E534 - Big Data Applications
Lecture Notes

Geoffrey C. Fox

Gregor von Laszewski

Editor

laszewski@gmail.com

https://cloudmesh-community.github.io/book/vonlaszewski-e534.epub

September 03, 2019 - 02:27 PM

Created by Cloudmesh & Cyberaide Bookmanager, https://github.com/cyberaide/bookmanager

E534 - Big Data Applications

Geoffrey C. Fox Gregor von Laszewski

E534 - BIG DATA APPLICATIONS

1 PREFACE
1.1 Disclaimer
1.1.1 Acknowledgment
1.1.2 Extensions
2 WEEK 1
2.1 Week 1
2.1.1 A. Summary of Course
2.1.2 B. Defining Clouds I
2.1.3 C. Defining Clouds II
2.1.4 D. Defining Clouds III: Cloud Market Share
2.1.5 E. Virtualization: Virtualization Technologies,
2.1.6 F. Cloud Infrastructure I
2.1.7 G. Cloud Infrastructure II
2.1.8 H. Cloud Software:
2.1.9 I. Cloud Applications I: Clouds in science where area called
2.1.10 J. Cloud Applications II: Characterize Applications using NIST
2.1.11 K. Parallel Computing
2.1.12 L. Real Parallel Computing: Single Program/Instruction Multiple Data SIMD SPMD
2.1.13 M. Storage: Cloud data
2.1.14 N. HPC and Clouds
2.1.15 O. Comparison of Data Analytics with Simulation:
2.1.16 P. The Future I
2.1.17 Q. other Issues II
2.1.18 R. The Future and other Issues III
3 REFERENCES

1 PREFACE

Tue Sep 3 14:27:49 EDT 2019

1.1 Disclaimer

This book has been generated with Cyberaide Bookmanager.

Bookmanager is a tool to create a publication from a number of sources on the internet. It is especially useful to create customized books, lecture notes, or handouts. Content is best integrated in markdown format as it is very fast to produce the output.

Bookmanager has been developed based on our experience over the last 3 years with a more sophisticated approach. Bookmanager takes the lessons from this approach and distributes a tool that can easily be used by others.

The following shields provide some information about it. Feel free to click on them.

pypi v0.2.20 | License Apache 2.0 | python 3.7 | format wheel | status stable | build unknown

1.1.1 Acknowledgment

If you use bookmanager to produce a document you must include the following acknowledgement.

> *"This document was produced with Cyberaide Bookmanager developed by Gregor von Laszewski available at https://pypi.python.org/pypi/cyberaide-bookmanager. It is in the responsibility of the user to make sure an author acknowledgement section is included in your document. Copyright verification of content included in a book is responsibility of the book editor."*

The bibtex entry is

```
@Misc{www-cyberaide-bookmanager,
  author =   {Gregor von Laszewski},
```

```
  title =    {{Cyberaide Book Manager}},
  howpublished = {pypi},
  month =    apr,
  year =     2019,
  url={https://pypi.org/project/cyberaide-bookmanager/}
}
```

1.1.2 Extensions

We are happy to discuss with you bugs, issues and ideas for enhancements. Please use the convenient github issues at

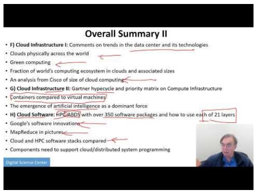
- https://github.com/cyberaide/bookmanager/issues

Please do not file with us issues that relate to an editors book. They will provide you with their own mechanism on how to correct their content.

2 WEEK 1

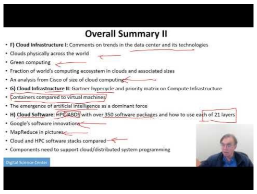
2.1 Week 1

2.1.1 A. Summary of Course



2.1.2 B. Defining Clouds I

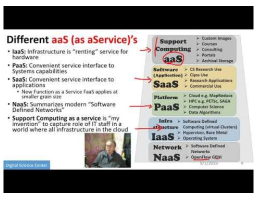
In this lecture we discuss the basic definition of cloud and two very simple examples of why virtualization is important.



How clouds are situated wrt HPC and supercomputers Why multicore chips are important Typical data center

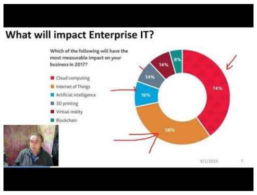
2.1.3 C. Defining Clouds II

Service-oriented architectures: Software services as Message-linked computing capabilities



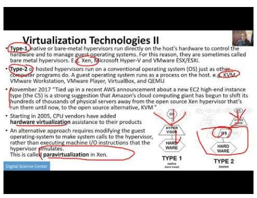
The different aaS's: Network, Infrastructure, Platform, Software The amazing services that Amazon AWS and Microsoft Azure have Initial Gartner comments on clouds (they are now the norm) and evolution of servers; serverless and microservices Gartner hypecycle and priority matrix on Infrastructure Strategies

2.1.4 D. Defining Clouds III: Cloud Market Share



How important are they? How much money do they make?

2.1.5 E. Virtualization: Virtualization Technologies,



Hypervisors and the different approaches KVM, Xen, Docker and Openstack

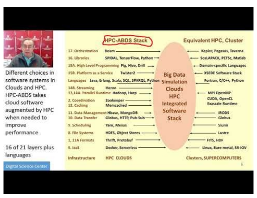
2.1.6 F. Cloud Infrastructure I

Comments on trends in the data center and its technologies Clouds physically across the world Green computing Fraction of world’s computing ecosystem in clouds and associated sizes An analysis from Cisco of size of cloud computing

2.1.7 G. Cloud Infrastructure II

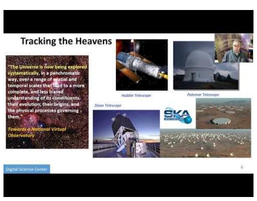
Gartner hypecycle and priority matrix on Compute Infrastructure Containers compared to virtual machines The emergence of artificial intelligence as a dominant force

2.1.8 H. Cloud Software:



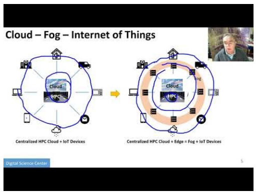
HPC-ABDS with over 350 software packages and how to use each of 21 layers Google's software innovations MapReduce in pictures Cloud and HPC software stacks compared Components need to support cloud/distributed system programming

2.1.9 I. Cloud Applications I: Clouds in science where area called



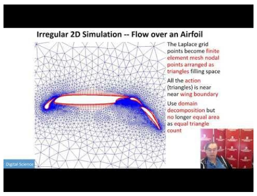
cyberinfrastructure; the science usage pattern from NIST Artificial Intelligence from Gartner

2.1.10 J. Cloud Applications II: Characterize Applications using NIST



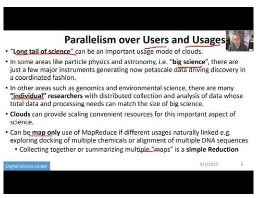
approach Internet of Things Different types of MapReduce

2.1.11 K. Parallel Computing



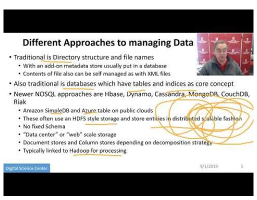
Analogies: Parallel Computing in pictures Some useful analogies and principles

2.1.12 L. Real Parallel Computing: Single Program/Instruction Multiple Data SIMD SPMD



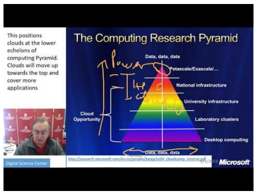
Big Data and Simulations Compared What is hard to do?

2.1.13 M. Storage: Cloud data



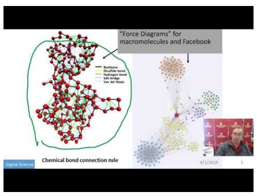
approaches Repositories, File Systems, Data lakes

2.1.14 N. HPC and Clouds



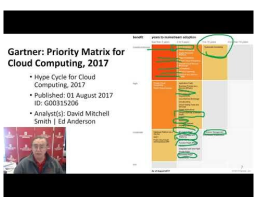
The Branscomb Pyramid Supercomputers versus clouds Science Computing Environments

2.1.15 O. Comparison of Data Analytics with Simulation:



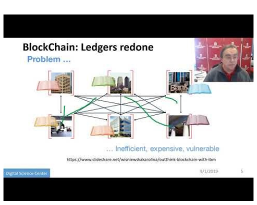
Structure of different applications for simulations and Big Data Software implications Languages

2.1.16 P. The Future I



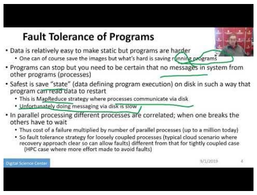
Gartner cloud computing hypecycle and priority matrix 2017 and 2019 Hyperscale computing Serverless and FaaS Cloud Native Microservices Update to 2019 Hypecycle

2.1.17 Q. other Issues II



Security Blockchain

2.1.18 R. The Future and other Issues III



Fault Tolerance

3 REFERENCES