MOBILE

BIG DATA
REPUBLIC | Transform Your Business With Data

ABOUT US   REGISTER   LOGIN

HOME   BLOGGERS   MESSAGES   POLLS   WEBINARS   RESOURCES   VIDEOS | STRATEGY   ANALYTICS   TECHNOLOGY   FB   TW   LI   G+   RSS

FINANCIAL SERVICES   HEALTHCARE   GOVERNMENT   EDUCATION   RETAIL

BIG DATA TECHNOLOGY

What Would Seymour Cray Think of Big Data?

Robert Plant, Associate Professor, School of Business Administration, University of Miami
12/4/2013
Comment
13 comments

Login

👍 👎
50%   50%

For many computer science students, especially those who were at University in the 1980s, the pinnacle of their programming experience was developing some code that would actually run on a supercomputer, but preferably "the" supercomputer, a Cray.

A pre-cursor to in-memory
Often jokingly referred to as the world's most expensive love seat, the X-MP 48 was capable of running at around 400Mflops, had a 16MB main memory, used a variant of UNIX, and in 1984, cost around $15M. In big data terms, the systems were capable of high velocity possessing, courtesy of their shared memory parallel vector processors. Volume was not a problem (in 1980s scale) as the systems had optional solid state drives, a pre-cursor to the in-memory data base storage solutions appearing in today's systems. These were available with 256, 512, or 1024 MB capacities each with a transfer rate of up to 1,000Mbp per channel. Where they would have suffered was on variety; vector processors were originally designed for high performance on numerical simulations involving parallel processing of highly similar tasks.

The legendary founder of Cray computing, Seymour Cray, was a true visionary and computing pioneer in all things speed, someone who would have relished the task at building a big data "hot rod." The decedents of his X-MP machines are still focused in many ways on helping solve certain tricky problem sets and tackling them in interesting ways. One problem that data scientists are constantly facing is the complexity of the interactions between data points.

By Jens Ayton (Ahruman) (Own work) [CC-BY-SA-2.5], via Wikimedia Commons

There are several ways to consider attacking this. As data scientists are often hybrid computer-data scientists, one natural thought process is to build a graph or tree structure to simplify the problem. Another is to look towards a Map Reduce type solution. This problem is not new, as research has been underway in this space for several decades. Pioneers such as Professor's Harrision and Reeve, of Imperial College, London, created a parallel graph reduction machine, known as ALICE, using Transputer in the early 1980s; and Professors Randell (Newcatle), Darlington (Imperial), Henderson (PRG, Oxford), and Turner (Kent) amongst others, created the basis of the functional languages upon which systems such as ALICE could run effectively, if not very efficiently, given the architectural constraints of the period.

The birth of rule-based pattern inference
The ALICE/functional language environment did, however, allow researchers to create rule-based pattern directed inference systems and the basis of methodologies for text mining and computational linguistics within a graph reduction architectural style.

Fast forward on the problem 20 years and Cray has built an interesting, purpose built graph analysis device known as Urika. Although the appliance is somewhat less stylish than the X-MP 48, with more than a passing resemblance to a drinks vending machine than an iconic piece of art, the system does in fact provide an effective graph analysis solution path.

Rather than following the common implementation strategy, where by graphs are partitioned across (low cost) commodity clusters, the device can scale to 512Tb of global shared data; utilizes a massively multithreaded specialist graph processor, of which there are up to 8,192, enabling it to run over 1M threads; and has data transfer rates of 350TB/hr.

This approach enables data scientists to overcome three major problems associated with partitioned clusters. Firstly, containing the whole graph in an appliance avoids the difficulties in effectively partitioning graphs across clusters; transferring data pertaining to nodes and edges across the cluster is a time-consuming and energy-inefficient process. Second, graph analyses requires significant compare

---

MORE BLOGS FROM ROBERT PLANT

A Review: Big Data in 2013
39 comments
Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

The 360 Degree View of a Customer
56 comments
Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

Lessons for Big Data From President Obama's Healthcare Implementation
24 comments
Three big data lessons for CEOs from HealthCare.gov's bumpy start.

Is Big Data Ruining Sports?
29 comments
From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

More from Robert Plant

FLASH POLL

All Polls

DIGITAL AUDIO

Big Data Technology Radio
Latest Archived Broadcast
Your Big Data Initiative Starts With a Great Team
The Big Data Conference and Nemertes Research have teamed up to provide expert advice on building a big data team and finding the right leader. Join us on Tuesday, Aug. 20.
Listen
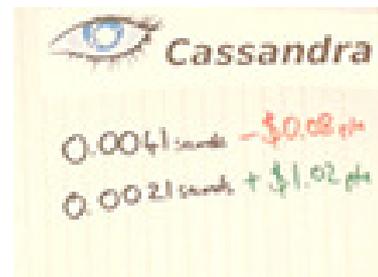
Full schedule   Archived shows   Download to calendar

BDR IN YOUR INBOX

Enter email   SUBMIT

FEATURED VIDEO

Big Data Explained: What Is Cassandra?   16

and evaluate processes, these can be handled through memory swapping routes but incur a high overhead, using multi-threading is more efficient use of processor-memory usage. Thirdly, graphs are highly dynamic and thus analysis requires either a very high I/O memory access speed or a more effective processor driven strategy.

Unfortunately, Seymour Cray (1925-1996) is no longer with us in body, but his spirit lives on to inspire; not only at the company he founded but in all data-computer scientists. So next time you have a big data problem, perhaps it would be fruitful to just ask yourself "what would Seymour Cray do?" The answer may surprise you.

Related posts:

- Could Watson Be Big Data's Killer App?
- Cognitive Computing & Big Data Analytics
- Weather: Big Data & Big Business
- What Would You Do With a $120M Supercomputer?

Email This   Print   Comment

COMMENTS

Newest First | Oldest First | Threaded View

PAGE 1 / 2   >   >>

Pradeepta Mishra, User Rank: Exabyte Executive
12/30/2013 | 3:37:58 AM

Login

Re: Was anyone thinking of big data then?
@Lara, I would say this is evolving as our technology grows. Cray's approach was more towards handling massive numerical data, so it is computation focused; however modern data big data also possesses variety which was a result of evolution of many platforms and applications.

50%   50%   Reply | Post Message | Edit/Delete | Messages List | Start a Board

James M. Connolly, User Rank: Blogger
12/13/2013 | 3:06:26 PM

Login

Re: Was anyone thinking of big data then?
I agree that Cray would have loved the science of big data but I'm not sure he would have been horrified by the commercialization of the technology. I dealt with the company (not the man) occasionally back in the 1980s, and they certainly did want to see supercomputing applied in the commercial space. In reality, it was used commercially in some areas such as drug discovery, which is one area where big data made its mark.

Selling toilet paper and Hello Kitty apparel? Seymour Cray would have looked at it as more ways to use computers.

BTW, I visited a supercomputer center at a university some years ago. It was about the time that researchers were realizing they could get the impact of a supercomputer by networking thousands of Intel machines in racks. The staff was in the process of dismantling their last Cray when I was there. It was kind of sad.

50%   50%   Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
12/10/2013 | 5:07:15 AM

Login

Re: What would Cray make of quality?
@dcawrey - probably a future scenario planning element there too. Yesterday's supercomputers are today's desktops.

50%   50%   Reply | Post Message | Edit/Delete | Messages List | Start a Board

dcawrey, User Rank: Exabyte Executive
12/7/2013 | 9:52:11 AM

Login

Re: What would Cray make of quality?
Interesting article. I don't think that we hear enough about supercomputing. Much of that probably has to do with a lack of access; this seems to me like more of an academic pursuit more than anything else.

But companies, especially larger ones, might want to think about the impact that supercomputers could have on big data in particular.

50%   50%   Reply | Post Message | Edit/Delete | Messages List | Start a Board

Mike Lata, User Rank: Blogger
12/6/2013 | 4:31:47 PM

Login

Re: What would Cray make of quality?
Yes, there is really no limit to how fast technology grows and the need for it to grow. At one point we thought hardware advancements in memory or CPU core speeds would be redundant outside of maybe gaming, but with the advent of big data there really is a need for hardware to continue to grow I to the current supercomputer range. The data we study, and analyze and manipulate is really staggering. There is definitely a need for quantum computing to take off: http://www.howstuffworks.com/quantum-computer.htm

50%   50%   Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
12/6/2013 | 10:38:39 AM

Login

Re: Was anyone thinking of big data then?
Ahhh @legalcio, at the end of the day doesn't all technological advancement end up felping to flog Hello Kitty gear?

I guess any scenario which used to feature a manager moving around with a clipboard was always on the road to big data.

50%   50%   Reply | Post Message | Edit/Delete | Messages List | Start a Board

Get to know cassandra, she's built for speed.
Watch This Video

More Video Blogs

FOLLOW US ON TWITTER

Big Data Republic@BigDataRepublic   15 Apr
AllJoyn: A Common Language For Internet Of Things - Open-source lingua franca aims to let IoT devices and services... ow.ly/2FHbY7
Retweeted by Yakamoto
Expand

Hajime Ozaki@hajozaki   21h
BDR Top 20 #BigData David Smith David Smith List of top tweeters on Big Data from @BigDataRepublic wp.me/p4Aavi-2
Show Summary

Marcus Borba@marcusborba   28 Apr
#BigData #Analytics #BusinessIntelligence #FS @informationweek @AllAnalytics @BigDataRepublic @BigDataNetwork @BigDataClub

LIKE US ON FACEBOOK

Big Data Republic
Like

4,168 people like Big Data Republic.

Facebook social plugin

ACCOLADES

Media Pioneer Award Winner

legalcio, User Rank: Exabyte Executive
12/5/2013 | 5:01:17 PM

Login

50% 50%

Re: Was anyone thinking of big data then?
Good point Lara. I think we have been aiming at the 360 degree view, but we haven't known it.
The development of market research, focus groups, samples on the grocery store floor, etc.,
have all been pointing toward some way to gather information. Databases helped but there
wasn't any real intelligence to them. In that respect technology pushed toward Big Data.

I think Cray would have been fascinated with the science but possibly mortified at using it to sell
toilet paper and Hello Kitty apparel.

Reply | Post Message | Edit/Delete | Messages List | Start a Board

---

Saul Sherry, User Rank: Blogger
12/5/2013 | 10:10:05 AM

Login

50% 50%

Re: Was anyone thinking of big data then?
Lara I dare say you are right... it would have taken a real visionary effort back then!

Reply | Post Message | Edit/Delete | Messages List | Start a Board

---

Lara, User Rank: Bit Player
12/5/2013 | 6:17:15 AM

Login

50% 50%

Re: Was anyone thinking of big data then?
On reflection - one or two visionaries perhaps - but I would say for the most part our solutions
have grown as our technologies have (or as our problems have... more data equals more
solutions, but more potential issues as well).

Reply | Post Message | Edit/Delete | Messages List | Start a Board

---

Saul Sherry, User Rank: Blogger
12/5/2013 | 5:02:55 AM

Login

50% 50%

Re: What would Cray make of quality?
Hmmm... probably at Cray's level yes... I think looking into discovery in the academic world there
would be some really clean and precise data.

But I take your point... maybe data has always been poorly managed when it comes to
businesses.

Reply | Post Message | Edit/Delete | Messages List | Start a Board

PAGE 1 / 2 > >>

BIG DATA REPUBLIC

ABOUT US | CONTACT US | HELP | REGISTER
TWITTER | FACEBOOK | LINKED IN | GOOGLE+ | RSS

in partnership with *InformationWeek*   UBM Tech   powered by UBM DEUSM

bigdatarepublic : /bigdatarepublic/section/2635 : /bigdatarepublic/section/2635/270177

UBM Tech

UBM TECH

OUR MARKETS: **Business Technology** | **Electronics** | **Game & App Development**

Working With Us: Advertising Contacts | Event Calendar | Tech Marketing Solutions | Corporate Site | Contact Us / Feedback