

Proposal: Addition of GetFlightSchema

Motivation: Currently, Flight's GetFlightInfo returns a set of endpoints and a schema. For certain use cases only the schema is needed and sending endpoints is premature. An example situation is the lazy evaluation paradigm in Spark where the schema is needed to perform the operations but the resulting data set and endpoints are dependent on further operations in Spark and a second call to GetFlightInfo.

We are proposing to add a new RPC endpoint called GetFlightSchema to Flight which would return the Schema only for a particular dataset.

Alternatives:

- Call GetFlightInfo twice and with a custom bytes cmd to differentiate the purpose of each call
- Add an argument to GetFlightInfo to tell it its being called only for the schema
- Use DoAction and wrap the expected FlightInfo in a Result

Design:

We propose adding a SchemaResult message and new RPC call as follows:

```
message SchemaResult {
  bytes schema = 1;
}

/*
 * For a given FlightDescriptor, get the Schema as described in
 Schema.fbs::Schema
 * This is used when a consumer needs the Schema of flight stream. Similar to
 * GetFlightInfo this interface may generate a new flight that was not
 previously
 * available in ListFlights.
 */
rpc GetSchema(FlightDescriptor) returns (SchemaResult) {}
```

The addition of the new protobuf message and rpc call should not drastically change current usage of Flight and will enable more flexible usage of the protocol. The GetFlightSchema rpc call can be delegated to GetFlightInfo by default, thereby minimising server changes.

One remaining question is if we should leave the Schema object on GetFlightInfo. Since the schema could still change between a GetFlightSchema and GetFlightInfo and not all users will want to make separate calls we believe it is best to not change the signature of GetFlightInfo.