

Big Data Applications

e534

Geoffrey C. Fox
Gregor von Laszewski

Editor

laszewski@gmail.com

<https://cloudmesh-community.github.io/book/vonlaszewski-523.epub>

September 03, 2019 - 12:02 PM

BIG DATA APPLICATIONS

Geoffrey C. Fox Gregor von Laszewski

(c) Indiana University, Gregor von Laszewski, Geoffrey Fox, 2018, 2019

BIG DATA APPLICATIONS

[1 PREFACE](#)

[2 WEEK 1](#)

[2.1 Week 1](#) 

[2.1.0.0.1 A\). Summary of Course](#)

[2.1.0.0.2 B\). Defining Clouds I: Basic definition of cloud and two very simple examples of why virtualization is important.](#)

[2.1.0.0.3 C\). Defining Clouds II: Service-oriented architectures: Software services as Message-linked computing capabilities](#)

[2.1.0.0.4 Motivation](#)

[2.1.0.0.5 00\) Mechanics of Course, Summary, and overall remarks on course](#)

[2.1.0.0.5.1 01A\) Technology Hypecycle I](#)

[2.1.0.0.5.2 01B\) Technology Hypecycle II](#)

[2.1.0.0.5.3 01C\) Technology Hypecycle III](#)

[2.1.0.0.6 01D\) Technology Hypecycle IV](#)

[2.1.0.0.7 02\)](#)

[2.1.0.0.7.1 02A\) Clouds/Big Data Applications I](#)

[2.1.0.0.7.2 02B\) Cloud/Big Data Applications II](#)

[2.1.0.0.7.3 02C\) Cloud/Big Data](#)

[2.1.0.0.8 03\) Jobs In areas like Data Science, Clouds and Computer Science and Computer](#)

[2.1.0.0.9 04\) Industry, Technology, Consumer Trends Basic trends 2018 Lectures 4A 4B have](#)

[2.1.0.0.10 05\) Digital Disruption and Transformation The Past displaced by Digital](#)

[2.1.0.0.11 06\)](#)

[2.1.0.0.11.1 06A\) Computing Model I Industry adopted clouds which are attractive for data](#)

[2.1.0.0.11.2 06B\) Computing Model II with 3 subsections is removed; please see 2018](#)

[2.1.0.0.12 07\) Research Model 4th Paradigm; From Theory to Data driven science?](#)

[2.1.0.0.13 08\) Data Science Pipeline DIKW: Data, Information, Knowledge, Wisdom, Decisions.](#)

[2.1.0.0.14 09\) Physics: Looking for Higgs Particle with Large Hadron Collider LHC Physics](#)

[2.1.0.0.15 10\) Recommender Systems I General remarks and Netflix example](#)

[2.1.0.0.16 11\) Recommender Systems II Exploring Data Bags and Spaces](#)

[2.1.0.0.17 12\) Web Search and Information Retrieval Another Big Data Example](#)

[2.1.0.0.18 13\) Cloud Applications in Research Removed Science Clouds, Internet of Things](#)

[2.1.0.0.19 14\) Parallel Computing and MapReduce Software Ecosystems](#)

[2.1.0.0.20 15\) Online education and data science education Removed.](#)

[2.1.0.0.21 16\) Conclusions](#)

[2.1.0.0.22 2018BDAAMotivation-1A\) Technology Hypecycle I General Remarks including Hype curves](#)

[2.1.0.0.23 2018BDAAMotivation-1B\) Technology Hypecycle II General Remarks including Hype curves](#)

[2.1.0.0.24 2018BDAAMotivation-2B\) Cloud/Big Data Applications II](#)

3 REFERENCES

1 PREFACE

Tue Sep 3 12:02:55 EDT 2019 

2 WEEK 1

2.1 WEEK 1

Initial One Test# Cloud Unit ## Bigdata Applications and Analytics ### Cloud Computing Unit for BDAA Course

2.1.0.0.1 A). Summary of Course

Overall Summary II

- 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
- 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
- H) Cloud Software: HPCABDS 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

Digital Science Center



2.1.0.0.2 B). Defining Clouds I: 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

2.1.0.0.3 C). Defining Clouds II: Service-oriented architectures: Software services as Message-linked computing capabilities

The differentaaS'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
D) Defining Clouds III: Cloud Market Share How important are they? How much money do they make?

E) Virtualization: Virtualization Technologies, Hypervisors and the different approaches KVM, Xen, Docker and Openstack #### 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 #### 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 #### 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 #### I) Cloud Applications I: Clouds in science where area called cyberinfrastructure; the science usage pattern from NIST Artificial Intelligence from Gartner #### J) Cloud Applications II: Characterize Applications using NIST approach Internet of Things Different types of MapReduce #### K) Parallel Computing Analogies: Parallel Computing in pictures Some useful analogies and principles #### L) Real Parallel Computing: Single Program/Instruction Multiple Data SIMD SPMD Big Data and Simulations Compared What is hard to do?

M) Storage: Cloud data approaches Repositories, File Systems, Data lakes #### N) HPC and Clouds: The Branscomb Pyramid Supercomputers versus clouds Science Computing Environments #### O) Comparison of Data Analytics with Simulation: Structure of different applications for simulations and Big Data Software implications Languages #### 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 #### Q) other Issues II: Security Blockchain #### R) The Future and other Issues III: Fault Tolerance

2.1.0.4 Motivation

Big Data Applications & Analytics: Motivation/Overview; Machine (actually Deep) Learning, Big Data, and the Cloud; Centerpieces of the Current and

Future Economy,

2.1.0.0.5 00) Mechanics of Course, Summary, and overall remarks on course

2.1.0.0.5.1 01A) Technology Hypecycle I

Today clouds and big data have got through the hype cycle (they have emerged) but features like blockchain, serverless and machine learning are on recent hype cycles while areas like deep learning have several entries (as in fact do clouds) Gartner's Hypecycles and especially that for emerging technologies in 2019 The phases of hypecycles Priority Matrix with benefits and adoption time Initial discussion of 2019 Hypecycle for Emerging Technologies

2.1.0.0.5.2 01B) Technology Hypecycle II

Today clouds and big data have got through the hype cycle (they have emerged) but features like blockchain, serverless and machine learning are on recent hype cycles while areas like deep learning have several entries (as in fact do clouds) Gartner's Hypecycles and especially that for emerging technologies in 2019 Details of 2019 Emerging Technology and related (AI, Cloud) Hypecycles

2.1.0.0.5.3 01C) Technology Hypecycle III

Today clouds and big data have got through the hype cycle (they have emerged) but features like blockchain, serverless and machine learning are on recent hype cycles while areas like deep learning have several entries (as in fact do clouds) Gartners Hypecycles and Priority Matrices for emerging technologies in 2018, 2017 and 2016 More details on 2018 will be found in Unit 1A of 2018 Presentation and details of 2015 in Unit 1B (Journey to Digital Business). 1A in 2018 also discusses 2017 Data Center Infrastructure removed as this hype cycle disappeared in later years.

2.1.0.0.6 01D) Technology Hypecycle IV

Today clouds and big data have got through the hype cycle (they have emerged) but features like blockchain, serverless and machine learning are on recent hype cycles while areas like deep learning have several entries (as in fact do clouds) Emerging Technologies hypecycles and Priority matrix at selected times 2008-

2015 Clouds start from 2008 to today They are mixed up with transformational and disruptive changes Unit 1B of 2018 Presentation has more details of this history including Priority matrices

2.1.0.0.7 02)

2.1.0.0.7.1 02A) Clouds/Big Data Applications I

The Data Deluge Big Data; a lot of the best examples have NOT been updated (as I can't find updates) so some slides old but still make the correct points Big Data Deluge has become the Deep Learning Deluge Big Data is an agreed fact; Deep Learning still evolving fast but has stream of successes!

2.1.0.0.7.2 02B) Cloud/Big Data Applications II

Clouds in science where area called cyberinfrastructure; The usage pattern from NIST is removed. See 2018 lectures 2B of the motivation for this discussion

2.1.0.0.7.3 02C) Cloud/Big Data

Usage Trends Google and related Trends Artificial Intelligence from Microsoft, Gartner and Meeker

2.1.0.0.8 03) Jobs In areas like Data Science, Clouds and Computer Science and Computer

Engineering

2.1.0.0.9 04) Industry, Technology, Consumer Trends Basic trends 2018 Lectures 4A 4B have

more details removed as dated but still valid See 2018 Lesson 4C for 3 Technology trends for 2016: Voice as HCI, Cars, Deep Learning

2.1.0.0.10 05) Digital Disruption and Transformation The Past displaced by Digital

Disruption; some more details are in 2018 Presentation Lesson 5

2.1.0.0.11 06)

2.1.0.0.11.1 06A) Computing Model I Industry adopted clouds which are attractive for data

analytics. Clouds are a dominant force in Industry. Examples are given

2.1.0.0.11.2 06B) Computing Model II with 3 subsections is removed; please see 2018

Presentation for this Developments after 2014 mainly from Gartner Cloud Market share Blockchain

2.1.0.0.12 07) Research Model 4th Paradigm; From Theory to Data driven science?

2.1.0.0.13 08) Data Science Pipeline DIKW: Data, Information, Knowledge, Wisdom, Decisions.

More details on Data Science Platforms are in 2018 Lesson 8 presentation

2.1.0.0.14 09) Physics: Looking for Higgs Particle with Large Hadron Collider LHC Physics

as a big data example

2.1.0.0.15 10) Recommender Systems I General remarks and Netflix example

2.1.0.0.16 11) Recommender Systems II Exploring Data Bags and Spaces

2.1.0.0.17 12) Web Search and Information Retrieval Another Big Data Example

2.1.0.0.18 13) Cloud Applications in Research Removed Science Clouds, Internet of Things

See 2018 Presentation (same as 2017 for lesson 13) and Cloud Unit 2019-I) this year

2.1.0.0.19 14) Parallel Computing and MapReduce Software Ecosystems

2.1.0.0.20 15) Online education and data science education Removed.

You can find it in the 2017 version

2.1.0.0.21 16) Conclusions

Motivation Archive Big Data Applications & Analytics: Motivation/Overview; Machine (actually Deep) Learning, Big Data, and the Cloud; Centerpieces of the Current and Future Economy. Backup Lectures from previous years referenced

in 2019 class

2.1.0.0.22 2018BDAAMotivation-1A) Technology Hypecycle I General Remarks including Hype curves

2.1.0.0.23 2018BDAAMotivation-1B) Technology Hypecycle II General Remarks including Hype curves

2.1.0.0.24 2018BDAAMotivation-2B) Cloud/Big Data Applications II

Clouds in science where area called cyberinfrastructure; the usage pattern from NIST Artificial Intelligence from Gartner and Meeker

2018BDAAMotivation-4A) Industry Trends I Lesson 4A Many Technology trends through end of 2014

2018BDAAMotivation-4B) Industry Trends II Lesson 4B 2015 onwards Many Technology Adoption trends

2017BDAAMotivation-4C) Industry Trends III (note 2017 not 2018) Lesson 4C 2015 onwards 3 Technology trends Voice as HCI Cars Deep Learning

2018BDAAMotivation-6B) Computing Model II Lesson 6B with 3 subsections Developments after 2014 mainly from Gartner Cloud Market share Blockchain

2017BDAAMotivation-8) Data Science Pipeline DIKW: Data, Information, Knowledge, Wisdom, Decisions. Data Science Platforms

2017BDAAMotivation-13) Cloud Applications in Research Science Clouds Internet of Things

2017BDAAMotivation-15) Data Science Education Opportunities at Universities

2019Cloud-I) Cloud Applications I: Clouds in science where area called cyberinfrastructure; the science usage pattern from NIST Artificial Intelligence from Gartner

3 REFERENCES

