**MANAV RACHNA INTERNATIONAL**

**INSTITUTE OF REASEARCH & STUDIES**

**FACULTY OF ENGINEERING AND TECHNOLOGY**



**PRACTICAL FILE**

**FOR**

**VIIth SEMESTER (ACADEMIC YEAR 2022-2023)**

**SUBJECT NAME: ITNS Lab**

**SUBJECT CODE: BCS-DS-730**

|  |  |
| --- | --- |
| **SUBMITTED BY:-** | **SUBMITTED TO:-** |
| **STUDENT NAME: Siddharth kapoor** | **FACULTY NAME: Ms. Rachna** |
| **ROLL NO.: 1/19/FET/BCS/025** | **DEPARTMENT: FET** |
|  |  |

**PRACTICAL –1**

**AIM-**To install VMware Workstation on your system

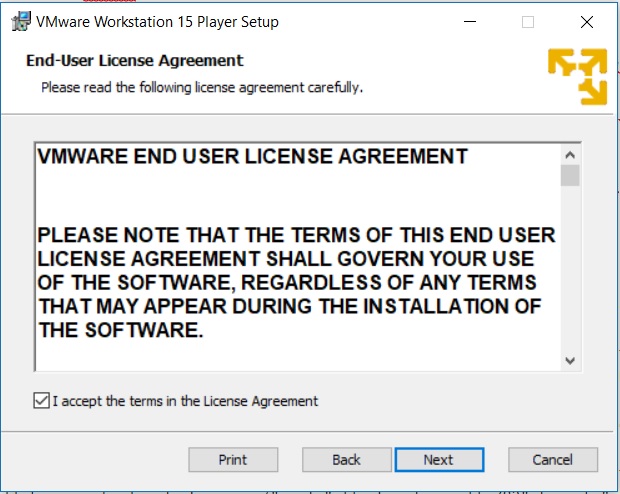
**PROCEDURE-**

VMware Player 15 Installation –

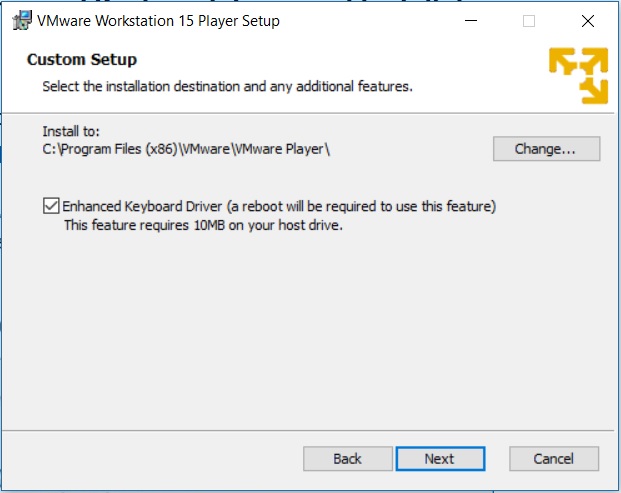
[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware-player-15-installation-setup-wizard/)

VMware Player 15 Installation – Setup Wizard

Click next and accept the license terms and click next again to move on to the next screen.

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_end_user_agreement_screenshot/)

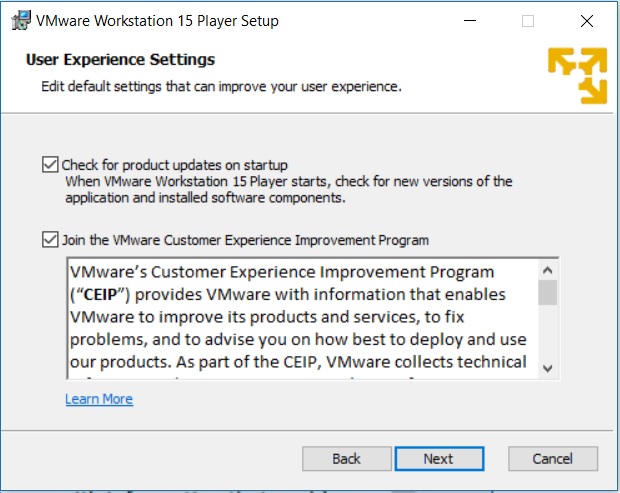
**Step 2 – Custom setup – Enhanced Keyboard driver and Installation directory**

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_custom_setup_enhanced_keyboard__screenshot/)

VMware Player 15 Installation – Custom Setup – Enhanced Keyboard Driver

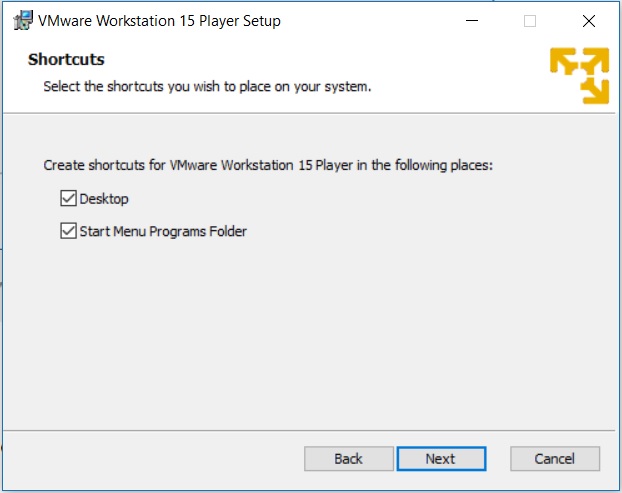
**Step 3 – User Experience Settings**

Check the options for Check the product update at Startup and Join the VMware Customer Program. I normally leave it as it is. You can unchecked it if you so desire. Click next

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_user_experience_settings/)

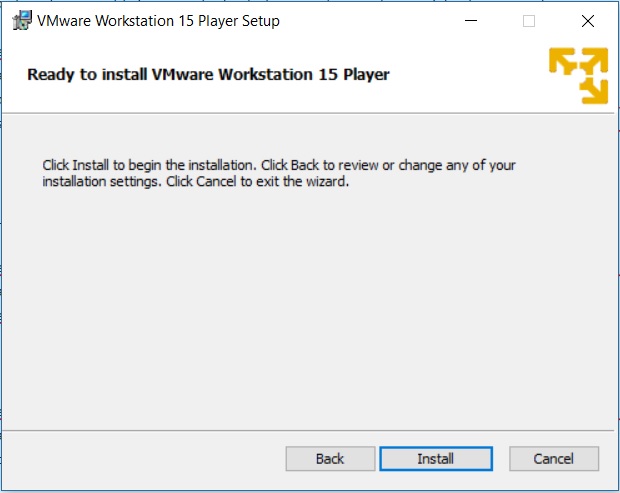
**Step 4 – Select where the shortcuts will be installed**

Check the box where the shortcut to run the application will be created. I leave it as it is. Click on next.

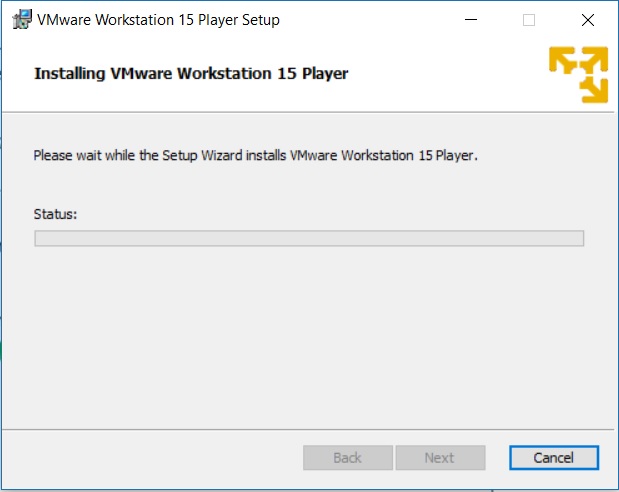
[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_shortcuts_screenshot/)

**Step 5 – Ready to install**

Now the installation wizard is ready to install. Click on install to begin the installation.

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_ready_to_install_screenshot/)

VMware Player 15 Installation – Ready to Install,Installation begins, wait for it to complete.

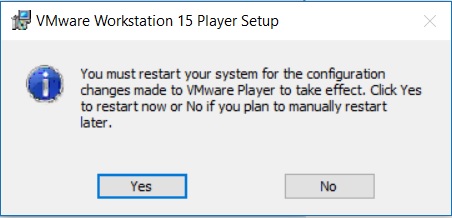
[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard__installation_screenshot/)

After sometime, you will see installation compete message. You are done.

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_installation_complete_screenshot/)

VMware Player 15 Installation – Installation Complete

Click on Finish to Complete the installation.

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_wizard_restart_screenshot/)

VMware Player 15 Installation – Reboot Required

**Step 6 – License**

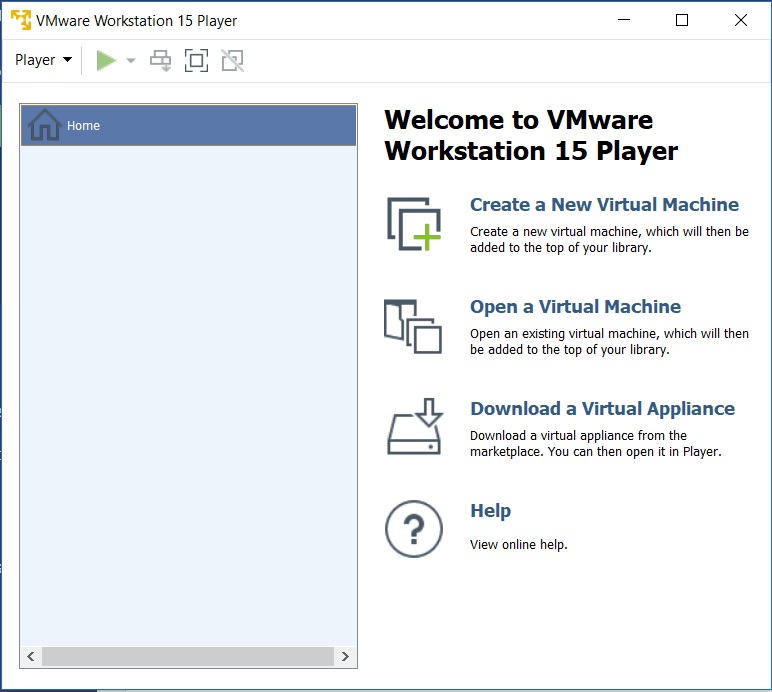
Once you run the application for the first time, you will be asked for licence. Select the option Use VMware Workstation Player 15 for for free for non commercial use.

Click continue.

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_license_screenshot/)[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_installation_setup_finish/)

VMware Player 15 Installation – License,Click on Finish.

Now you will see VMware Workstation Player 15 ready to be used for free for non-commercial purpose.

[](https://www.shaileshjha.com/step-by-step-install-vmware-workstation-player-12-in-windows-10/vmware_player_15_screenshot-2/)

VMware Player 15 – Home Screen

**PRACTICAL –2**

**AIM:-**Overview and getting started with Hadoop and Big Insights

**PROCEDURE:-**In this lab, we work on the InfoSphereBigInsights 3.0 Quick Start Edition VMware image.

The VMware image is set up in the following manner:

|  |  |  |
| --- | --- | --- |
|  | User | Password |
| VM Image root account | root | password |
| VM Image lab user account | biadmin | biadmin |
| BigInsights Administrator | biadmin | Biadmin |
| Big SQL Administrator | bigsql | Bigsql |
| Lab user | biadmin | Biadmin |

|  |  |
| --- | --- |
| Property | Value |
| Host name | bivm.ibm.com |
| BigInsights Web Console URL | http://bivm.ibm.com:8080 |
| Big SQL database name | Bigsql |
| Big SQL port number | 51000 |

Step: To install and launch the VMware image as well as start the required services.

1. Obtain a copy of the BigInsights 3.0 Quick Start Edition VMware image from [IBM’s external download site](http://www-01.ibm.com/software/data/infosphere/biginsights/quick-start/downloads.html)

(<http://www-01.ibm.com/software/data/infosphere/biginsights/quick-start/downloads.html>). Use the image for the single-node cluster.

1. Follow the instructions provided to decompress (unzip) the file and install the image on your laptop. Note that there is a README file with additional information.

Graphical user interface, application

Description automatically generated

1. Install VMware player or other required software to run VMware images. Details are in the README file provided with the BigInsights VMware image.

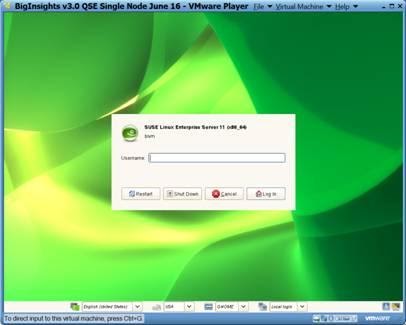


1. Launch the VMware image. When logging in for the first time, use the root ID (with a password of password). Follow the instructions to configure your environment, accept the licensing agreement, and enter the passwords for the root and biadmin IDs (root/password and biadmin/biadmin) when prompted. This is a one-time only requirement.

Text

Description automatically generated

1. When the one-time configuration process is completed, you will be presented with a SUSE Linux log in screen. Log in as biadmin with a password of biadmin.



1. Verify that your screen that appears.



1. Click Start BigInsights to start all required services. Wait until the operation completes. This may take several minutes, depending on your machine’s resources.



1. Verify that all required BigInsights services are up and running. From a terminal window, issue this command:$BIGINSIGHTS\_HOME/bin/status.sh.



1. Inspect the results, a subset of which are shown below. Verify that, at a minimum, the following components started successfully:hdm, zookeeper, hadoop, catalog, hive, bigsql, oozie, console, and httpfs.

**PRACTICAL –3**

**AIM:-Introduction to Hadoop.**

**PROCEDURE:-**Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed *storage* and *computation* across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

## **Hadoop Architecture**

At its core, Hadoop has two major layers namely −

* Processing/Computation layer (MapReduce), and
* Storage layer (Hadoop Distributed File System).



## **MapReduce**

MapReduce is a parallel programming model for writing distributed applications devised at Google for efficient processing of large amounts of data (multi-terabyte data-sets), on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. The MapReduce program runs on Hadoop which is an Apache open-source framework.

## **Hadoop Distributed File System**

The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. It is highly fault-tolerant and is designed to be deployed on low-cost hardware. It provides high throughput access to application data and is suitable for applications having large datasets.

Apart from the above-mentioned two core components, Hadoop framework also includes the following two modules −

* **Hadoop Common** − These are Java libraries and utilities required by other Hadoop modules.
* **Hadoop YARN** − This is a framework for job scheduling and cluster resource management.

**PRACTICAL –3**

**AIM:-**WAP to count words in a file using a Map Reduce application

**PROCEDURE:-Word Count MapReduce Program in Hadoop**

The first MapReduce program most of the people write after [installing Hadoop](https://netjs.blogspot.com/2018/02/installing-hadoop-single-node-pseudo-distributed-mode.html) is invariably the word count MapReduce program.

The detailed steps for writing word count MapReduce program in Java, IDE used is Eclipse.

**Creating and copying file**

1. If you already have a file in HDFS which you want to use as input then you can skip this step.
2. First thing is to create a file which will be used as input and copy it to HDFS.
3. Let’s say you have a file wordcount.txt with the following content.
4. Hello wordcount MapReduce Hadoop program.
5. This is my first MapReduce program.
6. You want to copy this file to /user/process directory with in HDFS. If that path doesn’t exist then you need to create those directories first.

hdfsdfs -mkdir -p /user/process

1. Then copy the file wordcount.txt to this directory.

hdfsdfs -put /netjs/MapReduce/wordcount.txt /user/process

**Word count example MapReduce code in Java**

1. Write your wordcount MapReduce code. WordCount example reads text files and counts the frequency of the words. Each mapper takes a line of the input file as input and breaks it into words. It then emits a key/value pair of the word (In the form of (word, 1)) and each reducer sums the counts for each word and emits a single key/value with the word and sum.
2. In the word count MapReduce code there is a Mapper class (MyMapper) with map function and a Reducer class (MyReducer) with a reduce function.
3. You will also need to add at least the following Hadoop jars so that your code can compile. You will find these jars inside the /share/hadoop directory of your Hadoop installation. With in /share/hadoop path look in hdfs, mapreduce and common directories for required jars.

hadoop-common-2.9.0.jar

hadoop-hdfs-2.9.0.jar

hadoop-hdfs-client-2.9.0.jar

hadoop-mapreduce-client-core-2.9.0.jar

hadoop-mapreduce-client-common-2.9.0.jar

hadoop-mapreduce-client-jobclient-2.9.0.jar

hadoop-mapreduce-client-hs-2.9.0.jar

hadoop-mapreduce-client-app-2.9.0.jar

commons-io-2.4.jar

**Creating jar of your wordcount MapReduce code**

Once you are able to compile your code you need to create jar file. In the eclipse IDE righ click on your Java program and select Export – Java – jar file.

**Running the code**

You can use the following command to run the program in your hadoop installation directory.

bin/hadoop jar /netjs/MapReduce/wordcount.jar org.netjs.WordCount  /user/process /user/out

/netjs/MapReduce/wordcount.jar is the path to your jar file.

org.netjs.WordCount is the fully qualified path to your Java program class.

/user/process – path to input directory.

/user/out – path to output directory.

One your word count MapReduce program is succesfully executed you can verify the output file.

hdfsdfs -ls /user/out

Found 2 items

-rw-r--r--   1 netjs supergroup          0 2018-02-27 13:37 /user/out/\_SUCCESS

-rw-r--r--   1 netjs supergroup         77 2018-02-27 13:37 /user/out/part-r-00000

As you can see Hadoop framework creates output files using part-r-xxxx format. Since only one reducer is used here so there is only one output file part-r-00000.

Graphical user interface, text, application

Description automatically generated

**PRACTICAL-4**

**AIM:-HBase shell command**

**PROCEDURE:-HBase Shell**

HBase contains a shell using which you can communicate with HBase. HBase uses the Hadoop File System to store its data. It will have a master server and region servers. The data storage will be in the form of regions (tables). These regions will be split up and stored in region servers.

The master server manages these region servers and all these tasks take place on HDFS. Given below are some of the commands supported by HBase Shell.

**Creating a Table using HBase Shell**

You can create a table using the create command, here you must specify the table name and the Column Family name. The syntax to create a table in HBase shell is shown on left page.

## Listing a Table using HBase Shell

list is the command that is used to list all the tables in HBase. Given on the left page is the syntax of the list command.

## Disabling a Table using HBase Shell

To delete a table or change its settings, you need to first disable the table using the disable command. You can re-enable it using the enable command.

Given below is the syntax to disable a table:

## Enabling a Table using HBase Shell

Syntax to enable a table:

## describe

This command returns the description of the table. Its syntax is as follows:

## alter

Alter is the command used to make changes to an existing table. Using this command, you can change the maximum number of cells of a column family, set and delete table scope operators, and delete a column family from a table.

### Changing the Maximum Number of Cells of a Column Family

Given below is the syntax to change the maximum number of cells of a column family.

## Deleting a Specific Cell in a Table

Using the delete command, you can delete a specific cell in a table. The syntax of delete command is as follows:

## Inserting Data using HBase Shell

This chapter demonstrates how to create data in an HBase table. To create data in an HBase table, the following commands and methods are used:

* put command,
* add() method of Put class, and
* put() method of HTable class.

### Inserting the First Row

Let us insert the first-row values into the emp table as shown on the left page.

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

A picture containing table

Description automatically generated

Graphical user interface, application

Description automatically generated

**PRACTICAL-5**

**AIM:-JAQL shell commands**

**PROCEDURE:-Jaql basics**

**Statement, Assignment and Comments**

Double and single quotes are treated the same. Semicolon terminates a statement.

jaql> "Hello world";  
"Hello world"

jaql> a = 10\*2;  
jaql> a;  
20

jaql> // This is a comment  
jaql> /\* and this is also  
         a comment \*/

Graphical user interface, text, application

Description automatically generatedShape, rectangle

Description automatically generated

**Data Types**

Jaql is a loosely typed functional language, with lazy evaluation. Type is usually inferred by how a value is provided. Many types have a function of the same name to force conversions of a value or variable (e.g. string(), double()).

null – null

boolean – true, false

string – “hi”

long – 10

double – 10.2, 10d, 10e-2

array – [1, 2, 3]

record – {a : 1, b : 2}

others as jaql extensions – decfloat, binary, date, schema, function, comparator, regex

jaql>num = 10;  
jaql> array = [1, 2, 3, "hello", 4, {color: "red"}];  
jaql> array;  
[  
  1,  
  2,  
  3,  
  "hello"  
  ...

**Operators**

arithmetic (+, -, /, \*)

boolean (and, or, not)

comparison (==, !=, <, >, in, isnull)

**Arrays**

Arrays can be accessed with the [] operator.

jaql> a = [1, 2, 3];  
jaql>a[1]; //retrieves start from zero  
2  
jaql> a[1:2]; //retrieve a range/subset  
[2,3]

But you can’t change a value, you need to use function replaceElement().

jaql> a = replaceElement(a, 1, 10);  
jaql> a[1];  
10

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generated

**Other array functions**

count() – count(['a', 'b', 'c']); returns 3

index() – index(['a', 'b', 'c'], 1); returns "b"

replaceElement() – replaceElement(['a', 'b', 'c'], 1, 'z'); returns [ "a", "z", "c" ]

slice() – index(['a', 'b', 'c', 'd'], 1, 2); returns [ "b", "c" ]

reverse() – reverse(['a', 'b', 'c']); returns [ "c", "b", "a" ]

range() – range(2, 5); returns [ 2, 3, 4, 5 ]

And many more – see [BigInsights Information Center](http://pic.dhe.ibm.com/infocenter/bigins/v2r1/topic/com.ibm.swg.im.infosphere.biginsights.jaql.doc/doc/c0057497.html)

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, Word

Description automatically generated

**Records**

Records are delineated by {} and contains a comma separated list of name:value pairs. Fields are then accessed by “.” operator.

jaql> a = { name : "scott", age : 42, children : ["jake", "sam"] };  
jaql> a.name;  
"scott"  
jaql>a.children[0];  
"jake"

Again you cannot change an existing record, but you can produce a new one:

jaql> a = { a.name, age : 37, a.children };

**The -> operator**

The -> operator “streams” an array through a function or core operator.

jaql>range(10) -> batch(5);  
[  
  [0, 1, 2, 3, 4],  
  [5, 6, 7, 8, 9]  
]

The array on the left is implicitly passed as the first argument to the function. This is identical to the above:

jaql> batch(range(10), 5);

Operator -> is just a “syntactic sugar” that can dramatically improve readability when multiple operations are involved on your data.

Graphical user interface, text, application, Word

Description automatically generatedGraphical user interface, application

Description automatically generated

**Input/Output with Jaql**

Input/Output operations are performed through I/O adapters. Adapters are a description of how to access and process a data source. I/O adapter is then passed into I/O function (e.g. read() or write()). Following is an example of writing array into a (comma-)delimited csv file where we used delimited file I/O adapter:

jaql> [1, 2, 3, 4, 5] -> write(del("test.csv"));  
jaql> read(del("test.csv"));  
[1, 2, 3, 4, 5]

Local files or HDFS are then accessed by specyfing the full path as URI:

jaql> read(del("file:///home/user/test.csv")); // for local file system  
jaql> read(del("hdfs://localhost:9000/user/test.csv")); // for hdfs file system

To read a JSON data from a URL, you can use jaglGet() function:

jaqlGet("file:///tmp/test.txt");

Text, letter

Description automatically generated

**Data manipulation (Core operators).**

Core operators manipulate streams (arrays) of data, much in the way SQL clauses interact with data.

**Filter**

$ represents the current array value being evaluated.

jaql> read(del("file:///path/to/people.txt")) -> filter $.name == "Fred";  
[  
  { fname: "Fred", lname: "Johnson", age: 20 }  
]

Other example could be:

jaql> data = [1, 2, 3, 4, 5, 6, 7, 8, 9];  
jaql> data -> filter 3 <= $ <= 6;  
[ 3, 4, 5, 6 ]

Alternatively, the each clause can be used to provide a name different than $:

jaql> data = [1, 2, 3, 4, 5, 6, 7, 8, 9];  
jakq> data -> filter each num (3 <= num<= 6);  
[ 3, 4, 5, 6 ]

Text, letter

Description automatically generatedA picture containing text

Description automatically generated

**Transform**

The transform operator allows you to manipulate the values in an array. An expression is applied to each element in the array:

jaql> recs = [ {a: 1, b: 4}, {a: 2, b: 5}, {a: -1, b: 4} ];  
jaql> recs -> transform $.a + $.b;  
[ 5, 7, 4 ]  
jaql> recs -> transform { sum: $.a + $.b };  
[ { sum: 5 }, { sum: 7 }, { sum: 3 } ]

A picture containing graphical user interface

Description automatically generated       Text

Description automatically generated with medium confidence

**Sort**

jaql> read(del("file:///path/to/people.txt")) -> sort by [$.ageasc];

Graphical user interface, text, application

Description automatically generatedText

Description automatically generated with medium confidence

**Other data manipulation operators**

Other data manipulation operators are expand, group, join, top.

**Join**

Text, letter

Description automatically generated

**Top**

Text

Description automatically generated

**Group**

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated