

# **XALT, OGRT & Related Technologies: Job-Level Usage Data on Today's Supercomputers**

ISC 2017

21.06.2017

Frankfurt, Germany

**What is happening  
on our machine?**

# What would you like to know?

- What software do my users run?
  - Unoptimized Python from the home directory?
  - Binary build of some software?
- Some algebra library had a bug - are my users affected?
  - Which jobs used that library?
- Is the environment "sane"?
  - Are there problematic environment variables set?
- What is killing my parallel filesystem?
- Can I reduce the number of applications?
- How does the next machine look like?

# History

- In 2015 XALT needed a launcher
- HPC is not Life Sciences/Bioinformatics is not HPC
  - lots of user built software
  - that interacts with the provided software
- Wanted to capture everything
- This lead to a different design with similar capabilities

# OGRT tracks

program execution

shared libraries

environment

loaded modules

**How?**

# LD\_PRELOAD

The loader "preloads" a shared object when loading a dynamic executable.

...combine with a "constructor":

**No Launcher/Wrapper**

# Design Principles

- as little computation as possible
- fail silently and without side effects
- configuration not changeable after compile
- no runtime dependencies



# Watermarking (Signature)

- Link in an object file at compile time
- Creates a note section in ELF (GCC does this too)
  - gets loaded into memory on execution
    - is read **out of memory** by OGRT
  - embeds UUID
  - can be read by readelf/OGRT

# Signature Example

```
georg.rath@willikins:~/devel/ogrt/demo$ readelf -n sw/R/R
```

```
Displaying notes found at file offset 0x00000254 with length 0x0000003c:
```

Owner	Data size	Description
OGRT	0x00000026	Unknown note type: (0x4f475254)

```
Displaying notes found at file offset 0x00000290 with length 0x00000020:
```

Owner	Data size	Description
GNU	0x00000010	NT_GNU_ABI_TAG (ABI version tag)
OS: Linux, ABI: 2.6.24		

```
Displaying notes found at file offset 0x000002b0 with length 0x00000024:
```

Owner	Data size	Description
GNU	0x00000014	NT_GNU_BUILD_ID (unique build ID bitstring)
Build ID: bf5f01b3f828588c901c32c997787e0064b8f8b9		

```
georg.rath@willikins:~/devel/ogrt/demo$ ogrt --show-signature sw/R/R
```

```
OGRT signature in section .note.ogrt.info
```

```
allocatable:  yes
version:      1
name:         OGRT
uuid:         123eb26a-6a79-49cd-bfc1-db9f0a9791e6
georg.rath@willikins:~/devel/ogrt/demo$
```

# Configuration

- support for sending environment
  - some/all variables
- support for sending loaded modules
- send *every* execution
  - or blacklist /usr, /bin, etc.
- configurable at compile-time

# The Transport



# The Server

- written in Golang
- receive and decode
- forward to configurable outputs
  - aggregation
  - batching
- buffers data
- embedded web-server for metrics

**The big picture...**

libogrt.so

R

**Server**

Receive & Decode

Output  
Elastic

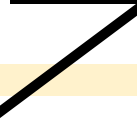
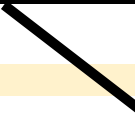
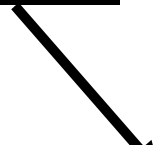
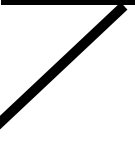
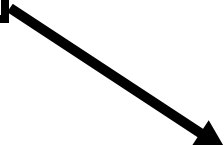
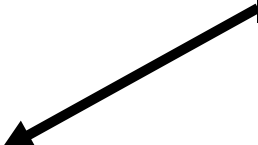
Output  
Debug

Worker 1

Worker 2

Worker 1

Elastic



# The Data

```
{
  "binpath": "/usr/bin/gcc-4.8",
  "cmdline": "/usr/bin/gcc --help",
  "hostname": "ogrtest",
  "job_id": "TESTJOB",
  "loaded_modules": [
    {
      "name": "gcc/4.6.2"
    }
  ],
  "parent_pid": 3342,
  "pid": 12089,
  "shared_objects": [
    {
      "path": "/tmp/install/lib/libogrt.so",
      "signature": "708e1ffd-4ced-45d3-81f3-52e059ea3128"
    },
    {
      "path": "/lib/x86_64-linux-gnu/libc-2.19.so"
    },
    {
      "path": "/lib/x86_64-linux-gnu/ld-2.19.so"
    },
    {
      "path": "/lib/x86_64-linux-gnu/libnss_files-2.19.so"
    }
  ],
  "time": 1475062870,
  "username": "georg.rath"
}
```



# ElasticSearch

- access to the whole ecosystem
- analytics with Kibana
- combine data with Grafana
  - job metrics and applications in a single dashboard
- stability and performance

# Example Query

## All jobs with broken library

```
{
  "size": 0,
  "query": {
    "nested": {
      "path": "shared_objects",
      "query": {
        "match": {
          "shared_objects.signature": "033c3dcc-e6ad-43df-9316-734b1b7e6693"
        }
      }
    }
  },
  "aggs": {
    "job_ids": {
      "terms": {
        "field": "job_id",
        "size": "100"
      }
    }
  }
}
```

# Example Query

## All execs of awk w/o locale

Story time!

```
{
  "size": 0,
  "query": {
    "bool" : {
      "must" :
        { "term" : { "binpath" : "/usr/bin/gawk" } },
      "must_not" :
        { "term" : { "environment_variables" : "LC_ALL=C" } }
    }
  },
  "aggs": {
    "job_ids": {
      "terms": { "field": "job_id", "size": "100" }
    }
  }
}
```

# Deployment in a Slide

## Server

```
wget -q https://github.com/georg-rath/ogrt/releases/download/v0.5.0/ogrt-server-v0.5.0.tar.bz2
tar xf ogrt-server-v0.5.0.tar.bz2
cd ogrt-server-v0.5.0
./ogrt-server
```

## Client

```
git clone https://github.com/georg-rath/ogrt.git
cd ogrt/client
./vendorize
./configure --prefix=/tmp/ogrt
make install
LD_PRELOAD=$(find /tmp/ogrt/ -name libogrt.so) OGRT_ACTIVE=1 bash
# every command you run in the spawned bash gets sent to the server
ls
```

# Discussion

<https://github.com/Fahey-McLay/xalt>

<https://github.com/georg-rath/ogrt> - [georg@rath.io](mailto:georg@rath.io)