

# **PROOF OF CONCEPT**

**BIG DATA AND ENGINEERING**

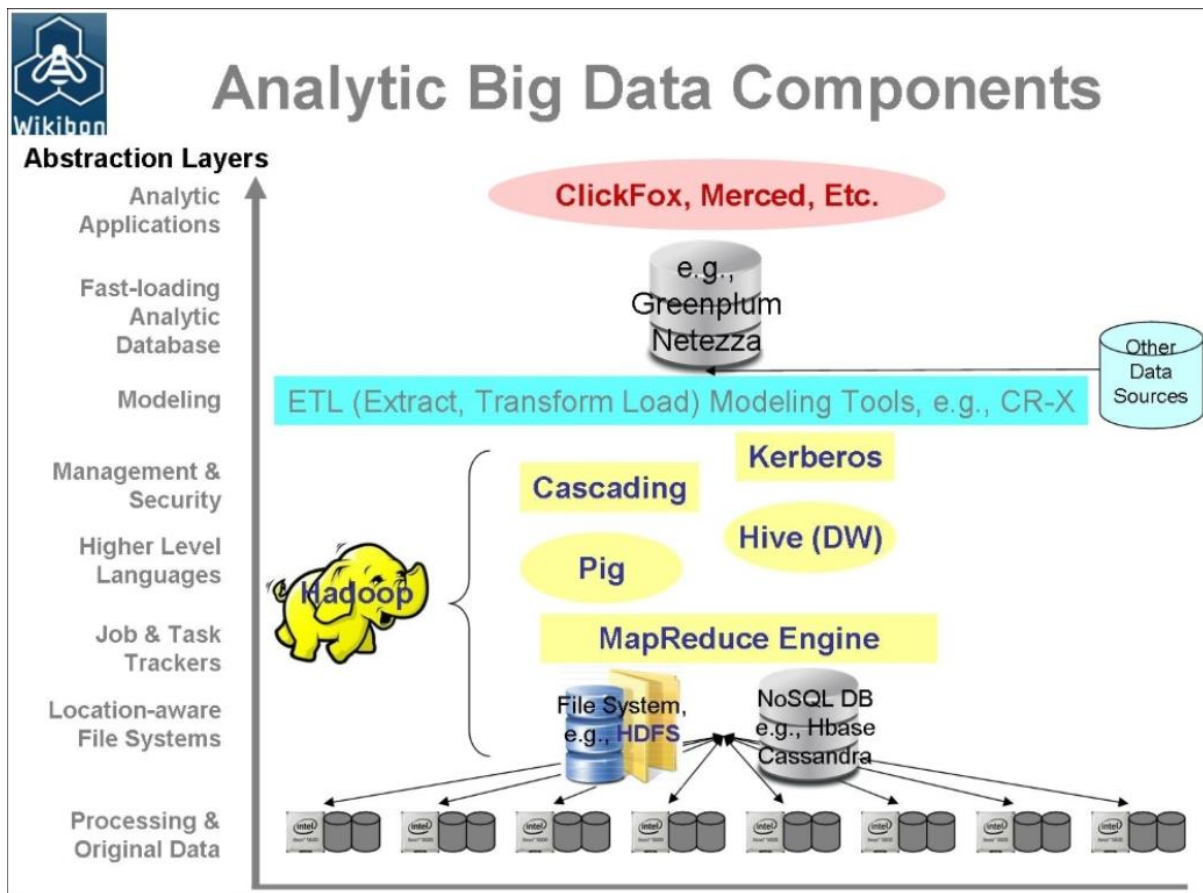
**GOOD BOOKS ANALYSIS**



**SUBMITTED TO: - Gurpreet Ma'am**

**SUBMITTED BY:-Harshit**

# INTRODUCTION TO BIG DATA TECHNOLOGIES



**Hadoop:** Hadoop is an open source software programming framework for storing a large amount of data and performing the computation. Its framework is based on Java programming with some native code in C and shell scripts.

**Sqoop:** It is used to import and export data to and from between HDFS and RDBMS.

**Pig:** It is a procedural language platform used to develop a script for Map Reduce operations.

**Hive:** It is a platform used to develop SQL type scripts to do Map Reduce operations.

**Map reduce:** It is a parallel programming model for processing large amounts of structured, semi-structured, and unstructured data on large clusters of commodity hardware.

## **COLUMN IN DATASET**

### **(Books.csv)**

- **Book ID** – ID no. of the book
- **Book Title** – name of the book
- **Author Name** - Name of the book writer
- **Rating** - average rating of the book
- **Language** – Book written in which language
- **Pages** - Number of pages in the book.
- **Rating count** – total count of the rating

## **PROBLEM STATEMENT**

### **HADOOP PROBLEM STATEMENT**

- TO MAKE THE DIRECTORY
- TO DISPLAY THE LIST OF DIRECTORIES
- LOADING FILES IN THE DIRECTORY
- CHECK FOR THE FILE EITHER LOADED IN OUR DATABASES OR NOT
- RETRIEVE CONTENT OF LOADED FILE
- CREATE THE FILE
- DISPLAY THE FILE
- TO SHOW THW HIDDEN FILES
- TO DISPLAY ALL THE PRESENT WORKING DIRECTORIES
- TO REMOVE THE DIRECTORY

## **HIVE PROBLEM STATEMENTS**

1. SHOW THE BOOK WHICH NAME START WITH “T”
2. SHOW THE BOOK OF A SPECIFIC WRITER
3. SHOW THE BOOK OF RATING ABOVE 3.5
4. SHOW THE BOOK HAVING RATING COUNT ABOVE 1lac.
5. DISPLAY THE BOOK WHICH IS IN ENG-US
6. PRINT ALL THE BOOKS OF ALL RATING.

## **PIG PROBLEM STATEMENTS**

1. LOADING DATA
2. DUMPING DATA
3. PRINT ALL THE RATING GREATER THAN 4
4. FIND OUT THE MAXIMUM AND MINIMUM RATING.

## **COMPARISON STUDY IN TIME OF EXECUTION**

1. DIFFERENCE BETWEEN THE EXECUTION TIME OF HIVE AND THE PIG.  
COMPARE THE PERFORMANCE OF HIVE AND PIG BY FINDING THE BOOK TITLE WITH RATING GREATER THAN 4.
2. COMPARISION BETWEEN MAP REDUCE AND PIG FOR FINDING WORD COUNT OF THE DOCUMENT.

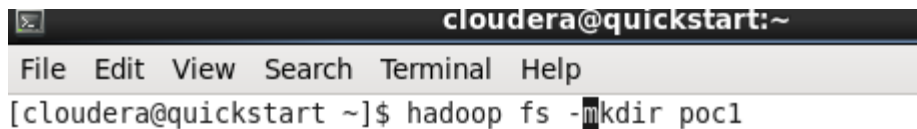
## **TOOLS USE IN PROOF OF CONCEPT (POC)**

1. Hadoop (HDFS Commands)
2. Hive
3. Pig
4. Java
5. Map Reduce

# HADOOP COMMANDS

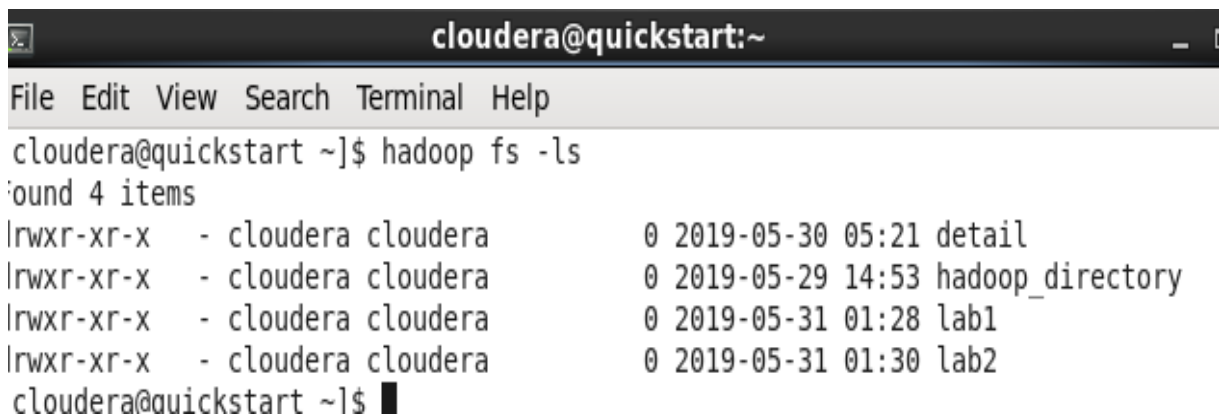
## HDFS COMMANDS

- TO MAKE THE DIRECTORY



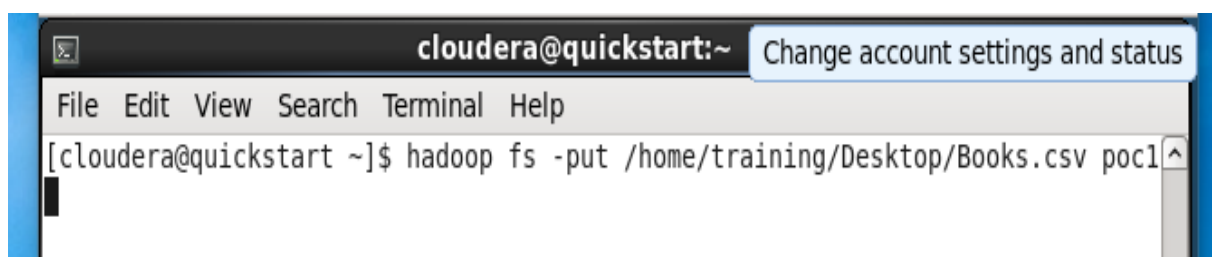
```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -mkdir poc1
```

- TO DISPLAY THE LIST OF THE DIRECTORIES



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
cloudera@quickstart ~]$ hadoop fs -ls  
Found 4 items  
lrwxr-xr-x  - cloudera cloudera          0 2019-05-30 05:21 detail  
lrwxr-xr-x  - cloudera cloudera          0 2019-05-29 14:53 hadoop_directory  
lrwxr-xr-x  - cloudera cloudera          0 2019-05-31 01:28 lab1  
lrwxr-xr-x  - cloudera cloudera          0 2019-05-31 01:30 lab2  
cloudera@quickstart ~]$
```

- LOADING THE FILE IN THE DIRECTORY



```
cloudera@quickstart:~ Change account settings and status  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -put /home/training/Desktop/Books.csv poc1
```

- CHECK FOR FILE EIHTER LOADED IN OUR DATABASE OR NOT

```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -ls /user/training/poc1
Found 1 items
-rw-r--r-- 1 training supergroup 2123441 2019-07-05 05:10 /user/training
/poc1/ Books.csv
```

- RETRIEVE THE CONTENT OF LOA DED FILE

```
22911,Amerika,Franz Kafka,3.75,805210644,9.78081E+12,eng,336,17134,497,
22912,Collections of Franz Kafka,Willi Muir-Edwin Muir-Gabriel Jospovic,4.35,679423836,9.78068E+12,eng,566,432,28,
22913,Kafka: Toward a Minor Literature,Gilles Deleuze-Félix Guattari-Franz Kafka,4.18,816615152,9.78082E+12,eng,136,1237,44,
22914,The Annotated Brothers Grimm,Jacob Grimm-Wilhelm Grimm-Maria Tatar-A.S. Byatt-George Cruikshank-Paul Hey-Walter Crane-Warwick Goble-Kay Nielsen-Arthur Rackham,4.46,393058484,9.78039E+12,eng,462,3838,180,
22915,The Complete Brothers Grimm Fairy Tales,Jacob Grimm-Wilhelm Grimm,4.31,517229250,9.78052E+12,eng,680,4398,165,
22916,The Complete Fairy Tales of The Brothers Grimm,Jacob Grimm-Wilhelm Grimm-Jack D. Zipes-Johnny Gruelle,4.31,553382160,9.78055E+12,eng,762,863,55,
22931,Beyond Innocence (Beyond Duet #1),Enna Holly,3.62,515130990,9.78052E+12,eng,US,295,1405,65,
22932,Enna Vol. 01,Kaoru Mori-森 薫,3.83,1401211321,9.7814E+12,eng,183,3414,331,
22934,Enna Vol. 03,Kaoru Mori-森 薫,4.11,1401211348,9.7814E+12,eng,187,1522,77,
22938,The Black Company: Glittering Stone II (The Chronicles of the Black Company #8-9),Glen Cook,4.31,739413007,9.78074E+12,eng,886,54,3,
22944,Drum Into Silence (Drums of Chaos #3),Jo Clayton-Kevin Andrew Murphy,3.36,812551249,9.78081E+12,eng,GB,399,1,0,
22945,Dance Down the Stars (Duel of Sorcery: Dancer #3),Jo Clayton,3.73,886776171,9.78089E+12,eng,368,33,0,
22947,Drum Calls (Drums of Chaos #2),Jo Clayton,3.68,812551230,9.78081E+12,eng,341,28,1,
22948,Fire in the Sky (Diaden: Shadowson #1),Jo Clayton,3.87,886776503,9.78089E+12,eng,352,61,2,
22950,Quester's Endgame,Jo Clayton,3.97,886771382,9.78089E+12,eng,372,153,6,
22951,Shadowkill (Diaden: Shadith's Quest #3),Jo Clayton-Jo Clayton,3.98,886774675,9.78089E+12,eng,400,64,1,
22966,Lidia's Family Table,Lidia Matticchio Bastianich-David Nussbaum-Christopher Hirsheimer,4.22,1400040353,9.7814E+12,eng,448,521,23,
22967,Lidia's Italy,Lidia Matticchio Bastianich-Tanya Bastianich Manuali-David Nussbaum-Christopher Hirsheimer,4.21,1400040361,9.7814E+12,eng,US,384,619,24,
22969,La Cucina Di Lidia: Recipes and Memories from Italy's Adriatic Coast,Lidia Matticchio Bastianich-Jay Jacobs,4.1,767914228,9.78077E+12,eng,GB,288,45,3,
22991,Blood Wedding and Yerma,Federico Garcia Lorca-W.S. Merwin-Langston Hughes,3.83,1559368001,9.78156E+12,eng,US,160,842,31,
22992,The Selected Poems,Federico Garcia Lorca-Francisco Garcia Lorca-Donald H. Allen,4.31,811216225,9.78081E+12,eng,186,6163,82,
22993,The House of Bernarda Alba and Other Plays,Federico Garcia Lorca-Michael Dewell-Carmen Zapata-Christopher Maurer,4.07,141185759,9.78014E+12,eng,GB,169,140,8,
22994,Selected Verse,Federico Garcia Lorca-Christopher Maurer,4.43,374528551,9.78037E+12,eng,432,184,13,
23000,Operation Wandering Soul,Richard Powers,3.62,806097611X,9.78006E+12,eng,352,366,32,
23001,Galatea 2.2,Richard Powers,3.72,312423136,9.78031E+12,eng,329,2863,183,
23002,Three Farmers on Their Way to a Dance,Richard Powers,3.84,60975091,9.78006E+12,eng,352,709,45,
23003,Christian Theology: An Introduction,Alister E. McGrath,4.04,1405153601,9.78141E+12,eng,534,1147,48,
23007,The Time of Our Singing,Richard Powers,4.24,99453835,9.7801E+12,eng,631,1710,230,
23009,The Christian Theology Reader,Alister E. McGrath,4.13,140515358X,9.78141E+12,eng,744,234,11,
23010,Prisoner's Dilemma,Richard Powers,3.87,60977086,9.78006E+12,eng,US,352,517,44,
23011,In the Beginning: The Story of the King James Bible and How it Changed a Nation a Language and a Culture,Alister E. McGrath,3.96,385722168,9.78039E+12,eng,354,491,70,
23012,The Gold Bug Variations,Richard Powers,4.13,60975088,9.78006E+12,eng,640,1383,126,
23017,Presidential Power and the Modern Presidents: The Politics of Leadership from Roosevelt to Reagan,Richard E. Neustadt,3.79,29227968,9.78003E+12,eng,384,277,17,
23022,The Grotesque,Patrick McGrath,3.63,679776214,9.78068E+12,eng,178,785,55,
23023,Port Mungo,Patrick McGrath,3.32,1400075483,9.7814E+12,eng,242,395,34,
23031,Time's Arrow,Martin Amis,3.74,679735720,9.78068E+12,eng,165,11360,796,
23032,Entre Amis: An Interactive Approach,Michael D. Oates-Larbi Oukada,3.52,618506918,9.78062E+12,eng,528,21,4,
23036,Experience,Martin Amis,3.96,99285827,9.7801E+12,eng,401,70,5,
23037,Heavy Water and Other Stories,Martin Amis,3.39,837570115X,9.78038E+12,eng,288,1056,44,
23040,Night Train,Martin Amis,3.24,375701141,9.78038E+12,eng,176,3444,279,
23042,Visiting Mrs Nabokov and Other Excursions,Martin Amis,3.71,99461870,9.7801E+12,eng,288,536,16,
23045,Intimacy And Other Stories,Hanif Kureishi,3.58,857121200X,9.78057E+12,eng,205,47,3,
23046,The Buddha of Suburbia,Hanif Kureishi,3.73,571200435,9.78057E+12,eng,US,284,362,22,
23053,Love In A Blue Time,Hanif Kureishi,3.48,057119222X,9.78057E+12,eng,212,768,32,
23054,Mythologies,Roland Barthes-Annette Lavers,4.1,99972204,9.7801E+12,eng,GB,159,220,15,
23056,Image-Music-Text,Roland Barthes-Stephen Heath,4.1,6861350,9.78001E+12,eng,220,2974,39,
23057,Camera Lucida: Reflections on Photography,Roland Barthes-Richard Howard,3.95,795000375,9.7800E+12,eng,119,75,6,
23060,The Fashion System,Roland Barthes-Matthew Ward-Richard Howard,3.57,520071778,9.78052E+12,eng,351,219,10,
23061,The Responsibility of Forns: Critical Essays on Music Art and Representation,Roland Barthes-Richard Howard,3.94,520072383,9.78052E+12,eng,320,92,3,
23063,The Pleasure of the Text,Roland Barthes-Richard Miller,4.02,224011871,9.78022E+12,eng,80,19,1,
23065,How to Travel with a Salmon & Other Essays,Umberto Eco-William Weaver,3.85,99428636,9.7801E+12,eng,229,186,11,
23066,How to Travel with a Salmon and Other Essays,Umberto Eco-William Weaver,3.85,151001367,9.78015E+12,eng,248,3396,148,
23079,Der Name der Rose,Umberto Eco-Burkhard Kroeber,4.12,3423105518,9.78342E+12,eng,657,646,19,
23080,Le Nom de la rose,Umberto Eco-Jean-Noël Schifano,4.12,2253033138,9.78225E+12,eng,755,665,43,
23082,Nachschrift zum Namen der Rose,Umberto Eco-Burkhard Kroeber,4.05,3446140379,9.78345E+12,eng,95,9,1,
```



22602,Carly's Sound,Ali Vali,4.21,1933110457,9.78193E+12,eng,212,441,11,  
22611,Die Valis Trilogie,Philip K. Dick-Thomas Ziegler-Alexander Martin,4.13,3453217276,9.78345E+12,ger,925,45,4,  
22613,Simulacra and Simulation,Jean Baudrillard-Sheila Faria Glaser,3.98,472865211,9.78047E+12,eng,164,8919,297,  
22614,América,Jean Baudrillard,3.9,8433925059,9.78843E+12,spa,172,12,1,  
22616,The System of Objects,Jean Baudrillard-James Benedict,4,1844678538,9.78184E+12,eng,224,957,26,  
22618,The Intelligence of Evil or the Lucidity Pact,Jean Baudrillard-Chris Turner,3.86,1845283348,9.78185E+12,eng,215,239,10,  
22620,Introducing Baudrillard,Chris Horrocks-Zoran Jevtić,3.31,1848468873,9.78184E+12,eng,176,155,24,  
22623,Existentialism and Human Emotions,Jean-Paul Sartre,3.84,886589823,9.78881E+12,eng,96,4898,146,  
22624,Nausea,Jean-Paul Sartre-James Wood-Robert Baldick,3.92,814118549X,9.78814E+12,eng,253,1459,120,  
22626,From Socrates to Sartre: The Philosophic Quest,T.Z. Lavine,3.93,553251619,9.78055E+12,eng,426,893,66,  
22630,Hey Rube: Blood Sport the Bush Doctrine and the Downward Spiral of Dumbness: Modern History from the ESPN.com Sports Desk,Hunter S. Thompson-John A. Walsh,3.68,684873286,9.78868E+12,eng,272,2242,128,  
22632,The Trial and Death of Socrates (Euthyphro Apology Crito Phaedo (death scene only)),Plato-G.M.A. Grube-John M. Cooper,4.07,872285541,9.78887E+12,eng,58,24698,326,  
22633,The Trial of Queen Caroline: The Scandalous Affair that Nearly Ended a Monarchy,Jane Robins,3.37,743255989,9.78874E+12,eng,384,51,9,  
22649,Bookends,Jane Green,3.71,767987817,9.78877E+12,eng,368,36227,812,  
22655,Bookends,Jane Green,3.71,148276521,9.78814E+12,eng,394,313,38,  
22656,Jemima J.: For those who Love Faking Friends and My Sweet Revenge by Jane Fallon,Jane Green,3.64,148276984,9.78814E+12,eng,450,669,68,  
22659,The Lost Boy,Dave Pelzer,4.09,929899836,9.78893E+12,eng,283,218,28,  
22669,Open House,Elizabeth Berg,3.7,613997638,9.78861E+12,eng,8,0,0,0,  
22675,A Morbid Taste for Bones (Chronicles of Brother Cadfael #1),Ellis Peters,3.98,751517496,9.78875E+12,eng,250,60,4,  
22677,Feeling Better Getting Better Staying Better: Profound Self-Help Therapy for Your Emotions,Albert Ellis,3.67,1886238358,9.78189E+12,eng,272,49,6,  
22680,American Sphinx: The Character of Thomas Jefferson,Joseph J. Ellis,3.91,679764410,9.78868E+12,eng,440,19892,569,  
22681,Line Of Vision,David Ellis,3.8,425183769,9.78843E+12,en-US,448,572,76,  
22686,The Other Woman,Jane Green,3.7,452287146,9.78845E+12,eng,389,28444,644,  
22687,No Other Woman (No Other #2),Shannon Drake,3.78,388781368,9.78838E+12,eng,391,296,7,  
22689,Speculum of the Other Woman,Luce Irigaray-Gillian C. Gill,4.81,881493387,9.7888E+12,eng,416,434,19,  
22690,The Other Woman (Dundee Idaho #7),Brenda Novak,4.11,373713444,9.78837E+12,eng,299,528,32,  
22691,The Other Woman,Ann O'Leary,3.64,1562882348,9.78156E+12,eng,213,73,5,  
22697,Sins of a Shaker Summer (Sister Rose Callahan #3),Deborah Woodworth,3.5,388792844,9.78838E+12,eng,272,55,5,  
22715,Inuyasha Ani-Manga Vol. 1 (Inuyasha Ani-Manga #1),Rumiko Takahashi,4.36,1591162825,9.78159E+12,eng,286,5787,54,  
22716,Sushi for Beginners,Marian Keyes,3.69,718144465,9.78872E+12,eng,448,269,28,  
22718,Pustebłume,Marian Keyes,3.8,3453189345,9.78345E+12,ger,568,55,0,  
22724,¿Quién te lo ha contado?,Marian Keyes,3.78,8497936736,9.7885E+12,spa,648,152,9,  
22725,Une vie de rêve,Marian Keyes,3.69,271443813X,9.78271E+12,fre,593,11,4,  
22726,Por los pelos,Marian Keyes,3.8,1488882214,9.7814E+12,spa,592,33,5,  
22727,Under the Duvet,Marian Keyes,3.51,141887478,9.78814E+12,eng,384,386,43,  
22729,Rachel's Holiday (Walsh Family #2),Marian Keyes,3.95,148271791,9.78814E+12,en-GB,626,598,53,  
22733,Rachel se va de viaje (La familia Walsh #2),Marian Keyes,3.95,8497599179,9.7885E+12,spa,595,288,20,  
22736,Yeats ist tot!,Joseph O'Connor-Diethard Herles-Pauline McLynn-Gerard Stenbridge-Frank McCourt-Conor McPherson-Gene Kerrigan-Gina Moxley-Marian Keyes-Anthony Cronin-Owen O'Neill-Hugo Hamilton-Charlie O'Neill-Tom Humphries-Donal O'Kelly,3.39,3548683254,9.78355E+12,ger,332,11,2,  
22739,Le Club de la dernière chance,Marian Keyes,3.8,2266138689,9.78227E+12,fre,688,6,0,  
22741,Les vacances de Rachel (Walsh Family #2),Marian Keyes-Roxane Azimi,3.95,2266188271,9.78227E+12,fre,563,6,1,  
22742,¿Quién te lo ha contado?,Marian Keyes,3.78,8481315824,9.7884E+12,spa,648,12,2,  
22743,Rachel se va de viaje (La familia Walsh #2),Marian Keyes,3.95,848132954X,9.7884E+12,spa,594,7,0,  
22745,Auszeit für Engel Walsh Family #3,Marian Keyes,3.75,3453868579,9.78345E+12,ger,477,15,1,  
22747,Melancia,Marian Keyes,3.78,8528689162,9.78853E+12,por,489,323,12,  
22748,Vattenmelonon,Marian Keyes-Lina Erkelius,3.78,9176435245,9.78918E+12,swe,518,94,3,  
22750,Auszeit für Engel (Walsh Family #3),Marian Keyes,3.75,3453877616,9.78345E+12,ger,488,34,0,  
22751,Férias! (Walsh Family #2),Marian Keyes-Heloisa Maria Leal,3.95,8528618855,9.78853E+12,por,568,179,4,  
22763,Open House,Elizabeth Berg,3.7,99461269,9.7881E+12,en-GB,314,61,7,  
22774,Slightly Settled (Slightly #2),Wendy Markham,3.66,373250479,9.78837E+12,en-US,312,4681,82,  
22784,犬夜叉 8,Rumiko Takahashi,4.26,4891252887,9.78489E+12,jpn,192,14,0,  
22786,犬夜叉 7,Rumiko Takahashi,4.27,4891252879,9.78489E+12,jpn,192,13,0,  
22787,犬夜叉 25,Rumiko Takahashi,4.28,4891256457,9.78489E+12,jpn,185,12,2,  
22789,犬夜叉 24,Rumiko Takahashi,4.27,4891256449,9.78489E+12,jpn,192,18,1,  
22799,犬夜叉 28,Rumiko Takahashi,4.29,4891256481,9.78489E+12,jpn,192,12,1,

22286,Clown Girl,Monica Drake-Chuck Palahniuk,3.42,976631156,9.78098E+12,eng,297,5305,328,  
22287,Choke,Chuck Palahniuk,3.69,99422689,9.7801E+12,en-GB,293,1457,108,  
22288,Haunted,Chuck Palahniuk,3.59,1400032822,9.7814E+12,eng,419,75255,3663,  
22289,Fugitives and Refugees: A Walk in Portland Oregon,Chuck Palahniuk,3.51,1400047838,9.7814E+12,en-US,176,8326,422,  
22290,Invisible Monsters,Chuck Palahniuk,4,99285444,9.7801E+12,eng,297,54799,2252,  
22292,Haunted,Chuck Palahniuk-Erik Darling-Arthur Morey-Scott Brick-Lorna Raver-Marc Cashman-Kimberly Farr-Renee Randman,3.59,739302868,9.78074E+12,eng,14,54,23,  
22293,Nonfiction,Chuck Palahniuk,3.57,224063022,9.78022E+12,eng,233,212,15,  
22295,Der Simulant,Chuck Palahniuk-Werner Schmitz,3.69,3442541662,9.78344E+12,ger,320,89,1,  
22296,Haunted,Chuck Palahniuk,3.59,224064452,9.78022E+12,eng,416,166,15,  
22297,Survivant,Chuck Palahniuk-Freddy Michalski,3.92,2070756270,9.78207E+12,fre,365,14,0,  
22298,Fantasmas,Chuck Palahniuk-Javier Calvo,3.59,9879397509,9.78908E+12,spa,442,112,11,  
22301,Fantasmas,Chuck Palahniuk,3.59,843972005X,9.78844E+12,spa,442,31,4,  
22302,Error humano,Chuck Palahniuk-Javier Calvo,3.57,8439711867,9.78844E+12,spa,253,10,1,  
22304,Life After God,Douglas Coupland,3.83,743231511,9.78074E+12,eng,291,9167,332,  
22306,City of Glass: Douglas Coupland's Vancouver,Douglas Coupland,3.83,1550548182,9.78155E+12,en-US,152,794,40,  
22307,Terry: Terry Fox and His Marathon of Hope,Douglas Coupland,4.34,1553651529,9.78155E+12,eng,176,254,22,  
22311,All Families Are Psychotic,Douglas Coupland,3.69,7117531,9.78001E+12,eng,279,350,32,  
22312,Shampoo Planet,Douglas Coupland,3.5,743231538,9.78074E+12,eng,282,8182,200,  
22313,Miss Wyoming,Douglas Coupland,3.48,3499232642,9.7835E+12,ger,352,6,3,  
22316,Miss Wyoming,Douglas Coupland,3.48,2259834,9.78E+12,eng,311,160,13,  
22317,Miss Wyoming,Douglas Coupland,3.48,7179820,9.78001E+12,eng,320,110,5,  
22320,Pattern Recognition (Blue Ant #1),William Gibson,3.87,425190685,9.78043E+12,eng,367,37313,1767,  
22321,All Tomorrow's Parties (Bridge #3),William Gibson,3.9,425190447,9.78043E+12,eng,339,12637,306,  
22322,Spook Country (Blue Ant #2),William Gibson,3.68,399154302,9.7804E+12,eng,371,15943,1246,  
22323,Burning Chrome (Sprawl #0),William Gibson-Bruce Sterling,4.06,00539028,9.78006E+12,eng,224,29648,533,  
22325,Idoru (Bridge #2),William Gibson,3.79,425190455,9.78043E+12,eng,308,17197,375,  
22326,Virtual Light (Bridge #1),William Gibson,3.84,140157727,9.78014E+12,eng,304,17658,340,  
22328,Neuromancer (Sprawl #1),William Gibson,3.89,441569595,9.78044E+12,eng,271,185418,5313,  
22331,Stable Strategies and Others,Eileen Gunn-Howard Waldrop-William Gibson,3.75,189239118X,9.78109E+12,eng,208,98,19,  
22332,Cyberpunk and Cyberculture: Science Fiction and the Work of William Gibson,Dani Cavallaro,3.79,485006073,9.78049E+12,eng,258,59,4,  
22337,The Neil Gaiman Audio Collection,Neil Gaiman,4.26,60732989,9.78006E+12,eng,1,1539,190,  
22338,The Day I Swapped My Dad for Two Goldfish,Neil Gaiman-Dave McKean,4.03,60587032,9.78006E+12,en-US,64,200,30,  
22339,Death: The Time of Your Life,Neil Gaiman-Chris Bachalo-Mark Buckingham-Mark Pennington-Claire Danes,4.22,1563893339,9.78156E+12,eng,95,12206,195,  
22341,The Wolves in the Walls,Neil Gaiman-Dave McKean,4.07,300810956,9.78038E+12,eng,56,13382,894,  
22344,Creatures of the Night,Neil Gaiman-Michael Zulli,3.84,1569719365,9.78157E+12,eng,46,1627,163,  
22345,New X-Men Volume 4: Riot at Xavier's,Grant Morrison-Frank Quitely,4.05,785110674,9.78079E+12,eng,120,2036,46,  
22347,The Invisibles Volume 1: Say You Want a Revolution,Grant Morrison-Steve Yeowell-Jill Thompson-Dennis Cramer-Peter Milligan,3.95,1563892677,9.78156E+12,eng,224,0669,364,  
22348,The Mystery Play,Grant Morrison-Jon J. Muth,3.44,1563891891,9.78156E+12,en-US,80,664,37,  
22350,Kid Eternity,Grant Morrison-Duncan Fegredo,3.4,1401209335,9.7814E+12,eng,144,659,39,  
22351,New X-Men Volume 1: E Is for Extinction,Grant Morrison-Frank Quitely,4.06,785108114,9.78079E+12,eng,144,7371,118,  
22354,Seven Soldiers of Victory Volume 1,Grant Morrison-Ryan Sook-Mick Gray-Frazer Irving-J.H. Williams III-Simone Bianchi-Cameron Stewart,3.79,1401209254,9.7814E+12,eng,224,2017,96,  
22355,Doom Patrol Vol. 1: Crawling from the Wreckage,Grant Morrison-Richard Case-Doug Braithwaite-Scott Hanna-Carlos Garzon-John Nyberg,4.09,1563890348,9.78156E+12,eng,192,5158,218,  
22356,The Filth,Grant Morrison-Chris Weston-Gary Erskine,3.75,1401200133,9.7814E+12,en-US,320,3805,258,  
22357,The Invisibles Vol. 2: Apocalipstick,Grant Morrison-Jill Thompson-Chris Weston-John Ridgway-Steve Parkhouse-Paul Johnson,4.22,1563897024,9.78156E+12,eng,208,5791,122,  
22360,Doom Patrol Vol. 5: Magic Bus,Grant Morrison-Richard Case-Ken Steacy-Stan Woch-Philip Bond-Mark McKenna-Scott Hanna,4.26,1401212026,9.7814E+12,eng,203,1308,50,  
22362,JLA Vol. 1: New World Order,Grant Morrison-Howard Porter-John Dell,3.97,156389369X,9.78156E+12,en-US,93,3601,91,  
22363,Batman: Gothic,Grant Morrison-Klaus Janson,3.79,1563890283,9.78156E+12,eng,128,4168,114,  
22364,JLA: Earth 2,Grant Morrison-Frank Quitely,3.96,1563896311,9.78156E+12,eng,96,5019,140,  
22365,New X-Men: Omnibus,Grant Morrison-Marc Silvestri-Chris Bachalo-John Paul Leon-Frank Quitely-Leinil Francis Yu-Igor Kordey-Ethan Van Sciver-Keron Grant-Ton Dericenick-Phil Jimenez,4.22,785123261,9.78079E+12,eng,1096,2861,95,  
22366,Superfolks,Robert Mayer-Grant Morrison,3.52,312339925,9.78031E+12,eng,240,409,56,  
22373,Kill Your Boyfriend,Grant Morrison-Philip Bond-D'Israeli-Daniel Vozzo,3.56,156389453X,9.78156E+12,en-US,55,1251,66,  
22374,Batman: Arkham Asylum - A Serious House on Serious Earth,Grant Morrison-Dave McKean,4.1,1401204252,9.7814E+12,eng,216,50214,1392,  
22375,New X-Men: New Worlds (New X-Men #3),Grant Morrison-Igor Kordey-Ethan Van Sciver,3.85,1904159087,9.7819E+12,eng,160,2,0,  
22380,JLA: Tierra 2,Grant Morrison-Frank Quitely-Rafael De la Iglesia,3.96,907224975X,9.78907E+12,spa,96,6,2,  
22381,JLA Vol. 4: Strength in Numbers,Grant Morrison-Mark Waid-Howard Porter-John Dell,3.91,1417658584,9.78142E+12,eng,222,15,2,



23060, The Fashion System, Roland Barthes-Matthew Ward-Richard Howard, 3.57, 520071778, 9.78052E+12, eng, 351, 219, 10,  
23061, The Responsibility of Forms: Critical Essays on Music Art and Representation, Roland Barthes-Richard Howard, 3.94, 520072383, 9.78052E+12, eng, 320, 92, 3,  
23063, The Pleasure of the Text, Roland Barthes-Richard Miller, 4.02, 224011871, 9.78022E+12, eng, 80, 19, 1,  
23065, How to Travel with a Salmon & Other Essays, Umberto Eco-William Weaver, 3.85, 99428636, 9.7801E+12, eng, 229, 106, 11,  
23066, How to Travel with a Salmon and Other Essays, Umberto Eco-William Weaver, 3.85, 151001367, 9.78015E+12, eng, 248, 3396, 148,  
23079, Der Name der Rose, Umberto Eco-Burkhard Kroeber, 4.12, 3423105518, 9.78342E+12, ger, 657, 646, 19,  
23080, Le Nom de la rose, Umberto Eco-Jean-Noël Schifano, 4.12, 2253033138, 9.78225E+12, fre, 755, 665, 43,  
23082, Nachschrift zum »Namen der Rose«, Umberto Eco-Burkhard Kroeber, 4.05, 3446140379, 9.78345E+12, ger, 95, 9, 1,  
23084, Der Name Der Rose: Königs Erläuterungen Und Materialien 00391, Cerstin Urban-Umberto Eco, 3.3804416276, 9.7838E+12, ger, 86, 3, 0,  
23089, The Carpenter's Pencil, Manuel Rivas-Jonathan Dunne, 3.6, 99440467, 9.7801E+12, eng, 165, 628, 53,  
23090, ¿Qué me quieres amor?, Manuel Rivas, 3.62, 8420402277, 9.78042E+12, glg, 216, 450, 19,  
23100, O lapis do carpinteiro, Manuel Rivas, 3.6, 8483022613, 9.78048E+12, glg, 152, 181, 13,  
23112, A man dos paños, Manuel Rivas, 3.46, 8483025914, 9.78048E+12, glg, 142, 24, 0,  
23114, Os libros arden mal, Manuel Rivas, 3.35, 849782461X, 9.7805E+12, glg, 741, 33, 2,  
23129, El perquè de tot plegat, Quim Monzó, 3.8, 8477271054, 9.78048E+12, cat, 176, 333, 22,  
23139, Ochenta y Seis Cuentos, Quim Monzó, 4.03, 8433924788, 9.78043E+12, spa, 504, 44, 4,  
23147, Guadalupe, Quim Monzó, 3.77, 847727164X, 9.78048E+12, cat, 199, 28, 2,  
23155, Self Service, Quim Monzó-Biel Mesquida, 2, 8473110102, 9.78047E+12, cat, 256, 1, 0,  
23158, The Age of Access: The New Culture of Hypercapitalism Where All of Life is a Paid-For Experience, Jeremy Rifkin-Ralph Fowler, 3.84, 1585420824, 9.78159E+12, eng, 320, 215, 12,  
23183, The Book of Questions, Pablo Neruda-William O'Daly, 4.05, 1556591608, 9.78156E+12, eng, 96, 2086, 213,  
23186, Bugs for Lunch/Insectos para el almuerzo, Margery Facklan-Sylvia Long-Lilliana Valenzuela, 3.78, 1570915067, 9.78157E+12, spa, 32, 4, 1,  
23194, My Story as told by Water: Confessions Druidic Rants Reflections Bird-watchings Fish-stalkings Visions Songs and Prayers Refracting Light from Living Rivers in the Age of the Industrial Dark , David James Duncan, 4.00, 1578050499, 9.78158E+12, eng, 304, 665, 52,  
23201, Pilgrims and Other Stories, Elizabeth Gilbert, 3.12, 330351745, 9.78033E+12, eng, 288, 1352, 173,  
23202, The Last American Man, Elizabeth Gilbert, 3.8, 142002836, 9.78014E+12, eng, 271, 7634, 934,  
23203, Pilgrims Pa, Elizabeth Gilbert, 3.12, 395924855, 9.7804E+12, eng, 224, 10, 2,  
23204, Stern Men, Elizabeth Gilbert, 3.39, 061812733X, 9.78062E+12, eng, 304, 4065, 521,  
23206, A Writer's Workbook: Daily Exercises for the Writing Life, Caroline Sharp-Elizabeth Gilbert, 3.71, 031228621X, 9.78031E+12, eng, 176, 148, 15,  
23217, Size 14 Is Not Fat Either (Heather Wells #2), Meg Cabot, 3.83, 60525126, 9.78006E+12, eng, 344, 26923, 1015,  
23218, Missing You (1-800-Where-R-You #5), Meg Cabot, 4.17, 60874309, 9.78006E+12, eng, 268, 10886, 365,  
23219, Queen of Babble (Queen of Babble #1), Meg Cabot, 3.69, 60851988, 9.78006E+12, eng, 309, 59613, 1675,  
23220, Size 12 Is Not Fat (Heather Wells #1), Meg Cabot, 3.74, 60525118, 9.78006E+12, eng, 345, 73592, 2360,  
23221, How to Be Popular, Meg Cabot, 3.54, 60880120, 9.78006E+12, eng, 288, 19183, 879,  
23222, Valentine Princess (The Princess Diaries #4.25), Meg Cabot, 3.67, 60847182, 9.78006E+12, eng, 96, 9330, 211,  
23223, Every Boy's Got One (Boy #3), Meg Cabot, 3.77, 330418882, 9.78033E+12, eng, 328, 20562, 734,  
23224, Darkest Hour (The Mediator #4), Meg Cabot, 4.17, 60725141, 9.78006E+12, eng, 316, 30018, 657,  
23225, Haunted (The Mediator #5), Meg Cabot, 4.15, 60751649, 9.78006E+12, eng, 263, 31501, 652,  
23226, She Went All the Way, Meg Cabot, 3.69, 60085444, 9.78006E+12, eng, 357, 12297, 432,  
23227, Are We There Yet?, David Levithan, 3.5, 037582846X, 9.78038E+12, eng, 215, 3048, 277,  
23228, Boy Meets Boy, David Levithan, 3.84, 375832998, 9.78038E+12, eng, 185, 57087, 3242,  
23229, This Is Push: New Stories from the Edge, David Levithan-Patricia McCormick-Matthew Roth-Kevin Waltman-Samantha Schutz-Coe Booth-Eddie de Oliveira-Tanuja Desai Hidier-Kevin Brooks-Chris Wooding-Markus Zusak-Brian James-Kristen Kemp-Eireann Corrigan-Christopher Krovatin-Billy Merrell, 3.62, 439890284, 9.78044E+12, eng, 240, 222, 23,  
23230, Naomi and Eli's No Kiss List, Rachel Cohn-David Levithan, 3.44, 375844406, 9.78038E+12, eng, 230, 13620, 978,  
23232, The Realm of Possibility, David Levithan, 3.93, 375836578, 9.78038E+12, eng, 210, 14438, 1176,  
23233, Ten Things I Hate about You, David Levithan, 4.11, 439087309, 9.78044E+12, eng, 170, 1414, 62,  
23235, Where We Are What We See: The Best Young Writers and Artists in America, David Levithan, 3.73, 439736463, 9.78044E+12, eng, 288, 41, 0,  
23236, Aphrodite's Passion (Superhero Central #2), Julie Kenner, 3.62, 505524740, 9.78051E+12, eng, 352, 524, 33,  
23237, The Manolo Matrix (Codebreaker Trilogy #2), Julie Kenner, 3.61, 743496140, 9.78074E+12, eng, 352, 635, 34,  
23238, The Cat's Fancy, Julie Kenner, 3.50, 505523973, 9.78051E+12, eng, 391, 243, 27,  
23239, Demons Are Forever (Demon-Hunting Soccer Mom #3), Julie Kenner, 3.89, 425215385, 9.78043E+12, eng, 292, 2343, 164,  
23240, Aphrodite's Secret (Superhero Central #3), Julie Kenner, 3.7, 505525097, 9.78051E+12, eng, 369, 405, 20,  
23242, The Good Ghouls' Guide to Getting Even (Beth Frasier #1), Julie Kenner, 3.58, 425213919, 9.78043E+12, eng, 256, 876, 71,  
23243, Aphrodite's Flame (Superhero Central #4), Julie Kenner, 3.67, 505525755, 9.78051E+12, eng, 338, 379, 16,  
23244, Carpe Demon (Demon-Hunting Soccer Mom #1), Julie Kenner, 3.63, 515142212, 9.78052E+12, eng, 320, 4750, 581,  
23249, Perfect Circle, Sean Stewart, 3.8, 1931520119, 9.78193E+12, eng, 243, 666, 67,  
23257, The Great Philosophers: An Introduction to Western Philosophy, Bryan Magee, 4.06, 019289322X, 9.78019E+12, eng, 352, 333, 16,

- **CREATE THE FILE**

```
harshit@dell-Inspiron-15-3567:~$ cat> file.txt
novel
story book
fiction
harry potter
J.K. Rowling
Me Before You^C
harshit@dell-Inspiron-15-3567:~$
```

- **DISPLAY THE FILE**

```
harshit@dell-Inspiron-15-3567:~$ cat file.txt
novel
story book
fiction
harry potter
J.K. Rowling
harshit@dell-Inspiron-15-3567:~$
```

- **TO SHOW THE HIDDEN FILES**

```
harshit@dell-Inspiron-15-3567:~$ ls -a
.          .dbus          ~/.lock.POC_GROUP39.docx#  scalpel-output
..         Downloads  .mozilla                  snap
.bash_history .eclipse       Pictures                  .ssh
.bash_logout examples.desktop .pki                      .sudo_as_admin_successful
.bashrc      file.txt       pocdocx.docx              .swt
books.csv    .gnupg        POC_GROUP39.docx          .tooling
.cache       .ICEauthority .profile                  'Untitled Document 1'
.config      .local        .scala_history
harshit@dell-Inspiron-15-3567:~$
```

- **TO DISPLAY THE PRESENT WORKING DIRECTORY**

```
harshit@dell-Inspiron-15-3567:~$ pwd
/home/harshit
harshit@dell-Inspiron-15-3567:~$
```

- **TO REMOVE THE DIRECTORY**

```
harshit@dell-Inspiron-15-3567:~$ hadoop fs -rm -r poc1
Deleted hdfs://localhost/user/training/poc1
```

## **SOLUTION**

#Hive

- **CREATE THE TABLE**

```
hive > create table poccl (Book_id int, Book title String, Book rating float
> Book writer String, Avg_rating float, language String, Pages int)
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> stored as textfile;
OK
Time taken: 4.25 seconds
```

- **LOAD THE DATA TO THE TABLE FROM THE DATASET**

```
hive> LOAD DATA LOCAL INPATH '/home/training/Desktop/dataset.csv' OVERWRITE INTO
TABLE poccl;
Copying data from file:/home/training/Desktop/dataset.csv
Copying file: file:/home/training/Desktop/dataset.csv
Loading data to tabel default.poccl
Deleted hdfs://localhost/user/hive/warehouse/poccl
OK
Time taken: 1.32 seconds
```

# QUERIES

1. Show the book with the maximum pages

```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$
select max(Pages) from poccl;

2019-07-10 05:21:25, 520 Stage-1 map = 100%, reduce = 100%
Ended Job = Job_3012547301311_13001
OK
3342
Time taken: 9.25 seconds
```

```
hive> select Book title from poccl where Pages = 3342;
```

# OUTPUT

```
2019-07-10 05:23:51, 321 Stage-1 map = 100%, reduce = 100%
Ended Job = job_20364030102010_0021
OK
Harry Potter Collection (Harry Potter #1-6)
Time taken: 8.14 seconds
```

2. Select Book which is having rating greater than “3.5”



```
hive> select Book title from poccl WHERE Avg_rating > 3.5;
```

```
Harry Potter and the Order of the Phoenix (Harry Potter #5)
Harry Potter and the Sorcerer's Stone (Harry Potter #1)
Harry Potter and the Chamber of Secrets (Harry Potter #2)
Harry Potter and the Prisoner of Azkaban (Harry Potter #3)
Harry Potter Boxed Set Books 1-5 (Harry Potter #1-5)
Unauthorized Harry Potter Book Seven News: "Half-Blood Prince" Analysis and Speculation
Harry Potter Collection (Harry Potter #1-6)
The Ultimate Hitchhiker's Guide: Five Complete Novels and One Story (Hitchhiker's Guide to the Galaxy #1-5)
The Ultimate Hitchhiker's Guide to the Galaxy
The Hitchhiker's Guide to the Galaxy (Hitchhiker's Guide to the Galaxy #1)
The Hitchhiker's Guide to the Galaxy (Hitchhiker's Guide to the Galaxy #1)
The Ultimate Hitchhiker's Guide (Hitchhiker's Guide to the Galaxy #1-5)
A Short History of Nearly Everything
Bill Bryson's African Diary
Bryson's Dictionary of Troublesome Words: A Writer's Guide to Getting It Right
In a Sunburned Country
I'm a Stranger Here Myself: Notes on Returning to America After Twenty Years Away
The Lost Continent: Travels in Small Town America
Neither Here nor There: Travels in Europe
Notes from a Small Island
The Mother Tongue: English and How It Got That Way
J.R.R. Tolkien 4-Book Boxed Set: The Hobbit and The Lord of the Rings
The Lord of the Rings (The Lord of the Rings #1-3)
The Lord of the Rings (The Lord of the Rings #1-3)
```

### 3. Select the book title starting With "H"

```
hive> SELECT Book title FROM poccl WHERE TITLE(Book title,1,1)=="H";
```

#### #OUTPUT

```
Harry Potter and the Half-Blood Prince
Harry Potter and the order of the Phoenix
Harry Potter and the Sorcerer's Stone
Harry Potter and the Chamber of Secerets
Harry Potter and the Prisoner of Azkaban
```

#### 4. PRINT THE BOOKS WITH MAXIMUM AND MINIMUM AVERAGE RATING IN THE DATASET

##### MAXIMUM RATING

```
hive> select max(Avg_rating) from poccl;
```

# OUTPUT

```
Ended Job = job_201903250102_0012  
OK  
4.9  
Time taken 10.21 seconds
```

##### MINIMUM RATING

```
hive> select min(Avg_rating) from poccl;
```

# OUTPUT

```
hive> select min(Avg_rating) from poccl;  
  
Ended Job = job_210246010242_0036  
OK  
2.78  
Time taken: 5.06 seconds
```

# PIG

## Loading data

```
grunt> doc1 = LOAD '/home/training/Desktop/doc1.csv' USING PigStorage(',')
>> AS (Book title:chararray, Book_id:int, Avg_rating:float, Writer:chararray);
grunt> doc2 = LOAD '/home/training /Desktop/doc2.csv' USING PigStorage(',')
>> AS (Book title:chararray, Book_id:int, Avg_rating:float, writer:chararray);
grunt> doc3 = LOAD 'user/training/poc/doc3.csv' USING PigStorage(',')
>>AS (Book title:chararray, Avg_rating:float, Pages:int, writer:chararray,Book_id:int, language:String);
```

## Dumping Data

```
grunt> dump doc1;
```

## # Output

```
22286,Clown Girl,Monica Drake-Chuck Palahniuk,3.42,976631156,9.78098E+12,eng,297,5305,328,
22287,Choke,Chuck Palahniuk,3.69,99422689,9.7801E+12,en-GB,293,1457,100,
22288,Haunted,Chuck Palahniuk,3.59,1400032022,9.7814E+12,eng,419,75255,3663,
22289,Fugitives and Refugees: A Walk in Portland Oregon,Chuck Palahniuk,3.51,1400047838,9.7814E+12,en-US,176,8326,422,
22290,Invisible Monsters,Chuck Palahniuk,4.99285444,9.7801E+12,eng,297,54799,2252,
22292,Haunted,Chuck Palahniuk-Erik Darling-Arthur Morey-Scott Brick-Lorna Raver-Marc Cashman-Kimberly Farr-Renee Randman,3.59,7393802868,9.78074E+12,eng,14,54,23,
22293,Nonfiction,Chuck Palahniuk,3.57,224063022,9.78022E+12,eng,233,212,15,
22295,Der Simulant,Chuck Palahniuk-Werner Schmitz,3.69,3442541662,9.78344E+12,ger,320,89,1,
22296,Haunted,Chuck Palahniuk,3.59,224064452,9.78022E+12,eng,416,166,15,
22297,Survivant,Chuck Palahniuk-Fredy Michalski,3.92,2070756270,9.78207E+12,fre,365,14,0,
22298,Fantasmas,Chuck Palahniuk-Javier Calvo,3.59,9879397509,9.78988E+12,spa,442,112,11,
22301,Fantasmas,Chuck Palahniuk,3.59,843972005X,9.78044E+12,spa,442,31,4,
22302,Error humano,Chuck Palahniuk-Javier Calvo,3.57,8439711867,9.78044E+12,spa,253,10,1,
22304,Life After God,Douglas Coupland,3.83,743231511,9.78074E+12,eng,291,9167,332,
22306,City of Glass: Douglas Coupland's Vancouver,Douglas Coupland,3.83,1550548182,9.78155E+12,en-US,152,794,40,
22307,Terry: Terry Fox and His Marathon of Hope,Douglas Coupland,4.34,1553651529,9.78155E+12,eng,176,254,22,
22311,All Families Are Psychotic,Douglas Coupland,3.69,7117531,9.78001E+12,eng,279,350,32,
22312,Shampoo Planet,Douglas Coupland,3.5,743231538,9.78074E+12,eng,282,8182,200,
22313,Miss Wyoming,Douglas Coupland,3.48,3499232642,9.7835E+12,ger,352,6,3,
22316,Miss Wyoming,Douglas Coupland,3.48,2259834,9.78E+12,eng,311,160,13,
22317,Miss Wyoming,Douglas Coupland,3.48,7179820,9.78001E+12,eng,320,110,5,
22320,Pattern Recognition (Blue Ant #1),William Gibson,3.87,425198685,9.78043E+12,eng,367,37313,1767,
22321,All Tomorrow's Parties (Bridge #3),William Gibson,3.9,425198447,9.78043E+12,eng,339,12637,306,
22322,Spook Country (Blue Ant #2),William Gibson,3.68,399154302,9.7804E+12,eng,371,15943,1246,
22323,Burning Chrome (Sprawl #0),William Gibson-Bruce Sterling,4.06,60539828,9.78006E+12,eng,224,29648,533,
22325,Idoru (Bridge #2),William Gibson,3.79,425198455,9.78043E+12,eng,308,17197,375,
22326,Virtual Light (Bridge #1),William Gibson,3.84,140157727,9.78014E+12,eng,304,17658,340,
22328,Neuromancer (Sprawl #1),William Gibson,3.89,441569595,9.78044E+12,eng,271,185418,5313,
22331,Stable Strategies and Others,Eileen Gunn-Howard Waldrop-William Gibson,3.75,189239118X,9.78189E+12,eng,208,98,19,
22332,Cyberpunk and Cyberculture: Science Fiction and the Work of William Gibson,Dani Cavallaro,3.79,485006073,9.78049E+12,eng,258,59,4,
22337,The Neil Gaiman Audio Collection,Neil Gaiman,4.26,60732989,9.78006E+12,eng,1,1539,190,
22338,The Day I Swapped My Dad for Two Goldfish,Neil Gaiman-Dave McKean,4.03,60587032,9.78006E+12,en-US,64,280,30,
22339,Death: The Time of Your Life,Neil Gaiman-Chris Bachalo-Mark Buckingham-Mark Pennington-Claire Danes,4.22,1563893339,9.78156E+12,eng,95,12206,195,
22341,The Wolves in the Walls,Neil Gaiman-Dave McKean,4.07,380010956,9.78038E+12,eng,56,13382,894,
22344,Creatures of the Night,Neil Gaiman-Michael Zull,3.84,1569719365,9.78157E+12,eng,46,1627,163,
22345,New X-Men Volume 4: Riot at Xavier's,Grant Morrison-Frank Quitely,4.85,785118674,9.78079E+12,eng,120,2036,46,
22347,The Invisibles Volume 1: Say You Want a Revolution,Grant Morrison-Steve Yeowell-Jill Thompson-Dennis Cramer-Peter Milligan,3.95,1563892677,9.78156E+12,eng,224,8669,364,
22348,The Mystery Play,Grant Morrison-Jon J. Muth,3.44,1563891891,9.78156E+12,en-US,80,664,37,
22350,Kid Eternity,Grant Morrison-Duncan Fegredo,3.4,1401209335,9.7814E+12,eng,144,659,39,
22351,New X-Men Volume 1: E Is for Extinction,Grant Morrison-Frank Quitely,4.06,785108114,9.78079E+12,eng,144,7371,118,
22354,Seven Soldiers of Victory Volume 1,Grant Morrison-Ryan Sook-Mick Gray-Frazer Irving-J.H. Williams III-Simone Bianchi-Cameron Stewart,3.79,1401209254,9.7814E+12,eng,224,2017,96,
22355,Doom Patrol Vol. 1: Crawling from the Wreckage,Grant Morrison-Richard Case-Doug Braithwaite-Scott Hanna-Carlos Garzon-John Nyberg,4.09,1563890348,9.78156E+12,eng,192,5158,218,
22356,The Filth,Grant Morrison-Chris Weston-Gary Erskine,3.75,1401200133,9.7814E+12,en-US,320,3805,258,
22357,The Invisibles Vol. 2: Apocaliptick,Grant Morrison-Jill Thompson-Chris Weston-John Ridgway-Steve Parkhouse-Paul Johnson,4.22,1563897024,9.78156E+12,eng,200,5791,122,
22360,Doom Patrol Vol. 5: Magic Bus,Grant Morrison-Richard Case-Ken Steacy-Stan Moch-Phillip Bond-Mark McKenna-Scott Hanna,4.26,1401212026,9.7814E+12,eng,203,1308,50,
22362,JLA Vol. 1: New World Order,Grant Morrison-Howard Porter-John Dell,3.97,156389369X,9.78156E+12,en-US,93,3601,91,
22363,Batman: Gothic,Grant Morrison-Klaus Janson,3.79,1563890283,9.78156E+12,eng,128,4168,114,
22364,JLA: Earth 2,Grant Morrison-Frank Quitely,3.96,1563896311,9.78156E+12,eng,96,5019,140,
22365,New X-Men: Omnibus,Grant Morrison-Marc Silvestri-Chris Bachalo-John Paul Leon-Frank Quitely-Leinil Francis Yu-Igor Kordey-Ethan Van Sciver-Keron Grant-Tom Derenick-Phil Jinenez,4.22,785123261,9.78074E+12,eng,1096,2861,95,
22366,Superfolks,Robert Mayer-Grant Morrison,3.52,312339925,9.78031E+12,eng,240,409,56,
22373,Kill Your Boyfriend,Grant Morrison-Phillip Bond-D'Israeli-Daniel Vozzo,3.56,156389453X,9.78156E+12,en-US,55,1251,66,
22374,Batman: Arkham Asylum - A Serious House on Serious Earth,Grant Morrison-Dave McKean,4.1,1401204252,9.7814E+12,eng,216,50214,1392,
22375,New X-Men: New Worlds (New X-Men #3),Grant Morrison-Igor Kordey-Ethan Van Sciver,3.85,1904159087,9.7819E+12,eng,160,2,0,
22380,JLA: Tierra 2,Grant Morrison-Frank Quitely-Rafael De la Iglesia,3.96,987224975X,9.78987E+12,spa,96,6,2,
22381,JLA Vol. 4: Strength in Numbers,Grant Morrison-Mark Wald-Howard Porter-John Dell,3.91,1417658584,9.78142E+12,eng,222,15,2,
```

## Union of two documents.

```
run> un = UNION doc1, doc2;
```

## # OUTPUT

```
23060,The Fashion System,Roland Barthes-Matthew Ward-Richard Howard,3.57,520071778,9.78052E+12,eng,351,219,10,
23061,The Responsibility of Forms: Critical Essays on Music Art and Representation,Roland Barthes-Richard Howard,3.94,520072383,9.78052E+12,eng,320,92,3,
23063,The Pleasure of the Text,Roland Barthes-Richard Miller,4.02,224011871,9.78022E+12,eng,80,19,1,
23065,How to Travel with a Salmon & Other Essays,Umberto Eco-William Weaver,3.85,99428636,9.7801E+12,eng,229,106,11,
23066,How to Travel with a Salmon and Other Essays,Umberto Eco-William Weaver,3.85,151001367,9.78015E+12,eng,248,3396,148,
23079,Der Name der Rose,Umberto Eco-Burkhard Kroeber,4.12,3423105518,9.78342E+12,ger,657,646,19,
23080,Le Nom de la rose,Umberto Eco-Jean-Noël Schifano,4.12,2253033138,9.78225E+12,fre,755,665,43,
23082,Nachschrift zum »Namen der Rose«,Umberto Eco-Burkhard Kroeber,4.05,3446140379,9.78345E+12,ger,95,9,1,
23084,Der Name Der Rose: Königs Erläuterungen Und Materialien 00391,Cerstin Urban-Umberto Eco,3.3804416276,9.7838E+12,ger,86,3,0,
23089,The Carpenter's Pencil,Manuel Rivas-Jonathan Dunne,3.6,99448467,9.7801E+12,eng,165,628,53,
23090,¿Qué me quieres amor?,Manuel Rivas,3.62,8420482277,9.78842E+12,glg,216,450,19,
23108,O lapis do carpinteiro,Manuel Rivas,3.6,8483022613,9.78848E+12,glg,152,181,13,
23112,A man dos painos,Manuel Rivas,3.46,8483025914,9.78848E+12,glg,142,24,0,
23114,Os libros arden mal,Manuel Rivas,3.35,849782461X,9.7885E+12,glg,741,33,2,
23129,El perquè de tot plegat,Quim Monzó,3.8,8477271054,9.78848E+12,cat,176,333,22,
23139,Ochenta y Seis Cuentos,Quim Monzó,4.03,8433924788,9.78843E+12,spa,584,44,4,
23147,Guadalajara,Quim Monzó,3.77,847727164X,9.78848E+12,cat,199,28,2,
23155,Self Service,Quim Monzó-Biel Mesquida,2,8473110102,9.78847E+12,cat,256,1,0,
23158,The Age of Access: The New Culture of Hypercapitalism Where All of Life is a Paid-For Experience,Jeremy Rifkin-Ralph Fowler,3.84,1585420824,9.78159E+12,eng,320,215,12,
23183,The Book of Questions,Pablo Neruda-William O'Daly,4.05,1556591608,9.78156E+12,eng,96,2086,213,
23186,Bugs for Lunch/Insectos para el almuerzo,Margery Facklam-Sylvia Long-Liliana Valenzuela,3.78,1570915067,9.78157E+12,spa,32,4,1,
23194,My Story as told by Water: Confessions Druidic Rants Reflections Bird-watchings Fish-stalkings Visions Songs and Prayers Refracting Light from Living Rivers in the Age of the Industrial Dark
,David James Duncan,4.08,1578050499,9.78158E+12,eng,304,665,52,
23201,Pilgrims and Other Stories,Elizabeth Gilbert,3.12,330351745,9.78033E+12,eng,288,1352,173,
23202,The Last American Man,Elizabeth Gilbert,3.8,142002836,9.78014E+12,eng,271,7634,934,
23203,Pilgrims Pa,Elizabeth Gilbert,3.12,395924055,9.7804E+12,eng,224,10,2,
23204,Stern Men,Elizabeth Gilbert,3.39,061812733X,9.78062E+12,eng,304,4065,521,
23206,A Writer's Workbook: Daily Exercises for the Writing Life,Caroline Sharp-Elizabeth Gilbert,3.71,031228621X,9.78031E+12,eng,176,148,15,
23217,Size 14 Is Not Fat Either (Heather Wells #2),Meg Cabot,3.83,60525126,9.78006E+12,eng,344,26923,1015,
23218,Missing You (1-800-Where-R-You #5),Meg Cabot,4.17,60874309,9.78006E+12,eng,268,10886,365,
23219,Queen of Babble (Queen of Babble #1),Meg Cabot,3.69,60851988,9.78006E+12,eng,309,59613,1675,
23220,Size 12 Is Not Fat (Heather Wells #1),Meg Cabot,3.74,60525118,9.78006E+12,eng,345,73592,2360,
23221,How to Be Popular,Meg Cabot,3.54,60880120,9.78006E+12,eng,288,19183,879,
23222,Valentine Princess (The Princess Diaries #4.25),Meg Cabot,3.67,60847182,9.78006E+12,eng,96,9330,211,
23223,Every Boy's Got One (Boy #3),Meg Cabot,3.77,330418882,9.78033E+12,eng,328,20562,734,
23224,Darkest Hour (The Mediator #4),Meg Cabot,4.17,60725141,9.78006E+12,eng,316,30018,657,
23225,Haunted (The Mediator #5),Meg Cabot,4.15,60751649,9.78006E+12,eng,263,31501,652,
23226,She Went All the Way,Meg Cabot,3.69,60805444,9.78006E+12,eng,357,12297,432,
23227,Are We There Yet?,David Levithan,3.5,037582846X,9.78038E+12,eng,215,3048,277,
23228,Boy Meets Boy,David Levithan,3.84,375832990,9.78038E+12,eng,185,57087,3242,
23229,This Is Push: New Stories from the Edge,David Levithan-Patricia McCormick-Matthue Roth-Kevin Waltman-Samantha Schutz-Coe Booth-Eddie de Oliveira-Tanuja Desai Hidler-Kevin Brooks-Chris Wooding-Markus
Zusak-Brian James-Kristen Kemp-Eireann Corrigan-Christopher Krovatin-Billy Merrell,3.62,439890284,9.78044E+12,eng,240,222,23,
23230,Naomi and Ely's No Kiss List,Rachel Cohn-David Levithan,3.44,375844406,9.78038E+12,eng,230,13620,978,
23232,The Realm of Possibility,David Levithan,3.93,375836578,9.78038E+12,eng,210,14438,1176,
23233,Ten Things I Hate about You,David Levithan,4.11,439087309,9.78044E+12,eng,170,1414,62,
23235,Where We Are What We See: The Best Young Writers and Artists in America,David Levithan,3.73,439736463,9.78044E+12,eng,288,41,0,
23236,Aphrodite's Passion (Superhero Central #2),Julie Kenner,3.62,505524740,9.78051E+12,eng,352,524,33,
23237,The Manolo Matrix (Codebreaker Trilogy #2),Julie Kenner,3.61,743496140,9.78074E+12,eng,352,635,34,
23238,The Cat's Fancy,Julie Kenner,3.58,505523973,9.78051E+12,eng,391,243,27,
23239,Demons Are Forever (Demon-Hunting Soccer Mom #3),Julie Kenner,3.89,425215385,9.78043E+12,eng,292,2343,164,
23240,Aphrodite's Secret (Superhero Central #3),Julie Kenner,3.7,505525097,9.78051E+12,eng,369,405,20,
23242,The Good Chouls' Guide to Getting Even (Beth Frasier #1),Julie Kenner,3.58,425213919,9.78043E+12,eng,256,876,71,
23243,Aphrodite's Flame (Superhero Central #4),Julie Kenner,3.67,505525755,9.78051E+12,eng,338,379,16,
23244,Carpe Demon (Demon-Hunting Soccer Mom #1),Julie Kenner,3.63,515142212,9.78052E+12,eng,320,4750,581,
23249,Perfect Circle,Sean Stewart,3.8,1931520119,9.78193E+12,eng,243,666,67,
23257,The Great Philosophers: An Introduction to Western Philosophy,Bryan Magee,4.06,019289322X,9.78019E+12,eng,352,333,16,
```



## CMPARISION STUDY

- **Difference between execution time between hive and pig. Comparison between performance of hive and pig while finding the Book title starting with the letter 'H'.**

Therefore Hive perform the query in 3.745 seconds and pig perform it in the 4 seconds. Therefore instead of being the most suitable language for large dataset, it take more time as compared to hive.

```
Job Stats (time in seconds):
JobId  Maps    Reduces MaxMapTime    MinMapTime    AvgMapTime    MaxReduceTime    MinReduceTime    A
job_201903280644_0005  1      0      4      4      4      0      0      0      doc1,f  MAP_ONLY

Input(s):
Successfully read 182 records (6351 bytes) from: "/user/training/poc2/P0C2.csv"

Output(s):
Successfully stored 31 records (1280 bytes) in: "hdfs://localhost/tmp/temp-294678962/tmp-1265818768"
```

- **Comparison between map reduce and pig for finding word count of the document**

```
[training@localhost ~]$ hadoop fs -mkdir word
[training@localhost ~]$ hadoop fs -put '/home/training/Desktop/word.txt' word
[training@localhost ~]$ pig
```

```
grunt> l =load '/user/training/word/word.txt' as(l:chararray);
grunt> t = foreach l generate flatten(tokenize(l)) as t:chararray;
2019-03-31 11:00:03,390 [main] ERROR org.apache.pig.tools.grunt.Grun
70: Could not resolve tokenize using imports: [, org.apache.pig.buil
ache.pig.impl.builtin.]
Details at logfile: /home/training/pig_1554055059668.log
grunt> t = FOREACH l GENERATE flatten(TOKENIZE(l)) as t:chararray;
grunt> letter = FOREACH t GENERATE SUBSTRING(t,0,1) AS letter:charar
grunt> lgroup = GROUP letter by letter;
grunt> c= FOREACH lgroup GENERATE group, COUNT(letter);
grunt> order = ORDER c BY $1 DESC;
```

(A, 100)	
(B, 125)	(K, 11)
(C, 121)	(L, 15)
(D, 60)	(m, 11)
(E, 22)	(h, 21)
(F, 20)	(r, 20)
(G, 18)	(u, 10)
(H, 22)	
(I, 31)	(g, 5)
(J, 30)	(a, 8)
(K, 11)	(t, 12)
(L, 15)	(o, 15)
(m, 11)	
(h, 21)	(p, 12)
(r, 20)	(x, 6)
(u, 10)	(h, 19)
(g, 5)	
(a, 8)	(q, 9)
(t, 12)	(z, 11)
(o, 15)	(b, 16)
(p, 12)	(k, 15)
(x, 6)	
(h, 19)	(S, 12)
(q, 9)	(X, 0)

# MAP REDUCE STATEMENTS

- WORD COUNT OF THE FILE

## COMMAND:

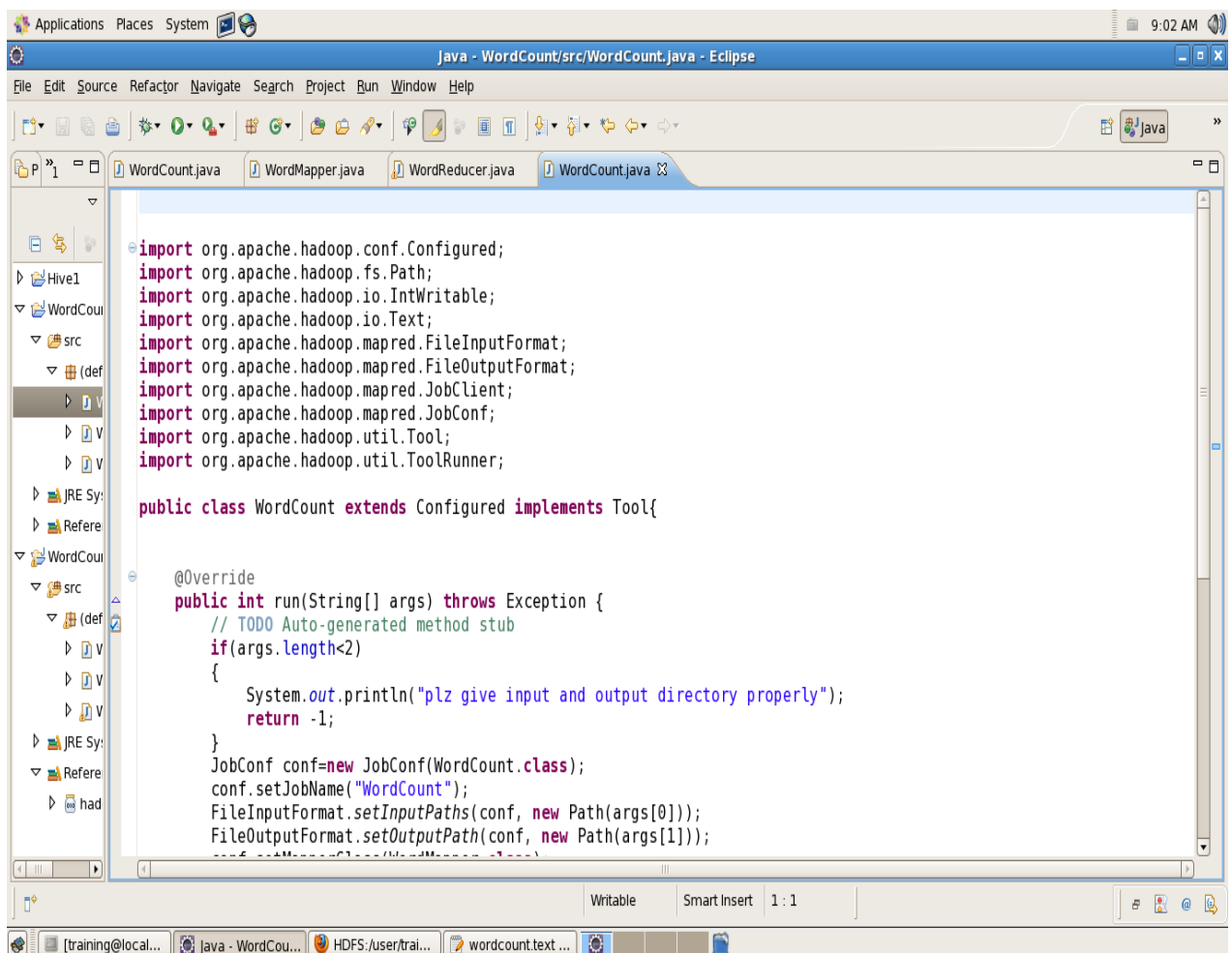
1. Create a Java Project :Word Count
2. Add External Jar files:

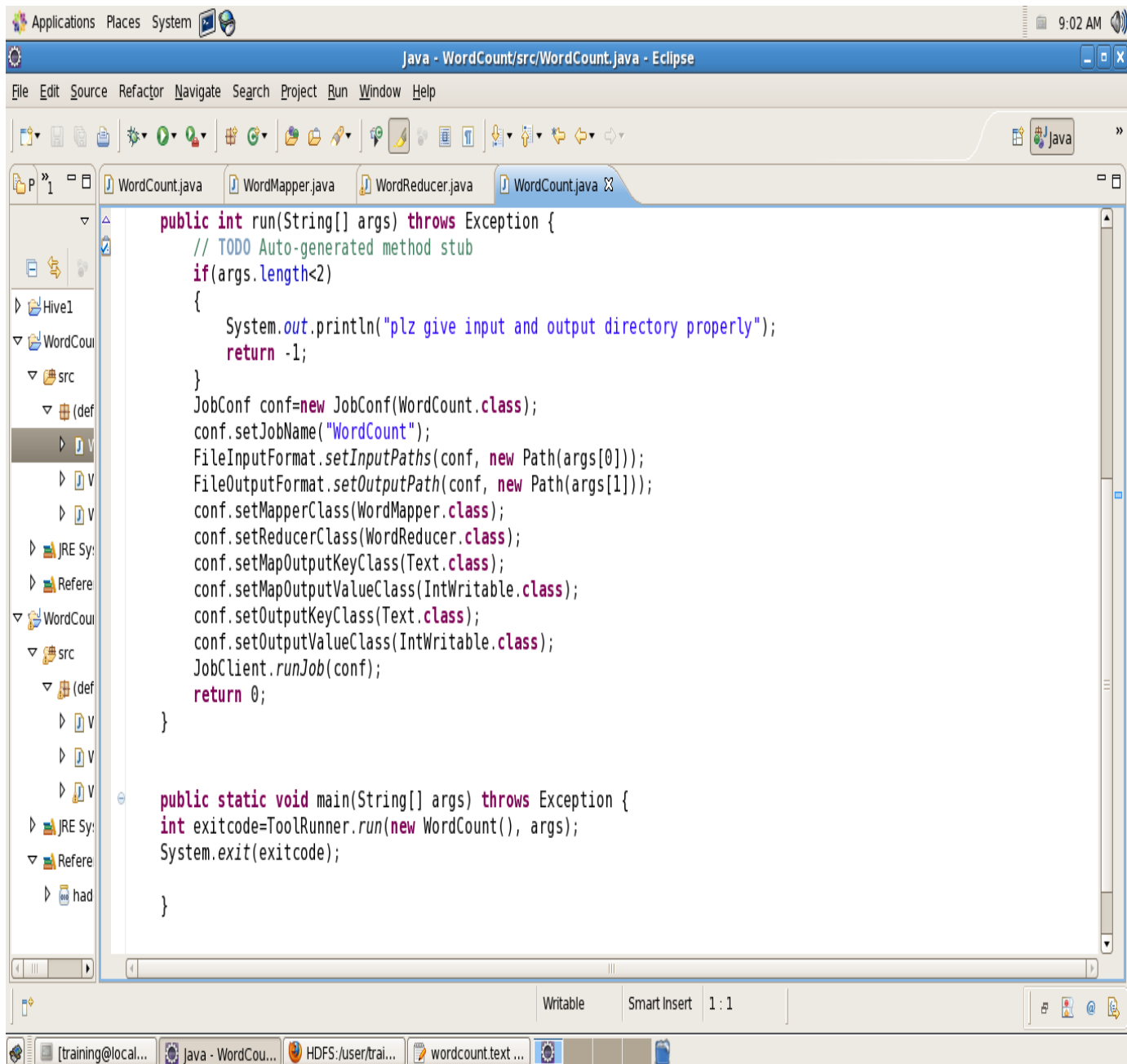
Right click on project → Properties → Java BuildPath → Add External jar Files →

Then:

Click on File System → usr → lib → hadoop-0.20 → hadoop-core.jar → ok → ok

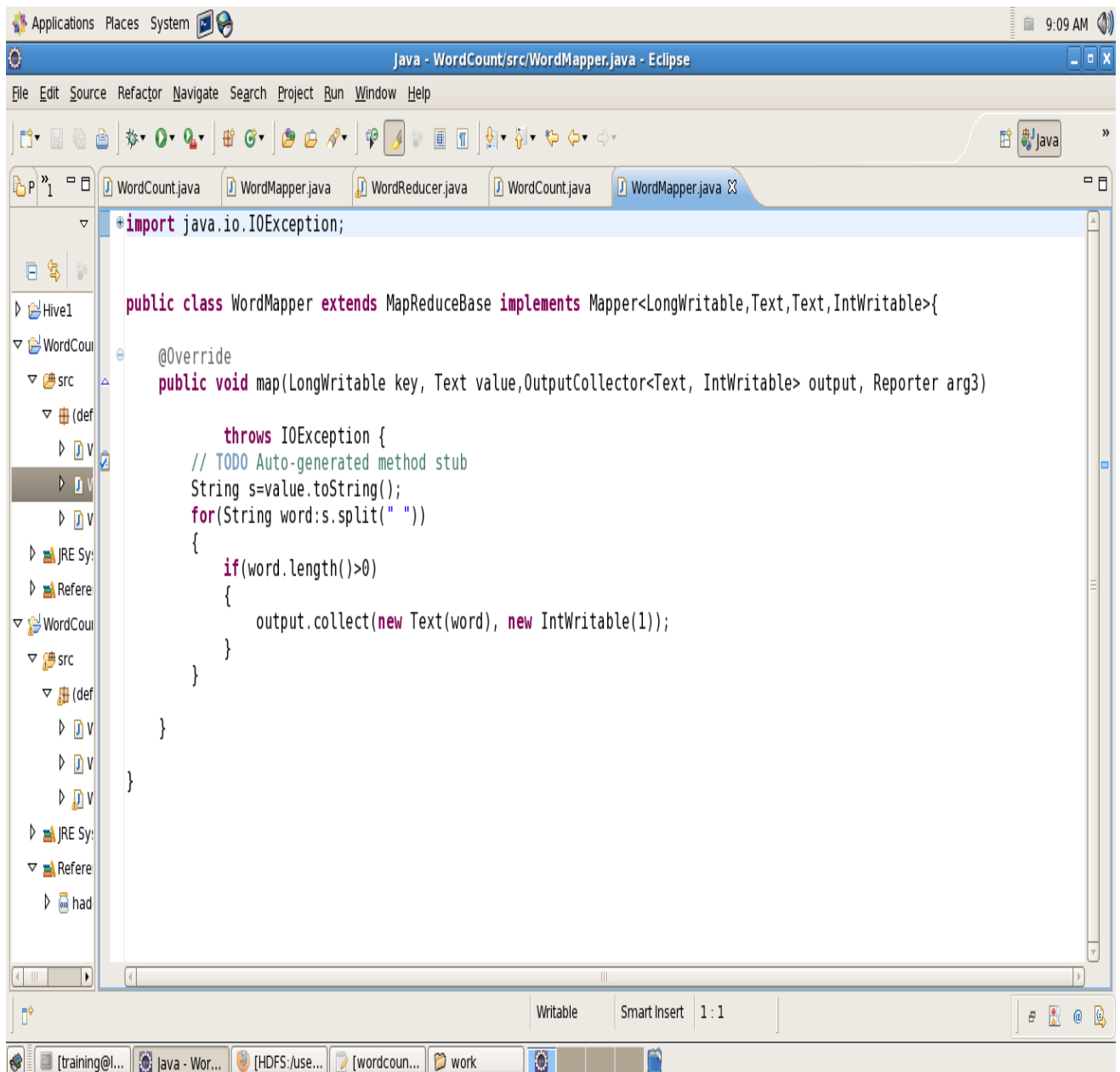
3. Create a class WordCount.java i.e. your Driver Code
4. Create a class Mapper i.e. WordMapper.java
5. Create a class Reducer i.e. WordReducer.java







## REDUCER CLASS

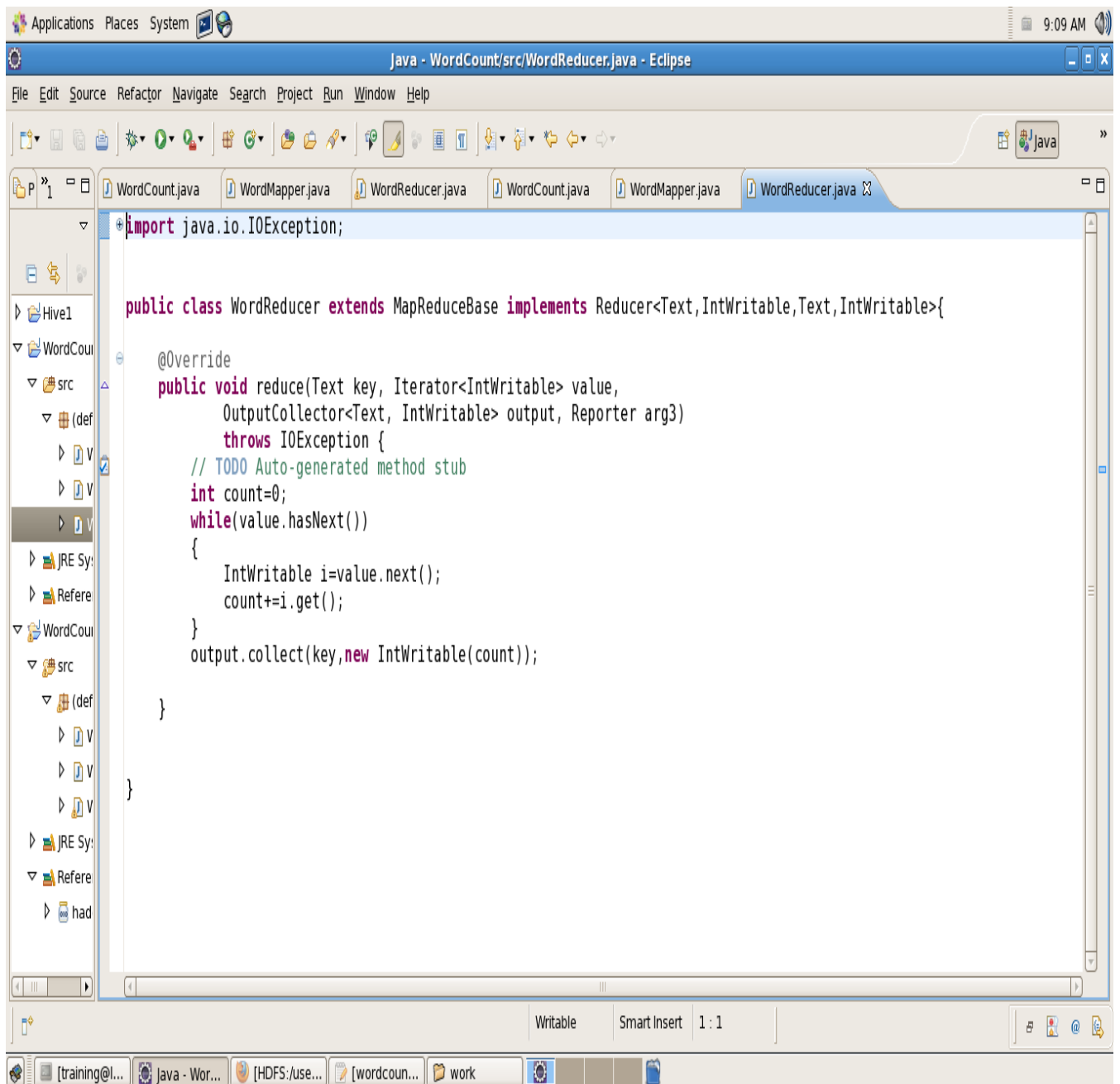


```
import java.io.IOException;

public class WordMapper extends MapReduceBase implements Mapper<LongWritable,Text,Text,IntWritable>{

    @Override
    public void map(LongWritable key, Text value,OutputCollector<Text, IntWritable> output, Reporter arg3)
        throws IOException {
        // TODO Auto-generated method stub
        String s=value.toString();
        for(String word:s.split(" "))
        {
            if(word.length(>0)
            {
                output.collect(new Text(word), new IntWritable(1));
            }
        }
    }
}
```

# MAPPER CLASS



```
import java.io.IOException;

public class WordReducer extends MapReduceBase implements Reducer<Text,IntWritable,Text,IntWritable>{

    @Override
    public void reduce(Text key, Iterator<IntWritable> value,
        OutputCollector<Text, IntWritable> output, Reporter arg3)
        throws IOException {
        // TODO Auto-generated method stub
        int count=0;
        while(value.hasNext())
        {
            IntWritable i=value.next();
            count+=i.get();
        }
        output.collect(key,new IntWritable(count));
    }
}
```

## #OUTPUT

(A, 100)	(N, 3)	
(B, 125)	(P, 4)	
(C, 121)	(R, 2)	
(D, 60)	(S, 3)	
(E, 22)	(T, 4)	
(F, 20)	(a, 25)	.
(G, 18)	(b, 2)	
(H, 22)	(c, 14)	
(I, 31)	(d, 6)	
(J, 30)	(e, 7)	
(K, 11)	(f, 22)	
(L, 15)	(g, 5)	
(m, 11)	(h, 4)	
(n, 21)	(i, 9)	
(o, 20)	(j, 1)	
(p, 10)	(k, 4)	
(q, 5)	(l, 2)	
(r, 8)	(m, 3)	
(s, 12)	(n, 16)	
(t, 15)	(o, 7)	
(u, 5)	(p, 4)	
(v, 8)	(q, 8)	
(w, 12)	(r, 29)	
(x, 15)	(s, 4)	
(y, 12)	(t, 3)	
(z, 6)	(u, 10)	
(h, 19)	(v, 1)	
(q, 9)	(w, 1)	
	(x, 0)	■