Towards Arrow-native and Coffea Analysis Support in SkyhookDM

Jayjeet Chakraborty, Aaron Chu, Ivo Jimenez, Jeff LeFevere, Carlos Maltzahn, Esmaeil Mirvakili

UC Santa Cruz
It has been understood loosely that ServiceX feeds into tools being developed by Analysis Systems group. (SkyHook has not a part of Analysis Systems discussions)

Analysis Facility is understood loosely as where the compute and services will be loaded, but the analysis model and capabilities of the facility not yet concrete.
DOMA
• It has been understood loosely that ServiceX feeds into tools being developed by Analysis Systems group. (SkyHook has not a part of Analysis Systems discussions)

Analysis Facility is understood loosely as where the compute and services will be loaded, but the analysis model and capabilities of the facility not yet concrete.
ARCHITECTURAL COMPONENTS

**RGW**
A web services gateway for object storage, compatible with S3 and Swift

**RBD**
A reliable, fully-distributed block device with cloud platform integration

**CEPHFS**
A distributed file system with POSIX semantics and scale-out metadata

**LIBRADOS**
A library allowing apps to directly access RADOS (C, C++, Java, Python, Ruby, PHP)

**RADOS**
A software-based, reliable, autonomous, distributed object store comprised of self-healing, self-managing, intelligent storage nodes and lightweight monitors
ARCHITECTURAL COMPONENTS

**RGW**
A web services gateway for object storage, compatible with S3 and Swift

**RBD**
A reliable, fully-distributed block device with cloud platform integration

**CEPHFS**
A distributed file system with POSIX semantics and scale-out metadata

**LIBRADOS**
A library allowing apps to directly access RADOS (C, C++, Java, Python, Ruby, PHP)

**RADOS**
A software-based, reliable, autonomous, distributed object store comprised of self-healing, self-managing, intelligent storage nodes and lightweight monitors
ACCESSING A RADOS CLUSTER

APPLICATION
LIBRADOS

OBJECT

socket

RADOS CLUSTER
RADOS ACCESS FOR APPLICATIONS

LIBRADOS
- Direct access to RADOS for applications
- C, C++, Python, PHP, Java, Erlang
- Direct access to storage nodes
- No HTTP overhead
- API for extending base object functionality
LIBRADOS

• Direct access to RADOS for applications
• C, C++, Python, PHP, Java, Erlang
• Direct access to storage nodes
• No HTTP overhead

• API for extending base object functionality
ACCESSING A RADOS CLUSTER

APPLICATION

LIBRADOS

OBJECT

RADOS CLUSTER
ACCESSING A RADOS CLUSTER

- Tabular data support
- Compute offloading
  - Filtering
  - Projection
  - User-defined functions
  - Compression

APPLICATION

LIBRADOS

OBJECT

RADOS CLUSTER

SkyhookDM
### Traditional Memory Buffer

<table>
<thead>
<tr>
<th>session_id</th>
<th>timestamp</th>
<th>source_ip</th>
</tr>
</thead>
<tbody>
<tr>
<td>1331246351</td>
<td>3/8/2012 2:38PM</td>
<td>65.87.165.114</td>
</tr>
<tr>
<td>1331244570</td>
<td>3/8/2012 2:09PM</td>
<td>71.10.106.181</td>
</tr>
<tr>
<td>1331261196</td>
<td>3/8/2012 6:46PM</td>
<td>76.102.156.138</td>
</tr>
</tbody>
</table>

### Arrow Memory Buffer

```
SELECT * FROM clickstream
WHERE session_id = 1331246351
```

- **session_id**: 1331246660, 1331246351, 1331244570, 1331261196
- **source_ip**: 99.155.155.225, 65.87.165.114, 71.10.106.181, 76.102.156.138

---

![Diagram](image.png)
Arrow C++ development platform

- Allocators and Buffers
- Columnar Data Structures and Builders
- Gandiva: LLVM Expr Compiler
- Data Frame Interface
- Embeddable Query Engine
- Binary IPC Protocol
- Compute Kernels
- Datasets Framework
- Multithreading Runtime
- CUDA Interop
- Plasma: Shared Mem Object Store
- Flight RPC
- File Format Interfaces
  - PARQUET
  - AVRO
  - CSV
  - JSON
  - ORC
- IO / Filesystem Platform
  - localfs
  - mmap
  - Azure
  - AWS S3
  - HDFS
  - GCP
- Compressor Interfaces

Red means planned / under construction work... and much more
SkyhookDM

Computational Storage API

File Format Interfaces
- PARQUET
- AVRO
- CSV
- JSON
- ORC

Filesystem API
- localfs
- mmap
- Azure
- AWS S3
- HDFS
- GCP

Datasets API

DataFrame API
- Common Analytics Functions

Query Engine
SkyhookDM

Computational Storage API

RADOS

File Format Interfaces
- PARQUET
- AVRO
- CSV
- JSON
- ORC

Filesystem API
- localfs
- mmap
- Azure
- AWS S3
- HDFS
- GCP

Datasets API

DataFrame API
Common Analytics Functions
Query Engine
```python
import pyarrow as pa
from pyarrow import dataset as ds

schema = pa.schema([pa.field('id', pa.int32()), pa.field('cost', pa.float64()), pa.field('cost_components', pa.list_(pa.float64()))])

dataset = ds.dataset(
    source=['obj.0', 'obj.1', 'obj.2', 'obj.3'],
    format=ds.RadosFormat("test-pool", "/etc/ceph/ceph.conf"),
    schema=schema)

dataset.to_table(columns=['cost_components', "cost"], filter=(ds.field('cost') > 9)).to_pandas()
```
Coffea Example
Coffea Example
Coffea Example

IPC files
ROOT files
Parquet files
...

map

coffeac.log
reduce
Histograms
Event lists
...

coffe executor
Roadmap

• Tidy up our fork; open PR against Arrow upstream repository.

• Documentation; hello world, Coffea and ServiceX examples.

• Support for large table fragments (>128MB).

• Higher-level services: expression caching, async/scheduled offloading, and others (in collaboration with Arrow community).

• Support other Arrow features: dataset discovery, user-defined functions, tensor fragments, Gandiva (in-OSD GPU acceleration).