

Example HTTP Tablespace

Tajo provides an example HTTP tablespace. This tablespace allows for Tajo to directly read and process JSON data on the web. This section briefly shows how the example tablespace is implemented. The full code can be found in the [tajo-tablespace-example](#) module.

Code

ExampleHttpFileFragment

ExampleHttpFileFragment is very simple and same with FileFragment.

ExampleHttpFileFragment

```
public class ExampleHttpFileFragment extends AbstractFileFragment {

    public ExampleHttpFileFragment(Uri uri,
                                   String inputSourceId,
                                   long startKey,
                                   long endKey) {
        super("EXAMPLE-HTTP", uri, inputSourceId, startKey, endKey, endKey - startKey, null);
    }
}
```

ExampleHttpFileFragmentSerde

ExampleHttpFileFragmentSerde simply serializes / deserializes ExampleHttpFileFragment into / from a protocol buffer message. The following code shows the protocol buffer message definition, and Serde class is omitted.

ExampleHttpFileFragmentProto

```
message ExampleHttpFileFragmentProto {
    required string uri = 1;
    required string table_name = 2;
    required int64 start_key = 3;
    required int64 end_key = 4;
}
```

ExampleHttpFileTablespace

The one of most important methods of ExampleHttpFileTablespace is *getSplits()*. This method generally returns multiple fragments for distributed processing of a large data. However, in this example, it returns a single fragment for simplicity.

ExampleHttpFileTablespace

```
@Override
public List<Fragment> getSplits(String inputSourceId,
                                TableDesc tableDesc,
                                boolean requireSort,
                                @Nullable EvalNode filterCondition)
    throws IOException, TajoException {

    long tableVolume = getTableVolume(tableDesc, Optional.empty());
    return Lists.newArrayList(new ExampleHttpFileFragment(tableDesc.getUri(), inputSourceId, 0, tableVolume));
}
```

ExampleHttpJsonScanner is omitted here because its implementation is almost same with DelimitedTextFileScanner. Appender is not provided for the example HTTP tablespace.

Configuration

The example http tablespace is disabled for HTTP scheme by default because it is not proper for real applications, so it is preserved for user-defined tablespaces. To use this example tablespace, the following lines need to be added to storage-site.json.

storage-site.json

```
"http": {  
  "handler": "org.apache.tajo.storage.http.ExampleHttpFileTablespace",  
  "default-format": "json"  
}
```