

HBase Hive MetaStore Mapping for Nutch 2.x

When you need to map your HBase table which is used by Nutch 2.x, You may use below query in order to map it to Hive. Please fill in <crawlId> tags for your owns. This query can be used for all the sections which use Hive metastore. i.e. Impala

```
CREATE EXTERNAL TABLE <crawlId>_webpage (  
key string, baseUrl string, status int, prevFetchTime bigint, fetchTime bigint, fetchInterval bigint, retriesSinceFetch int, reprUrl string, content string,  
contentType string, protocolStatus string, modifiedTime bigint, prevModifiedTime bigint, batchId string, title string, text string, parseStatus int, signature  
string, prevSignature string, score int, headers map<string,string>, inlinks map<string,string>, outlinks map<string,string>, metadata map<string,string>,  
markers map<string,string>  
) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES (  
"hbase.columns.mapping" = ":key,f:bas,f:st,f:pts#b,f:ts#b,f:fi#b,f:rsf,f:rpr,f:cnt,f:typ,f:prot,f:mod#b,f:pmod#b,f:bid,p:t,p:c,p:st,p:sig,p:psig,s:s,h:,il:,ol:,mtdt:,  
mk:"  
) TBLPROPERTIES (  
"hbase.table.name" = "<crawlId>_webpage"  
);
```