OAK RIDGE National Laboratory

Real-Time GPU Telemetry Collection and Posting with Authentication

Daniel Moon

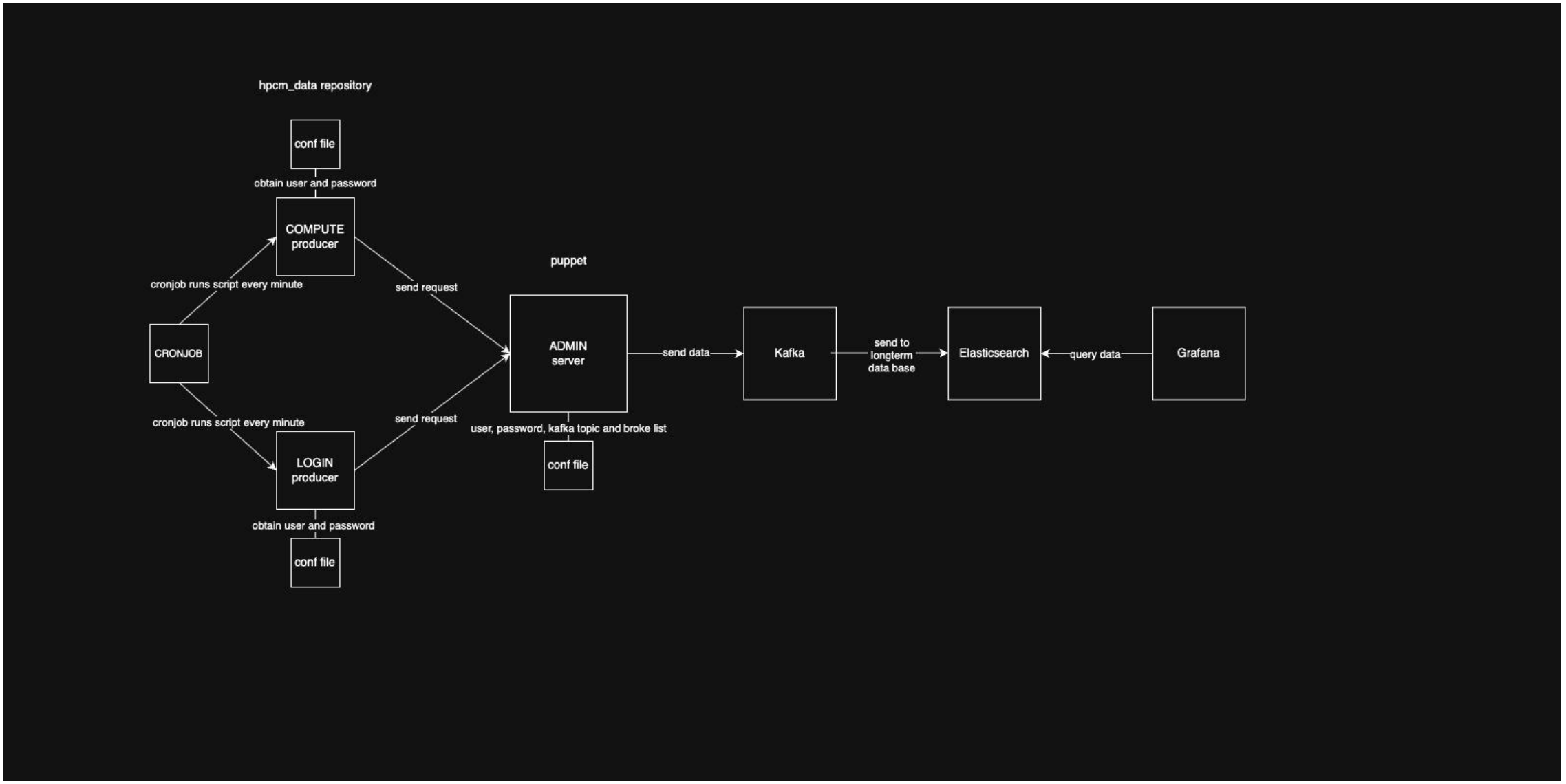
ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Purpose

- Monitor and log GPU metrics on HPC systems
- Designed for internal use on the ARCH cluster
  - NVIDIA GH200 144G HBM3e
- Enables system insights during SLURM job execution

Data Pipeline Overview



Use Cases

- Real-time GPU load monitoring for SLURM jobs
- Feeding data into logging or alerting systems
- Used on Grafana for visualization
- Logs important information
  - Memory utilization, power draws, temperatures, etc.

Expanding the Framework

- Framework supports modular metric types
- Extend: add CPU, RAM queries in producer
  - Minimal changes needed to server/client structure

```
41 # ---------- Run nvidia-smi and capture GPU metrics ----------
42 result = subprocess.run([
43     "nvidia-smi",
44     "--query-gpu=index,name,timestamp,memory.total,memory.used,memory.free,driver_version,vbios_version,g
   ks.mem,memory.reserved",
45     "--format=csv,noheader,nounits"
```

```
59 # ---------- Parse each GPU entry ----------
60 for i, line in enumerate(result.stdout.strip().split("\n")):
61     fields = [f.strip() for f in line.split(",")]
62     (
63         index, name, timestamp, mem_total, mem_used, mem_free,
64         driver, vbios_version, gpu_uuid, gpu_serial, gpu_temp,
65         mem_temp, mem_util, gpu_util, pci_id, module_id,
66         power_limit, power_draw, pstate, gpu_clock, mem_clock,
67         mem_reserved
68     ) = fields
69
70     gpu_info.update({
71         "name_" + index: name,
72         "memory_total_MB_" + index: int(mem_total),
73         "memory_used_MB_" + index: int(mem_used),
74         "memory_free_MB_" + index: int(mem_free),
75         "driver_version_" + index: driver,
76         "vbios_version_" + index: vbios_version,
77         "gpu_uuid_" + index: gpu_uuid,
78         "gpu_serial_" + index: gpu_serial,
79         "gpu_temp_" + index: int(gpu_temp),
80         "mem_temp_" + index: int(mem_temp),
81         "mem_utilization_" + index: int(mem_util),
82         "gpu_utilization_" + index: int(gpu_util),
83         "pci_id_" + index: pci_id,
84         "module_id_" + index: module_id,
85         "power_limit_" + index: float(power_limit),
86         "power_draw_" + index: float(power_draw),
87         "pstate_" + index: pstate,
88         "gpu_clock_" + index: int(gpu_clock),
89         "mem_clock_" + index: int(mem_clock),
90         "mem_reserved_" + index: int(mem_reserved)
91     })
92
```

What I've Learned:

- nvidia-smi
- Deployment Process of a System
- Daily struggles of a System Administrator
- OSI Model
- Different Repositories
- RPM build
- DHCP, PXE booting
- VM (Virtual Box)

Questions