Introduction to Database Systems CSE 414

Lecture 25: Parallel Databases

Announcements

- Webquiz out later today, due next Friday last WQ! ☺
- HW7 due on Wednesday
- HW8 will be posted soon
 - Will take more hours than other HWs (complex setup, queries run for many hours)
 - Deadline: Last Friday of the qtr; NO LATE DAYS
 - Plan ahead!
 - Help each other out as you learn AWS
- Next four lectures: parallel databases
 - Traditional, MapReduce+PigLatin

Parallel Computation Today

Two Major Forces Pushing towards Parallel Computing:

- Change in Moore's law
- Cloud computing

Parallel Computation Today

- Change in Moore's law* (exponential growth in transistors per chip density) no longer results in increased clock speeds
 - Increased hw performance available only through parallelism
 - Think multicore: 4 cores today, 64 possible soon, but who can figure out how to program it?

Moore's law says that the number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years [Intel co-founder Gordon E. Moore described the trend in his 1965 paper and predicted that it will last for at least 10 years]

Parallel Computation Today

- 2. Cloud computing commoditizes access to large clusters
 - Ten years ago, only Google could afford 1000 servers;
 - Today you can rent this from Amazon Web Services (AWS)

Science is Facing a Data Deluge!

- Astronomy: Large Synoptic Survey Telescope LSST: 30TB/night (high-resolution, high-frequency sky surveys)
- Physics: Large Hadron Collider 25PB/year
- Biology: lab automation, high-throughput sequencing
- Oceanography: high-resolution models, cheap sensors, satellites
- Medicine: ubiquitous digital records, MRI, ultrasound

Science is Facing a Data Deluge!











CSE 414 - Spring 2013

Industry is Facing a Data Deluge!

Clickstreams, search logs, network logs, social networking data, RFID data, etc.

- Facebook:
 - 15PB of data in 2010
 - 60TB of new data every day
- Google:
 - In May 2010 processed 946PB of data using MapReduce
- Twitter, Google, Microsoft, Amazon, Walmart, etc.

Industry is Facing a Data Deluge!



Computers Tree X	Application on TO	RIMOD 3	System on Tr	ORNADO	Security on TORNADI	Setvert.De	app Let		
8 🔛 TORNADO (luca)	H () H (A South Hearth) O H								
Separate Application	Tree	Contra Contra	Tere	(Count)		Canana .			
System	And Servers	13.05 2000	0.33-45	1000	Secrety	Lowel confi	topseo dischart	LOSS ACC.	Lour
SQL Servers Support Diterret (www, frg, nat) Otto: (pel Deco) State-ret (server, frg, nat) Deco: (pel Deco) State-ret (server, frg, Application	Audit Surveys	12.05.2005	16/33-47	882	Seculty	Loopel.com	topppoint-had	1085600	Loopo
	Audit Surveys	12.05.2005	16.29-65	882	Security	Lowellow	toppapoint-had	1085400	Lown
	Audit Success	12 06 2008	14,73,79	887	(interfect	Longe & canif	Company (address)	1088400	- Lane
	Audit Transme	12 00 2000	14,78.76	887	Second .	Lossed and	TORNOO(Rhibad	TORMOO	Langer
	Audit Tucceus	17.05.2000	17.05.00	100	Jacobia Contraction	Logon & could	TORNAL CONTRACTOR CONTRACT	TORNADO.	Scell.
	Audit Income	17 05 2000	11.06.00	8.90	Second .	Loose & coold	TORNAL CONTRACTOR	TORNADO.	Sec.
	Audit Success	12 05 2000	17.05.00	101	Security	Looved coold	PERSONAL	TORNADO.	10000
Security	Aude Surrent	12 05 2000	12:05:00	100	Seculty	Account Loopo	TORNAL CONTRACTOR	TORNAGO,	10000
States	Aude German	12.05.2000	11-04-59	540	Seculty	Loose A cool!	NT & THEFT CARGO	TOPRADO.	Same
Sandar and Sandar Sa Sandar Sandar Sandar Sandar Sandar Sandar Sandar Sandar	Aude Garness	12.05.2000	11-04-55	SM.	Seculty	Looved cool!	NEW DRIVETYLOCK	0084900	Sheric
	Aude Garrans	12.05.2000	12:02:55	530	Seculty	Looved cool!	AT M THORSTON OF M	0000000	Game
	Audt Surran	12.05.2000	12-04-55	SM.	Security	Looped cool?	AT & DESCRIPTION	000000	Static
	Aude Garrans	12 05 2000	12-08-55	530	Security	Looped cool!	AT & THOST VIET DAT	000000	Same
	Auret German	12.05.2000	12:08:55	520	Security	Looved cool!	10/SPM	1089400	2000
	Audi Daman	13.05.3000	12.00.00	513	STO DOT	Cogory Cogor	provident (1000000	and a
	Audi Duran	13.05.2000	12.02.24	830	Ferender	Lowed and	TORNAL CONTRACTOR	1000000	Then I
	Audi Danes	12.00.2000	12.00.01	1000	Feasier	Langelogon	TORNEOUPRONEN	TORNADO	They b
	Audit Durante	12.00.2000	11.475.78	874	Tecality	Logor J. Cogoli	TORNADOPROVAD	1000000	Dieda
	Audit Ducant	17.05.2000	11.03.30	110	Telefit	Loose Logali	TORNADCION IN AN	2008000	Darra -
	Con Man	12.00.2000	11.00.00		January .	Contraction of the	Contractory so that	1011000	
	Legged an user Uber M Domai Logon Logon User whose oree Target Target	ane: Mid ii 100 ID: (2x) GLEx - entials vere user User Name: Add Downin: 100	hadi INMDO 0,0:200140 é ninistrator INMDO						



Big Data

- Companies, organizations, scientists have data that is too big, too fast, and too complex to be managed without changing tools and processes
- Relational algebra and SQL are easy to parallelize and parallel DBMSs have already been studied in the 80's!

Data Analytics Companies

As a result, we are seeing an explosion of and a huge success of db analytics companies

- Greenplum founded in 2003 acquired by EMC in 2010; A parallel shared-nothing DBMS (this lecture)
- Vertica founded in 2005 and acquired by HP in 2011; A parallel, column-store shared-nothing DBMS (see 444 for discussion of column-stores)
- DATAllegro founded in 2003 acquired by Microsoft in 2008; A parallel, shared-nothing DBMS
- Aster Data Systems founded in 2005 acquired by Teradata in 2011; A parallel, shared-nothing, MapReduce-based data processing system (next lecture). SQL on top of MapReduce
- Netezza founded in 2000 and acquired by IBM in 2010. A parallel, shared-nothing DBMS.

Great time to be in the data management, data mining/statistics, or machine learning!

Two Approaches to Parallel Data Processing

- Parallel databases, developed starting with the 80s (this lecture)
 - OLTP (Online Transaction Processing)
 - OLAP (Online Analytic Processing, or Decision Support)
- MapReduce, first developed by Google, published in 2004 (next lecture)

– Only for Decision Support Queries

Today we see convergence of the two approaches (Greenplum, Dremmel)

Parallel DBMSs

Goal

Improve performance by executing multiple operations in parallel

- Key benefit
 - Cheaper to scale than relying on a single increasingly more powerful processor
- Key challenge
 - Ensure overhead and contention do not kill performance

Performance Metrics for Parallel DBMSs

- P = the number of nodes (processors, computers)
- Speedup:
 - More nodes, same data \rightarrow higher speed
- Scaleup:
 - More nodes, more data \rightarrow same speed
- OLTP: "Speed" = transactions per second (TPS)
- **Decision Support**: "Speed" = query time

Linear v.s. Non-linear Speedup



CSE 414 - Spring 2013

Linear v.s. Non-linear Scaleup



Challenges to Linear Speedup and Scaleup

• Startup cost

- Cost of starting an operation on many nodes

• Interference

Contention for resources between nodes

• Skew

Slowest node becomes the bottleneck

Architectures for Parallel Databases

- Shared memory
- Shared disk
- Shared nothing

Shared Memory



Shared Memory

- Nodes share both RAM and disk
- Dozens to hundreds of processors

Example: SQL Server runs on a single machine and can leverage many threads to get a query to run faster (see query plans)

- Easy to use and program
- But very expensive to scale: last remaining cash cows in the hardware industry

Shared Disk



Shared Disk

- All nodes access the same disks
- Found in the largest "single-box" (noncluster) multiprocessors

Oracle dominates this class of systems.

Characteristics:

 Also hard to scale past a certain point: existing deployments typically have fewer than 10 machines

Shared Nothing



Shared Nothing

- Cluster of machines on high-speed network
- Called "clusters" or "blade servers"
- Each machine has its own memory and disk: lowest contention.

NOTE: Because all machines today have many cores and many disks, then shared-nothing systems typically run many "nodes" on a single physical machine.

Characteristics:

- Today, this is the most scalable architecture.
- Most difficult to administer and tune.

We discuss only Shared Nothing in class

A Professional Picture...

Figure 1 - Types of database architecture



From: Greenplum Database Whitepaper

SAN = "Storage Area Network"

DB

Disk

In Class

- You have a parallel machine. Now what?
- How do you speed up your DBMS?

Approaches to Parallel Query Evaluation

- Inter-query parallelism
 - Transaction per node
 - OLTP



Approaches to Parallel Query Evaluation

- Inter-query parallelism
 - Transaction per node
 - OLTP
- Inter-operator parallelism
 - Operator per node
 - Both OLTP and Decision Support





Approaches to Parallel Query Evaluation

- Inter-query parallelism
 - Transaction per node
 - OLTP
- Inter-operator parallelism
 - Operator per node
 - Both OLTP and Decision Support
- Intra-operator parallelism
 - Operator on multiple nodes
 - Decision Support



Approaches to Parallel Query Evaluation

- Inter-query parallelism
 - Transaction per node
 - OLTP
- Inter-operator parallelism
 - Operator per node
 - Both OLTP and Decision Support
- Intra-operator parallelism
 - Operator on multiple nodes
 - Decision Support



We study only intra-operator parallelism: most scalable