1. File & Directory Management (Handling Large Datasets)

Command	Description	Example
`ls -lh`	List files in human-readable size	`ls -lh /data/`
`du -sh`	Check directory size	`du -sh /data/`
`df -h`	Check available disk space	`df -h`
`find`	Locate files based on criteria	`find /data -name "*.csv"`
`xargs`	Execute command on multiple files	`` `find /data -name "*.log" ``
`rsync`	Sync large data files	`rsync -av /data/ backup/`
`split`	Split large files into smaller chunks	`split -b 500M bigfile.csv part_`

2. File Processing (CSV, JSON, Logs, etc.)

Command	Description	Example
`cat`	Display file contents	`cat data.csv`
`head`	Show first N lines of a file	`head -n 10 data.csv`
`tail`	Show last N lines of a file	`tail -n 10 data.csv`
`less`	View large files interactively	`less biglog.log`
`awk`	Extract and process data from files	`awk -F',' '{print $1, $3}' data.csv`
`sed`	Find and replace text in a file	`sed 's/old/new/g' data.csv`
`grep`	Search for patterns in files	`grep "error" logfile.log`
`sort`	Sort a file's contents	`sort -k2 -n data.csv`
`uniq`	Remove duplicate lines	`uniq sorted_data.csv`
`cut`	Extract specific columns	`cut -d',' -f2,3 data.csv`
`jq`	Parse JSON files	`` `cat data.json ``

3. Data Transfer & Networking

Command	Description	Example
`scp`	Secure copy between machines	`scp file.csv user@server:/data/`
`wget`	Download data from URLs	`wget http://example.com/data.csv`
`curl`	Fetch data from an API	`curl -X GET https://api.example.com/data`
`netstat`	Monitor network connections	`netstat -tulnp`
`ss`	Show active network connections	`ss -tulnp`
`ping`	Check server connectivity	`ping google.com`

3. Data Transfer & Networking

Command	Description	Example
`scp`	Secure copy between machines	`scp file.csv user@server:/data/`
`wget`	Download data from URLs	`wget http://example.com/data.csv`
`curl`	Fetch data from an API	`curl -X GET https://api.example.com/data`
`netstat`	Monitor network connections	`netstat -tulnp`
`ss`	Show active network connections	`ss -tulnp`
`ping`	Check server connectivity	`ping google.com`

4. Database Management (PostgreSQL, MySQL, etc.)

Command	Description	Example
`mysql -u user -p`	Login to MySQL	`mysql -u root -p`
`psql -U user -d database`	Login to PostgreSQL	`psql -U postgres -d mydb`
`mysqldump`	Backup MySQL database	`mysqldump -u root -p dbname > backup.sql`
`pg_dump`	Backup PostgreSQL database	`pg_dump -U postgres mydb > backup.sql`
`mongo`	Connect to MongoDB	`mongo --host localhost --port 27017`
`sqlite3`	Open SQLite database	`sqlite3 mydatabase.db`

5. Parallel Processing (For Handling Big Data)

Command	Description	Example
`nohup`	Run command in the background	`nohup python script.py &`
`screen`	Run a persistent session	`screen -S mysession`
`parallel`	Execute tasks in parallel	`` `cat urls.txt ``
`xargs -P`	Run multiple processes at once	`` `cat files.txt ``

6. System Monitoring & Performance Optimization

Command	Description	Example
`top`	Show running processes	`top`
`htop`	Interactive process monitor	`htop`
`vmstat`	Show CPU and memory usage	`vmstat 5`
`iostat`	Show disk read/write speed	`iostat -dx 5`
`free -m`	Show memory usage	`free -m`
`ulimit`	Set file limits for large datasets	`ulimit -n 100000`

7. Compression & Archiving

Command	Description	Example
`tar -czvf`	Compress folder to `.tar.gz`	`tar -czvf data.tar.gz /data/`
`tar -xzvf`	Extract `.tar.gz` archive	`tar -xzvf data.tar.gz`
`zip`	Create a `.zip` file	`zip -r data.zip /data/`
`unzip`	Extract `.zip` file	`unzip data.zip`

8. Scheduling & Automation (Cron Jobs)

Command	Description	Example
`crontab -e`	Edit cron jobs	`crontab -e`
`crontab -l`	List scheduled tasks	`crontab -l`
`echo "0 2 * * * /path/to/script.sh"	crontab -`	Run a script daily at 2 AM

9. Data Streaming (Kafka, Hadoop, Spark, etc.)

Command	Description	Example
`kafka-console-producer.sh`	Send data to Kafka topic	`kafka-console-producer.sh --topic my_topic --broker-list localhost:9092`
`kafka-console-consumer.sh`	Read data from Kafka topic	`kafka-console-consumer.sh --topic my_topic --bootstrap-server localhost:9092`
`hdfs dfs -ls /`	List Hadoop HDFS directories	`hdfs dfs -ls /user/data`
`spark-submit`	Run a Spark job	`spark-submit --master local script.py`

10. ETL (Extract, Transform, Load) Operations

Command	Description	Example
`python`	Run Python ETL scripts	`python etl_script.py`
`awk -F',' '{print $1, $3}' file.csv`	Extract specific columns	
`sed 's/old/new/g' file.csv`	Transform data in a file	
`pgloader`	Load data into PostgreSQL	`pgloader csv_to_pg.load`
`sqoop import`	Import data from SQL to Hadoop	`sqoop import --connect jdbc:mysql://host/db --table employees --target-dir /hadoop/employees`