Stefan Radtke's Blog

12 May 2015

Comparing Hadoop performance on DAS and Isilon and why disk locality is irrelevant

In a previous blog [3] I discussed how Isilon enables you to create a Data Lake that can serve multiple Hadoop compute clusters with different Hadoop versions and distributions simultaneously. I stated that many workloads run faster on Isilon than on traditional Hadoop clusters that use DAS storage. This statement has recently been confirmed by IDC [2] who ran various Hadoop benchmarks against Isilon and a native Hadoop cluster on DAS. Although I will show their results right at the beginning, the main purpose is to discuss *why* Isilon is capable of delivering such good results and what the differences are in regard to data distribution and balancing within the clusters.

The test environment

- Cloudera Hadoop Distribution CDH5.
- The Hadoop DAS cluster contained 7 nodes with one master and six worker nodes with eight 10k RPM 300 GB disks.
- The Isilon Cluster was build out of four x410 nodes, each with 57 TB disks and 3,2 TB SSDs and 10 GBE connections.
- For more details see the IDC validation report [2].

NFS access

First of all, IDC tested NFS read and write access and with no surprise, Isilon provides MUCH more throughput even with just 4 nodes.
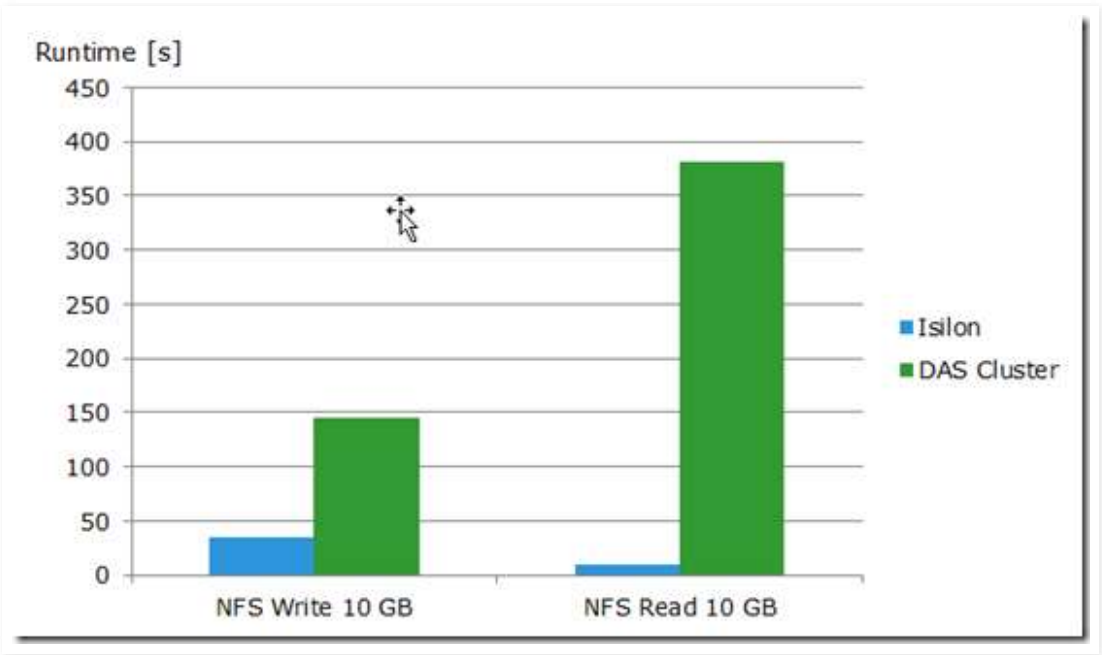

Figure 1: Runtimes for NFS write and read copy jobs while copying a 10GB file (blocksize is not mentioned but I would assume 1MB or larger)

NFS write turned out to be 4.2 times faster. This is quite important if you want to ingest data via NFS. Read performance is almost 37 times faster.

Hadoop workloads

Three Hadoop workload types have been run and compared by using standard Hadoop benchmarks:

1. Sequential write using TeraGen
2. Mixed read/write using TeraSort
3. Sequential read using TeraValidate
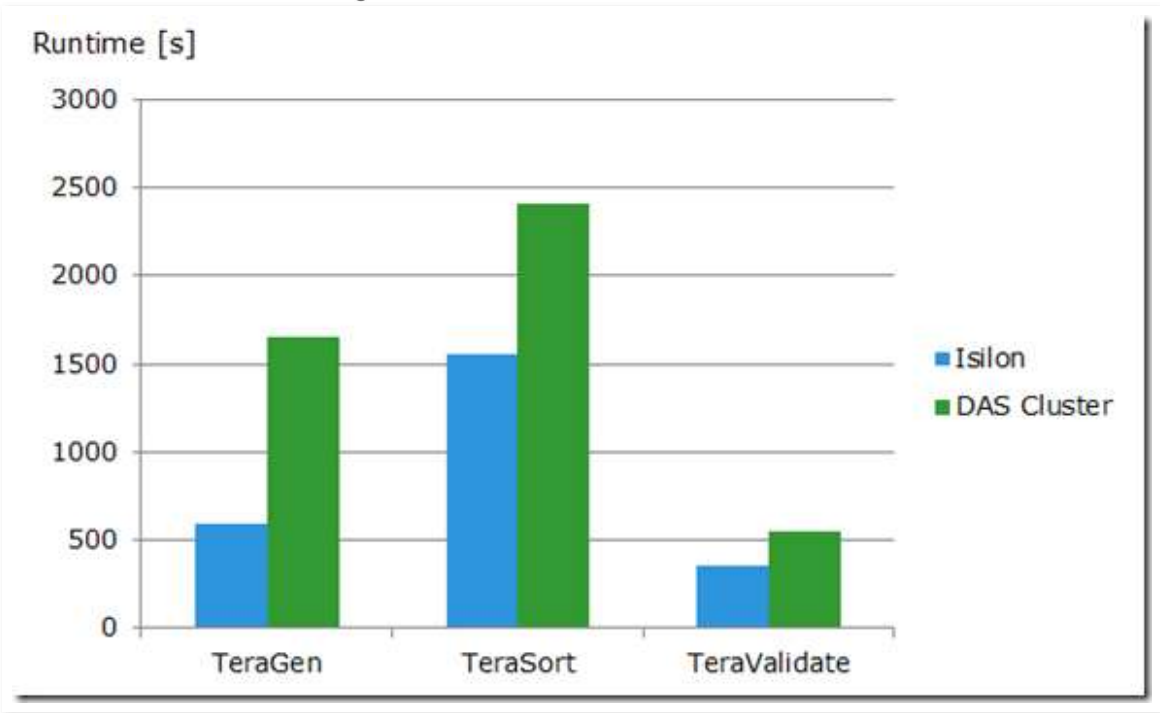
The results are illustrated in figure 2.

Figure 2: Runtimes for three different workloads using TeraGen, TeraSort and TeraValidate

It turns out that the runtimes for the write performance were about 2.6 shorter on Isilon and 1.5 times shorter for the other two workload types. The related throughputs are outlined in the following table.

Job	Compute + Isilon	Hadoop DAS Cluster
TeraGen	1681 MB/s	605 MB/s
TeraSort	642 MB/s	416 MB/s
TeraValidate	2832 MB/s	1828 MB/s

Table 1: Throughput for all three workloads on Isilon vs. DAS cluster with similar compute node configuration (results rounded).

The results speak for themselves but let's look at what techniques OneFS provides to achieve this level of performance advantages over a DAS cluster.

The anatomy of file reads on Isilon

Although IOs on a DAS cluster are distributed across all nodes, an individual 64 MB block is served by a single node in the cluster. This is different on Isilon where the load distribution works more granular. The steps for a read on Isilon can be described as follows.

1. The compute node sends HDFS metadata request to the Name Node service which runs on all Isilon nodes (no SPoF)
2. The Name Node service will return the IP addresses and block numbers of any 3 Isilon nodes in the same rack as the compute node. This provides effective rack locality.
3. The Compute node sends HDFS 64 MB block read request to the Data Node service on the first Isilon node returned.
4. The contacted Isilon node will retrieve through the internal Infiniband network all 128 KB Isilon blocks that comprise the 64 MB HDFS block. The blocks will be read from disks if they are not already stored in the L2 cache. As said above, this is fundamentally different than on a DAS cluster where the whole 64 MB block is read from one node only. That means the IO on Isilon is served by much more disks and CPUs than on the DAS cluster.
5. The contacted Isilon node will return the entire HDFS block to the calling compute node.

The anatomy of file writes on Isilon

When a client requests that a file be written to the cluster, the node to which the client is connected is the node that receives and processes the file.

1. That node creates a write plan for the file including calculating FEC (this is much more space efficient compared to a DAS cluster where we typically do 3 copies of each block for data protection)
2. Data blocks assigned to the node are written to the NVRAM of that node. The NVRAM cards are special for Isilon and not available on DAS clusters.
3. Data blocks assigned to other nodes travel through the Infiniband network to their L2 cache, and then to their NVRAM.
4. Once all nodes have all the data and FEC blocks in NVRAM a commit is returned to the client. That means, we do not need to wait until the data is written to disks as all IOs are securely buffered by NVRAM on Isilon.
5. Data block(s) assigned to this node stay cached in L2 for future reads of that file.
6. Data is then written onto the spindles.

The myth of disk locality importance for Hadoop

We sometimes hear objections from admins who claim that disk locality is critical for Hadoop. But remember that traditional Hadoop was designed for slow star networks which typically operated at 1 Gb/s. The only way to effectively deal with slow networks was to strive to keep all IO local to the server (disk locality).
There are several facts that make disk locality irrelevant:

I. Fast networks are standard today.

- Today, a single non-blocking 10 Gbps switch port (up to 2500 MB/sec full duplex) can provide more bandwidth than a typical disk subsystem with 12 disks (360 – 1200 MB/sec).
- We are no longer constrained to maintain data locality in order to provide adequate I/O bandwidth.
- Isilon provides rack-locality, not disk-locality. This reduces the Ethernet traffic between racks.

By looking at the following illustration of the IO path, it is obvious that the bottleneck in the path is the disks, not the network (as long as it is a 10 GBE network).
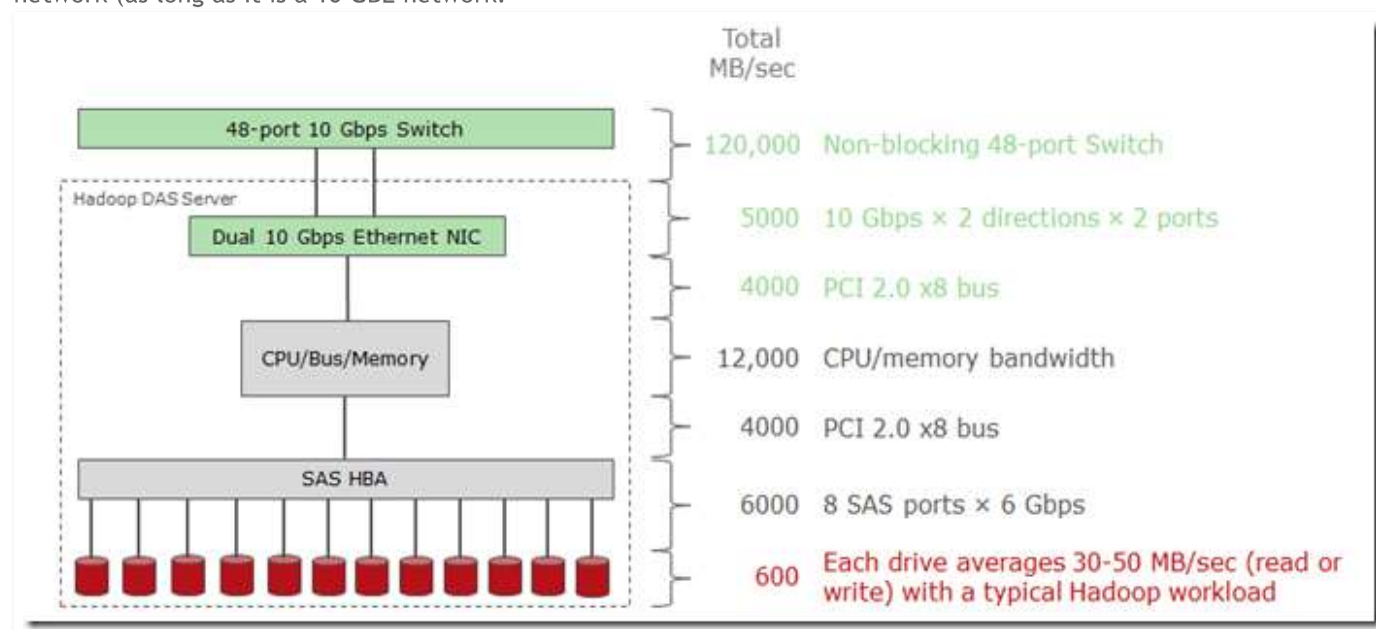

Figure 3: IO path in a DAS architecture. Considering a non-blocking 10 Gbps network it is obvious that the network is not the bottleneck. Even if we would double the number of disks in the system, the disks remain the bottleneck. As a result, disk locality is irrelevant for most workloads.

II. Disk locality is lost under several common situations:

- All nodes of a DAS cluster with a replica of the block are running the maximum number of tasks. This is very common for busy clusters!
- Input files are compressed with a non-splitable codec such as gzip.
- "Analysis of Hadoop jobs from Facebook [1] underscores the difficulty in attaining disk-locality: overall, only 34% of tasks run on the same node that has the input data."
- Disk locality provides very low latency IO, however this latency has very little effect for batch operations such as MapReduce.

III. Data replication for performance

- For very busy traditional clusters, a high replication count may be needed for hot files that are used often by many concurrent tasks. This is required for data locality and high concurrent reads.
- On Isilon, a high replication count is not required because:
  a) Data locality is not required and
  b) Reads are split evenly over many Isilon nodes with a globally coherent cache, providing very high concurrent read performance

Other Isilon performance relevant technologies

As mentioned earlier, OneFS is very mature and has been designed for more than a decade for high throughput and low latency for multi-protocol access. You can google a number of articles and papers describing relevant features. I'll just give some keywords here:

- All writes are buffered by redundant NVRAM. This makes writes extremely fast
- OneFS provide a L1 cache, a globally coherent L2 cache and L3 caches on SSDs for accelerated reads
- Access patterns can be configured per cluster, pool or even on directory level to optimize and balance pre-fetching. Patterns are random, concurrent or streaming.
- Meta data acceleration is provided by the L3 cache or can be configured alternatively. OneFS will store all filesystem meta data on SSDs

Summary

Isilon is a scale-out NAS system with a distributed filesystem that has been built for massive throughput requirements and workloads like Hadoop. HDFS is implemented as a protocol and Name Node as well as Data Node services are delivered in a highly available manner by all Isilon nodes. IDCs performance validation [2] showed up to 2.5 times higher performance compared to a DAS cluster. Due to modern networking technologies, the often referenced disks locality is irrelevant for Hadoop on Isilon. Besides the better performance, there are many other advantages that Isilon provides, such as the much higher capacity efficiency and many enterprise storage features. Furthermore, storage and compute nodes can be scaled independently and you can access the same data with different Hadoop versions and distributions simultaneously.

References

[1] Disk-Locality in Datacenter Computing Considered Irrelevant, Ganesh Ananthanarayanan, University of California, Berkeley
[2]   EMC Isilon Scale-out Data Lake Foundation - Essential Capabilities for Building Big Data Infrastructure, IDC White Paper, October 2014
[3] How to access data with different Hadoop versions and distributions simultaneously, Stefan Radtke, Blog post 2015
[4] EMC Isilon OneFS - A Technical Overview; White Paper, November 2013
The White Papers mentioned here are all available for download at https://support.emc.com

Acknowledgement

I have stolen several aspects and topics of the discussion from the excellent training material which my colleagues Claudio Fahey put together. Thanks to Matthias Radtke for improving my non-native language writing.

Posted by Stefan Radtke at 1:48 AM　　　　　　G+

Labels: EMC, HDFS, Isilon, NAS, NFS, Protocols

57 comments:

NDM Solution June 29, 2015 at 4:57 AM

Wow that's a wonderfull blog having all details & helpful. Hadoop cluster NJ

Reply

Sai Santosh October 26, 2015 at 12:40 AM

There are many institutes for hadoop allover, however many people and from countries like Russia are preferring hadoop online training in India.

Reply

savitha singh April 23, 2016 at 12:34 AM

Hi Admin, I went through your article and it's totally awesome. You can consider including RSS feed for easy content sharing, So that you can drive huge traffic to your blog. Hadoop Training in Chennai | Big Data Training in Chennai

Reply

Ahab July 7, 2016 at 9:37 AM

Good read. Stefan, have you ever tried implementing running HDFS on top of an existing NFS? Any performance metrics for that? I'm aware of concepts like NFS gateway, but was just curious if you've ever tried it.

Reply

> Replies
>
> **Stefan Radtke**　　　February 8, 2017 at 9:39 AM
>
> Hi Ahab, no, never tried and for the technology described here, it is never required because we provide native access via NFS *and* HDFS at the same time, max speed. No need for a gateway layer or the like.

Reply

Unknown August 4, 2016 at 6:16 AM

This article describes the Hadoop Software, All the modules in Hadoop are designed with a fundamental assumption that hardware failures are common and should be automatically handled by the framework. This post gives great idea on Hadoop Certification for beginners. Also find best Hadoop Online Training in your locality at StaygreenAcademy.com

Reply

mythily mythu November 1, 2016 at 4:59 AM

Good read.
weblogic training in chennai

Reply

deepak singh November 18, 2016 at 3:23 AM

Thanks for providing this informative information you may also refer.
http://www.s4techno.com/blog/2016/08/13/installing-a-storm-cluster/

Reply

yasar March 16, 2017 at 12:30 AM

very nice and informative blog
arduino projects in chennai
final year dotnet projects chennai
final year ns2 projects chennai
final year mba projects chennai

Reply

RS Trainings March 31, 2017 at 4:12 AM

Hadoop Training in Hyderabad

Reply

A1 trainings April 3, 2017 at 11:09 PM

Thx for sharing information

Hadoop Training in Hyderabad

Reply

jazz April 18, 2017 at 2:52 AM

Finding the time and actual effort to create a superb article like this is great thing. I'll learn many new stuff right here! Good luck for the next post buddy..
PHP training in chennai

Reply

Emergers Technologies April 26, 2017 at 5:25 AM

Just found your post by searching on the Google, I am Impressed and Learned Lot of new thing from your post. I am new to blogging and always try to learn new skill as I believe that blogging is the full time job for learning new things day by day.
"Emergers Technologies"

Reply

sai venkat May 8, 2017 at 10:08 PM

Thank you so much for sharing this worth able content with us. The concept taken here will be useful for my future programs and i will surely implement them in my study. Keep blogging article like this.

Hadoop Training In Chennai

Reply

sai venkat May 31, 2017 at 10:31 PM

Good Post! Thank you so much for sharing this pretty post, it was so good to read and useful to improve my knowledge as updated one, keep blogging…
Hadoop Online Training

Reply

farman wxi June 16, 2017 at 3:44 AM

Automotive tools suppliers in Singapore, automotive tools traders in Singapore, Hand Tools, Shop Equipment, Auto Lifts, Lift Parts and Lift Accessories Complete coverage with parts breakdowns for easy ordering. VIEW MORE - Automotive tools

Reply

Dubai Raju June 19, 2017 at 2:32 AM

Great Blog Thanks for sharing ...... such a helpful information keep sharing these type of blogs.
Hadoop Training in Hyderabad

Reply

sathya June 21, 2017 at 6:48 AM

actually we want to know about the cloud computing. Because it is much necessary subject which will helps us to groove in IT industry. So we want to know more about this. So please keep update like this.

Hadoop Training in Chennai

Base SAS Training in Chennai

Reply

sathya June 21, 2017 at 6:49 AM

actually we want to know about the cloud computing. Because it is much necessary subject which will helps us to groove in IT industry. So we want to know more about this. So please keep update like this.

Hadoop Training in Chennai

Base SAS Training in Chennai

Reply

sunitha vishnu June 30, 2017 at 10:56 PM

It is amazing and wonderful to visit your site.Thanks for sharing this information,this is useful to me...
Android Training in Chennai
Ios Training in Chennai

Reply

soumya Reddy July 26, 2017 at 5:50 AM

It's Amazing content thanks for providing for us..........

Reply

Karthi Keyan August 14, 2017 at 3:34 AM

Interesting blog post.This blog shows that you have a great future as a content writer.waiting for more updates...
Digital Marketing Company in India

Reply

john stany August 21, 2017 at 10:27 PM

You have provided an nice article, Thank you very much for this one. And i hope this will be useful for many people.. and i am waiting

for your next post keep on updating these kinds of knowledgeable things...
iOS App Development Company
Android App Development Company
Best Mobile app Development company
Android App Development Company in chennai
iOS App Development Company in chennai

Reply

rose August 29, 2017 at 8:59 PM

That is very interesting; you are a very skilled blogger. I have shared your website in my social networks! A very nice guide. I will definitely follow these tips. Thank you for sharing such detailed article.

Hadoop Training in Marathahalli|
Hadoop Training in Bangalore|
Data science training in Marathahalli|
Data science training in Bangalore|

Reply

sathya September 1, 2017 at 7:09 AM

Thank you for sharing the information here. Its much informative and really i got some valid information. You had posted the amazing article.

Dataware Housing Training in Chennai | Hadoop Training in Chennai

Reply

santhosh k September 7, 2017 at 10:38 PM

Thanks for one marvelous posting! I enjoyed reading it; you are a great author. I will make sure to bookmark your blog and may come back someday. I want to encourage that you continue your great posts
Hadoop Training in Bangalore

Reply

Unknown September 15, 2017 at 3:19 AM

Thanks for sharing well blog Hadoop training in training
AWS training in bangalore
Tableau training in bangalore
PHP training in bangalore
Android training in bangalore
Digital marketing training in bangalore

Reply

venkatesan September 15, 2017 at 4:51 AM

Nice blog Devops training in bangalore
Machine leraning training in Bangalore
Iot Training in Bangalore

Reply

sathya September 26, 2017 at 7:39 AM

Great articles, first of all Thanks for writing such lovely Post! Earlier I thought that posts are the only most important thing on any blog. But here a Shout me loud found how important other elements are for your blog.Keep update more posts..

MSBI Training in Chennai

Informatica Training in Chennai

Dataware Housing Training in Chennai

Reply

vanessa Oconnell October 5, 2017 at 6:00 AM

Excellent post on hadoop Technologies Please makes more post on this tech to make us update in this.
Hadoop Training in Bangalore

Reply

Unknown October 11, 2017 at 5:43 AM

very nice blog hadoop and bigdata training in bangalore

Reply

Unknown October 17, 2017 at 4:37 AM

Great information. Thank you for sharing

Reply

Unknown October 17, 2017 at 4:38 AM

Very nice post. Awesome article... Really helpful

Reply

Unknown October 30, 2017 at 6:01 AM

Excellent blog on Comparing Hadoop performance on DAS and Isilon and why disk locality is irrelevant
keep blogging more thank you
Artificial Intelligence Training in Bangalore
Devops Training in Bangalore
Informatica interview questions

Reply

mojah October 30, 2017 at 9:39 PM

Nice Blog
Iot Training in Bangalore
Iteanz
Artifiacial Intelligence Training in Bangalore

Reply

Unknown November 3, 2017 at 8:00 PM

usefull article. Thanks for sharing

Reply

louis philip November 5, 2017 at 8:31 PM

Pretty article! I found some useful information in your blog, it was awesome to read, thanks for sharing this great content
to my vision, keep sharing..
Sms marketing
Text message marketing
Fitness SMS

Reply

Papumoni Gogoi November 6, 2017 at 9:16 PM

Nice Interview Questions
SQL Interview Questions
BPO Interview Questions
C Interview Questions
Spring Interview Questions
Iot Interview Questions
HTML5 Interview Questions

Reply

sujay November 8, 2017 at 9:51 PM

Nice blog python training
artificial intelligence training

Reply

Unknown November 9, 2017 at 4:12 AM

nice blog Machine learning training

Reply

Raju November 22, 2017 at 3:04 AM

Real Estate Research in Chennai
Real Estate Tax Advisor in Chennai
Portfolio Management Services in Chennai

Reply

Gokul Ravi November 23, 2017 at 1:10 AM

very nice blog it was useful

Reply

Gokul Ravi November 23, 2017 at 1:10 AM

very nice blog It was useful

Reply

Thanuja Sri December 7, 2017 at 11:20 AM

Really useful post about hadoop, i have to know information about hadoop online training institute in india.

Reply

Amar G December 14, 2017 at 1:08 AM

very helpfull blog it was a pleasure reading your blog
would love to read it more
knowldege is not found but earned through hardwork and good teaching
that being said click here to join us the next best thing in bangalore
devops online training
Devops Training in Bangalore

Reply

performancecarparts February 21, 2018 at 3:15 AM

Hello,
This is nice blog and sharing informative content .Thankyou
performance parts dubai

Reply

Ancy merina February 23, 2018 at 9:23 PM

This is incredible posting! I quite enjoyed reading it, you happen to be a great author. I will remember to bookmark your blog and will eventually come back very soon. Also share with my community and friends about this.
Web development company in bangalore
Outsource magento ecommerce services india
ECommerce Website developers in bangalore

Reply

Attiqa Shad February 28, 2018 at 8:37 AM

Very Nice Blog. Keep it up. I have also some good blogs check it. **Web Development in Pakistan**
Insurance Terms
Tech Free Tricks
Everything Is Here Online
Health Fitness Tricks
Insurance NEW YORK
Books World Online

Reply

Anuj Singh March 13, 2018 at 4:03 AM

Nice blog and absolutely outstanding. You can do something much better but i still say this perfect.Keep trying for the best. Hadoop development services in India

Reply

Janvi Sanvi March 28, 2018 at 7:22 AM

nice information keep updating the post Hadoop Admin Online Course

Reply

sathyatech April 9, 2018 at 11:20 PM

awesome post,keep it update post..Hadoop Online Training In USA!
Hadoop Online Training In India!
Hadoop Online Training In Hyderabad!

Reply

parveen kumar April 10, 2018 at 3:00 AM

Paper Angle Board Manufacturer in Gurgaon
Paper Angle Board Manufacturer in Bawal
Paper Angle Board Manufacturer in Behror
Paper Angle Board Manufacturer in Bhiwadi
Paper Angle Board Manufacturer in Neemrana
Paper Angle Board Manufacturer in Dharuhera
Paper Angle Board Manufacturer in Binola
Paper Angle Board Manufacturer in Binola
Paper Angle Board Manufacturer in Sohna
Thankyou..

Reply

Shailendra April 10, 2018 at 5:06 AM

This information you provided in the blog that is really unique I love it!! Thanks for sharing such a great blog. Keep posting..
Hadoop training
Hadoop Course
Hadoop training institute

Reply

gkr ragini April 13, 2018 at 11:36 PM

I simply wanted to thank you so much again. I am not sure the things that I might have gone through without the type of hints revealed by you regarding that situation.
Authorized Dot Net training in chennai
Advance Digital Marketing Training in chennai- 100% Job Guarantee

Reply

Rstrainings April 16, 2018 at 3:25 AM

nice blog...Hadoop training in Hyderabad!
Hadoop training in Madhapur!
Hadoop training in India!

Reply

Kausal Vikash April 17, 2018 at 12:05 AM

very interesting , good job and thanks for sharing such a good blog. artificial intelligence

Reply

Careen joseph April 20, 2018 at 10:40 PM

I simply wanted to write down a quick word to say thanks to you for those wonderful tips and hints you are showing on this site.
Best Hadoop Training Institute In chennai

amazon-web-services-training-institute-in-chennai

Reply

Enter your comment...

Comment as: Select profile... ▼

Publish   Preview

Newer Post   Home   Older Post

Subscribe to: Post Comments (Atom)

Powered by Blogger.