

ScreenShots:

Single Node Setup

Step 1: Choose an Amazon Machine Image (AMI)

Quick Start

AMI Name	Description	Select
Amazon Linux AMI 2017.09.1 (HVM), SSD Volume Type - ami-bf4193c7	The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.	Select
Red Hat Enterprise Linux 7.4 (HVM), SSD Volume Type - ami-9fa343e7	Red Hat Enterprise Linux version 7.4 (HVM), EBS General Purpose (SSD) Volume Type	Select
SUSE Linux Enterprise Server 12 SP3 (HVM), SSD Volume Type - ami-e3ef329b	SUSE Linux Enterprise Server 12 Service Pack 3 (HVM), EBS General Purpose (SSD) Volume Type. Public Cloud, Advanced Systems Management, Web and Scripting, and Legacy modules enabled.	Select
Ubuntu Server 16.04 LTS (HVM), SSD Volume Type - ami-0a0ce72	Ubuntu Server 16.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (http://www.ubuntu.com/cloud/services).	Select
Microsoft Windows Server 2016 Base - ami-343ef84c		Select

Feedback English (US)

© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 6:42 PM

Step 2: Choose an Instance Type

Name	Memory (GiB)	Storage (GiB)	Price (USD)	Network
d2.xlarge	4	30.5	3 x 2048	Yes Moderate Yes
d2.2xlarge	8	61	6 x 2048	Yes High Yes
d2.4xlarge	16	122	12 x 2048	Yes High Yes
d2.8xlarge	36	244	24 x 2048	Yes 10 Gigabit Yes
i2.xlarge	4	30.5	1 x 800 (SSD)	Yes Moderate Yes
i2.2xlarge	8	61	2 x 800 (SSD)	Yes High Yes
i2.4xlarge	16	122	4 x 800 (SSD)	Yes High Yes
i2.8xlarge	32	244	8 x 800 (SSD)	- 10 Gigabit Yes
i3.large	2	15.25	1 x 475 (SSD)	Yes Up to 10 Gigabit Yes
i3.xlarge	4	30.5	1 x 950 (SSD)	Yes Up to 10 Gigabit Yes
i3.2xlarge	8	61	1 x 1900 (SSD)	Yes Up to 10 Gigabit Yes
i3.4xlarge	16	122	2 x 1900 (SSD)	Yes Up to 10 Gigabit Yes

Cancel Previous Review and Launch Next: Configure Instance Details

Feedback English (US)

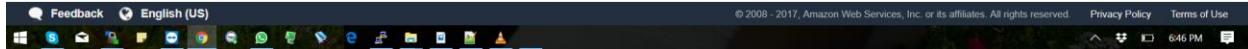
© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 6:45 PM

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot Instances to take advantage of the lower pricing, assign an access management role to the instance, and more.

Number of Instances	1	Launch into Auto Scaling Group
Purchasing option	<input type="checkbox"/> Request Spot Instances	
Network	vpc-ab82eacd (default)	<input type="button"/> Create new VPC
Subnet	No preference (default subnet in any Availability Zone)	<input type="button"/> Create new subnet
Auto-assign Public IP	Use subnet setting (Enable)	
Placement group	No placement group	
IAM role	None	<input type="button"/> Create new IAM role
Shutdown behavior	Stop	
Enable termination protection	<input type="checkbox"/> Protect against accidental termination	
Monitoring	<input type="checkbox"/> Enable CloudWatch detailed monitoring <small>Additional charges apply.</small>	
EBS-optimized instance	<input checked="" type="checkbox"/> Launch as EBS-optimized instance	
Tenancy	Shared - Run a shared hardware instance	<small>Additional charges will apply for dedicated tenancy.</small>

Review and Launch



Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and Instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more about storage options in Amazon EC2.](#)

Volume Type	Device	Snapshot	Size (GiB)	Volume Type	IOPS	Throughput (MB/s)	Delete on Termination	Encrypted
Root	/dev/sda1	snap-0b9c16d670f4e2685	128	General Purpose SSD (GP2)	384 / 3000	N/A	<input checked="" type="checkbox"/>	Not Encrypted
Instance Store 0	/dev/nvme*n1	N/A	N/A	N/A	N/A	N/A	<input type="checkbox"/>	Not Encrypted

Add New Volume

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.



Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver.

A copy of a tag can be applied to volumes, instances or both.

Tags will be applied to all instances and volumes. Learn more about tagging your Amazon EC2 resources.

Key	(127 characters maximum)	Value	(255 characters maximum)	Instances	Volumes
Name		Hadoop		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Add another tag (Up to 50 tags maximum)

Cancel Previous Review and Launch Next: Configure Security Group

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. Learn more about Amazon EC2 security groups.

Type	Protocol	Port Range	Source	Description
SSH	TCP	22	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50030	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50070	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	54310	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	54311	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50040	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50060	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
All ICMP - IPv4	ICMP	0 - 65535	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. Learn more about Amazon EC2 security groups.

Assign a security group:

- Create a new security group
- Select an existing security group

Security group name: launch-wizard-4

Description: launch-wizard-4 created 2017-11-23T18:48:00.997-06:00

Type	Protocol	Port Range	Source	Description
SSH	TCP	22	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50030	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50070	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	54310	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	54311	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50040	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
Custom TCP	TCP	50060	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop
All ICMP - IPv4	ICMP	0 - 65535	Anywhere	0.0.0.0/0, ::/0 e.g. SSH for Admin Desktop

Add Rule Cancel Previous Review and Launch

Feedback English (US) © 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 6:55 PM

Screenshot of the AWS Launch Instance Wizard Step 7: Review Instance Launch.

AMI Details:

- Ubuntu Server 16.04 LTS (HVM), SSD Volume Type - ami-0a00ce72**
- Free tier eligible**
- Description:** Ubuntu Server 16.04 LTS (HVM) EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>)
- Root Device Type:** ebs **Virtualization type:** hvm

Instance Type:

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
i3 large	9	2	15.25	EBS only	Yes	Up to 10 Gigabit

Security Groups:

Type	Protocol	Port Range	Source	Description
SSH	TCP	22	0.0.0.0	
SSH	TCP	22	::/0	
Custom TCP Rule	TCP	50030	0.0.0.0	
Custom TCP Rule	TCP	50030	::/0	
Custom TCP Rule	TCP	50070	0.0.0.0	
Custom TCP Rule	TCP	50070	::/0	

Buttons: Cancel, Previous, Launch

Screenshot of the AWS Launch Instance Wizard Step 7: Review Instance Launch.

AMI Details:

- Ubuntu Server 16.04 LTS (HVM), SSD Volume Type - ami-0a00ce72**
- Free tier eligible**
- Description:** Ubuntu Server 16.04 LTS (HVM) EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>)
- Root Device Type:** ebs **Virtualization type:** hvm

Instance Type:

Instance Type	ECUs	vCPUs	Memory (GiB)
i3 large	9	2	15.25

Security Groups:

Type	Protocol	Port Range	Source	Description
SSH	TCP	22	0.0.0.0	
SSH	TCP	22	::/0	
Custom TCP Rule	TCP	50030	0.0.0.0	
Custom TCP Rule	TCP	50030	::/0	
Custom TCP Rule	TCP	50070	0.0.0.0	
Custom TCP Rule	TCP	50070	::/0	

Select an existing key pair or create a new key pair:

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about [removing existing key pairs from a public AMI](#).

Choose an existing key pair:

Select a key pair:

I acknowledge that I have access to the selected private key file (ansanAWS.pem), and that without this file, I won't be able to log into my instance.

Buttons: Cancel, Launch Instances

The screenshot shows the AWS EC2 Management Console interface. On the left, there's a sidebar with navigation links like EC2 Dashboard, Instances, Images, and Network & Security. The main area displays a table of instances. One instance is selected, showing its details: Name (Cluster), Instance ID (i-05ea7b957a005e0ce), Instance Type (i3.large), Availability Zone (us-west-2c), Status (running), and Public DNS (ec2-52-24-88-102.us-west-2.compute.amazonaws.com). The instance also has a private IP (172.31.9.151) and a key name (ahsanAWS).

Cluster Setup:

This screenshot shows the 'Create Image' dialog box overlaid on the EC2 Management Console. The dialog box has fields for 'Instance ID' (i-0b1dfbec3983763d6), 'Image name' (Cluster), and 'Image description' (Cluster image for Hadoop & Spark). Below these, there's a table for 'Instance Volumes' showing one volume (Root) with a size of 512 GiB and a snapshot ID of snap-06b86396bc2b1df78. At the bottom right of the dialog box is a 'Create Image' button.

Screenshot of the AWS EC2 Management Console showing the creation of an EBS image from an instance.

EC2 Dashboard

- Instances (selected)
- Launch Templates
- Spot Requests
- Reserved Instances
- Dedicated Hosts
- Scheduled Instances

Images

- AMIs
- Bundle Tasks

Elastic Block Store

- Volumes
- Snapshots

Network & Security

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces

Feedback English (US)

Create Image

Instance: i-0b1dfbec3983763d6 Public DNS: ec2-34-216-131-210.us-west-2.compute.amazonaws.com

Description: Create Image request received.
View pending image ami-db5e65a3
Any snapshots backing your new EBS image can be managed on the snapshots screen after successful image creation.

Close

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Public IP	IPv6 IPs	Key Name
i-0dc0fa0cecd71c9c	m5.2xlarge	us-west-2a	stopped							ahsanAWS
i-0b1dfbec3983763d6	i3.large	us-west-2c	running	2/2 checks ...	None		ec2-34-216-131-210.us...	34.216.131.210	-	ahsanAWS
i-aae727880967a0a43	i3.4xlarge	us-west-2c	stopped							ahsanAWS

© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 8:32 PM

Screenshot of the AWS EC2 Management Console showing a list of instances and the details for a specific terminated instance.

EC2 Dashboard

- Instances (selected)
- Launch Templates
- Spot Requests
- Reserved Instances
- Dedicated Hosts
- Scheduled Instances

Images

- AMIs
- Bundle Tasks

Elastic Block Store

- Volumes
- Snapshots

Network & Security

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces

Feedback English (US)

Instance: i-02109471659f62d45 Public DNS: ec2-35-164-219-178.us-west-2.compute.amazonaws.com

Description: Instance ID: i-02109471659f62d45
Instance state: running
Instance type: i3.large
Elastic IPs:
Availability zone: us-west-2c
Secondary private IPs:
Launch wizard-15: view inbound rules
Scheduled events:
AMI ID: ubuntu/images/hvm-ssd/ubuntu-xenial-16-04-amd64-server-20171121.1 (ami-0def3275)
Platform:
IAM role:
Key pair name: sadique

Details: Public DNS (IPv4): ec2-35-164-219-178.us-west-2.compute.amazonaws.com
IPv4 Public IP: 35.164.219.178
IPv6 IPs:
Private DNS: ip-172-31-4-219.us-west-2.compute.internal
Private IPs: 172.31.4.219
VPC ID: vpc-0a3dfe6d
Subnet ID: subnet-3623536e
Network interfaces: eth0
Source/dest. check: True
T2 Unlimited: -

© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 9:01 PM

Hadoop

ubuntu@ip-172-31-4-201: ~

```
ubuntu@ip-172-31-4-201:~$ ./gensort -a 10737418 input.txt
ubuntu@ip-172-31-4-201:~$ ls -lah input.txt
-rwxrwxr-x 1 ubuntu ubuntu 1.0G Nov 24 01:20 input.txt
ubuntu@ip-172-31-4-201:~$
```

ubuntu@ip-172-31-4-201: ~

```
-rwxrwxr-x 1 ubuntu ubuntu 1.0G Nov 24 01:20 input.txt
ubuntu@ip-172-31-4-201:~$ hdfs dfs -copyFromLocal input.txt /input
ubuntu@ip-172-31-4-201:~$ hadoop jar hadoop /input /output
hadoop/          hadoop-2.8.1/      hadoop-2.8.1.tar.gz hadoopTeraSort.jar
ubuntu@ip-172-31-4-201:~$ hadoop jar hadoopTeraSort.jar /input /output
17/11/24 01:23:41 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/11/24 01:23:41 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/11/24 01:23:41 INFO InputFileInputFormat: Total input files to process : 1
17/11/24 01:23:41 INFO mapreduce.JobSubmitter: number of splits:8
17/11/24 01:23:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1046445237_0001
17/11/24 01:23:42 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
17/11/24 01:23:42 INFO mapreduce.Job: Running job: job_local1046445237_0001
17/11/24 01:23:42 INFO mapred.LocalJobRunner: OutputCommitter set in config null
17/11/24 01:23:42 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/11/24 01:23:42 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/11/24 01:23:42 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
17/11/24 01:23:42 INFO mapred.LocalJobRunner: Waiting for map tasks
17/11/24 01:23:42 INFO mapred.LocalJobRunner: Starting task: attempt_local1046445237_0001_m_000000_0
17/11/24 01:23:42 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/11/24 01:23:42 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/11/24 01:23:42 INFO mapred.Task: Using ResourceCalculatorProcessTree: []
17/11/24 01:23:42 INFO mapred.MapTask: Processing split: hdfs://172.31.201:input+0+134217728
17/11/24 01:23:42 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/11/24 01:23:42 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/11/24 01:23:42 INFO mapred.MapTask: soft limit at 83886080
17/11/24 01:23:42 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/11/24 01:23:42 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/11/24 01:23:42 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/11/24 01:23:43 INFO mapreduce.Job: Job job_local1046445237_0001 running in uber mode : false
17/11/24 01:23:43 INFO mapreduce.Job: map 0% reduce 0%
17/11/24 01:23:44 INFO mapred.MapTask: Spilling map output
17/11/24 01:23:44 INFO mapred.MapTask: bufstart = 0; bufend = 72315600; bufvoid = 104857600
17/11/24 01:23:44 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/11/24 01:23:44 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192) kvi 18078904(72315616)
17/11/24 01:23:47 INFO mapred.MapTask: Finished spill 0
17/11/24 01:23:48 INFO mapred.LocalJobRunner: 
17/11/24 01:23:48 INFO mapred.MapTask: Starting flush of map output
17/11/24 01:23:48 INFO mapred.MapTask: Spilling map output
17/11/24 01:23:48 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252808; bufvoid = 104857600
17/11/24 01:23:48 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325964(65303856); length = 2476085/6553600
17/11/24 01:23:49 INFO mapred.Merger: Finished spill 1
17/11/24 01:23:49 INFO mapred.Merger: Merging 2 sorted segments
17/11/24 01:23:49 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902142 bytes
17/11/24 01:23:50 INFO mapred.Task: Task:attempt_local1046445237_0001_m_000000_0 is done. And is in the process of committing
17/11/24 01:23:50 INFO mapred.LocalJobRunner: map > sort
17/11/24 01:23:50 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000000_0' done.
17/11/24 01:23:50 INFO mapred.LocalJobRunner: Finishing task: attempt_local1046445237_0001_m_000000_0
17/11/24 01:23:50 INFO mapred.LocalJobRunner: Starting task: attempt_local1046445237_0001_m_000001_0
17/11/24 01:23:50 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
```

```

ubuntu@ip-172-31-4-201: ~
17/11/24 01:23:50 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/11/24 01:23:50 INFO mapred.Task: Using ResourceCalculatorProcessTree: [ ]
17/11/24 01:23:50 INFO mapred.MapTask: Processing split: hdfs://172.31.4.201/input:134217728+134217728
17/11/24 01:23:50 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/11/24 01:23:50 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/11/24 01:23:50 INFO mapred.MapTask: soft limit at 83886080
17/11/24 01:23:50 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/11/24 01:23:50 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/11/24 01:23:50 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/11/24 01:23:51 INFO mapred.Task: map 100% reduce 0%
17/11/24 01:23:52 INFO mapred.MapTask: Spilling map output
17/11/24 01:23:52 INFO mapred.MapTask: bufstart = 72315600; bufend = 104857600
17/11/24 01:23:52 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/11/24 01:23:52 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/11/24 01:23:52 INFO mapred.MapTask: Finished spill 0
17/11/24 01:23:54 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/11/24 01:23:55 INFO mapred.LocalJobRunner: 
17/11/24 01:23:55 INFO mapred.MapTask: Starting flush of map output
17/11/24 01:23:55 INFO mapred.MapTask: Spilling map output
17/11/24 01:23:55 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252708; bufvoid = 104857600
17/11/24 01:23:55 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325968(65303872); length = 2476081/6553600
17/11/24 01:23:55 INFO mapred.Task: map 13% reduce 0%
17/11/24 01:23:57 INFO mapred.MapTask: Finished spill 1
17/11/24 01:23:57 INFO mapred.Merger: Merging 2 sorted segments
17/11/24 01:23:57 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902040 bytes
17/11/24 01:23:58 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000001_0' is done. And is in the process of committing
17/11/24 01:23:58 INFO mapred.LocalJobRunner: map > sort
17/11/24 01:23:58 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000001_0' done.
17/11/24 01:23:58 INFO mapred.LocalJobRunner: Finishing task: attempt_local1046445237_0001_m_000001_0
17/11/24 01:23:58 INFO mapred.LocalJobRunner: Starting task: attempt_local1046445237_0001_m_000002_0
17/11/24 01:23:58 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/11/24 01:23:58 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/11/24 01:23:58 INFO mapred.Task: Using ResourceCalculatorProcessTree: [ ]
17/11/24 01:23:58 INFO mapred.MapTask: Processing split: hdfs://172.31.4.201/input:268435456+134217728
17/11/24 01:23:58 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/11/24 01:23:58 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/11/24 01:23:58 INFO mapred.MapTask: soft limit at 83886080
17/11/24 01:23:58 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/11/24 01:23:58 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/11/24 01:23:58 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/11/24 01:23:58 INFO mapred.Task: map 100% reduce 0%
17/11/24 01:23:59 INFO mapred.MapTask: Spilling map output
17/11/24 01:23:59 INFO mapred.MapTask: bufstart = 72315600; bufend = 104857600
17/11/24 01:23:59 INFO mapred.MapTask: kvstart = 18802048(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/11/24 01:23:59 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/11/24 01:23:59 INFO mapred.MapTask: Finished spill 0
17/11/24 01:24:00 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/11/24 01:24:02 INFO mapred.LocalJobRunner: 
17/11/24 01:24:02 INFO mapred.MapTask: Starting flush of map output
17/11/24 01:24:02 INFO mapred.MapTask: Spilling map output

```

```

ubuntu@ip-172-31-4-201: ~
17/11/24 01:24:02 INFO mapred.MapTask: Spilling map output
17/11/24 01:24:02 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252708; bufvoid = 104857600
17/11/24 01:24:02 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325968(65303872); length = 2476081/6553600
17/11/24 01:24:03 INFO mapred.Task: map 25% reduce 0%
17/11/24 01:24:04 INFO mapred.MapTask: Finished spill 1
17/11/24 01:24:04 INFO mapred.Merger: Merging 2 sorted segments
17/11/24 01:24:04 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902040 bytes
17/11/24 01:24:05 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000002_0' is done. And is in the process of committing
17/11/24 01:24:05 INFO mapred.LocalJobRunner: map > sort
17/11/24 01:24:05 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000002_0' done.
17/11/24 01:24:05 INFO mapred.LocalJobRunner: Finishing task: attempt_local1046445237_0001_m_000003_0
17/11/24 01:24:05 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/11/24 01:24:05 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/11/24 01:24:05 INFO mapred.Task: Using ResourceCalculatorProcessTree: [ ]
17/11/24 01:24:05 INFO mapred.MapTask: Processing split: hdfs://172.31.4.201/input:402653184+134217728
17/11/24 01:24:05 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/11/24 01:24:05 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/11/24 01:24:05 INFO mapred.MapTask: soft limit at 83886080
17/11/24 01:24:05 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/11/24 01:24:05 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/11/24 01:24:05 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/11/24 01:24:05 INFO mapred.Task: map 100% reduce 0%
17/11/24 01:24:06 INFO mapred.MapTask: Spilling map output
17/11/24 01:24:06 INFO mapred.MapTask: bufstart = 72315600; bufend = 104857600
17/11/24 01:24:06 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/11/24 01:24:06 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/11/24 01:24:06 INFO mapred.MapTask: Finished spill 0
17/11/24 01:24:08 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/11/24 01:24:09 INFO mapred.LocalJobRunner: 
17/11/24 01:24:09 INFO mapred.MapTask: Starting flush of map output
17/11/24 01:24:09 INFO mapred.MapTask: Spilling map output
17/11/24 01:24:09 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252808; bufvoid = 104857600
17/11/24 01:24:09 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325964(65303856); length = 2476085/6553600
17/11/24 01:24:10 INFO mapred.Task: map 38% reduce 0%
17/11/24 01:24:10 INFO mapred.MapTask: Finished spill 1
17/11/24 01:24:10 INFO mapred.Merger: Merging 2 sorted segments
17/11/24 01:24:10 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902142 bytes
17/11/24 01:24:11 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000003_0' is done. And is in the process of committing
17/11/24 01:24:11 INFO mapred.LocalJobRunner: map > sort
17/11/24 01:24:11 INFO mapred.Task: Task 'attempt_local1046445237_0001_m_000003_0' done.
17/11/24 01:24:11 INFO mapred.LocalJobRunner: Finishing task: attempt_local1046445237_0001_m_000004_0
17/11/24 01:24:11 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/11/24 01:24:11 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/11/24 01:24:11 INFO mapred.Task: Using ResourceCalculatorProcessTree: [ ]
17/11/24 01:24:11 INFO mapred.MapTask: Processing split: hdfs://172.31.4.201/input:536870912+134217728
17/11/24 01:24:11 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/11/24 01:24:11 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/11/24 01:24:11 INFO mapred.MapTask: soft limit at 83886080

```

```

ubuntu@ip-172-31-4-201:~ 
17/12/24 01:24:11 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/24 01:24:11 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/12/24 01:24:11 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/24 01:24:12 INFO mapred.MapTask: map 100% reduce 0%
17/12/24 01:24:13 INFO mapred.MapTask: Spilling map output
17/12/24 01:24:13 INFO mapred.MapTask: bufstart = 0; bufend = 72315600; bufvoid = 104857600
17/12/24 01:24:13 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/12/24 01:24:15 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/12/24 01:24:15 INFO mapred.MapTask: Finished spill 0
17/12/24 01:24:15 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/12/24 01:24:16 INFO mapred.LocalJobRunner:
17/12/24 01:24:16 INFO mapred: Starting flush of map output
17/12/24 01:24:16 INFO mapred.MapTask: Spilling map output
17/12/24 01:24:16 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252708; bufvoid = 104857600
17/12/24 01:24:16 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325968(65303872); length = 2476081/6553600
17/12/24 01:24:16 INFO mapred.MapTask: map 50% reduce 0%
17/12/24 01:24:17 INFO mapred.MapTask: Finished spill 1
17/12/24 01:24:17 INFO mapred.Merger: Merging 2 sorted segments
17/12/24 01:24:17 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902040 bytes
17/12/24 01:24:18 INFO mapred.Task: Task@attempt_local1046445237_0001_m_000004 is done. And is in the process of committing
17/12/24 01:24:18 INFO mapred.LocalJobRunner: map > sort
17/12/24 01:24:18 INFO mapred.Task: Task@attempt_local1046445237_0001_m_000004 0% done.
17/12/24 01:24:18 INFO mapred.LocalJobRunner: Finishing task: attempt_local1046445237_0001_m_000004
17/12/24 01:24:18 INFO mapred.LocalJobRunner: Starting task: attempt_local1046445237_0001_m_000005
17/12/24 01:24:18 INFO output.FileOutputCommitter: File output committer Algorithm version is 1
17/12/24 01:24:18 INFO output.FileOutputCommitter: Fileoutputcommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/24 01:24:18 INFO mapred.Task: Using ResourcecalculatorProcessTree : []
17/12/24 01:24:18 INFO mapred.MapTask: Processing split: hdfs://172.31.4.201/input:671088640+134217728
17/12/24 01:24:18 INFO mapred.MapTask: (EQUATOR) 0 kvi 16214396(104857584)
17/12/24 01:24:18 INFO mapred.MapTask: map 100% reduce 0%
17/12/24 01:24:18 INFO mapred.MapTask: soft limit = 83866000
17/12/24 01:24:18 INFO mapred.MapTask: Bufstart = 0; bufend = 104857600
17/12/24 01:24:18 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/12/24 01:24:18 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/24 01:24:19 INFO mapred.MapTask: map 100% reduce 0%
17/12/24 01:24:19 INFO mapred.MapTask: Spilling map output
17/12/24 01:24:19 INFO mapred.MapTask: bufstart = 0; bufend = 72315600; bufvoid = 104857600
17/12/24 01:24:19 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/12/24 01:24:19 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/12/24 01:24:21 INFO mapred.MapTask: Finished spill 0
17/12/24 01:24:21 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/12/24 01:24:22 INFO mapred.LocalJobRunner:
17/12/24 01:24:22 INFO mapred: Starting flush of map output
17/12/24 01:24:22 INFO mapred.MapTask: Spilling map output
17/12/24 01:24:22 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252708; bufvoid = 104857600
17/12/24 01:24:22 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325968(65303872); length = 2476081/6553600
17/12/24 01:24:23 INFO mapred.MapTask: map 63% reduce 0%
17/12/24 01:24:24 INFO mapred.MapTask: Finished spill 1
17/12/24 01:24:24 INFO mapred.Merger: Merging 2 sorted segments

```

Configuration 1:-

13.large cpu info

```

ubuntu@ip-172-31-1-36:~ 
ubuntu@ip-172-31-1-36:~$ cat /proc/cpuinfo/
cat: /proc/cpuinfo/: Not a directory
ubuntu@ip-172-31-1-36:~$ cat /proc/cpuinfo
processor       : 0
vendor_id      : GenuineIntel
cpu family     : 6
model          : 79
model name     : Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.30GHz
stepping        : 1
microcode      : 0xb0000024
cpu MHz         : 2029.750
cache size     : 46080 KB
physical id    : 0
siblings        : 2
core id         : 0
cpu cores      : 1
apicid          : 0
initial apicid : 0
fpu             : yes
fpu_exception   : yes
cpuid level    : 13
wp              : yes
flags           : fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush mmx fxsr sse ss
top tsc aperfmpfperf eagerfpu pni pclmulqdq sse3 fma cx16 pcid sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_time
il Hle avx2 smep bm12 erms invpcid rtm rdseed adx xsaveopt
bugs            :
bogomips       : 4600.13
clflush size   : 64
cache alignment : 64
address sizes   : 46 bits physical, 48 bits virtual
power management:

processor       : 1
vendor_id      : GenuineIntel
cpu family     : 6
model          : 79
model name     : Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.30GHz
stepping        : 1
microcode      : 0xb0000024
cpu MHz         : 2700.703
cache size     : 46080 KB
physical id    : 0
siblings        : 2
core id         : 0
cpu cores      : 1
apicid          : 1
initial apicid : 1
fpu             : yes
fpu_exception   : yes
cpuid level    : 13

```

128GB Input file generation

```
ubuntu@ip-172-31-1-36: ~/msadique
ubuntu@ip-172-31-1-36:~/msadique$ time ./gensort -a 1374389535 input.txt

real    28m44.192s
user    7m57.184s
sys     0m0.534s

ubuntu@ip-172-31-1-36:~/msadique$ ls -lash input.txt
128.1G -rwxrwxr-x 1 ubuntu ubuntu 128.0G Dec  3 13:36 input.txt
ubuntu@ip-172-31-1-36:~/msadique$
```

a) Shared-Memory TeraSort

b) 1 Thread : Size 128GB

```
ubuntu@ip-172-31-1-36: ~/msadique/CloudPthread
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 1

***      Share Memory Terasort using Pthread and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of threads: 1

Input File Size      : 128 GB
Total Memory(RAM)   : 14.93 GB
Available Memory(RAM) : 14.54 GB

Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 32
***** Thread 1 Started Sorting *****
Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 2582.33 sec
          Sorting Time for chunks       = 19897.65 sec
          Merging Time for Sorted chunks = 2947.80 sec

          Total Shared Memory Tera Sorting Time   = 25427.78 sec
-----
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread$
```

2 Threads 128GB

```
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 2

***      Share Memory Terasort using Pthread and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of threads: 2

Input File Size      : 128 GB
Total Memory(RAM)    : 14.93 GB
Available Memory(RAM) : 14.55 GB

Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads   : 48

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks          = 2608.15 sec
Sorting Time for chunks           = 15633.87 sec
Merging Time for Sorted chunks    = 2888.8 sec

-----  
Total Shared Memory Tera Sorting Time      = 21130.87 sec
-----
```

4 Threads 128GB

```
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 4

***      Share Memory Terasort using Pthread and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of threads: 4

Input File Size      : 128 GB
Total Memory(RAM)    : 14.93 GB
Available Memory(RAM) : 14.55 GB

Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 64

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****
***** Thread 3 Started Sorting *****
***** Thread 4 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 2634.23 sec
          Sorting Time for chunks        = 11749.09 sec
          Merging Time for Sorted chunks = 2831.06 sec

          Total Shared Memory Tera Sorting Time   = 17214.39 sec
-----
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread$
```

8 Threads 128GB

```
ubuntu@ip-172-31-1-36:~/msadique/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 8

***      Share Memory Terasort using Pthread and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of threads: 8

Input File Size      : 128 GB
Total Memory(RAM)   : 14.93 GB
Available Memory(RAM) : 14.55 GB

Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 96

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****
***** Thread 3 Started Sorting *****
***** Thread 4 Started Sorting *****
***** Thread 5 Started Sorting *****
***** Thread 6 Started Sorting *****
***** Thread 7 Started Sorting *****
***** Thread 8 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 2556.76 sec
          Sorting Time for chunks       = 9475.07 sec
          Merging Time for Sorted chunks = 3007.96 sec

          Total Shared Memory Tera Sorting Time    = 15039.80 sec
-----
```

ubuntu@ip-172-31-1-36:~/msadique/CloudPthread\$

b. Hadoop TeraSort:

Screenshots:

```
ubuntu@ip-172-31-1-36: ~$ jps
16241 ResourceManager
20551 Jps
15752 NameNode
16088 SecondaryNameNode
16361 NodeManager
15903 DataNode
ubuntu@ip-172-31-1-36:~$
```

Time taken to generate Input file

```
ubuntu@ip-172-31-1-36: ~/msadique
ubuntu@ip-172-31-1-36:~/msadique$ time ./gensort -a 1374389535 input.txt
real    28m44.192s
user    7m57.184s
sys     0m0.534s

ubuntu@ip-172-31-1-36:~/msadique$ ls -lash input.txt
128.1G -rwxrwxr-x 1 ubuntu ubuntu 128.0G Dec  3 13:36 input.txt
```

Time it takes to transfer 128GB file to Hdfs

```
ubuntu@ip-172-31-1-36: ~
ubuntu@ip-172-31-1-36:~$ time hdfs dfs -copyFromLocal input.txt /input3
real    35m42.565s
user    26m36.724s
sys     0m0.920s
ubuntu@ip-172-31-1-36:~$ ls -lash input.txt
128GB -rwxrwxr-x 1 ubuntu ubuntu 128GB Dec  3 18:18 input.txt
ubuntu@ip-172-31-1-36:~$
```

Hadoop 128GB

```
ubuntu@ip-172-31-1-36:~/msadique$ time hadoop jar hadoopTeraSort.jar /input /output
17/12/03 18:43:48 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/12/03 18:43:48 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/12/03 18:43:48 INFO input.FileInputFormat: Total input files to process : 1
17/12/03 18:43:48 INFO mapreduce.JobSubmitter: number of splits:8
17/12/03 18:43:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local17214462_0001
17/12/03 18:43:48 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
17/12/03 18:43:48 INFO mapreduce.Job: Running job: job_local17214462_0001
17/12/03 18:43:48 INFO mapred.LocalJobRunner: OutputCommitter set in config null
17/12/03 18:43:48 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 18:43:48 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 18:43:48 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
17/12/03 18:43:49 INFO mapred.LocalJobRunner: Waiting for map tasks
17/12/03 18:43:49 INFO mapred.LocalJobRunner: Starting task: attempt_local17214462_0001_m_000000_0
17/12/03 18:43:49 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 18:43:49 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 18:43:49 INFO mapred.Task: Using ResourceCalculatorProcessTree: []
17/12/03 18:43:49 INFO mapred.MapTask: Processing split: hdfs://172.31.1.36/input:4:0+134217728
17/12/03 18:43:49 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/12/03 18:43:49 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/12/03 18:43:49 INFO mapred.MapTask: soft limit at 83886080
17/12/03 18:43:49 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/03 18:43:49 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/12/03 18:43:49 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/12/03 18:43:49 INFO mapred.MapTask: Finished spill 0
17/12/03 18:43:49 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/03 18:43:49 INFO mapred.MapTask: Job: job_local17214462_0001 running in uber mode : false
17/12/03 18:43:49 INFO mapred.MapTask: map 0% reduce 0%
17/12/03 18:43:50 INFO mapred.MapTask: Spilling map output
17/12/03 18:43:50 INFO mapred.MapTask: bufstart = 72315600; bufvoid = 104857600
17/12/03 18:43:50 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/12/03 18:43:50 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/12/03 18:43:53 INFO mapred.MapTask: Finished spill 0
17/12/03 18:43:53 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/12/03 18:43:54 INFO mapred.LocalJobRunner:
17/12/03 18:43:54 INFO mapred.MapTask: Starting flush of map output
17/12/03 18:43:54 INFO mapred.MapTask: Spilling map output
17/12/03 18:43:54 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252808; bufvoid = 104857600
17/12/03 18:43:54 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325964(65303856); length = 2476085/6553600
17/12/03 18:43:56 INFO mapred.MapTask: Finished spill 1
17/12/03 18:43:56 INFO mapred.Merger: Merging 2 sorted segments
17/12/03 18:43:56 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902142 bytes
17/12/03 18:43:56 INFO mapred.Merger: Task: attempt_local17214462_0001_m_000000_0 is done. And is in the process of committing
17/12/03 18:43:57 INFO mapred.LocalJobRunner: map > sort
17/12/03 18:43:57 INFO mapred.Task: Task 'attempt_local17214462_0001_m_000000_0' done.
17/12/03 18:43:57 INFO mapred.LocalJobRunner: Finishing task: attempt_local17214462_0001_m_000000_0
17/12/03 18:43:57 INFO mapred.LocalJobRunner: Starting task: attempt_local17214462_0001_m_000001_0
17/12/03 18:43:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 18:43:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 18:43:57 INFO mapred.Task: Using ResourceCalculatorProcessTree: []
17/12/03 18:43:57 INFO mapred.MapTask: Processing split: hdfs://172.31.1.36/input:134217728+134217728
```

```
ubuntu@ip-172-31-1-36:~/msadique
17/12/03 18:43:57 INFO mapred.LocalJobRunner: map > sort
17/12/03 18:43:57 INFO mapred.Task: Task 'attempt_local17214462_0001_m_000000_0' done.
17/12/03 18:43:57 INFO mapred.LocalJobRunner: Finishing task: attempt_local17214462_0001_m_000000_0
17/12/03 18:43:57 INFO mapred.LocalJobRunner: Starting task: attempt_local17214462_0001_m_000001_0
17/12/03 18:43:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 18:43:57 INFO mapred.Task: Using ResourceCalculatorProcessTree: []
17/12/03 18:43:57 INFO mapred.Task: Processing split: hdfs://172.31.1.36/input:134217728+134217728
17/12/03 18:43:57 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/12/03 18:43:57 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/12/03 18:43:57 INFO mapred.MapTask: soft limit at 83886080
17/12/03 18:43:57 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/03 18:43:57 INFO mapred.MapTask: kvstart = 26214396(104857584); length = 6553600
17/12/03 18:43:57 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/03 18:43:57 INFO mapred.MapTask: map 100% reduce 0%
17/12/03 18:43:58 INFO mapred.MapTask: Spilling map output
17/12/03 18:43:58 INFO mapred.MapTask: bufstart = 72315600; bufvoid = 104857600
17/12/03 18:43:58 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/12/03 18:43:58 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/12/03 18:43:58 INFO mapred.MapTask: Finished spill 0
17/12/03 18:44:00 INFO mapred.MapTask: Finished spill 1
17/12/03 18:44:01 INFO mapred.LocalJobRunner: Starting flush of map output
17/12/03 18:44:01 INFO mapred.MapTask: Spilling map output
17/12/03 18:44:01 INFO mapred.MapTask: bufstart = 32252708; bufend = 32252708; bufvoid = 104857600
17/12/03 18:44:01 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325968(65303872); length = 2476085/6553600
17/12/03 18:44:01 INFO mapred.MapTask: map 1% reduce 0%
17/12/03 18:44:01 INFO mapred.MapTask: Finished spill 1
17/12/03 18:44:01 INFO mapred.Merger: Merging 2 sorted segments
17/12/03 18:44:01 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 136902040 bytes
17/12/03 18:44:03 INFO mapred.Task: Task: attempt_local17214462_0001_m_000001_0 is done. And is in the process of committing
17/12/03 18:44:03 INFO mapred.LocalJobRunner: map > sort
17/12/03 18:44:03 INFO mapred.Task: Task 'attempt_local17214462_0001_m_000001_0' done.
17/12/03 18:44:03 INFO mapred.LocalJobRunner: Finishing task: attempt_local17214462_0001_m_000001_0
17/12/03 18:44:03 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 18:44:03 INFO mapred.Task: Using ResourceCalculatorProcessTree: []
17/12/03 18:44:03 INFO mapred.Task: Processing split: hdfs://172.31.1.36/input:268435456+134217728
17/12/03 18:44:03 INFO mapred.MapTask: map 100% reduce 0%
17/12/03 18:44:03 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/12/03 18:44:03 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/12/03 18:44:03 INFO mapred.MapTask: soft limit at 83886080
17/12/03 18:44:03 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/03 18:44:03 INFO mapred.MapTask: kvstart = 26214396(104857584); length = 6553600
17/12/03 18:44:03 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/03 18:44:03 INFO mapred.MapTask: Spilling map output
17/12/03 18:44:05 INFO mapred.MapTask: bufstart = 72315600; bufend = 104857600
17/12/03 18:44:05 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93287104); length = 2892621/6553600
17/12/03 18:44:05 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
```

```
ubuntu@ip-172-31-1-36: ~/msadique$
```

```
17/12/03 22:08:51 INFO mapred.LocalJobRunner: reduce task executor complete.
17/12/03 22:08:51 INFO mapreduce.Job: map 100% reduce 100%
17/12/03 22:08:52 INFO mapreduce.Job: Job job_local717214462_0001 completed successfully
17/12/03 22:08:52 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=4427452297268
    FILE: Number of bytes written=7085441208461
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=3183200336624
    HDFS: Number of bytes written=578746830200
    HDFS: Number of read operations=63602
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=5929
  Map-Reduce Framework
    Map input records=343597376
    Map output records=343597376
    Map output bytes=34359737600
    Map output materialized bytes=39427800624
    Input split bytes=25600
    Combine input records=343597376
    Combine output records=343597376
    Reduce input groups=343597376
    Reduce shuffle bytes=39427800624
    Reduce input records=343597376
    Reduce output records=343597376
    Spilled Records=323513376
    Shuffled Maps =256
    Failed Shuffles=0
    Merged Map outputs=256
    GC time elapsed (ms)=213890
    Total committed heap usage (bytes)=136918859776
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=137438953472
  File Output Format Counters
    Bytes Written=137438953472

real    203m52.612s
user    198m19.888s
sys     1m23.728s
```

```
ubuntu@ip-172-31-1-36: ~/msadique$
```

c. Spark TeraSort:

```
root@ip-172-31-23-238 ~]$ spark-1.6.0-bin-hadoop2.6/bin/spark-shell
log4j:WARN No appenders could be found for logger (org.apache.hadoop.metrics2.lib.MutableMetricsFactory).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties
To adjust logging level use sc.setLogLevel("INFO")
Welcome to

    \____/ \
   / \  \_ \_ \_ \_ \_ \_ \_ \
  / \_ \_ \_ \_ \_ \_ \_ \_ \_ \
 / \_ \_ \_ \_ \_ \_ \_ \_ \_ \
version 1.6.0

Using Scala version 2.10.5 (OpenJDK 64-Bit Server VM, Java 1.7.0_151)
Type in expressions to have them evaluated.
Type :help for more information.
Spark context available as sc.
17/12/03 07:12:40 WARN Connection: BoneCP specified but not present in CLASSPATH (or one of dependencies)
17/12/03 07:12:40 WARN Connection: BoneCP specified but not present in CLASSPATH (or one of dependencies)
17/12/03 07:12:47 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not enabled so recording th
17/12/03 07:12:47 WARN ObjectStore: Failed to get database default, returning NoSuchObjectException
17/12/03 07:12:50 WARN Connection: BoneCP specified but not present in CLASSPATH (or one of dependencies)
17/12/03 07:12:50 WARN Connection: BoneCP specified but not present in CLASSPATH (or one of dependencies)
SQL context available as sqlContext.

scala> :load /root/scode.scala
Loading /root/scode.scala...
lines: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[1] at textFile at <console>:27
file Size(bytes):137438953472
sort: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[7] at map at <console>:29
Sorting time for splitted file : 1301.054232767
Total Spark Tera Sort Time (Sort, Merge and Transfer To Local) : 5209.548903017

root@ip-172-31-23-238 ~$
```

d. MPI TeraSort

Install MPI on I3.large

```
ubuntu@ip-172-31-1-36:~/msadique/CloudMPI$ sudo apt install libmpich-dev
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  gfortran gfortran-5 hwloc-nox libcr-dev libcr0 libgfortran-5-dev libgfortran3 libhwloc-plugins libhwloc5 libltdl7 libmpich12 mpich ocl-icd-libopencl1
Suggested packages:
  gfortran-multilib gfortran-doc gfortran-5-multilib gfortran-5-doc libgfortran3-dbg blcr-dkms libhwloc-contrib-plugins blcr-util mpich-doc opencl-icd
The following NEW packages will be installed:
  gfortran gfortran-5 hwloc-nox libcr-dev libcr0 libgfortran-5-dev libgfortran3 libhwloc-plugins libhwloc5 libltdl7 libmpich-dev libmpich12 mpich ocl-icd-libopencl1
0 upgraded, 14 newly installed, 0 to remove and 5 not upgraded.
Need to get 11.4 MB of archives.
After this operation, 40.3 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/main amd64 libgfortran3 amd64 5.4.0-6ubuntu1~16.04.5 [260 kB]
Get:2 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/main amd64 libgfortran-5-dev amd64 5.4.0-6ubuntu1~16.04.5 [291 kB]
Get:3 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial-updates/main amd64 gfortran-5 amd64 5.4.0-6ubuntu1~16.04.5 [8,161 kB]
Get:4 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/main amd64 gfortran amd64 4:5.3.1-1ubuntu1 [1,288 B]
Get:5 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 libcr0 amd64 0.8.5-2.2 [18.8 kB]
Get:6 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 libcr-dev amd64 0.8.5-2.2 [24.5 kB]
Get:7 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/main amd64 libltdl7 amd64 2.4.6-0.1 [38.3 kB]
Get:8 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 libhwloc5 amd64 1.11.2-3 [99.5 kB]
Get:9 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 hwloc-nox amd64 1.11.2-3 [151 kB]
Get:10 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/main amd64 ocl-icd-libopencl1 amd64 2.2.8-1 [29.7 kB]
Get:11 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 libhwloc-plugins amd64 1.11.2-3 [13.2 kB]
Get:12 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 libmpich12 amd64 3.2-6build1 [886 kB]
Get:13 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 libmpich-dev amd64 3.2-6build1 [1,193 kB]
Get:14 http://us-west-2.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 mpich amd64 3.2-6build1 [197 kB]
Fetched 11.4 MB in 0s (40.7 MB/s)
```

Number of Processing Node = 1

File Size :128GB

```
ubuntu@ip-172-31-1-36: ~/msadique/CloudMPI
ubuntu@ip-172-31-1-36:~/msadique/CloudMPI$ make
mpicc -o run TeraMPSort.c -lm
mpirun -np 1 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 1

Input File Size      : 128 GB
Total Memory(RAM)   : 14.93 GB
Available Memory(RAM) : 14.54 GB
Total No. chunks Created Based on Available Memory & Nodes : 32

***** Processor - 1 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 2784.84 sec
          Sorting Time for chunks       = 29316.39 sec
          Merging Time for Sorted chunks = 3070.27 sec

          Total MPI Tera Sorting Time   = 35171.50 sec
-----
ubuntu@ip-172-31-1-36:~/msadique/CloudMPI$
```

Number of Processing Node = 2

File Size :128GB

```
ubuntu@ip-172-31-1-36: ~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 2 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 2

Input File Size      : 128 GB
Total Memory(RAM)   : 14.93 GB
Available Memory(RAM) : 14.53 GB
Total No. chunks Created Based on Available Memory & Nodes : 48

***** Processor - 1 Started Sorting *****
***** Processor - 2 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3002.05 sec
Sorting Time for chunks        = 20475.89 sec
Merging Time for Sorted chunks = 2978.16 sec

Total MPI Tera Sorting Time   = 26456.12 sec
-----
```

ubuntu@ip-172-31-1-36:~/msadique/CloudMPI\$

Number of Processing Node = 4

File Size :128GB

```
ubuntu@ip-172-31-1-36: ~/msadique/CloudMPI
ubuntu@ip-172-31-1-36:~/msadique/CloudMPI$ make
mpicc -o run TeraMPSort.c -lm
mpirun -np 4 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 4

Input File Size      : 128 GB
Total Memory(RAM)   : 14.93 GB
Available Memory(RAM) : 14.53 GB
Total No. chunks Created Based on Available Memory & Nodes : 64

***** Processor - 1 Started Sorting *****
***** Processor - 2 Started Sorting *****
***** Processor - 3 Started Sorting *****
***** Processor - 4 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 3236.21 sec
          Sorting Time for chunks       = 16612.62 sec
          Merging Time for Sorted chunks = 2888.81 sec

          Total MPI Tera Sorting Time    = 22737.65 sec
-----
```

Number of Processing Node = 8

File Size :128GB

```
ubuntu@ip-172-31-1-36:~/msadique/CloudMPI
ubuntu@ip-172-31-1-36:~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 8 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 8

Input File Size      : 128 GB
Total Memory(RAM)   : 14.93 GB
Available Memory(RAM) : 14.53 GB
Total No. chunks Created Based on Available Memory & Nodes : 96

***** Processor - 1 Started Sorting *****
***** Processor - 8 Started Sorting *****
***** Processor - 4 Started Sorting *****
***** Processor - 6 Started Sorting *****
***** Processor - 7 Started Sorting *****
***** Processor - 5 Started Sorting *****
***** Processor - 2 Started Sorting *****
***** Processor - 3 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 2583.34 sec
          Sorting Time for chunks       = 9772.13 sec
          Merging Time for Sorted chunks = 3165.23 sec

          Total MPI Tera Sorting Time    = 15520.71 sec
-----
```

Configuration 2

a. Shared-Memory TeraSort:

```
ubuntu@ip-172-31-4-162:~/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 1
***      Share Memory Terasort using Pthread and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of threads: 1

Input File Size      : 1024 GB
Total Memory(RAM)    : 120.00 GB
Available Memory(RAM) : 117.50 GB
Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 32

***** Thread 1 Started Sorting *****

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3883.56 sec
Sorting Time for chunks        = 79428.49 sec
Merging Time for Sorted chunks = 5530.92 sec

Total Shared Memory Tera Sorting Time    = 88842.98 sec
-----
```

ubuntu@ip-172-31-4-162:~/CloudPthread\$

No of threads : 2

File size:1024 GB

ubuntu@ip-172-31-4-162:~/CloudPthread\$

```
ubuntu@ip-172-31-4-162:~/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 2
***      Share Memory Terasort using Pthread and MergeSort          ***

Input file name: input.txt
Output file name: output.txt
Number of threads: 2

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.50 GB
Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 48

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3796.25 sec
Sorting Time for chunks        = 51061.17 sec
Merging Time for Sorted chunks = 5701.98 sec

Total Shared Memory Tera Sorting Time    = 60559.41 sec
-----
```

ubuntu@ip-172-31-4-162:~/CloudPthread\$

No of threads : 4

File size:1024 GB

```
ubuntu@ip-172-31-4-162:~/CloudPthread
ubuntu@ip-172-31-4-162:~/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 4
***      Share Memory Terasort using Pthread and MergeSort          ***
Input file name: input.txt
Output file name: output.txt
Number of threads: 4

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.50 GB
Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 64

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****
***** Thread 3 Started Sorting *****
***** Thread 4 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3710.90 sec
Sorting Time for chunks        = 22693.85 sec
Merging Time for Sorted chunks = 5878.33 sec

Total Shared Memory Tera Sorting Time    = 32283.09 sec
-----
ubuntu@ip-172-31-4-162:~/CloudPthread$
```

No of threads : 8
File size:1024 GB

```
ubuntu@ip-172-31-4-162:~/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 8
***      Share Memory Terasort using Pthread and MergeSort      ***
Input file name: input.txt
Output file name: output.txt
Number of threads: 8

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.50 GB
Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 96

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****
***** Thread 3 Started Sorting *****
***** Thread 4 Started Sorting *****
***** Thread 5 Started Sorting *****
***** Thread 6 Started Sorting *****
***** Thread 7 Started Sorting *****
***** Thread 8 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 3627.47 sec
          Sorting Time for chunks       = 14892.84 sec
          Merging Time for Sorted chunks = 6060.13 sec

          Total Shared Memory Tera Sorting Time    = 24580.45 sec
-----
```

ubuntu@ip-172-31-4-162:~/CloudPthread\$

No of Threads : 16

File Size: 1024 GB

```
ubuntu@ip-172-31-4-162:~/CloudPthread$ make
gcc -g -o run TeraSharedSort.c -lpthread -lm
./run input.txt output.txt 16
***      Share Memory Terasort using Pthread and MergeSort      ***
Input file name: input.txt
Output file name: output.txt
Number of threads: 16

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.50 GB
Initializing Sorting of Chunks
Completed splitting files !

Total No. chunks Created Based on Available Memory & Threads : 192

***** Thread 1 Started Sorting *****
***** Thread 2 Started Sorting *****
***** Thread 3 Started Sorting *****
***** Thread 4 Started Sorting *****
***** Thread 5 Started Sorting *****
***** Thread 6 Started Sorting *****
***** Thread 7 Started Sorting *****
***** Thread 8 Started Sorting *****
***** Thread 9 Started Sorting *****
***** Thread 10 Started Sorting *****
***** Thread 11 Started Sorting *****
***** Thread 12 Started Sorting *****
***** Thread 13 Started Sorting *****
***** Thread 14 Started Sorting *****
***** Thread 15 Started Sorting *****
***** Thread 16 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3545.91 sec
Sorting Time for chunks        = 7091.83 sec
Merging Time for Sorted chunks = 6247.56 sec

Total Shared Memory Tera Sorting Time    = 16885.31 sec
-----
```

b. Hadoop TeraSort:

File 1TB generation Time

```
ubuntu@ip-172-31-4-162: ~/msadique
ubuntu@ip-172-31-4-162:~/msadique$ time ./gensort -a 10995116278 input.txt

real    105m26.442s
user    65m31.224s
sys     0m02.534s
ubuntu@ip-172-31-4-162:~/msadique$ ls -lash input.txt
1T -rwxrwxr-x 1 ubuntu ubuntu 1T Dec  3 13:36 input.txt
ubuntu@ip-172-31-4-162:~/msadique$
```

1 TB File Transfer To HDFS

```
ubuntu@ip-172-31-4-162: ~/msadique
ubuntu@ip-172-31-4-162:~/msadique$ time hdfs dfs -copyFromLocal input.txt /input3

real    132m12.905s
user    97m14.714s
sys     0m1.920s
ubuntu@ip-172-31-4-162:~/msadique$ ls -lash input.txt
1T -rwxrwxr-x 1 ubuntu ubuntu 1T Dec  3 18:18 input.txt
ubuntu@ip-172-31-4-162:~/msadique$
```

```
ubuntu@ip-172-31-4-162: ~/msadique
ubuntu@ip-172-31-4-162:~/msadique$ time hadoop jar hadoop-terasort.jar /input /output
17/12/03 16:40:48 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/12/03 16:40:48 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/12/03 16:40:48 INFO input.FileInputFormat: Total input files to process : 1
17/12/03 16:40:48 INFO mapreduce.JobSubmitter: number of splits:3
17/12/03 16:40:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1422652445_0001
17/12/03 16:40:48 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
17/12/03 16:40:48 INFO mapreduce.Job: Running job: job_local1422652445_0001
17/12/03 16:40:48 INFO mapred.LocalJobRunner: outputCommitter set in config null
17/12/03 16:40:48 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 16:40:48 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 16:40:48 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
17/12/03 16:40:49 INFO mapred.LocalJobRunner: Waiting for map tasks
17/12/03 16:40:49 INFO mapred.LocalJobRunner: Starting task: attempt_local1422652445_0001_m_000000
17/12/03 16:40:49 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 16:40:49 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 16:40:49 INFO mapred.Task: Using ResourceCalculatorForTree : []
17/12/03 16:40:49 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/12/03 16:40:49 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/12/03 16:40:49 INFO mapred.MapTask: soft limit at 83886080
17/12/03 16:40:49 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/03 16:40:49 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/12/03 16:40:49 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/03 16:40:50 INFO mapreduce.Job: Job: job_local1422652445_0001 running in uber mode : false
17/12/03 16:40:50 INFO mapreduce.Job: map 0% reduce 0%
17/12/03 16:40:50 INFO mapred.MapTask: Spilling map output
17/12/03 16:40:50 INFO mapred.MapTask: bufstart = 0; bufend = 57801564; bufvoid = 104857600
17/12/03 16:40:50 INFO mapred.MapTask: Kvstart = 26214396(104857584); kvend = 1693268(78773072); length = 6521129/6553600
17/12/03 16:40:50 INFO mapred.MapTask: (EQUATOR) 64380844 kvi 16095204(64380816)
17/12/03 16:40:55 INFO mapred.MapTask: Finished spill 0
17/12/03 16:40:55 INFO mapred.MapTask: (RESET) equator 64380844 kv 16095204(64380816) kvi 14471724(57886896)
17/12/03 16:40:55 INFO mapred.MapTask: Spilling map output
17/12/03 16:40:55 INFO mapred.MapTask: bufstart = 64380844; bufend = 17323553; bufvoid = 104857597
17/12/03 16:40:55 INFO mapred.MapTask: kvstart = 16095204(64380816); kvend = 9573768(38295072); length = 6521437/6553600
17/12/03 16:40:55 INFO mapred.MapTask: (EQUATOR) 23902833 kvi 5975704(23902816)
17/12/03 16:40:59 INFO mapred.MapTask: Finished spill 1
17/12/03 16:40:59 INFO mapred.MapTask: (RESET) equator 23902833 kv 5975704(23902816) kvi 4351872(17407488)
17/12/03 16:40:59 INFO mapred.LocalJobRunner:
17/12/03 16:40:59 INFO mapred.MapTask: Starting flush of map output
17/12/03 16:40:59 INFO mapred.MapTask: Spilling map output
17/12/03 16:40:59 INFO mapred.MapTask: bufstart = 23902833; bufend = 55015455; bufvoid = 104857600
17/12/03 16:40:59 INFO mapred.MapTask: kvstart = 5975704(23902816); kvend = 2465396(9861584); length = 3510309/6553600
17/12/03 16:41:01 INFO mapred.LocalJobRunner: map > sort
17/12/03 16:41:01 INFO mapreduce.Job: map 2% reduce 0%
17/12/03 16:41:01 INFO mapred.MapTask: Finished spill 2
17/12/03 16:41:01 INFO mapred.Merger: Merging 3 sorted segments
17/12/03 16:41:01 INFO mapred.Merger: Down to the last merge-pass, with 3 segments left of total size: 154654215 bytes
17/12/03 16:41:04 INFO mapred.Task: Task@attempt_local1422652445_0001_m_000000_0 is done. And is in the process of committing
17/12/03 16:41:04 INFO mapred.LocalJobRunner: map > sort
```

ubuntu@ip-172-31-4-162: ~/msadique

```
17/12/03 16:43:16 INFO mapred.MapTask: Finished spill 1
17/12/03 16:43:16 INFO mapred.MapTask: (RESET) equator 23904332 kv 5976076(23904304) kvi 4352044(17408176)
17/12/03 16:43:17 INFO mapred.LocalJobRunner:
17/12/03 16:43:17 INFO mapred.MapTask: Starting flush of map output
17/12/03 16:43:17 INFO mapred.MapTask: Spilling map output
17/12/03 16:43:17 INFO mapred.MapTask: bufstart = 23904332; bufend = 55015258; bufvoid = 104857600
17/12/03 16:43:17 INFO mapred.MapTask: kvstart = 5976076(23904304); kvend = 2465576(9862304); length = 3510501/6553600
17/12/03 16:43:17 INFO mapreduce.Job: map 31% reduce 0%
17/12/03 16:43:18 INFO mapred.MapTask: Finished spill 2
17/12/03 16:43:18 INFO mapred.Merger: Merging 3 sorted segments
17/12/03 16:43:18 INFO mapred.MapTask: Down to the last merge-pass, with 3 segments left of total size: 154654299 bytes
17/12/03 16:43:20 INFO mapred.LocalJobRunner: map > sort >
17/12/03 16:43:20 INFO mapreduce.Job: map 34% reduce 0%
17/12/03 16:43:22 INFO mapred.Task: Task@attempt_local1422652445_0001_m_000010_0 is done. And is in the process of committing
17/12/03 16:43:22 INFO mapred.LocalJobRunner: map > sort
17/12/03 16:43:22 INFO mapred.Task: Task@attempt_local1422652445_0001_m_000010_0 done.
17/12/03 16:43:22 INFO mapred.LocalJobRunner: Finishing task: attempt_local1422652445_0001_m_000010_0
17/12/03 16:43:22 INFO mapred.LocalJobRunner: Starting task: attempt_local1422652445_0001_m_000011_0
17/12/03 16:43:22 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 16:43:22 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 16:43:22 INFO mapred.MapTask: Using ResourceCalculatorProcessTree: []
17/12/03 16:43:22 INFO mapred.MapTask: Processing split: hdfs://172.31.6.220/input2:1476395008+134217728
17/12/03 16:43:22 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
17/12/03 16:43:22 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/12/03 16:43:22 INFO mapred.MapTask: soft limit at 8388000
17/12/03 16:43:22 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/03 16:43:22 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/12/03 16:43:22 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/03 16:43:22 INFO mapreduce.Job: map 100% reduce 0%
17/12/03 16:43:22 INFO mapred.MapTask: Spilling map output
17/12/03 16:43:22 INFO mapred.MapTask: bufstart = 0; bufend = 57802706; bufvoid = 104857600
17/12/03 16:43:22 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 19693560(78774240); length = 6520837/6553600
17/12/03 16:43:22 INFO mapred.MapTask: (EQUATOR) 64382002 kvi 16095496(64381984)
17/12/03 16:43:26 INFO mapred.MapTask: Finished spill 0
17/12/03 16:43:26 INFO mapred.MapTask: (RESET) equator 64382002 kv 16095496(64381984) kvi 14471160(57884640)
17/12/03 16:43:27 INFO mapred.MapTask: Spilling map output
17/12/03 16:43:27 INFO mapred.MapTask: bufstart = 64382002; bufend = 17322710; bufvoid = 104857599
17/12/03 16:43:27 INFO mapred.MapTask: kvstart = 16095496(64381984); kvend = 9573556(38294224); length = 6521941/6553600
17/12/03 16:43:27 INFO mapred.MapTask: (EQUATOR) 23901990 kvi 5975492(23901968)
17/12/03 16:43:30 INFO mapred.MapTask: Finished spill 1
17/12/03 16:43:30 INFO mapred.MapTask: (RESET) equator 23901990 kv 5975492(23901968) kvi 4351724(17406896)
17/12/03 16:43:30 INFO mapred.LocalJobRunner:
17/12/03 16:43:30 INFO mapred.MapTask: Starting flush of map output
17/12/03 16:43:30 INFO mapred.MapTask: Spilling map output
17/12/03 16:43:30 INFO mapred.MapTask: bufstart = 23901990; bufend = 55015258; bufvoid = 104857600
17/12/03 16:43:30 INFO mapred.MapTask: kvstart = 5975492(23901968); kvend = 2465372(9861488); length = 3510121/6553600
17/12/03 16:43:31 INFO mapreduce.Job: map 34% reduce 0%
17/12/03 16:43:32 INFO mapred.Merger: Merging 3 sorted segments
17/12/03 16:43:32 INFO mapred.Merger: Down to the last merge-pass, with 3 segments left of total size: 154654683 bytes
```

```
ubuntu@ip-172-31-4-162: ~/msadique
17/12/03 20:13:26 INFO mapred.LocalJobRunner: Finishing task: attempt_local1422652445_0001_r_000000_0
17/12/03 20:13:26 INFO mapred.LocalJobRunner: reduce task executor complete.
17/12/03 20:13:27 INFO mapreduce.Job: Job job_local1422652445_0001 completed successfully
17/12/03 20:13:27 INFO mapreduce.Job: Counters: 35
    File System Counters
        FILE: Number of bytes read=50640525811980
        FILE: Number of bytes written=81042198608835
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=36408961178640
        HDFS: Number of bytes written=6619618197000
        HDFS: Number of read operations=727470
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=67815
    Map-Reduce Framework
        Map input records=2748779008
        Map output records=2748779008
        Map output bytes=42748779008
        Map output materialized bytes=2748779008
        Input split bytes=291232
        Combine input records=26323542701
        Combine output records=26231952875
        Reduce input groups=2748779008
        Reduce shuffle bytes=49446296224
        Reduce input records=1323830316
        Reduce output records=1230223174
        Spilled Records=4949923899
        Shuffled Maps =2048
        Failed Shuffles=0
        Merged Map outputs=2048
        GC time elapsed (ms)=245160
        Total committed heap usage (bytes)=1095350878208
    Shuffle Errors
        BAD_ID=0
        CONNECTION=0
        IO_ERROR=0
        WRONG_LENGTH=0
        WRONG_MAP=0
        WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=1095350878208
    File Output Format Counters
        Bytes Written=1095350878208
real    223m46.612s
user    201m21.128s
sys     1m23.728s
```

```
ubuntu@ip-172-31-4-162: ~/msadique$
```

c. Spark TeraSort:

```
ubuntu@ip-172-31-71-212:~/msadique
root@ip-172-31-71-212 ~]$ spark-1.6.0-bin-hadoop2.6/bin/spark-shell
log4j:WARN No appenders could be found for logger (org.apache.hadoop.metrics2.lib.MutableMetricsFactory).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Using Spark's repl log4j profile: org/apache/spark/log4j-defaults-repl.properties
To adjust logging level use sc.setLogLevel("INFO")
Welcome to

    \   _ _   _ _ _ 
    / \ / \ / \ / \ 
  /_ /_ /_ /_ /_ /_ \
  version 1.6.0

Using Scala version 2.10.5 (OpenJDK 64-Bit Server VM, Java 1.7.0_151)
Type in expressions to have them evaluated.
Type :help for more information.
Spark context available as sc.
17/12/03 14:46:40 WARN Connection: BoneCP specified but not present in CLASSPATH (or one of dependencies)
17/12/03 14:46:47 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.verification is not
17/12/03 14:46:47 WARN ObjectStore: Failed to get database default, returning NoSuchObjectException
17/12/03 14:46:50 WARN Connection: BoneCP specified but not present in CLASSPATH (or one of dependencies)
SQL context available as sqlContext.

scala> :load /root/SparkTera.scala
Loading /root/SparkTera.scala...
lines: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[1] at textFile at <console>:27
file Size(bytes):109511627776
sort: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[7] at map at <console>:29
Sorting time for splitted file : 1781.0234645
Total Spark Tera Sort Time (Sort, Merge and Transfer To Local) : 6704.2445635

root@ip-172-31-71-212 ~]$
```

d. MPI TeraSort:

No of Nodes :1

File Size: 1024 Gb

```
ubuntu@ip-172-31-4-162: ~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 1 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 1

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.50 GB
Total No. chunks Created Based on Available Memory & Nodes : 32

***** Processor - 1 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 4091.06 sec
Sorting Time for chunks        = 88901.92 sec
Merging Time for Sorted chunks = 5826.44 sec

Total MPI Tera Sorting Time    = 98819.43 sec
-----
ubuntu@ip-172-31-4-162:~/msadique/CloudMPI$
```

No of Nodes:2

File Size: 1024 GB

```
ubuntu@ip-172-31-4-162: ~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 2 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 2

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.50 GB
Total No. chunks Created Based on Available Memory & Nodes : 48

***** Processor - 1 Started Sorting *****
***** Processor - 2 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3999.08 sec
Sorting Time for chunks        = 41836.2 sec
Merging Time for Sorted chunks = 6006.64 sec

Total MPI Tera Sorting Time    = 51841.93 sec
-----
ubuntu@ip-172-31-4-162:~/msadique/CloudMPI$
```

No of Nodes:4

File Size: 1024 GB

```
ubuntu@ip-172-31-4-162:~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 4 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 4

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.49 GB
Total No. chunks Created Based on Available Memory & Nodes   : 64

***** Processor - 1 Started Sorting *****
***** Processor - 3 Started Sorting *****
***** Processor - 2 Started Sorting *****
***** Processor - 4 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 3909.17 sec
          Sorting Time for chunks       = 21665.17 sec
          Merging Time for Sorted chunks = 6192.41 sec

          Total MPI Tera Sorting Time   = 31766.76 sec
-----
```

ubuntu@ip-172-31-4-162:~/msadique/CloudMPI\$

No of Nodes:8

File Size: 1024 GB

```
ubuntu@ip-172-31-4-162:~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 8 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 8

Input File Size      : 1024 GB
Total Memory(RAM)    : 120.00 GB
Available Memory(RAM) : 117.47 GB
Total No. chunks Created Based on Available Memory & Nodes : 96

***** Processor - 1 Started Sorting *****
***** Processor - 2 Started Sorting *****
***** Processor - 4 Started Sorting *****
***** Processor - 5 Started Sorting *****
***** Processor - 7 Started Sorting *****
***** Processor - 8 Started Sorting *****
***** Processor - 3 Started Sorting *****
***** Processor - 6 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
          Splitting Time for chunks      = 3821.28 sec
          Sorting Time for chunks       = 14119.71 sec
          Merging Time for Sorted chunks = 6383.93 sec

          Total MPI Tera Sorting Time   = 24324.93 sec
-----
```

ubuntu@ip-172-31-4-162:~/msadique/CloudMPI\$

```
ubuntu@ip-172-31-4-162: ~/msadique/CloudMPI
ubuntu@ip-172-31-4-162:~/msadique/CloudMPI$ make
mpicc -o run TeraMPISort.c -lm
mpirun -np 16 ./run input.txt output.txt

***      MPI Terasort using MPI and MergeSort      ***

Input file name: input.txt
Output file name: output.txt
Number of Nodes: 16

Input File Size      : 1024 GB
Total Memory(RAM)   : 120.00 GB
Available Memory(RAM) : 117.45 GB
Total No. chunks Created Based on Available Memory & Nodes : 192

***** Processor - 1 Started Sorting *****
***** Processor - 7 Started Sorting *****
***** Processor - 3 Started Sorting *****
***** Processor - 6 Started Sorting *****
***** Processor - 8 Started Sorting *****
***** Processor - 4 Started Sorting *****
***** Processor - 15 Started Sorting *****
***** Processor - 9 Started Sorting *****
***** Processor - 10 Started Sorting *****
***** Processor - 5 Started Sorting *****
***** Processor - 14 Started Sorting *****
***** Processor - 16 Started Sorting *****
***** Processor - 2 Started Sorting *****
***** Processor - 11 Started Sorting *****
***** Processor - 12 Started Sorting *****
***** Processor - 13 Started Sorting *****

In Memory Sorting Complete

Main Thread : Merging Started for Thread Chunks

-----
Splitting Time for chunks      = 3735.37 sec
Sorting Time for chunks        = 7470.75 sec
Merging Time for Sorted chunks = 6581.37 sec

Total MPI Tera Sorting Time    = 17787.51 sec
-----
ubuntu@ip-172-31-4-162:~/msadique/CloudMPI$
```

```
ubuntu@ip-172-31-1-36:~/msadique$  
17/12/03 23:22:31 INFO mapred.LocalJobRunner: reduce task executor complete.  
17/12/03 23:22:31 INFO mapreduce.Job: map 100% reduce 100%  
17/12/03 23:22:32 INFO mapreduce.Job: Job job_local1717214462_0001 completed successfully  
17/12/03 23:22:32 INFO mapreduce.Job: Counters: 35  
File System Counters  
FILE: Number of bytes read=50526155135500  
FILE: Number of bytes written=93661908367875  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=42078483114000  
HDFS: Number of bytes written=7650410325000  
HDFS: Number of read operations=840750  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=78375  
Map-Reduce Framework  
Map input records=3350074416  
Map output records=3350074416  
Map output bytes=335007441600  
Map output materialized bytes=45457800624  
Input split bytes=512000  
Combine input records=3350074416  
Combine output records=3350074416  
Reduce input groups=323435932453  
Reduce shuffle bytes=45427802345  
Reduce input records=34543597376  
Reduce output records=34543597376  
Spilled Records=4523513376  
Shuffled Maps =2048  
Failed Shuffles=0  
Merged Map outputs=2048  
GC time elapsed (ms)=243890  
Total committed heap usage (bytes)=136918859776  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
File Input Format Counters  
Bytes Read=1099511627776  
File Output Format Counters  
Bytes Written=1099511627776  
  
real 274m34.512s  
user 223m19.888s  
sys 0m23.728s
```

```
ubuntu@ip-172-31-1-36:~/msadique$
```

```
ubuntu@ip-172-31-1-36:~/msadique$ time hadoop jar hadoopTeraSort.jar /input /output
17/12/03 17:21:48 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
17/12/03 17:21:48 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/12/03 17:21:48 INFO input.FileInputFormat: Total input files to process : 1
17/12/03 17:21:48 INFO mapreduce.JobSubmitter: number of splits:3
17/12/03 17:21:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1717214462_0001
17/12/03 17:21:48 INFO mapreduce.Job: The url to track the job: http://localhost:9080/
17/12/03 17:21:48 INFO mapreduce.Job: Running job: job_local1717214462_0001
17/12/03 17:21:48 INFO mapred.LocalJobRunner: OutputCommitter set in config null
17/12/03 17:21:48 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 17:21:48 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 17:21:48 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
17/12/03 17:21:49 INFO mapred.LocalJobRunner: Waiting for map tasks
17/12/03 17:21:49 INFO mapred.LocalJobRunner: Starting task: attempt_local1717214462_0001_m_000000_0
17/12/03 17:21:49 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 17:21:49 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 17:21:49 INFO mapred.Task: Using ResourceCalculatorProcessTree : []
17/12/03 17:21:49 INFO mapred.MapTask: Processing split: hdfs://172.31.1.36/input4:0+134217728
17/12/03 17:21:49 INFO mapred.MapTask: (EQUATOR) 0 kvi 20214396(104857584)
17/12/03 17:21:49 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/12/03 17:21:49 INFO mapred.MapTask: soft limit at 83886080
17/12/03 17:21:49 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/12/03 17:21:49 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/12/03 17:21:49 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/12/03 17:21:49 INFO mapreduce.Job: Job job_local1717214462_0001 running in uber mode : false
17/12/03 17:21:49 INFO mapreduce.Job: map 0% reduce 0%
17/12/03 17:21:50 INFO mapred.MapTask: Spilling map output
17/12/03 17:21:50 INFO mapred.MapTask: bufstart = 0; bufend = 72315600; bufvoid = 104857600
17/12/03 17:21:50 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 23321776(93207104); length = 389261/6553600
17/12/03 17:21:50 INFO mapred.MapTask: (EQUATOR) 75208208 kvi 18802048(75208192)
17/12/03 17:21:53 INFO mapred.MapTask: Finished spill 0
17/12/03 17:21:53 INFO mapred.MapTask: (RESET) equator 75208208 kv 18802048(75208192) kvi 18078904(72315616)
17/12/03 17:21:54 INFO mapred.LocalJobRunner:
17/12/03 17:21:54 INFO mapred.MapTask: Starting flush of map output
17/12/03 17:21:54 INFO mapred.MapTask: Spilling map output
17/12/03 17:21:54 INFO mapred.MapTask: bufstart = 75208208; bufend = 32252808; bufvoid = 104857600
17/12/03 17:21:54 INFO mapred.MapTask: kvstart = 18802048(75208192); kvend = 16325964(65303856); length = 2476085/6553600
17/12/03 17:21:56 INFO mapred.MapTask: Finished spill 1
17/12/03 17:21:56 INFO mapred.Merger: Merging 2 sorted segments
17/12/03 17:21:56 INFO mapred.Merger: Down to the last merge-pass, with 2 segments left of total size: 13690141 bytes
17/12/03 17:21:57 INFO mapred.Task: Task:attempt_local1717214462_0001_m_000000_0 is done. And is in the process of committing
17/12/03 17:21:57 INFO mapred.LocalJobRunner: map > sort
17/12/03 17:21:57 INFO mapred.Task: Task 'attempt_local1717214462_0001_m_000000_0' done.
17/12/03 17:21:57 INFO mapred.LocalJobRunner: Finishing task attempt_local1717214462_0001_m_000000_0
17/12/03 17:21:57 INFO mapred.LocalJobRunner: Starting task: attempt_local1717214462_0001_m_000001_0
17/12/03 17:21:57 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/12/03 17:21:57 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
17/12/03 17:21:57 INFO mapred.Task: Using ResourceCalculatorProcessTree : []
17/12/03 17:21:57 INFO mapred.MapTask: Processing split: hdfs://172.31.1.36/input:134217728+134217728
17/12/03 17:21:57 INFO mapred.MapTask: (EQUATOR) 0 kvi 20214396(104857584)
```