



Department of Computer Science and Engineering (Data Science)

Subject: Big Data Engineering (DJ19DSL604)

AY: 2022-23

Experiment 7

(No SQL Data Store)

**Name: Suyash S Konduskar**

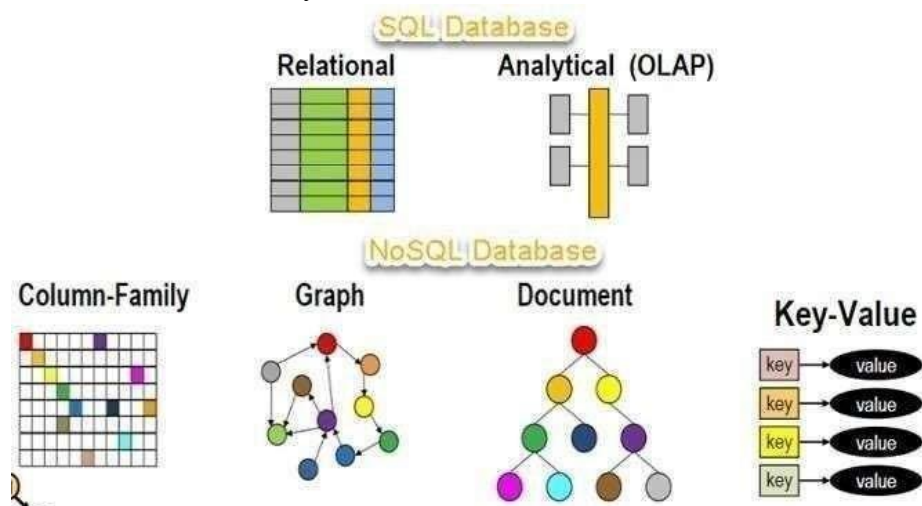
**Sap id: 60009220109**

Aim: Implement No SQL Data Store using HBase.

Theory:

NoSQL:

NoSQL Database is a non-relational Data Management System, that does not require a fixed schema. It avoids joins, and is easy to scale. The major purpose of using a NoSQL database is for distributed data stores with humongous data storage needs. NoSQL is used for Big data and realtime web apps. For example, companies like Twitter, Facebook and Google collect terabytes of user data every single day. NoSQL database stands for “Not Only SQL” or “Not SQL.”



Department of Computer Science and Engineering (Data Science) Difference between SQL and NoSQL data stores:

- SQL databases are relational, and NoSQL databases are non-relational.
- SQL databases use structured query language (SQL) and have a predefined schema. NoSQL



Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)

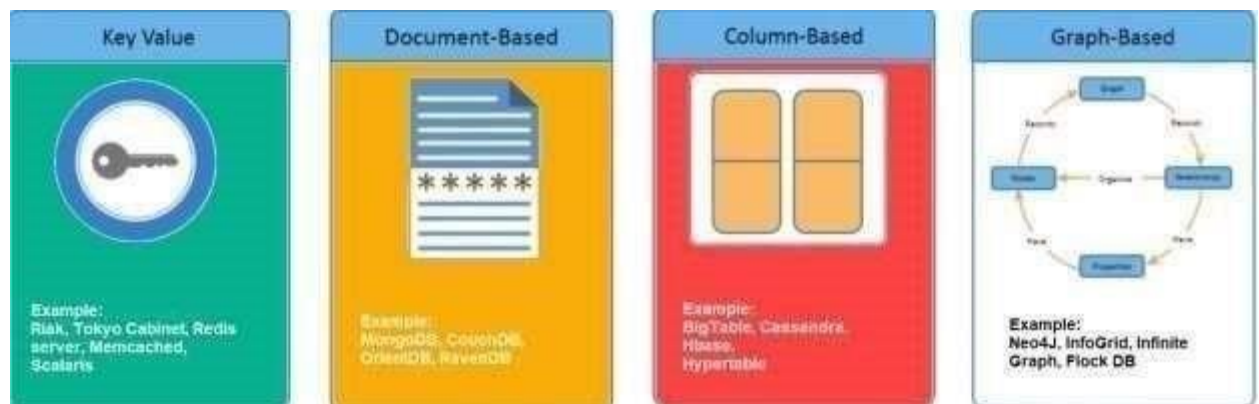


- databases have dynamic schemas for unstructured data.
- SQL databases are vertically scalable, while NoSQL databases are horizontally scalable.
- SQL databases are table-based, while NoSQL databases are document, key-value, graph, or wide-column stores.
- SQL databases are better for multi-row transactions, while NoSQL is better for unstructured data like documents or JSON.

### Types of NoSQL Databases

NoSQL Databases are mainly categorized into four types: Key-value pair, Column-oriented, Graph-based and Document-oriented. Every category has its unique attributes and limitations. None of the above-specified database is better to solve all the problems. Users should select the database based on their product needs. Types of NoSQL Databases:

- Key-value Pair Based
- Column-oriented Graph
- Graphs based
- Document-oriented



### Introduction to HBase

HBase is a distributed column-oriented database built on top of the Hadoop file system. It is an open-source project and is horizontally scalable.

Department of Computer Science and Engineering (Data Science)

HBase is a data model that is similar to Google's big table designed to provide quick random access to huge amounts of structured data. It leverages the fault tolerance provided by the Hadoop File System (HDFS).

It is a part of the Hadoop ecosystem that provides random real-time read/write access to data in the Hadoop File System.



Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

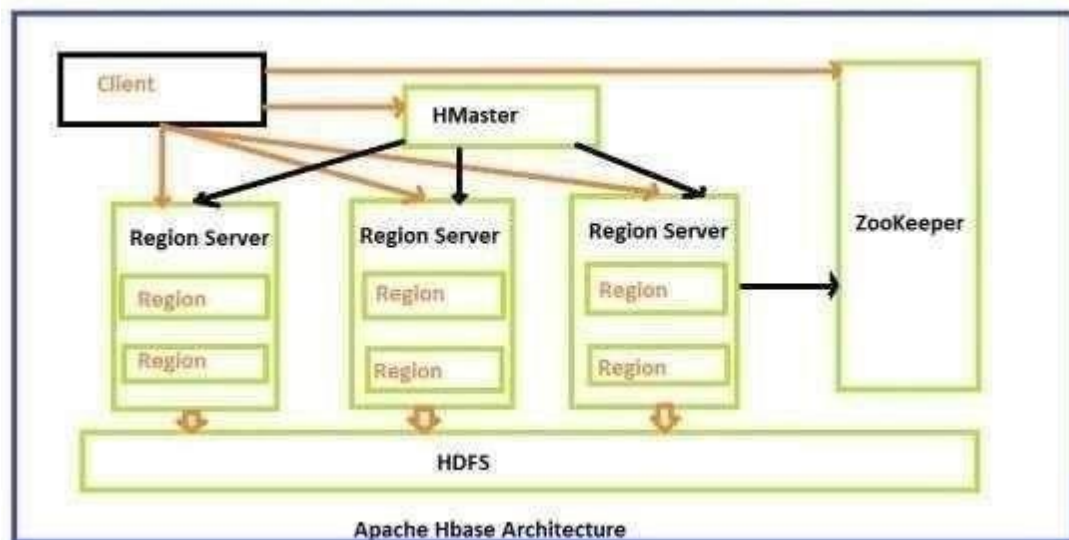
NAAC Accredited with "A" Grade (CGPA : 3.18)



One can store the data in HDFS either directly or through HBase. Data consumer reads/accesses the data in HDFS randomly using HBase. HBase sits on top of the Hadoop File System and provides read and write access.

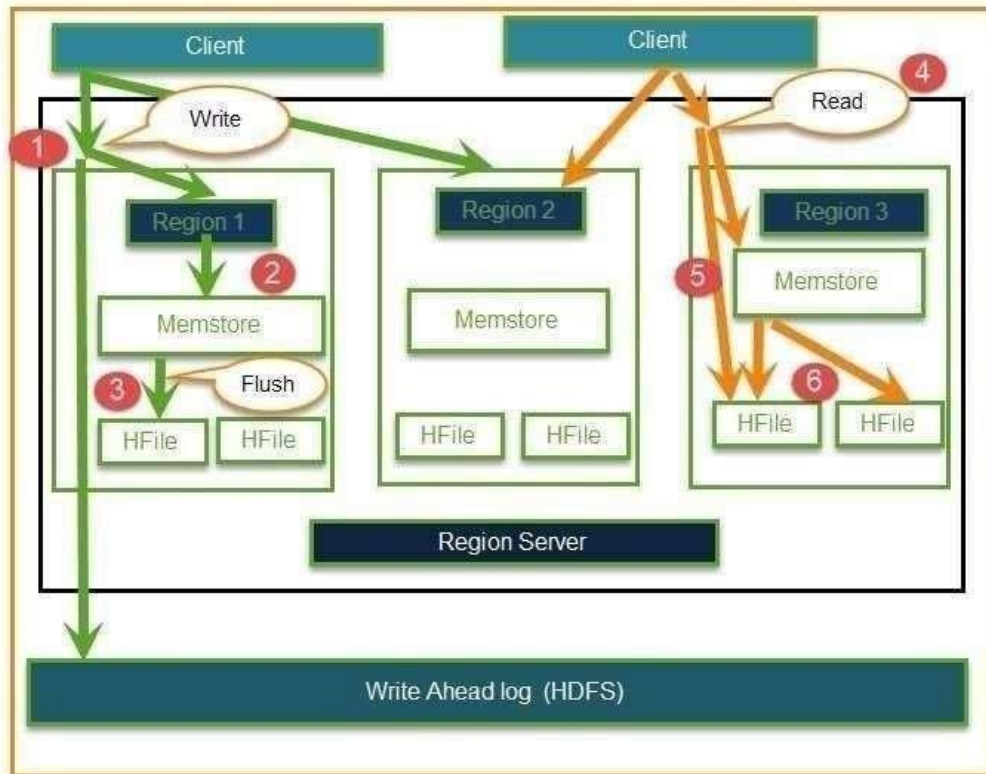
Department of Computer Science and Engineering (Data Science)

## HBASE Architecture



Department of Computer Science and Engineering (Data Science)

## HBase Read and Write Data



Department of Computer Science and Engineering (Data Science)

### Lab Assignment:

1. Installation of HBase on standalone mode.
2. Implementation of HBase Create Table with Java API & Shell.
3. Implement HBase Shell Commands and dynamic scaling:
  - a. General commands
  - b. Tables Managements commands
  - c. Data manipulation commands
  - d. Cluster Replication Commands

[illegible]



Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

The screenshot shows a laptop screen with a web browser displaying the 'HBase Releases' page. The page contains information about downloading HBase releases, including a table of versions and their release dates. Below the browser window, a terminal window is open, showing the user's shell prompt and commands being executed.

**HBase Releases**

Please make sure you're downloading from a [trusted mirror site](#), not from [www.apache.org](#)

We support downloading the current [stable release](#).

The 2.5.x series is the current stable release line, it supersedes earlier release lines.

If you are a user of HBase 1.x and can not yet upgrade to the stable release line, we suggest downloading the latest 1.x minor release.

From 0.9.4:

- 0.9.4 was EOL'd September 2013; 0.9.6 and 0.9.8 were EOL'd April 2017
- 1.0 was EOL'd January 2016; 1.1 was EOL'd December 2017; 1.2 was EOL'd June 2019; 1.3 was EOL'd in August 2020; 1.4 was EOL'd in October 2021; 1.5 was EOL'd in August 2022
- 2.0 was EOL'd September 2019; 2.1 was EOL'd May 2019; 2.2 was EOL'd April 2021; 2.3 was EOL'd Oct 2021

For older versions, check the [historical archive of release artifacts](#).

File	Download Date	Size	Download
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-
hbase-2.5.8-bin.tar.gz	2023-08-08 09:26	-	-

cs-ds@mum0923cpu0981: ~/Downloads/hbase-2.5.8-bin/hbase-2.5.8/conf

```
(base) cs-ds@mum0923cpu0981:~/Downloads/hbase-2.5.8-bin/hbase-2.5.8/conf$ gedit hbase-env.sh
sh
^C
(base) cs-ds@mum0923cpu0981:~/Downloads/hbase-2.5.8-bin/hbase-2.5.8/conf$ gedit ~/.bashrc
```





Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
Open  ~/Downloads/hbase-2.5.8-bin/hbase-2.5.8/conf  Save  *hbase-site.xml
*bashrc  *hbase-site.xml
20 */
21 -->
22 <configuration>
23 <!--
24   The following properties are set for running HBase as a single process on a
25   developer workstation. With this configuration, HBase is running in
26   "stand-alone" mode and without a distributed file system. In this mode, and
27   without further configuration, HBase and ZooKeeper data are stored on the
28   local filesystem, in a path under the value configured for 'hbase.tmp.dir'.
29   This value is overridden from its default value of '/tmp' because many
30   systems clean '/tmp' on a regular basis. Instead, it points to a path within
31   this HBase installation directory.
32
33   Running against the 'LocalFileSystem', as opposed to a distributed
34   filesystem, runs the risk of data integrity issues and data loss. Normally
35   HBase will refuse to run in such an environment. Setting
36   'hbase.unsafe.stream.capability.enforce' to 'false' overrides this behavior,
37   permitting operation. This configuration is for the developer workstation
38   only and __should not be used in production!__
39
40   See also https://hbase.apache.org/book.html#standalone\_dist
41 -->
42 <property>
43   <name>hbase.cluster.distributed</name>
44   <value>false</value>
45 </property>
46 <property>
47   <name>hbase.tmp.dir</name>
48   <value>./tmp</value>
49 </property>
50 <property>
51   <name>hbase.unsafe.stream.capability.enforce</name>
52   <value>false</value>
53 </property>
54 </configuration>
55
56
57 <property>
58
59 <name>hbase.rootdir</name>
60
61 <value>file:///home/hduser/HBASE/hbase</value>
62
63 </property>
64
65 <property>
66
67 <name>hbase.zookeeper.property.dataDir</name>
68
69 <value>/home/hduser/HBASE/zookeeper</value>
70
71 </property>
```

XML Tab Width: 8 Ln 71, Col 12 INS



Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)

```
Open  +  *.bashrc  Save

100 # You may want to put all your additions into a separate file like
101 # ~/.bash_aliases, instead of adding them here directly.
102 # See /usr/share/doc/bash-doc/examples in the bash-doc package.
103
104 if [ -f ~/.bash_aliases ]; then
105     . ~/.bash_aliases
106 fi
107
108 # enable programmable completion features (you don't need to enable
109 # this, if it's already enabled in /etc/bash.bashrc and /etc/profile
110 # sources /etc/bash.bashrc).
111 if ! shopt -oq posix; then
112     if [ -f /usr/share/bash-completion/bash_completion ]; then
113         . /usr/share/bash-completion/bash_completion
114     elif [ -f /etc/bash_completion ]; then
115         . /etc/bash_completion
116     fi
117 fi
118
119 export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
120 export HADOOP_HOME=/usr/local/hadoop
121 export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
122 export PDSH_RCMD_TYPE=ssh
123 source <(kubectl completion bash)
124
125 export NVM_DIR="$HOME/.nvm"
126 [ -s "$NVM_DIR/nvm.sh" ] && \. "$NVM_DIR/nvm.sh" # This loads nvm
127 [ -s "$NVM_DIR/bash_completion" ] && \. "$NVM_DIR/bash_completion" # This load
    bash_completion
128
129 # >>> conda initialize >>>
130 # !! Contents within this block are managed by 'conda init' !!
131 __conda_setup="$('/home/cs-ds/miniconda3/bin/conda' 'shell.bash' 'hook' 2> /dev
132 if [ $? -eq 0 ]; then
133     eval "$__conda_setup"
134 else
135     if [ -f "/home/cs-ds/miniconda3/etc/profile.d/conda.sh" ]; then
136         . "/home/cs-ds/miniconda3/etc/profile.d/conda.sh"
137     else
138         export PATH="/home/cs-ds/miniconda3/bin:$PATH"
139     fi
140 fi
141 # Set HIVE_HOME
142
143 export HIVE_HOME="/home/Downloads/apache-hive-3.1.2-bin"
144 export PATH=$PATH:$HIVE_HOME/bin
145
146 export HBASE_HOME="/home/Downloads/hbase-2.5.8-bin/hbase-2.5.8
147 export PATH=$PATH:$HBASE_HOME/bin
148
149 unset __conda_setup
150 # <<< conda initialize <<<
```





Shri Vile Parle Kelavani Mandal's

**DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**

(Autonomous College Affiliated to the University of Mumbai)

NAAC Accredited with "A" Grade (CGPA : 3.18)



```
cs-ds@mum0923cpu0981: ~/Downloads/hbase-2.5.8-bin/hbas...
(base) cs-ds@mum0923cpu0981:~/Downloads/hbase-2.5.8-bin/hbase-2.5.8/conf$ gedit
hbase-env.sh
```

```
Open  ~/Downloads/hbase-2.5.8-bin/hbase-2.5.8/conf  Save
1 #!/usr/bin/env bash
2 #
3 #/**
4 # * Licensed to the Apache Software Foundation (ASF) under one
5 # * or more contributor license agreements. See the NOTICE file
6 # * distributed with this work for additional information
7 # * regarding copyright ownership. The ASF licenses this file
8 # * to you under the Apache License, Version 2.0 (the
9 # * "License"); you may not use this file except in compliance
10 # * with the License. You may obtain a copy of the License at
11 # *
12 # * http://www.apache.org/licenses/LICENSE-2.0
13 # *
14 # * Unless required by applicable law or agreed to in writing, software
15 # * distributed under the License is distributed on an "AS IS" BASIS,
16 # * WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
17 # * See the License for the specific language governing permissions and
18 # * limitations under the License.
19 # */
20
21 # Set environment variables here.
22
23 # This script sets variables multiple times over the course of starting an hbase process,
24 # so try to keep things idempotent unless you want to take an even deeper look
25 # into the startup scripts (bin/hbase, etc.)
26
27 # The java implementation to use. Java 1.8+ required.
28 # export JAVA_HOME=/usr/java/jdk1.8.0/
29 export JAVA_HOME=/usr/lib/jvm/java-17-openjdk-amd64/jre
30
31 # Extra Java CLASSPATH elements. Optional.
32 # export HBASE_CLASSPATH=
33
34 # The maximum amount of heap to use. Default is left to JVM default.
35 # export HBASE_HEAPSIZE=1G
36
37 # Uncomment below if you intend to use off heap cache. For example, to allocate 8G of
38 # offheap, set the value to "8G".
39 # export HBASE_OFFHEAPSIZE=1G
40
41 # Extra Java runtime options.
42 # Default settings are applied according to the detected JVM version. Override these default
43 # settings by specifying a value here. For more details on possible settings,
44 # see http://hbase.apache.org/book.html#\_jvm\_tuning
45 # export HBASE_OPTS
46
```