Hewlett Packard
Enterprise

CLUSTERSTOR MONITORING API

Apr 2024

Dan Matthews, Tim Morneau

CLUSTERSTOR API OVERVIEW

- Redfish/Swordfish Brief Overview
- CS- (née Neo-, née RFSF-, née DP-) API Deployment and Access Details
- Request Data Flow
- Architectural Support for Extensions
- Resource Tree Representation of the Cluster
  - Discovering Resources
  - Cluster Component Resource Representation
- API EventService
  - Generic Event Subscriptions
  - MetricReport Event Subscriptions
  - MetricReport Data Flow Example
  - Client Side EventDestination “Receivers”
- API Client Side Libraries
- References

REDFISH/SWORDFISH BRIEF OVERVIEW

- Redfish (DTMF):
  - Standard and application programming interface (API) designed to deliver simple and secure management for converged, hybrid IT, and the Software Defined Data Center (SDDC)
  - Provides RESTful interface semantics to access schema-based data model to conduct management operations
  - Suitable for a wide range of devices, from stand-alone servers, to composable infrastructures, and to large-scale cloud environments.
  - Redfish Spec: https://www.dmtf.org/sites/default/files/standards/documents/DSP0266_1.11.1.pdf
  - Redfish Schemas: https://redfish.dmtf.org/schemas/v1/

- Swordfish (SNIA):
  - Extension to Redfish Scalable Platforms Management API
  - Defines comprehensive, RESTful API for storage management that addresses:
    - Block storage
    - File systems
    - Object storage
    - Storage network infrastructure
  - Swordfish Spec: https://www.snia.org/sites/default/files/technical-work/swordfish/release/v1.2.3/html/Specification/Swordfish_v1.2.3_Specification.html
  - Swordfish Schemas: https://redfish.dmtf.org/schemas/swordfish/

CS API DEPLOYMENT AND ACCESS DETAILS - CURRENT

- API deployed on mgmt nodes as set of systemd services managed by HA as an active/standby group
  - Active on mgmt=primary node, standby on mgmt=secondary node
  - Follows failover / failback of md64-group

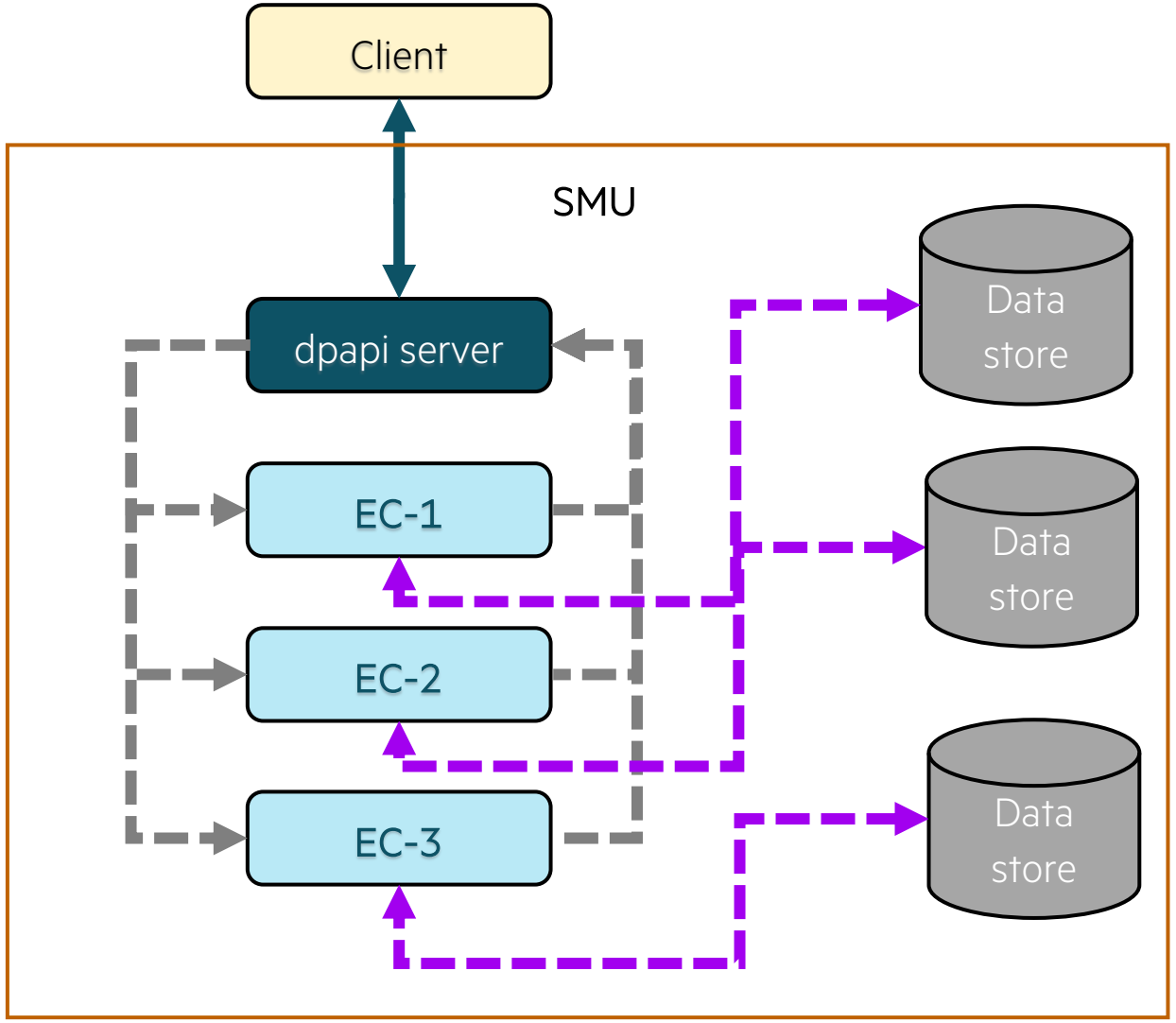
```
  * Resource Group: grp-rfsf-api:
    * prm-dp-api        (systemd:dp-api):        Started kjlmo1200
    * prm-hw-ec (systemd:hw-ec):         Started kjlmo1200
    * prm-dm-mgmt-ec    (systemd:dm-mgmt-ec):    Started kjlmo1200
    * prm-fs-mgmt-ec    (systemd:fs-mgmt-ec):    Started kjlmo1200
    * prm-raid-mgmt-ec  (systemd:raid-mgmt-ec):  Started kjlmo1200
    * prm-monitoring-ec (systemd:monitoring-ec):         Started kjlmo1200
```

- Accessible over HTTPS via port :8081 across the public, internal mgmt, and localhost networks
  - Use same self signed cert/key that other CS components behind TLS use
  - Users may upload their own CA signed certs if they choose
- Accessible for valid ClusterStor admins users that have been granted an access token
  - Token granted via a POST request to /redfish/v1/SessionService/Sessions to create a Session
  - POST request body contains user credentials
  - Credentials validated against local cluster LDAP instance
  - If accepted, response header contains valid auth token (JWT)
  - Token passed in subsequent request header in order to access other areas of resource tree

REQUEST DATA FLOW - CURRENT

- The below is a step by step overview of the flow of requests
  - EC = Element Controller which is responsible for one or more Redfish/Swordfish resources (usually on branch of the resource tree or a subset of a branch) including:
    – GETs, PUTs, POSTs, Events, Actions etc.

1. Client sends HTTPS RESTful request to dp-api-server on mgmt node over https
2. API routes request to appropriate EC over gRPC
3. EC fetches data from appropriate backend data source
4. EC returns response to dp-api-server over gRPC
5. dp-api-server returns response to client over HTTPS



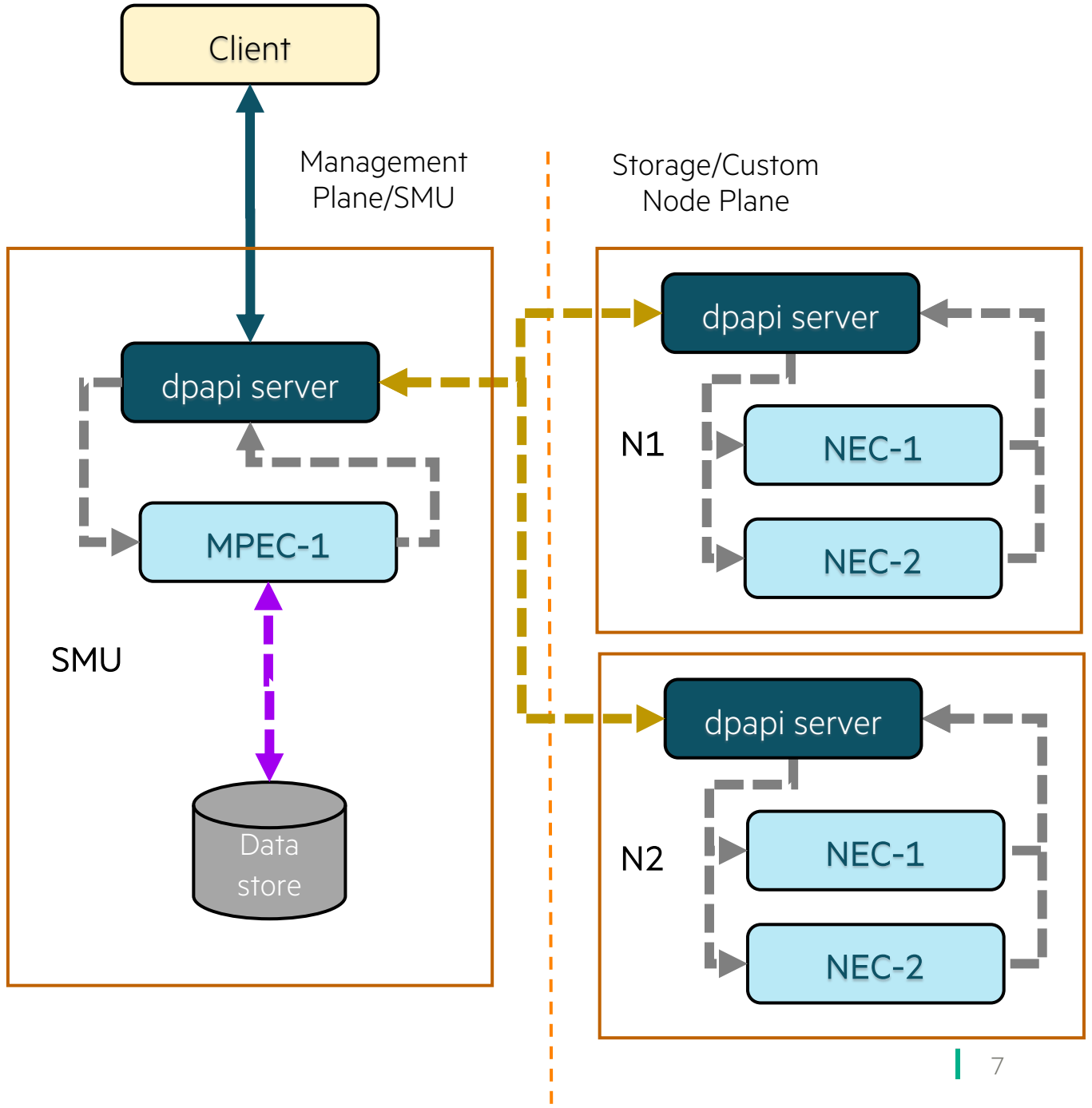
CS API DEPLOYMENT - FUTURE

- Dp-api-node instances exist on the nodes to handle requests coming in from mgmt node
  - Design is symmetric i.e. management node vs storage/custom node
  - Goes through exact same ReST router with different EC backend
- Element Controllers moving out to the nodes to provide node local resource reporting and control
  - Element Controllers on ClusterStor have up till now have been resident on the management plane
  - Remove hard to control data collection components to be replace by EC's
- Element controllers and backend data sources can be swapped in and out at node or mgmt level without changing interface / entry point
  - API doesn't change

REQUEST DATA FLOW - FUTURE

- The below is a step by step overview of the flow of requests
  - MPEC = Management Plane Element Controller
  - NEC = Node Element Controller
    Note: an EC is an EC is an EC – the added prefixes are just for clarity

1. Client sends RESTful request to dp-api-server In the Management Plane
2. API routes request to appropriate EC over gRPC or routes the HTTPS request to the Node dpapi server which routes the request to its local EC
3. EC fetches data from appropriate backend data source
4. EC returns response to dp-api-server over gRPC which will either be returned directly to the client or pass back to the Management dpapi server for return to the client



ARCHITECTURAL SUPPORT FOR EXTENSIONS

- API is designed in such a way that it is easily extensible
- API front end interface remains the same
- Backend ECs and data collection sources are modular
- API endpoints and routes are known and remain the same (Resource Tree)
- Data models are known and remain the same

Future

- Extend API instances onto Storage Nodes (in process)
- EC dynamic insertion (remove fixed gRPC ports)
  - The facilitates easy addition after the fact of new/updated EC's
- Proxying to BMC/iLO to provide single point of data delivery

RESOURCE TREE REPRESENTATION OF THE CLUSTER

- All Cluster Components and Collections represented as a Redfish/Swordfish Resource
  - Known Schema based data model (json format)
  - All Resources are uniquely identifiable
- All resources within the “Resource Tree” are discoverable from the ServiceRoot: /redfish/v1
- ServiceRoot provides path to “top-level” **Resource** collections:
  - /redfish/v1/Chassis
  - /redfish/v1/StorageSystems
  - /redfish/v1/StorageServices
  - /redfish/v1/Events
- ServiceRoot provides path to ”top-level” **Service** resources and their collections:
  - /redfish/v1/SessionService
    – /redfish/v1/SessionService/Sessions
  - /redfish/v1/EventService
    – /redfish/v1/EventService/Subscriptions
  - /redfish/v1/UpdateService
    – /redfish/v1/UpdateService/SoftwareInventory

RESOURCE TREE REPRESENTATION OF THE CLUSTER (CONT)

- Each resource in a collection may provide sub-collections
- Examples:
  - /redfish/v1/StorageSystems/{ComputerSystemId}/NetworkInterfaces
  - /redfish/v1/StorageServices/{StorageServiceId}/FileSystems
  - /redfish/v1/Chassis/{ChassisId}/Thermal/Fans
- More info of full resource tree available in documentation

EVENTING

- Redfish enables clients to receive messages outside of the normal request / response paradigm
- The EventService uses these messages, or events, to asynchronously notify the client of a state change or error condition
- Push-style Eventing
  - When the service detects the need to send an event, it calls HTTP POST to push the event message to the client.
  - Clients can enable reception of events by creating a subscription entry in the Event Service
- Two “categories” of Event subscriptions implemented in NEO RFSF API:
  - “Generic” Events
    - Similar to alerts
    - Sent asynchronously as they occur
    - Occur on an API resources
    - Can filter on specific resources
    - Subscribers are unique
  - MetricReport Events (telemetry and statistic-based events)
    - Sent at client-defined frequencies
    - Subscribe to specific data sets (defined in EventService)
    - Subscribers start or join a data stream HTTPS

GENERIC EVENTS

- Events that occur on API resources
- Examples:
  - status change of a FRU
  - Start or stop of lustre targets
  - Failovers
  - Etc.
- A subscription to receive Events is created via a POST request to /redfish/v1/EventService/Subscriptions/{EventDestinationId}
- Can specify “filtering” information in EventDestination resource request body
  - Filter on ResourceTypes
  - Filter on OriginResources
- Runtime Event history maintained in /redfish/v1/Events collection
  - Lookup by EventID
  - “Rotated” after a certain max number of events is reached
  - Lost on service restart or failover
  - Events meant to be stored off box if required
- More details provided in documentation

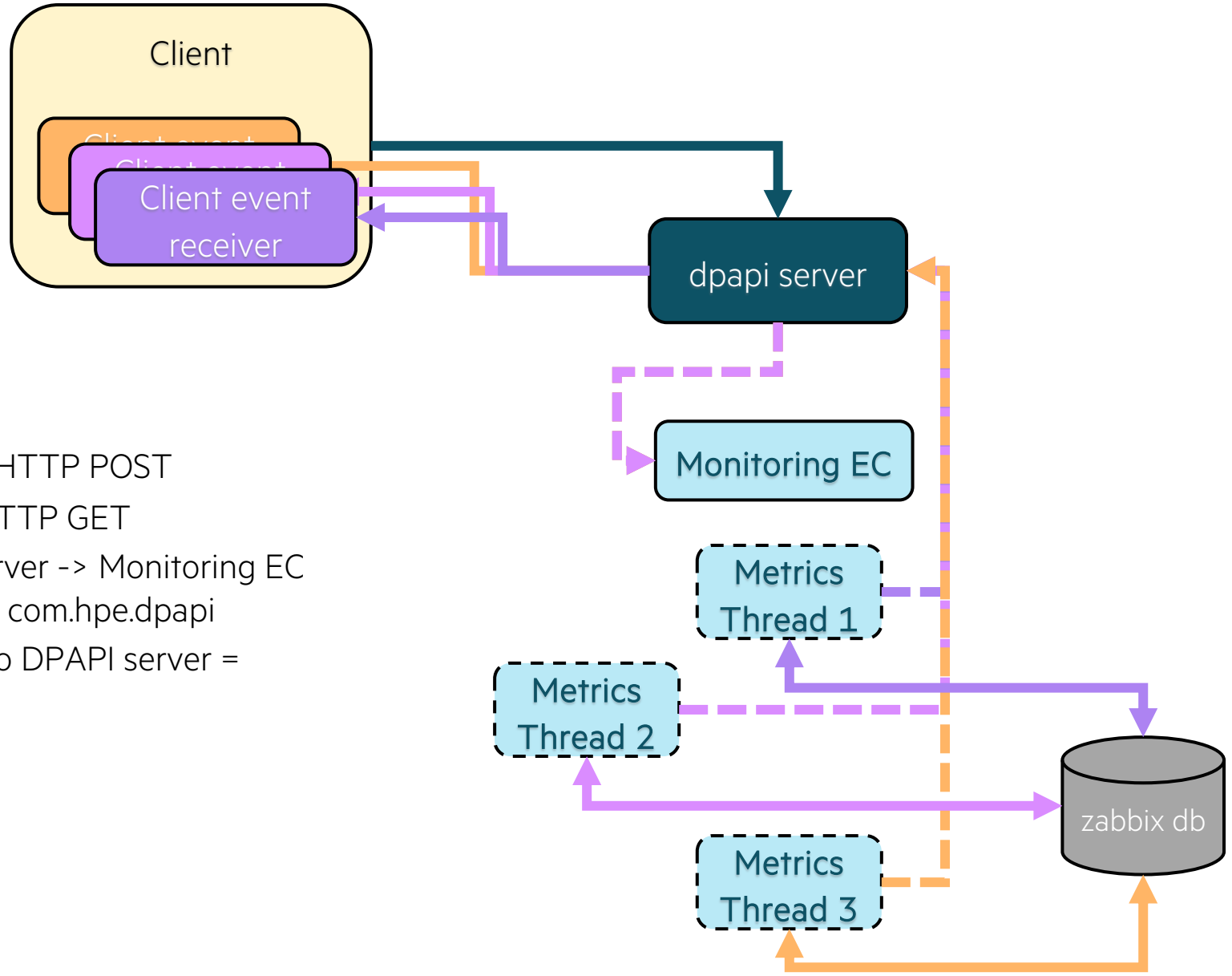
METRICREPORT EVENTS

- Special case of EventService notifications to receive metrics and telemetry data
- A subscription to receive Events is created via a POST request to /redfish/v1/EventService/Subscriptions/{EventDestinationId}
  - Must contain EventTypes[“MetricReport”] in request body
  - Must contain MetricReportDefinitions in request body
    – Specifies data_set, send_interval, etc.
- Register to receive statistics for a certain data set
  - LustreStats
  - LinuxStats
  - JobStats
  - EnvironmentalStats
  - TieringStats
- Register to receive data from that data set at a specified interval
- Initial subscription for a data set “creates” the data stream at defined interval
- Subsequent subscriptions for a data set “join” the data stream

CLIENT SIDE EVENTDESTINATION “RECEIVERS”

- A subscription to receive Events is created via a POST request to /redfish/v1/EventService/Subscriptions/{EventDestinationId}
- Request body passed is an EventDestination Resource
- The EventDestination.Destination field of the resource specifies a client side endpoint for API EventService to receive Events on
  - Events are sent from API via HTTP POST
  - Client side Event “Receiver” must implement an HTTP POST request handler
- EventService defines parameters to control retries to send Events to a subscriber who is not listening

METRICREPORT DATA FLOW EXAMPLE



Connectors Key:

- Solid Uni-directional = HTTP POST
- Solid Bi-directional = HTTP GET
- Dashed from DPAPI server -> Monitoring EC = dbus message across com.hpe.dpapi
- Dashed from EC back to DPAPI server = gRPC POST request

API CLIENT SIDE LIBRARIES

- Provided by rfsf-api-client package
  - Repo: https://github.hpe.com/hpe/hpc-sp-rfsf-api-client
- Provides libraries and config files to:
  - Connect to API
  - Authenticate a user
  - Subscribe to Receive Events
  - Subscriber to Receive MetricReports:
    – LustreStats, LinuxStats, JobStats
  - Discover full API resource tree
- Provides monitoring examples to:
  - Iterate through API resource tree
  - Start threaded client Event Receiver
    – Parse incoming Event messages
    – Send GET request to resource Event occurred on
- Provides example client side event receiver code

WHAT'S NEXT?!

- API instances out to all cluster nodes
- ClusterStor GUI and CLI moving to use API on backend everywhere
- Custom Metrics/Data Collection (PoC completed but hasn't been exposed)
- Composability – provision/teardown of mid to high level resource constructs
- Firmware / Software updates via API

REFERENCES

- Redfish Spec: https://www.dmtf.org/sites/default/files/standards/documents/DSP0266_1.11.1.pdf
- Redfish Schemas: https://redfish.dmtf.org/schemas/v1/
- Swordfish Spec: https://www.snia.org/sites/default/files/technical-work/swordfish/release/v1.2.3/html/Specification/Swordfish_v1.2.3_Specification.html
- Swordfish Schemas: https://redfish.dmtf.org/schemas/swordfish/
- Neo Redfish/Swordfish REST API documentation: https://hpe.sharepoint.com/:w:/r/teams/hpc_storage/clusterstorteam/_layouts/15/Doc.aspx?sourcedoc=%7B488c0c22-30ca-4389-afd4-2f3ef4159c40%7D&action=edit&wdPid=534d6646&cid=23406062-e3b4-428b-a60f-45a843dc7fd6

FIN

Thank-you!

Default (ssh)

```
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]# curl -Ssk https://kjlmo17n000.hpc.amslabs.hpecorp.net:8081/redfish/v1 | jq
{
  "@odata.context": "ServiceRoot.ServiceRoot",
  "@odata.id": "/redfish/v1",
  "@odata.type": "ServiceRoot.v1_10_0.ServiceRoot",
  "Chassis": {
    "@odata.id": "/redfish/v1/Chassis"
  },
  "EventService": {
    "@odata.id": "/redfish/v1/EventService"
  },
  "Id": "ServiceRoot",
  "Links": {},
  "Name": "NEO RedfishSwordfish API ServiceRoot",
  "Oem": {
    "Events": {
      "@odata.id": "/redfish/v1/Events"
    }
  },
  "ProtocolFeaturesSupported": {
    "DeepOperations": {}
  },
  "SessionService": {
    "@odata.id": "/redfish/v1/SessionService"
  },
  "StorageServices": {
    "@odata.id": "/redfish/v1/StorageServices"
  },
  "StorageSystems": {
    "@odata.id": "/redfish/v1/StorageSystems"
  },
  "TelemetryService": {
    "@odata.id": "/redfish/v1/TelemetryService"
  },
  "UUID": "cf5d0f9f-b946-4b77-8975-7994423a8031",
  "UpdateService": {
    "@odata.id": "/redfish/v1/UpdateService"
  }
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]# curl -Ssk https://kjlmo17n000.hpc.amslabs.hpecorp.net:8081/redfish/v1/StorageServic
es | jq
{
  "error": {
    "code": "401",
    "message": "No Auth Token found in request header. User is NOT AUTHORIZED to use the API."
  }
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]#
```

Default (ssh)

```
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]# curl -Ssk -H "X-Auth-Token: $AUTH" https://kjlmo17n000.hpc.amslabs.hpecorp.net:8081
/redfish/v1/StorageServices | jq
{
  "@odata.context": "StorageServiceCollection.StorageServiceCollection",
  "@odata.id": "/redfish/v1/StorageServices",
  "@odata.type": "StorageServiceCollection.v1_0_0.StorageServiceCollection",
  "Description": "Collection of references StorageService Resources.",
  "Members": [
    {
      "enclosure_id: 2": "/redfish/v1/StorageServices/MXE30000122VS009"
    },
    {
      "enclosure_id: 3": "/redfish/v1/StorageServices/MXE3000018NVS007"
    },
    {
      "enclosure_id: 1": "/redfish/v1/StorageServices/MXE3000018VVS001"
    },
    {
      "enclosure_id: 112": "/redfish/v1/StorageServices/MXQ3310CTY"
    },
    {
      "enclosure_id: 108": "/redfish/v1/StorageServices/MXQ3310CV4"
    },
    {
      "enclosure_id: 111": "/redfish/v1/StorageServices/MXQ3310CV6"
    },
    {
      "enclosure_id: 109": "/redfish/v1/StorageServices/MXQ3310CV7"
    },
    {
      "enclosure_id: 4": "/redfish/v1/StorageServices/SGFGDZ150648D94"
    },
    {
      "enclosure_id: 5": "/redfish/v1/StorageServices/SGFGDZ15064F171"
    },
    {
      "enclosure_id: 7": "/redfish/v1/StorageServices/SGFGDZ151648779"
    },
    {
      "enclosure_id: 6": "/redfish/v1/StorageServices/SGFGDZ21764EF63"
    },
    {
      "Lustre Filesystem": "/redfish/v1/StorageServices/lustre-kjlmo17"
    }
  ],
  "Members@odata.count": 12,
  "Name": "StorageServiceCollection"
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]#
```

Default (ssh)

```
      "Bay 23": "/redfish/v1/StorageServices/MXE30000122VS009/StoragePools/MXE30000122VS009/CapacitySources/MXE30000122VS009/ProvidingDrives/S4YPNG0R904671"
    },
    {
      "Bay 24": "/redfish/v1/StorageServices/MXE30000122VS009/StoragePools/MXE30000122VS009/CapacitySources/MXE30000122VS009/ProvidingDrives/S4YPNG0R904672"
    }
  ],
  "Members@odata.count": 24,
  "Name": "MXE30000122VS009 ProvidingDrives"
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]# curl -Ssk -H "X-Auth-Token: $AUTH" https://kjlmo17n000.hpc.amslabs.hpecorp.net:8081/redfish/v1/StorageServices/MXE30000122VS009/StoragePools/MXE30000122VS009/CapacitySources/MXE30000122VS009/ProvidingDrives/S4YPNG0R904738 | jq
{
  "@odata.context": "Drive.Drive",
  "@odata.id": "/redfish/v1/StorageServices/MXE30000122VS009/StoragePools/MXE30000122VS009/CapacitySources/MXE30000122VS009/ProvidingDrives/S4YPNG0R904738",
  "@odata.type": "Drive.v1_12_1.Drive",
  "Id": "S4YPNG0R904738",
  "Links": {},
  "Manufacturer": "SAMSUNG",
  "MediaType": "NVME",
  "Model": "MZWLJ3T8HBLS-00007",
  "Name": "Drive S4YPNG0R904738 in bay 12",
  "Oem": {
    "SmartData": {
      "SMART [nvme24]: Critical warning": "0",
      "SMART [nvme24]: Device model": "SAMSUNG MZWLJ3T8HBLS-00007",
      "SMART [nvme24]: Exit status": "0",
      "SMART [nvme24]: Media errors": "0",
      "SMART [nvme24]: Percentage used": "0%",
      "SMART [nvme24]: Power on hours": "15999",
      "SMART [nvme24]: Power_Cycle_Count": "5",
      "SMART [nvme24]: Serial number": "S4YPNG0R904738",
      "SMART [nvme24]: Smartctl error": "",
      "SMART [nvme24]: Temperature": "30 °C"
    }
  },
  "PhysicalLocation": {
    "Info": "Bay 12",
    "InfoFormat": "Bay",
    "PartLocation": {
      "LocationOrdinalValue": 12,
      "LocationType": "Bay",
      "ServiceLabel": "Bay 12"
    },
    "Placement": {},
    "PostalAddress": {}
  },
  "Protocol": "NVME",
  "SerialNumber": "S4YPNG0R904738",
  "Status": {
    "State": "Ok"
  }
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]#
```

Default (ssh)

```
  "Members": [
    {
      "@odata.id": "/redfish/v1/StorageServices/lustre-kjlmo17/FileSystems/kjlmo17"
    }
  ],
  "MembersOdataCount": 1,
  "Name": "kjlmo17 Lustre FileSystemCollection"
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]# curl -Ssk -H "X-Auth-Token: $AUTH" https://kjlmo17n000.hpc.amslabs.hpecorp.net:8081
/redfish/v1/StorageServices/lustre-kjlmo17/FileSystems/kjlmo17 | jq
{
  "OdataContext": "FileSystem.FileSystem",
  "OdataId": "/redfish/v1/StorageServices/lustre-kjlmo17/FileSystems/kjlmo17",
  "OdataType": "FileSystem.v1_2_2.FileSystem",
  "Description": "FileSystem resource to represent top-level Lustre FileSystem kjlmo17",
  "Id": "kjlmo17",
  "Links": {
    "Oem": {
      "kjlmo17-MDT0000": "/redfish/v1/StorageServices/MXE30000122VS009/FileSystems/kjlmo17-MDT0000",
      "kjlmo17-MDT0001": "/redfish/v1/StorageServices/MXE30000122VS009/FileSystems/kjlmo17-MDT0001",
      "kjlmo17-OST0000": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0000",
      "kjlmo17-OST0001": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0001",
      "kjlmo17-OST0002": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0002",
      "kjlmo17-OST0003": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0003",
      "kjlmo17-OST0004": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0004",
      "kjlmo17-OST0005": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0005",
      "kjlmo17-OST0006": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0006",
      "kjlmo17-OST0007": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0007",
      "kjlmo17-OST0008": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0008",
      "kjlmo17-OST0009": "/redfish/v1/StorageServices/MXE3000018NVS007/FileSystems/kjlmo17-OST0009",
      "nfs": "/redfish/v1/StorageServices/MXQ3310CV4/FileSystems/nfs"
    }
  },
  "Name": "FileSystem kjlmo17",
  "Oem": {
    "FsType": "Lustre",
    "Lustre": {
      "MgsNIDS": "172.20.0.5@tcp,172.20.0.6@tcp:172.20.0.7@tcp,172.20.0.8@tcp",
      "FsName": "kjlmo17",
      "Statistics": {
        "OST Total": "10",
        "Total Available FS Space Percentage": "98.96464392214341",
        "Total FS Space Available": "1406812765126656",
        "Total FS Space": "1421530669309952",
        "Total FS Read": "6423681429",
        "Total FS Space Used": "379007991808",
        "Total FS Write": "3789894447",
        "Avg (120 sec) aggregated all OST(s) read_bytes": "6057633139.416667",
        "Avg (120 sec) aggregated all OST(s) write_bytes": "3664619106.7083335",
        "Max (24 hour) aggregated all OST(s) read_bytes": "7833334559",
        "Max (24 hour) aggregated all OST(s) write_bytes": "5224110487",
        "Total FS MD ops": "67"
      }
    }
  }
}
[root@kjlmo17-oem storage_tech_forum_rfsf_demo]#
```