CS 340

#20: MapReduce and SaaS Introduction

Computer Systems November 1, 2022 · Wade Fagen-Ulmschneider

MapReduce

- Developed as a research project out of Google.
- OSDI'04: "MapReduce: Simplified Data Processing on Large Clusters"
- **Big Idea:** Create a framework for processing data based on functions that can be "automatically parallelized".
 - Allows many nodes to contribute to processing the data without human design/programming.

MapReduce: Simplified Data Processing on Large Clusters						
Jeffrey Dean and Sanjay Ghemawat						
jeff@google.com, sanjay@google.com						
Google	e, Inc.					
Abstract	given day, etc. Most such computations are conceptu-					
Mapfehese is a programming model and an acceleration of the processing and practice state of the processing of the procesing of the procesing of the processing of the processing of	ally assignificated. However, the input data is usually indicated by the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second of the second					
1 Introduction	with an implementation of this interface that achieves high performance on large clusters of commodity PCs. Section 2 describes the basic programming model and					
Over the past five years, the authors and many others at Google have implemental handbed or special-purpose computations that process large announts of raw data, such as crawled documents, web request logs, etc., to compute various kinds of derived data, such as inverted indices, various representations of the graph structure of web documents, summaries of the number of pages crawled per host, the set of most frequent queries in a	gives several examples. Section 3 describes an imple- mentation of the Mapdadoci intrifect and introduced and our chatter-based computing environment. Section 4 de- scribes several refunctions of the programming model that we have found useful. Section 5 has performance that we have found useful. Section 5 has performance and the section of the section of the section of the sec- tion of the section of the section of the section of the data section of coglose the use of Magdendace within Geogle including our experiences in using it as the basis					
To appear in OSDI 2004	1					

https://static.googleusercontent.com/media/research.google.com/en //archive/mapreduce-osdi04.pdf

MapReduce: Map Functions

- Input:
- Output:

Reduce Function:

- Input:
- Output:

Example #1: Word Count

The	quick	brown	fox	jumps	over	the	lazy	dog
[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]

Map:

Reduce:

Example #2: Mutual Friends

Through asking about your friends about their friends, you have identified who are friends of whom (\rightarrow means *"is friends with"*):

$\mathbf{A} \rightarrow$	Β,	С	
$\mathbf{B} \rightarrow$	A,	C,	D
$\mathbf{C} \rightarrow$	A,	Β,	D
$D \rightarrow$	Β,	С	

You want to identify all **mutual friends** to any set of two people. For example: $\{A, B\} \rightarrow C$, D.

Map:

Reduce:

Software as a Service (SaaS)

Q: What makes SaaS different from just software?

- (1):
- (2):
- (3):

Much like all of the other things talked about in the cloud, SaaS is a model of software that is independent of programming language and independent of architectural design.						
Q: What features differentiate SaaS from PaaS?						
(1):						
(2):						
(3):						

Some software tools have characteristics of both SaaS and PaaS – these are general terms to understand the level of abstraction that you, as a developer, are working with or providing.

Large Categories of SaaS Software

- Human Resource Management (HRM)
- Customer Relationship Management (CRM)
- Enterprise Resource Planning (ERP)
- Content Management Systems (CMS)
- Billing Platforms
- Project Management
- Bookkeeping/Accounting
- Ecomerence
- Communication

Advantages of SaaS

1)

2)

3)

- 4)

Legend:

Abstracted by Cloud Provider		Customer Managed Unit of Scale		Customer Managed
---------------------------------	--	-----------------------------------	--	------------------

laaS Infrastructure as a Service	CaaS Containers as a Service	PaaS Platform as a Service	FaaS Functions as a Service	SaaS Software as a Service
Data	Data	Data	Data	Data
Functions	Functions	Functions	Functions	Functions
Applications	Applications	Applications	Applications	Applications
Runtime	Runtime	Runtime	Runtime	Runtime
Containers*	Containers	Containers*	Containers*	Containers*
OS	OS	OS	OS	OS
Virtualization	Virtualization	Virtualization	Virtualization	Virtualization
Hardware	Hardware	Hardware	Hardware	Hardware

Examples of SaaS at Illinois