

Improvements to SAM

(so you'll like it better)

Adam Lyon

CD/REX

May 2011

Remember what SAM does?

SAM:

SAM:

Stores file metadata

SAM:

Stores file metadata

Maintains replica catalog for files

SAM:

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

SAM:

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

Delivers files (tape/cache) to jobs or you

SAM:

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

Delivers files (tape/cache) to jobs or you

Store files to tape

SAM:

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

Delivers files (tape/cache) to jobs or you

Store files to tape

Manage all of this stuff with some **sanity**

SAM:

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

Delivers files (tape/cache) to jobs or you

Store files to tape

Manage all of this stuff with some **sanity**

Used by CDF, DØ, and MINOS

SAM:

Stores file metadata

Maintains replica catalog for files

Maintains dataset definitions

Delivers files (tape/cache) to jobs or you

Store files to tape

Manage all of this stuff with some **sanity**

Used by CDF, DØ, and MINOS

Starting to be used by Minerva, NOvA

But adoption and use of
SAM has been slow by IF

But adoption and use of
SAM has been slow by IF

Why?

But adoption and use of
SAM has been slow by IF

Why? 3 reasons

<1> Won't have enough data to need it

<1> Won't have enough data to need it

<2> Complicated to set up

<1> Won't have enough data to need it

<2> Complicated to set up

<3> Complicated to use

<1> Won't have enough data to need it

But you may still have data on tape you want to bring in and out of disk. SAM will manage this.

You may want to run where the Bluearc cannot be mounted (remotely). SAM can do this too.

<2> Complicated to set up

You're right - it is. We do it for you!

Servers and DBs need to be set up.

The SAM client needs to be deployed everywhere it will be used

Steps I'll address in this talk:

- Interfacing with your framework

- Reducing the need of the SAM client

<3> Complicated to use

You're right - it is!

Though SAM deals with files, its interface is not familiar nor simple

sam commands are built for maximum functionality **at the expense of ease of use**

Learning curve is steep for causal users (you!)

Steps I'll address in this talk:

- Making the SAM interface more familiar

<3> Complicated to use

You're right - it is!

Though SAM deals with files, its interface is not familiar nor simple

sam commands are built for maximum functionality **at the expense of ease of use**

Learning curve is steep for casual users (you!)

Steps I'll address in this talk:

Making the SAM interface more familiar

[Note: The current interfaces will remain]

We've learned from CDF, DØ, and MINOS

We think that ease of use can be achieved, but **at the expense of generality**

We think that ease of use can be achieved, but at the expense of generality

But this is ok; **even a good thing!**

What do casual users do with SAM?

What do casual users do with SAM?

They submit jobs given a pre-defined dataset definition

What do casual users do with SAM?

They submit jobs given a pre-defined dataset definition

Look up files that satisfy a very limited set of meta-data

What do casual users do with SAM?

They submit jobs given a pre-defined dataset definition

Look up files that satisfy a very limited set of meta-data

Experiment experts make the dataset definitions using slightly more meta-data

What do casual users do with SAM?

They submit jobs given a pre-defined dataset definition

Look up files that satisfy a very limited set of meta-data

Experiment experts make the dataset definitions using slightly more meta-data

Maybe you want a particular file put on your disk for testing

What do casual users do with SAM?

They submit jobs given a pre-defined dataset definition

Look up files that satisfy a very limited set of meta-data

Experiment experts make the dataset definitions using slightly more meta-data

Maybe you want a particular file put on your disk for testing

Maybe you want to store files onto tape

What do casual users do with SAM?

They submit jobs given a pre-defined dataset definition

Look up files that satisfy a very limited set of meta-data

Experiment experts make the dataset definitions using slightly more meta-data

Maybe you want a particular file put on your disk for testing

Maybe you want to store files onto tape

The lesson - a simpler limited interface will be ok

<improvements>
2 big ones

<1> SamWeb

<1> SamWeb

A REST Web Interface to a limited set of SAM functions.

Everything that a job needs

Interact over HTTP

Scalable Python web servers are well developed

<http://samweb.fnal.gov/listMatchingFiles run=12334&subrun=45&stream=nu>

<http://samweb.fnal.gov/createProject?group=minerva&defintion=nu 123>

<1> SamWeb

Addresses two complications:

<1> SamWeb

Addresses two complications:

<A> Interfacing SAM to experiment's framework is time consuming

<1> SamWeb

Addresses two complications:

<A> Interfacing SAM to experiment's framework is time consuming

 The SAM client must be deployed everywhere SAM is used

<1> SamWeb

Addresses two complications:

<A> Interfacing SAM to experiment's framework is time consuming

Interface to SamWeb using a standard HTTP library (libCurl).
Your framework may already be able to do this.

Can be done more by the experimenters

Done similarly for all Frameworks

<1> SamWeb

Addresses two complications:

 The SAM client must be deployed everywhere SAM is used

No SAM client is needed on worker nodes, etc.

Deployment is now **much** easier for clients

<2> samfs

<2> samfs

You are used to dealing with files

<2> samfs

You are used to dealing with files with the filesystem

```
ls /home/lyon/myStuff/analyses/data/run134_5455.root
```

<2> samfs

You are used to dealing with files with the filesystem

```
ls /home/lyon/myStuff/analyses/data/run134_5455.root
```

SAM does not work this way

```
sam translate constraints --dim="run 134 subrun 5455 data_tier root"
```

<2> samfs

You are used to dealing with files with the filesystem

```
ls /home/lyon/myStuff/analyses/data/run134_5455.root
```

SAM does not work this way

```
sam translate constraints --dim="run 134 subrun 5455 data_tier root"
```

Getting a file on your disk requires more learning

<2> samfs

Solution: Make the SAM behave like a filesystem

<2> samfs

Solution: Make the SAM behave like a filesystem

Technology is FUSE

<2> samfs

Solution: Make the SAM behave like a filesystem

Technology is FUSE

Present **limited parts** of the SAM metadata catalog as a directory tree

<2> samfs

Solution: Make the SAM behave like a filesystem

Technology is FUSE

Present **limited parts** of the SAM metadata catalog as a directory tree

Use standard filesystem idioms to bring files to local disk and use them

<2> samfs

Solution: Make the SAM behave like a filesystem

Technology is FUSE

Present **limited parts** of the SAM metadata catalog as a directory tree

Use standard filesystem idioms to bring files to local disk and use them

Perhaps can hang off of SamWeb

<2> samfs

Solution: Make the SAM behave like a filesystem

Technology is FUSE

Present **limited parts** of the SAM metadata catalog as a directory tree

Use standard filesystem idioms to bring files to local disk and use them

Perhaps can hang off of SamWeb

<crude demo>

Conclusions:

With some improvements, we think we can make SAM more adoptable by IF

Conclusions:

With some improvements, we think we can make SAM more adoptable by IF

Not a huge development project -- mainly integrating what we have with some new tools

Conclusions:

With some improvements, we think we can make SAM more adoptable by IF

Not a huge development project -- mainly integrating what we have with some new tools

Details are still in the works

See <http://cd-docdb.fnal.gov/cgi-bin/ShowDocument?docid=4325>