Supporting Observer Reads from Routers

Author: Simbarashe Dzinamarira
Contributions from: Owen O’Malley

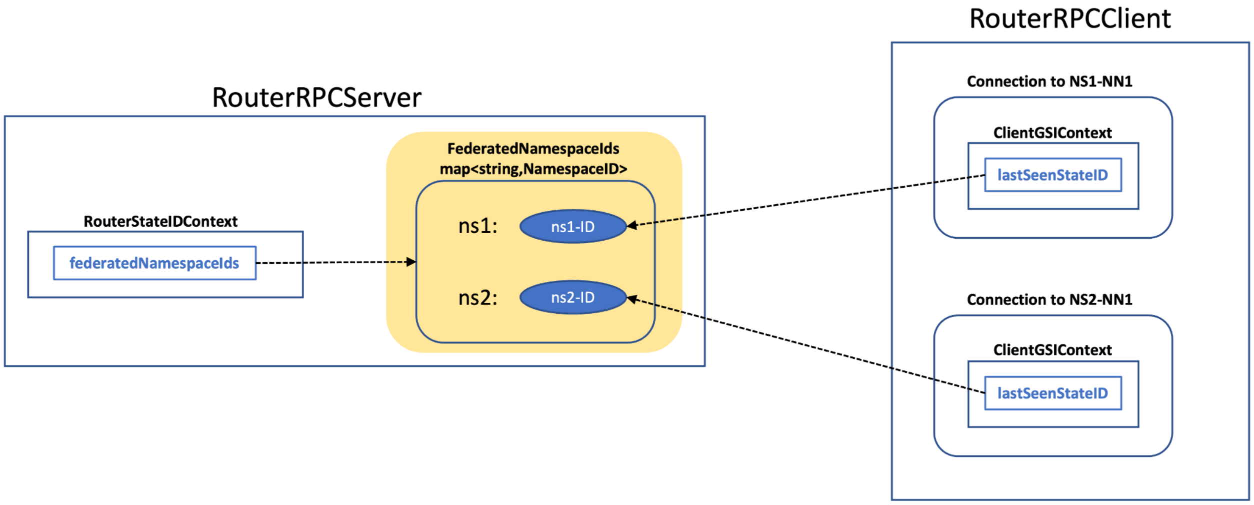
Summary of proposed changes

1. Add a nameservice to stateID mapping, called **nameserviceStateIds,** to the RPCHeader.
2. Add a composite alignment context, called the **FederatedNamespaceIds,** to the router. This contains a map from nameservices to NamespaceId objects specific to each nameservice.
3. Communication between routers and clients
    a. A router uses the **FederatedNamespaceIds** object to update the **nameserviceStateIds** map in the RPCResponseHeader sent to clients.
    b. A router updates the **FederatedNamespaceIds object** with information in the **nameserviceStateIds** map in the RPCRequestHeader.
4. Communication between routers and namenodes.
    a. When communicating with a Namenode, a router uses the a ClientGSIContext linked to a NamespaceId contained within the composite **FederatedNamespaceIds.**
        i. StateID updates received in RPCResponseHeaders from Namenodes are implicitly integrated into the **FederatedNamespaceIds** when applied to the nameservice specific ClientGSIContext.
        ii. Updates to the **FederatedNamespaceIds** will also be implicitly included in the RPCRequestHeaders sent to NameNodes.
5. When a client does an *msync* call to a router, the router fans out this call to all nameservices in order to fully update the **FederatedNamespaceIds.**
6. For old clients which do not have the **nameserviceStateIds** map**,** the router always does an *msync* before each read call so that it obtains the latestSeenStateID for that nameservice.

FederationNamespaceIds

The FederatedNamespaceIds onbject (highlighted in yellow below) stores the last seen stateIDs seen by a router. These stateIDs are received either from clients requests or namenodes responses to the router. Similarly, the map is used to set the RPC head when the router sends RPC requests to namenodes, or sends RPC responses to clients.

- For Router to Client communication, the FederatedNamespaceIds object is used directly by the RouterStateIDContext.
- For Router to Namenode communication, the FederationNamespaceIds object is used indirectly because the ClientGSIContexts in the routers reference elements of the FederationNamespaceIds' map.



Dotted lines represent object references

Implementation breakdown

1. IPC changes: https://github.com/apache/hadoop/pull/4311
   a. Modifies RPCHeader proto
   b. Creates classes to propagate FederatedState between clients and routers
2. Directing router reads to observers: https://github.com/apache/hadoop/pull/4127
   a. Select observed as target for read operations.
   b. For old clients without federated state, performs msync for every call.
   c. Implements router msync that fans across all namespaces.