tr>
<tr>
<td>Core</td>
</tr>
<tr>
<td>Catalog and Archive</td>
</tr>
<tr>
<td>Grid</td>
</tr>
</tbody>
</table>

Contributing to OODT

- Getting started with Apache OODT