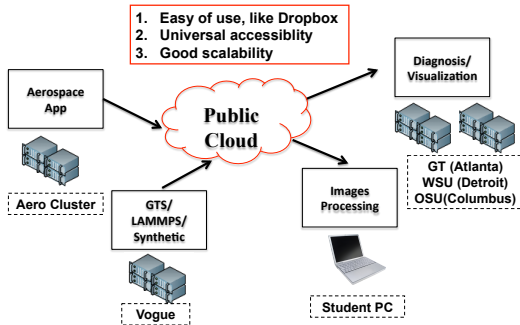


# Scibox: Online Sharing of Scientific Data via the Cloud

Jian Huang\*, Xuechen Zhang\*, Greg Eisenhauer\*, Karsten Schwan\*, Matt Wolf\$, Stephane Ethier^, Scott Klasky\$

\* Georgia Institute of Technology ^ Princeton Plasma Physics Laboratory \$ Oak Ridge National Laboratory

## Use Cases of Cloud Storage

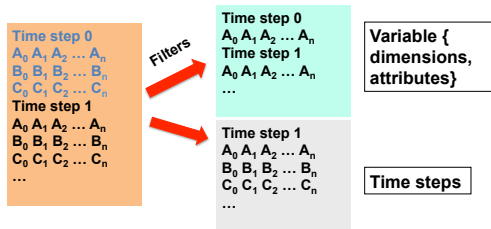


1. Easy of use, like Dropbox
2. Universal accessibility
3. Good scalability

## Challenges of Sharing Data via the Cloud

- Inadequate levels of ingress and egress bandwidth available to/from the remote cloud stores in which data is maintained.
- Cloud storage cost can quickly exceed science budgets, in lieu of the current storage pricing of standard storage provided by Amazon S3.

## Filtering Unnecessary Data at Producers by Exploring the Meta-data



Users of analytics know what data is needed.



## Our Approaches to Reduce the In-cloud Data Size

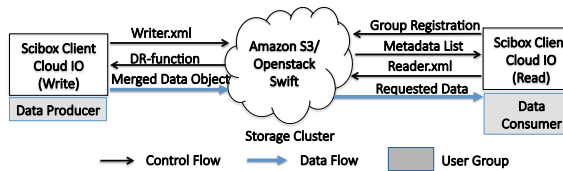
- Leveraging ADIOS' metadata-rich descriptions of scientific data;
- Using D(ata)R(eduction) functions to filter out redundant or irrelevant data;
- Merging data objects for overlapped subsets of original data;
- Supporting partial object access in cloud;

## Examples of DR-functions

```
<?xml version="1.0"?>
<adios-config host-language="Fortran">

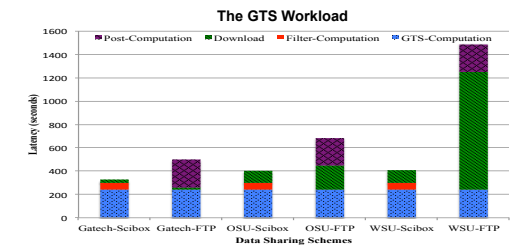
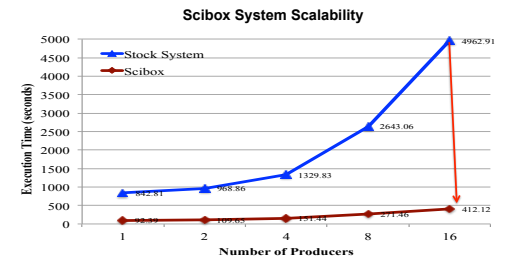
  <adios-group name="restart" coordination-communicator="comm">
    <var name="mype" type="integer"/>
    <var name="numberpe" type="integer"/>
    <var name="istep" type="integer"/>
    <var name="MIMAX_VAR" type="integer"/>
    <var name="NX" type="integer"/>
    <var name="NY" type="integer"/>
    <var name="zion0_1Darray" gwrite="zion0" type="double"
      dimensions="MIMAX_VAR"/>
    <var name="phi_2Darray" gwrite="phi" type="double" dimensions="NX, NY"/>
    <!-- for reader.xml -->
    <dr type="8" name="phi_2Darray" cod="int i;\n\ndouble sum = 0.0;\nfor(i = 0;
      i<input.count; i= i+1) {\nsum = sum + input.vals[i];\n}\nreturn sum;\n"/>
  </adios-group>
<method group="restart" method="CLOUD-IO"></method>
<buffer size-MB="20" allocate-time="now"/>
</adios-config>
```

## Scibox Architecture



## Performance Evaluation

- Software: Openstack and ADIOS
- Hardware: Clusters at GT, WSU, and OSU
- Workloads: Synthetic workloads, GTS, and others.



	GT->WSU	GT->OSU	GT->GT
Bandwidth	900 KB/s	4.4 MB/s	44 MB/s

