Session 9

Assignment 2

|  |  |
| --- | --- |
| **Prepared For:** | AcadGild |
|  |  |
| **Document Approval:** | **AcadGild** |
|  |  |
|  |  |
|  |  |
|  |  |
| **Project Title:** | Session 9 – Assignment 2 |
|  |  |
| **Prepared By:** | Duncan Burgess |
|  |  |
|  | dburgess@duncb.com |
|  |  |
| **Primary Engineer:** | Duncan Burgess |
|  |  |
| **Document Reference:** | **Session 9 – Assignment 2** |
|  |  |
| **Start Date:** | 25/09/2017 |
|  |  |
|  |  |



# 

# Contents

[Contents 2](#_Toc494105319)

[Change History 3](#_Toc494105320)

[1. Problem Statement 4](#_Toc494105321)

[1.1. NoSQL Databases 5](#_Toc494105322)

[1.2. Types of NoSQL Databases 5](#_Toc494105323)

[1.3. CAP Theorem 6](#_Toc494105324)

[1.4. HBase Architecture 8](#_Toc494105325)

[1.5. HBase vs RDBMS 14](#_Toc494105326)

# Change History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Document Revision** | **Date** | **Authored By** | **Authorised By** | **Sections Affected** | **Reason for Change** |
| Rev 01 | 25/09/2017 | Duncan Burgess |  | All | Initial release. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# Problem Statement

**Problem 1:**

Create an HBase table named 'clicks' with a column family 'hits' such that it should be able to store last 5 values of qualifiers inside 'hits' column family.

**Problem 2:**

Add few records in the table and update some of them. Use IP Address as row-key. Scan the table to view if all the previous versions are getting displayed.

# Solutions

**Setting up the environment**

**\*\*\* Note I performed this on a MAPR cluster as the VM would not create the table.**

*[mapr@maprdemo bin]$ ./start-hbase.sh*

*The authenticity of host 'maprdemo (10.100.100.69)' can't be established.*

*RSA key fingerprint is 6a:24:76:81:7d:53:ab:4d:3e:b5:29:0a:cb:ab:dd:9a.*

*Are you sure you want to continue connecting (yes/no)? yes*

*maprdemo: Warning: Permanently added 'maprdemo,10.100.100.69' (RSA) to the list of known hosts.*

*Password:*

*maprdemo: starting zookeeper, logging to /opt/mapr/hbase/hbase-1.1.8/logs/hbase-mapr-zookeeper-maprdemo.out*

*starting master, logging to /opt/mapr/hbase/hbase-1.1.8/logs/hbase-mapr-master-maprdemo.out*

*OpenJDK 64-Bit Server VM warning: ignoring option PermSize=128m; support was removed in 8.0*

*OpenJDK 64-Bit Server VM warning: ignoring option MaxPermSize=128m; support was removed in 8.0*

*starting regionserver, logging to /opt/mapr/hbase/hbase-1.1.8/logs/hbase-mapr-1-regionserver-maprdemo.out*

*[mapr@maprdemo bin]$ hbase shell*

*HBase Shell; enter 'help<RETURN>' for list of supported commands.*

*Type "exit<RETURN>" to leave the HBase Shell*

*Version 1.1.8-mapr-1703, rb3fd15bea43b4ca29e6ac92afa442ca6114c7635, Mon Apr 3 1 9:31:47 UTC 2017*

*Not all HBase shell commands are applicable to MapR tables.*

*Consult MapR documentation for the list of supported commands.*

*hbase(main):001:0> list*

TABLE

0 row(s) in 0.2520 second

## Create an HBase table named 'clicks' with a column family 'hits' such that it should be able to store last 5 values of qualifiers inside 'hits' column family.

To perform the required table and enable 5 versions the following command was run.

*hbase(main):002:0> create 'clicks', {NAME => 'visit', VERSIONS => 5}*

*0 row(s) in 0.1650 seconds*

*=> Hbase::Table - /tables/clicks*

*hbase(main):003:0> list*

*TABLE*

*clicks*

*1 row(s) in 0.0110 seconds*

*=> ["clicks"]*

*hbase(main):004:0> describe clicks*

*NameError: undefined local variable or method `clicks' for #<Object:0x2c7db926>*

*hbase(main):005:0> describe 'clicks'*

*Table clicks is ENABLED*

*clicks, {TABLE\_ATTRIBUTES => {MAX\_FILESIZE => '4294967296', METADATA => {'AUTOSP*

*LIT' => 'true', 'MAPR\_UUID' => '6536337c-08bb-f3c9-a7bd-0801f3c85900', 'MAX\_VALU*

*E\_SIZE\_IN\_MEM' => '100'}}*

*COLUMN FAMILIES DESCRIPTION*

*{NAME => 'visit', BLOOMFILTER => 'ROW', VERSIONS => '5', IN\_MEMORY => 'false', K*

*EEP\_DELETED\_CELLS => 'FALSE', DATA\_BLOCK\_ENCODING => 'NONE', TTL => 'FOREVER', C*

*OMPRESSION => 'NONE', MIN\_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '8*

*192', REPLICATION\_SCOPE => '0', METADATA => {'compression\_raw' => '31'}}*

*1 row(s) in 0.0590 seconds*

**Populate it with Data**

*hbase(main):006:0> put 'clicks','104.199.201.246','visit:date','04/Jan/2017'*

*0 row(s) in 0.0940 seconds*

*hbase(main):008:0> put 'clicks','104.199.201.246','visit:site','http://www.duncb.com/'*

*0 row(s) in 0.0030 seconds*

*hbase(main):010:0> put 'clicks','104.199.201.246','visit:browser','Mozilla/5.0'*

*0 row(s) in 0.0070 seconds*

*hbase(main):012:0> put 'clicks','104.199.201.246','visit:request','put'*

*0 row(s) in 0.0040 seconds*

*hbase(main):016:0> put 'clicks','104.199.201.246','visit:transport','HTTP/1.1'*

*0 row(s) in 0.0080 seconds*

*hbase(main):020:0> put 'clicks','77.88.47.100','visit:date','06/Jan/2017'*

*0 row(s) in 0.0020 seconds*

*hbase(main):022:0> put 'clicks','77.88.47.100','visit:site','https://www.githuub.com'*

*0 row(s) in 0.0030 seconds*

*hbase(main):024:0> put 'clicks','77.88.47.100','visit:browser','IE8'*

*0 row(s) in 0.0050 seconds*

*hbase(main):026:0> put 'clicks','77.88.47.100','visit:request','put'*

*0 row(s) in 0.0030 seconds*

*hbase(main):028:0> put 'clicks','77.88.47.100','visit:transport','HTTP/1.1'*

*0 row(s) in 0.0030 seconds*

**Check results**

*hbase(main):031:0\* scan 'clicks'*

*ROW COLUMN+CELL*

*104.199.201.246 column=visit:browser, timestamp=1506341770021, value=Mozil*

*la/5.0*

*104.199.201.246 column=visit:date, timestamp=1506341769783, value=04/Jan/2*

*017*

*104.199.201.246 column=visit:request, timestamp=1506341770088, value=put*

*104.199.201.246 column=visit:site, timestamp=1506341769884, value=http://w*

*ww.duncb.com/*

*104.199.201.246 column=visit:transport, timestamp=1506341770186, value=HTT*

*P/1.1*

*77.88.47.100 column=visit:browser, timestamp=1506341770386, value=IE8*

*77.88.47.100 column=visit:date, timestamp=1506341770261, value=06/Jan/2*

*017*

*77.88.47.100 column=visit:request, timestamp=1506341770438, value=put*

*77.88.47.100 column=visit:site, timestamp=1506341770301, value=https://*

*www.githuub.com*

*77.88.47.100 column=visit:transport, timestamp=1506341770481, value=HTT*

*P/1.1*

*2 row(s) in 0.0650 seconds*

*hbase(main):032:0> get 'clicks','77.88.47.100'*

*COLUMN CELL*

*visit:browser timestamp=1506341770386, value=IE8*

*visit:date timestamp=1506341770261, value=06/Jan/2017*

*visit:request timestamp=1506341770438, value=put*

*visit:site timestamp=1506341770301, value=https://www.githuub.com*

*visit:transport timestamp=1506341770481, value=HTTP/1.1*

*5 row(s) in 0.0530 second*

## Add few records in the table and update some of them. Use IP Address as row-key. Scan the table to view if all the previous versions are getting displayed.

**To test this I modified visit:request 5 times**

hbase(main):003:0> put 'clicks','77.88.47.100','visit:request','if'

0 row(s) in 0.0360 seconds

hbase(main):005:0\* put 'clicks','77.88.47.100','visit:request','what'

0 row(s) in 0.0130 seconds

hbase(main):006:0> put 'clicks','77.88.47.100','visit:request','when'

0 row(s) in 0.0080 seconds

hbase(main):008:0\* put 'clicks','77.88.47.100','visit:request','where'

0 row(s) in 0.0130 seconds

hbase(main):009:0> put 'clicks','77.88.47.100','visit:request','who'

0 row(s) in 0.0080 seconds

**To check Results**

*hbase(main):025:0> scan 'clicks',{COLUMN=>'visit:request',VERSIONS=>5}*

*ROW COLUMN+CELL*

*104.199.201.246 column=visit:request, timestamp=1506341770088, value=put*

*77.88.47.100 column=visit:request, timestamp=1506342962291, value=where*

*77.88.47.100 column=visit:request, timestamp=1506342564305, value=who*

*77.88.47.100 column=visit:request, timestamp=1506342545437, value=where*

*77.88.47.100 column=visit:request, timestamp=1506342531004, value=when*

*77.88.47.100 column=visit:request, timestamp=1506342520884, value=what*

*2 row(s) in 0.0220 seconds*

**\*\*\* Note** for 77.88.47.100 the value for visit:request does not show put thus we retain 5 versions.