



Cloud Computing

A Presentation by Ole
Methler



Contents

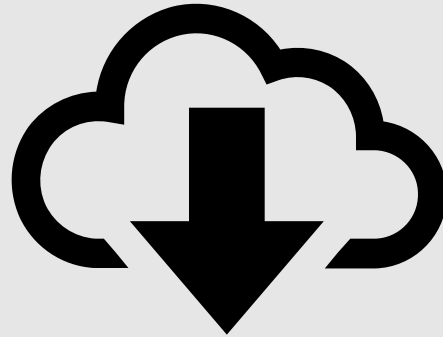
- What is Cloud Computing?
 - How does it work?
 - Where did it come from?
- What is MapReduce, how do we use it in Cloud Computing?
- What are Containers, how can we use them?
- What is the future of Cloud Computing?
- Bibliography



Cloud Computing

- “Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user”. (Wikipedia)
- Main Types of Cloud Computing:
 - Infrastructure as a Service, IaaS (computing power, data storage)
 - Platform as a Service, PaaS (databases, web servers)
 - Software as a Service, SaaS (email, game-streaming)
 - Mobile “backend” as a service and Serverless computing

Cloud Computing



- Data is Stored in a separate server
- When requested by the system, data is provided by the cloud
- Data can also be processed in the Cloud

Cloud Computing



- The cloud symbol/metaphor originates from diagrams where the cloud was used to represent “other” systems, that were not important to the main one

Cloud Computing

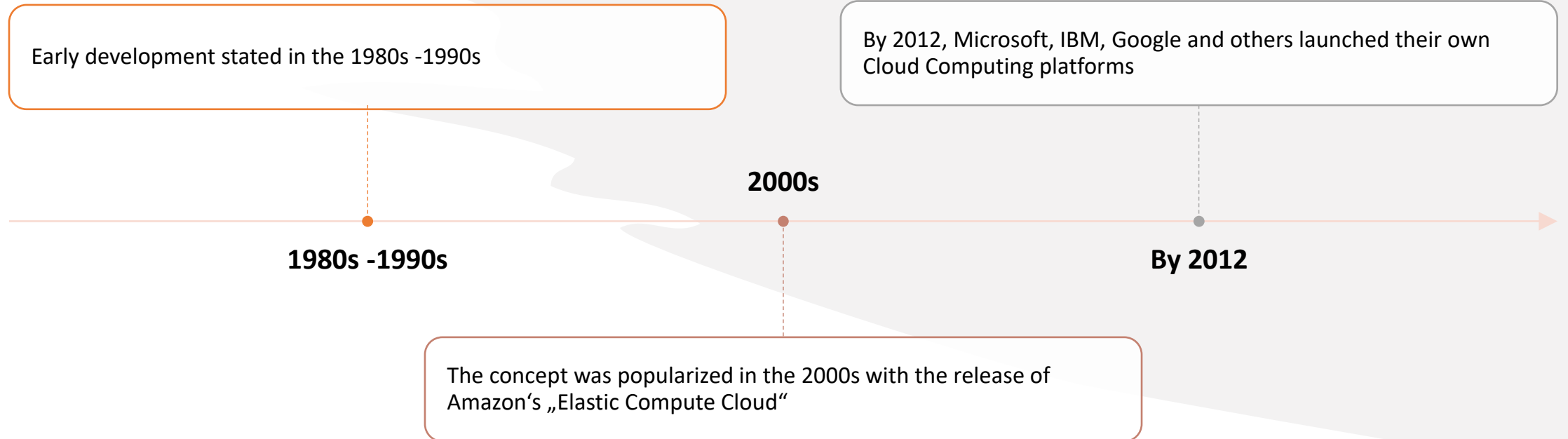
Advantages

- Allows much smaller enterprises to work on much larger scales

Disadvantages

- High frequency, stable internet connection required
- Additional cost

History of Cloud Computing



MapReduce

MapReduce is a processing technique that can be used to process huge amounts of data simultaneously

Sending the computer to the data instead of the other way around

Made up of two „stages“

- Mapping/Map stage – Data is processed by a „mapper“ and divided into several small chunks of data
- Reducing/Reduce stage – These chunks are first „shuffled“ and then processed individually

The Reduced data is stored in the same place

Hadoop, MapReduce and HDFS

Hadoop

- A framework for data analysis
- Developed by Apache

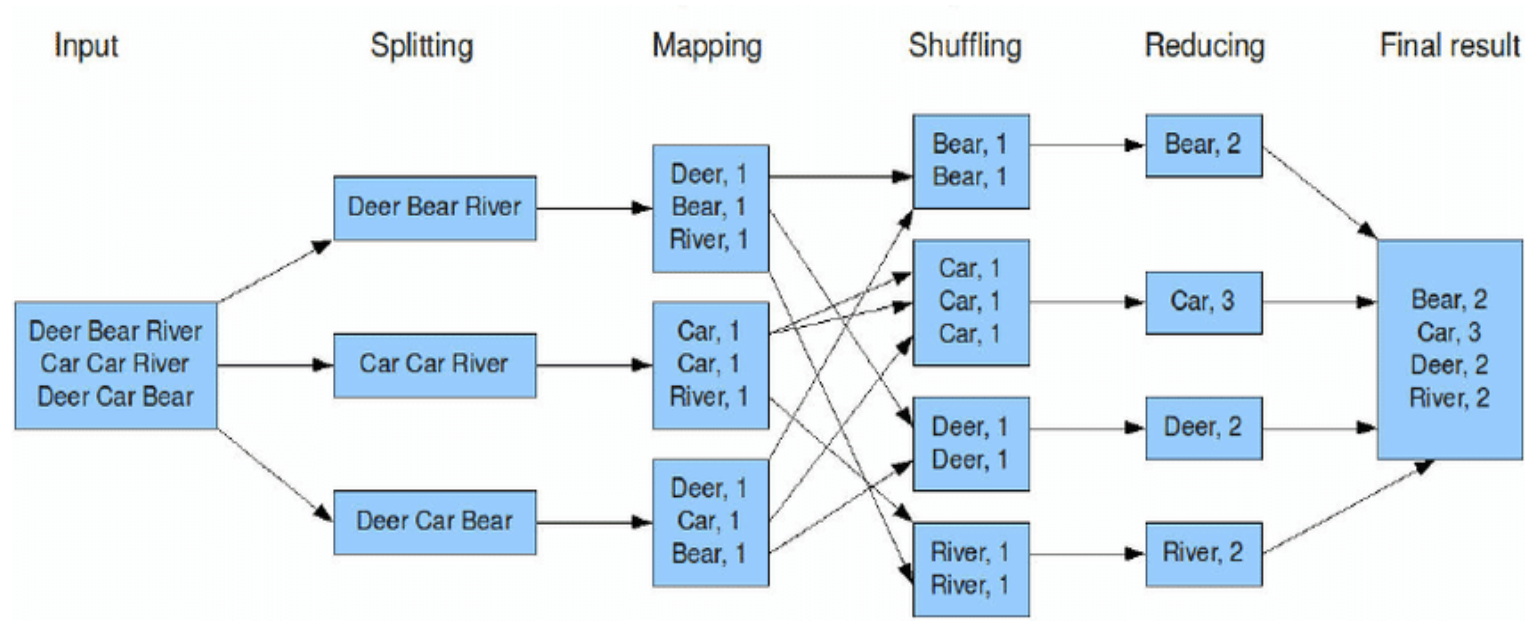
MapReduce

- A method that is implemented by hadoop
- Developed by Google

Hadoop File System (HDFS)

- The file system that makes access to the data possible

MapReduce wordcount example



Wordcount example

Youth is not a time of life; it is a state of mind; it is not a matter of rosy cheeks, red lips and supple knees;

it is a matter of the will, a quality of the imagination, a vigor of the emotions; it is the freshness of the deep springs of life.

Youth means a temperamental predominance of courage over timidity of the appetite, for adventure over the love of ease.

This often exists in a man of sixty more than a body of twenty. Nobody grows old merely by a number of years. We grow old

deserting our ideals.

Years may wrinkle the skin, but to give up enthusiasm wrinkles the soul. Worry, fear, self-distrust bows the heart and turns the spirit back to dust.

Whether sixty or sixteen, there is in every human being's heart the lure of wonder,

the unfailing child-like appetite of what's next, and the joy of the game of living.

In the center of your heart and my heart there is a wireless station; so long as it receives messages of beauty,

hope, cheer, courage and power from men and from the Infinite, so long are you young.

When the aerials are down, and your spirit is covered with snows of cynicism and the ice of pessimism,

then you are grown old, even at twenty, but as long as your aerials are up, to catch the waves of optimism, there is hope you may die young at eighty.

"Youth" by Samuel Ullman

Wordcount Code 1/2

```
package polyu.bigdata;

import java.io.IOException;
import java.util.StringTokenizer;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {

    //Mapper which implement the mapper() function
    public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable> {

        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
            while (itr.hasMoreTokens()) {

                word.set(itr.nextToken());
                context.write(word, one);
            }
        }
    }
}
```

Wordcount Code 2/2

```
//Reducer which implement the reduce() function
public static class IntSumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
    private IntWritable result = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        result.set(sum);
        context.write(key, result);
    }
}

//Driver class to specific the Mapper and Reducer
public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(WordCount.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setReducerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    job.setMapOutputKeyClass(Text.class);
    job.setMapOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}
```

Wordcount

```
~/Programs/hadoop/sbin/start-all.sh
```

- Starts the hadoop application

```
~/Programs/hadoop/bin/hadoop fs -mkdir -p  
/user/bigdata/wordcount/input
```

- Creates a directory for the input

```
~/Programs/hadoop/bin/hadoop fs -put ~/hadoop.txt  
/user/bigdata/wordcount/input
```

- Uploads the input text file into the file system

```
~/Programs/hadoop/bin/hadoop jar ~/wordcount.jar  
polyu.bigdata.WordCount /user/bigdata/wordcount/input  
/user/bigdata/wordcount/output
```

- Executes the Code within the hadoop framework

Wordcount result

All Applications

Cluster Metrics

| Apps Submitted | Apps Pending | Apps Running | Apps Completed | Containers Running | Memory Used | Memory Total | Memory Reserved | VCores Used | VCores Total | VCores Reserved | Active Nodes | Decommissioned Nodes | Lost Nodes |
|----------------|--------------|--------------|----------------|--------------------|-------------|--------------|-----------------|-------------|--------------|-----------------|--------------|----------------------|------------|
| 2 | 0 | 0 | 2 | 0 | 0 B | 8 GB | 0 B | 0 | 8 | 0 | 1 | 0 | 0 |

Show 20 entries

| ID | User | Name | Application Type | Queue | StartTime | FinishTime | State | FinalStatus | Progress | Trace |
|--|---------|------------|------------------|---------|-------------------------------|-------------------------------|----------|-------------|----------------------------------|----------------------|
| application_1453354809925_0002 | bigdata | word count | MAPREDUCE | default | Thu, 21 Jan 2016 06:34:03 GMT | Thu, 21 Jan 2016 06:34:14 GMT | FINISHED | SUCCEEDED | <div style="width: 100%;"></div> | Hist |
| application_1453354809925_0001 | bigdata | word count | MAPREDUCE | default | Thu, 21 Jan 2016 06:19:55 GMT | Thu, 21 Jan 2016 06:20:12 GMT | FINISHED | FAILED | <div style="width: 100%;"></div> | Hist |

Showing 1 to 2 of 2 entries

```
In 1
Infinite, 1
Nobody 1
This 1
We 1
When 1
Whether 1
Worry, 1
Years 1
Youth 2
a 11
adventure 1
aerials 2
and 8
```

Containers/Singularity

A container is a System that is packaged into a singular unit for simpler usage

Contained software can run on any system with minimal adjustments

Singularity creates Containers for scientific computing with the main goals being reproducibility and mobility

Uses of Containers

Publish

Publish software easily

Integrate

Integrate programs into an existing system

Run

Run Linux-based software on Windows

Simplify

Simplify the process of similar tasks

How to use Singularity

- **Running containers from the Cloud**
- `sudo singularity build (platform)://(path)`
 - `sudo singularity build docker://sequenciq/hadoop-docker`
 - `sudo singularity pull library://library/default/ubuntu:18.04`
- Either downloading, or running directly in the cloud

Sylabs.io Search Cloud Library...

Home Cloud Library Remote Builder Keystore Help Sign In

Cloud Library

Cloud Library is the official image registry provided by **Sylabs.io**. Users can share Singularity images through the Cloud Library, as well as pull/push SIF™ images through Singularity CLI. Email any feature requests or feedback to support@sylabs.io.

Share your awesome containers!
[Sign In](#)

ORTHEDEN/DEFAULT/IMAGE
Description: No Description
Download 378.36 MB
AMD64 STARS: ☆ 0 DOWNLOADS: 6

Signed With: 236794da9fcdd1f5b4e5b65725891393f940bb21

JAMES-S-SANTANGELO/PCANGSD/PCANGSD
Description: No Description
Download 325.06 MB
AMD64 STARS: ☆ 0 DOWNLOADS: 0

MOST DOWNLOADED

| | |
|---|--------|
| library/default/alpine | 401445 |
| sylabs-bot/smoke-test-collection/testimage0 | 210362 |
| sylabs-bot/smoke-test-collection/testimage1 | 210336 |
| library/default/busybox | 204255 |
| sylabs/tests/not-default | 106845 |
| sylabs/tests/unsigned | 60762 |
| library/default/ubuntu | 24856 |
| sylabs/tests/signed | 22482 |
| sylabsed/examples/olcow | 12082 |
| library/default/centos | 10263 |

docker hub Search for great content (e.g. mysql) Explore Pricing Sign In Sign Up

Docker Containers Plugins

Filters 1 - 25 of 6.935.956 available images. Most Popular

Images

- Verified Publisher
- Official Images
Official Images Published By Docker

Categories

- Analytics
- Application Frameworks
- Application Infrastructure
- Application Services
- Base Images
- Databases
- DevOps Tools
- Featured Images
- Messaging Services
- Monitoring
- Operating Systems
- Programming Languages

ubuntu OFFICIAL IMAGE
Updated 11 hours ago
10M+ Downloads 10K+ Stars
Ubuntu is a Debian-based Linux operating system based on free software.
Container Linux x86-64 PowerPC 64 LE 386 ARM ARM 64 IBM Z Base Images Operating Systems

redis OFFICIAL IMAGE
Updated 11 hours ago
10M+ Downloads 9.4K Stars
Redis is an open source key-value store that functions as a data structure server.
Container Linux Windows 386 ARM 64 IBM Z ARM x86-64 mips64le PowerPC 64 LE Databases

postgres OFFICIAL IMAGE
Updated 11 hours ago
10M+ Downloads 9.3K Stars
The PostgreSQL object-relational database system provides reliability and data integrity.

How to use Singularity

- **Building your own containers**
- Requires a definition file
 - Consists of a header and a few sections defining how the container
- `sudo singularity build (container.sif) (definition.def)`

```
Bootstrap: library
From: ubuntu:18.04

%setup
    touch /file1
    touch ${SINGULARITY_ROOTFS}/file2

%files
    /file1
    /file1 /opt

%environment
    export LISTEN_PORT=12345
    export LC_ALL=C

%post
    apt-get update && apt-get install -y netcat
    NOW=`date`
    echo "export NOW=\"${NOW}\"" >> $SINGULARITY_ENVIRONMENT

%runscript
    echo "Container was created $NOW"
    echo "Arguments received: $*"
    exec echo "$@"

%startscript
    nc -lp $LISTEN_PORT

%test
    grep -q NAME="\Ubuntu\" /etc/os-release
    if [ $? -eq 0 ]; then
        echo "Container base is Ubuntu as expected."
    else
        echo "Container base is not Ubuntu."
    fi

%labels
    Author d@sylabs.io
    Version v0.0.1

%help
    This is a demo container used to illustrate a def file that uses
    supported sections.
```

The Future of Cloud Computing



ARTIFICIAL INTELLIGENCE



GAMING



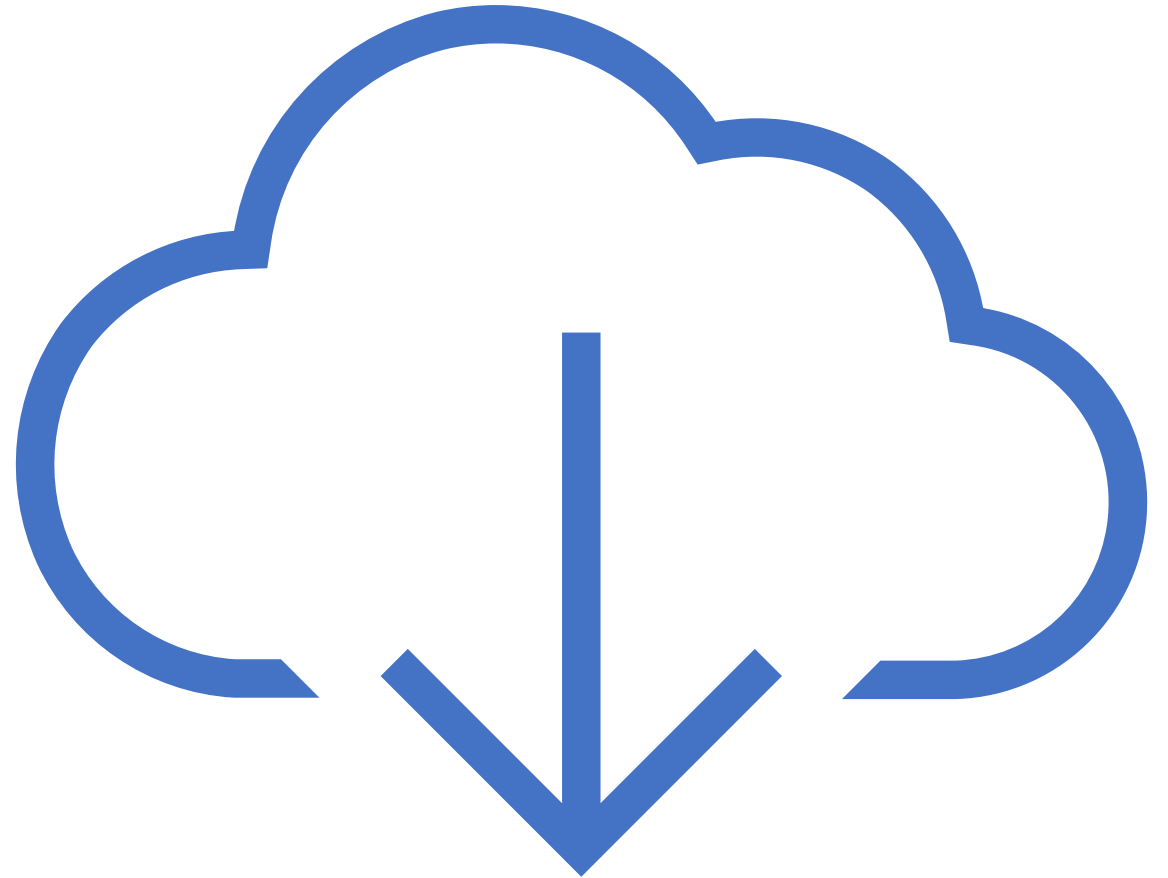
MORE INTEGRATION INTO
NON-TECH COMPANIES



GENERAL EXPANSION

Thank you for
watching

Final remarks



Bibliography

Info Sources

- <https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-about-the-cloud/>
- <https://docs.microsoft.com/en-us/learn/modules/cmu-analytics-engines-mapreduce/>
- https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
- https://sylabs.io/guides/3.0/user-guide/quick_start.html
- https://en.wikipedia.org/wiki/Amazon_Web_Services (24.04.21)
- https://en.wikipedia.org/wiki/Cloud_computing (24.04.21)
- <https://www.leadingedgetech.co.uk/it-services/it-consultancy-services/cloud-computing/what-are-the-types-of-cloud-computing/> (30.04.21)
- https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm (30.04.21)
- <https://www.forbes.com/sites/bernardmarr/2020/11/02/the-5-biggest-cloud-computing-trends-in-2021/?sh=3967f77212d9> (30.04.21)
- <https://www.netapp.com/devops-solutions/what-are-containers/> (08.05.21)

Picture Sources

- https://www.researchgate.net/figure/Word-count-program-flow-executed-with-MapReduce-5_fig6_270448794 (08.05.21)
- <http://www.cse.cuhk.edu.hk/~ericlo/teaching/bigdata/lab/2-HadoopMR/HadoopMR.html> (12.05.21)
- <https://sylabs.io/guides/3.0/user-guide/> (12.05.21)