

Render Farm Grid Data Optimization Solution

PRODUCTIVITY OF RENDER FARMS

CGI companies deploy CPUs in clusters or grids, going up to several 100s, in order to provide the computing power required to render computer generated imagery (CGI), required for film and television projects. Even with sophisticated computing, rendering 3D or 2D models into graphic images is a time consuming and compute intensive activity and poses several concerns for administrators and IT leaders. Efficiency or getting the best performance from the available resources is certainly a big one amongst them. There are other concerns driven by industry trends - data sets becoming larger, jobs becoming more complex, more collaboration in the projects than before, delivery deadlines becoming tighter and of course, budget spending becoming smaller. Small wonder then, that there has been a demand to make the render farms more efficient and productive.

THE PROBLEM

Typically, in an animation company, a group of animators collaborate over a movie project. Each one of them designs different frames and scenes. Over the course of the day and weeks, hundreds of jobs are deployed across the different servers in the grid cluster. The same files are repeatedly read creating a large read I/O load over the network and the network storage. At crunch times the large volumes of NAS read operations slow down the NAS. This not only affects the existing jobs, but also creates a ripple effect across the grid. The NAS speed started to fluctuate from a typical 70 MBps to 20 MBps.

THE CHALLENGE

Overcoming these performance bottlenecks by over provisioning millions of dollars' worth of NAS hardware, is the usual approach. Besides becoming prohibitively expensive, this approach introduces newer problems – that of load balancing. Data management continues to be a challenge. Data in the system is maintained far away from the jobs and has to be retrieved over the network. This increases IO latency, high cost IOPS, slowed applications down considerable. Faster caches can be used locally but the pay off is not always consistent and the cost of IOPS is always a question. For administrators to optimise their data delivery system for higher speeds, requires for them to know their data better. For instance, they would need to know which files were most frequently used and so could be cached for speedy access or which jobs or applications required frequent storage access and which didn't. Without this knowledge, optimising their data delivery systems for efficiency and productivity is never easy.

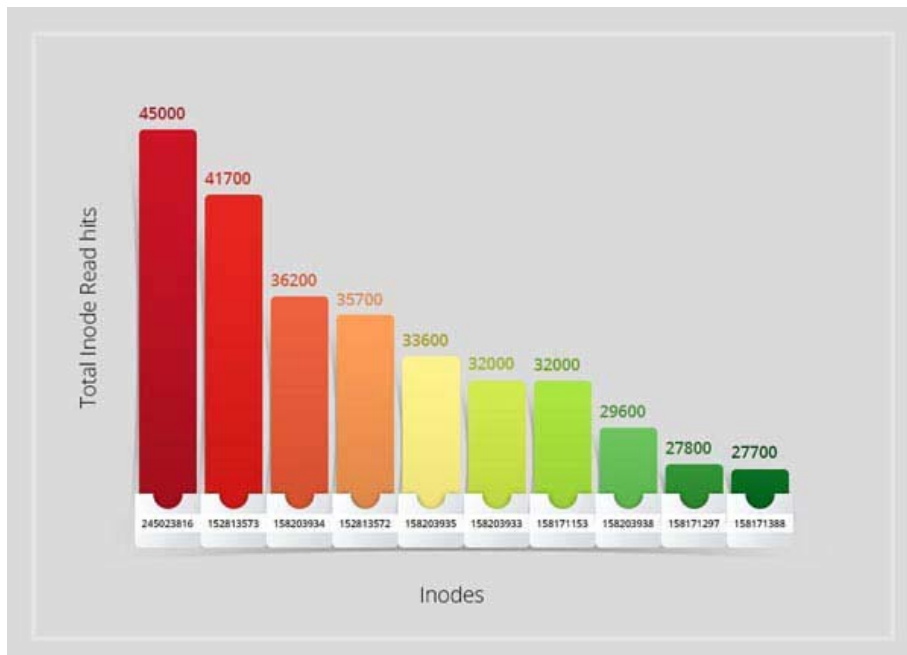
FLEXIBLE CACHE DEVICES

Cheap and expensive SSD options are also available for expansion. The cheap ones can be connected as SAS SSD or the expensive ones as PCI SSD. However, the acceleration is dependent on the type of caching device used. The table below provides a rough ballpark of actual numbers with different local and network devices.

Caching Device	Performance Characteristics
Single SATA Drive	80 MBps
Single SAS Drive	110 MBps
Single SAS SSD	240 MBps
Raid 0 two SAS SSD	460 MBps
PCIe SSD	900 MBps
NAS 1G	20-70 MBps
SAN	150 MBps

THE IO FINGER PRINT OF THE RENDER JOB

A principal problem for administrators in responding to these challenges is that of understanding what is happening at the IO level, in their grid environment. Administrators don't have any information available about the IO patterns of the workload. This is where Datagres provides value to administrators. PerfAccel software in the analytics mode provides insights to the IO finger print of each job. PerfAccel software accrues such information and presents the data in a graphical format that provides a view into an IO trace for the whole application across time.



Datagres software provides information about the different files that are used by the jobs and collects information on which parts of the files are hot or frequently used by the jobs. Over time in a particular movie project the same scene files are used again and again.

THE DATAGRES PERFACCEL SOLUTION

PerfAccel gives administrators, for the first time, a point of view of the active data and its dynamics in their grid environment. PerfAccel, not only gives visibility of active data, but, through an analytic mode, it gives administrators the power to understand, to make sense of these dynamics and also the power to control these dynamics, to increase efficiency and acceleration of the application.

BENEFITS

Understanding & View of the Active Data

Administrators could now:

- Track Macro and Micro level reporting of all I/O
- Track IO Timeline
- Identify Hot files and hot areas within each file, Over time, in a particular movie project, the same scene files were now used again and again, improving performance
- Track Latency and Throughput measurement at application level
- Troubleshoot and locate network bottlenecks

Control of the Active Data

Administrators could now:

- Cache only hot areas of the file or frequently accessed files
- Control which data to cache and which to not
- Sticky flag for performance critical data
- Pre-fetch files in cases where a large job was going to be deployed.

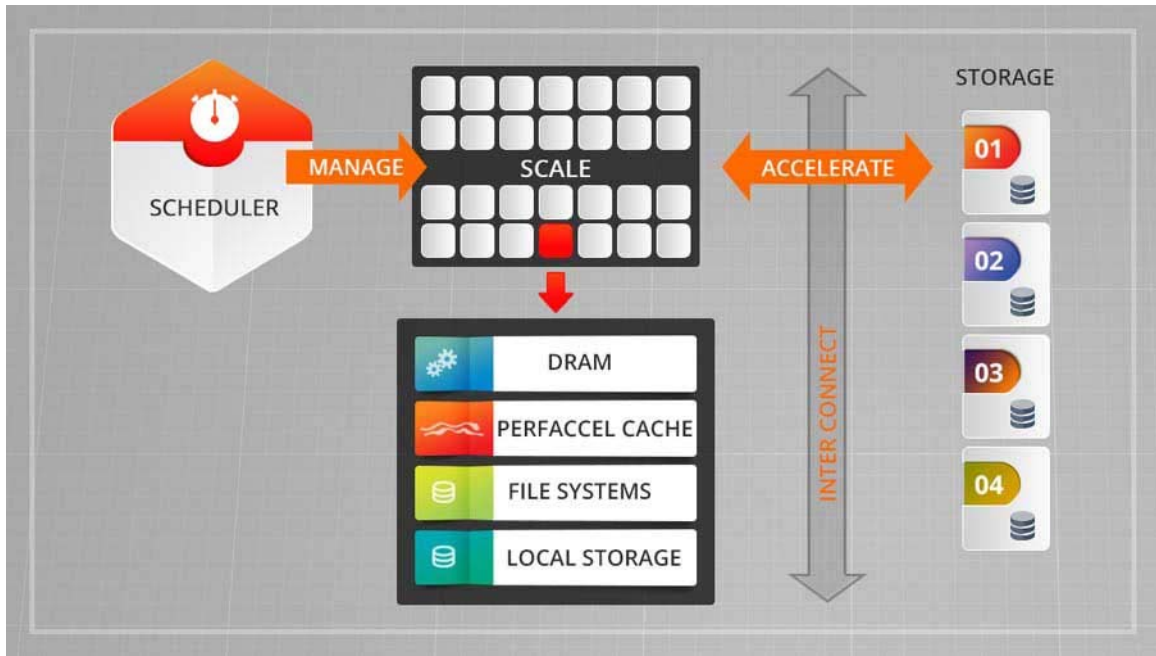
Accelerate Data

Administrators could now:

- Move data closer to CPU, Just-in-time and reduce IO latency
- Perform file-level caching on the server itself and reduce workload on NAS by 80%
- Use available storage resource intelligently

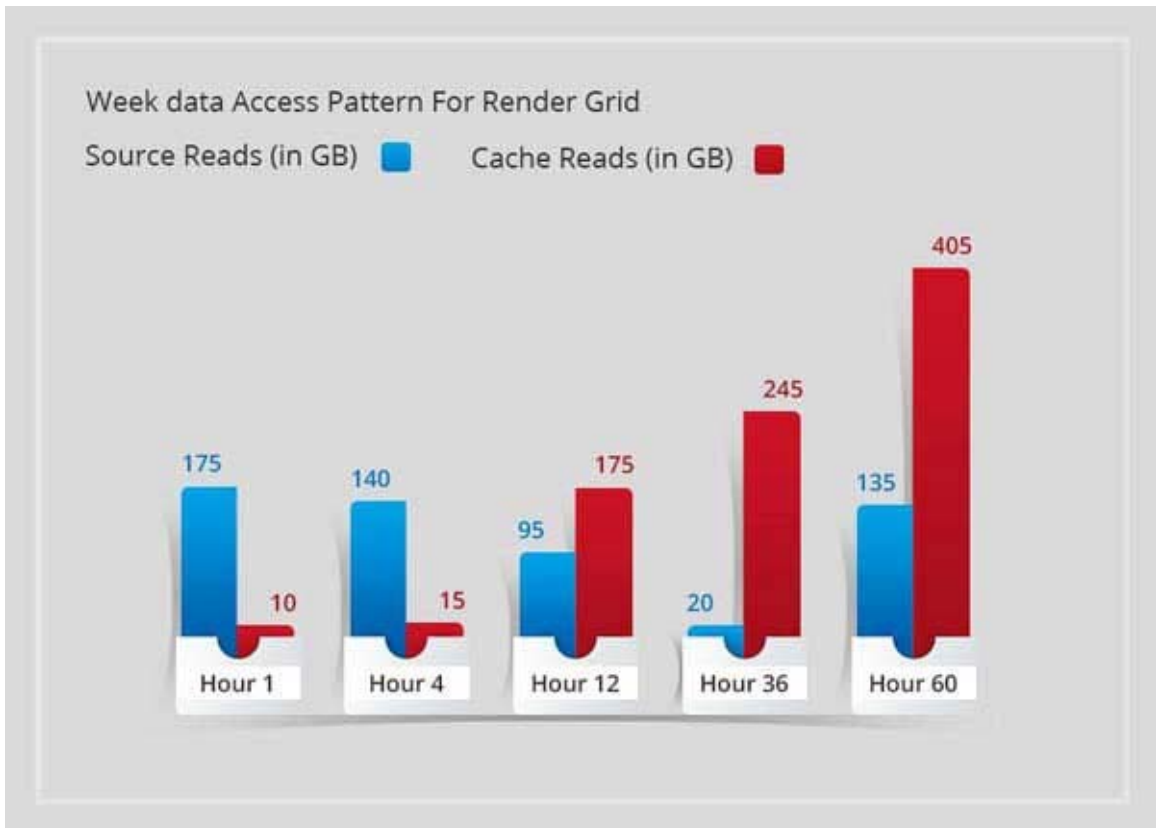
DEPLOYMENT ARCHITECTURE

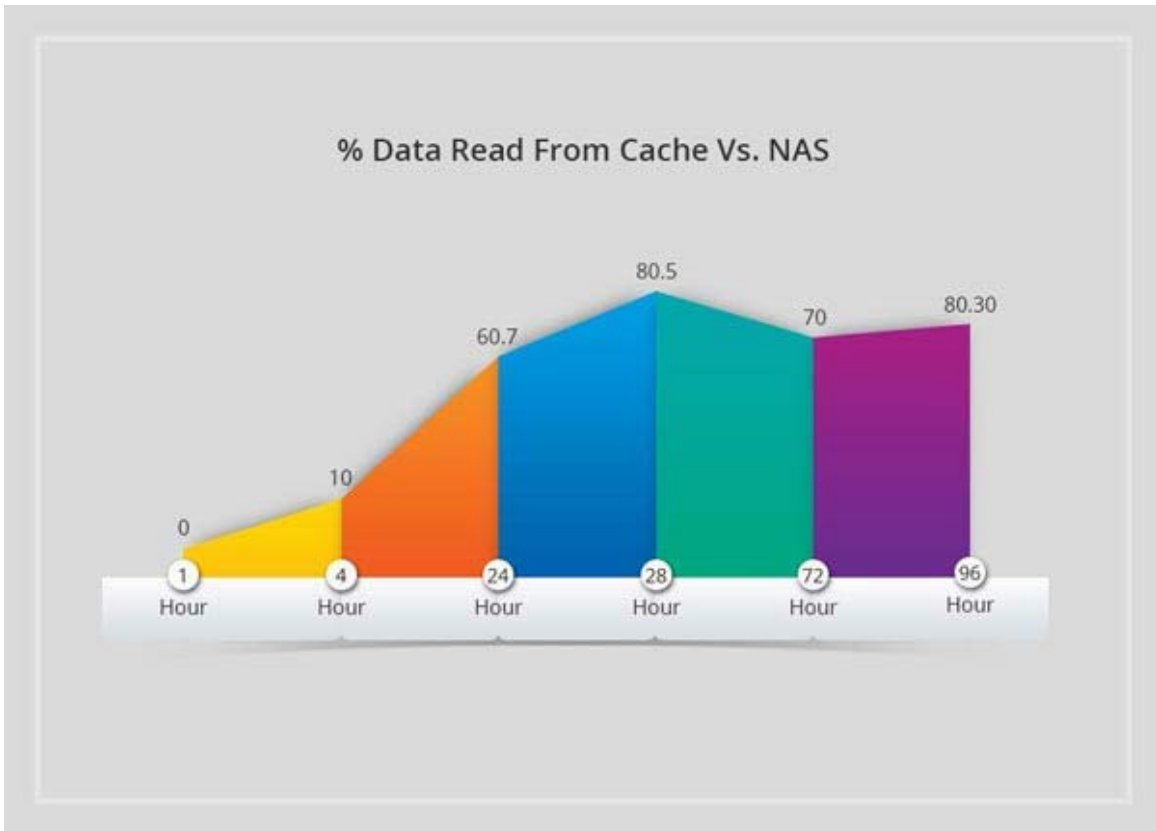
The deployment and configuration on the first batch of 60 servers takes less than 2 hours of time. Datagres also provides simple command line usage and a Single Pane of Management GUI with performance dashboards. The PerfAccel software deploys inside the Linux kernel. Data analytics and a single point of management is a great value proposition of the solution.



Datagres PerfAccel Deployed on each node of the grid

EFFECT ON CACHE BEHAVIOR





RETURN ON INVESTMENT

Initially the data is read from the source and stored in the cache devices. Over a period of three days, two-thirds of the data is serviced from the cache. This translates into network data savings within only three days of observation. More importantly, the extra IOPS on the backend NAS server increases the coverage ratio of number of servers per NAS.

“Once the performance bottleneck is alleviated from the backend NAS storage, the performance gains are measured in terms of price of local disk capacity. The need to continually upgrade expensive backend NAS storage for performance gains goes away.”

In some cases, savings are estimated to be ~\$200,000 of Tier 1 storage vendor per year including maintenance and management costs. Acceleration from the local SATA drives in some cases provides a gain of 40 servers per year estimated at ~\$200,000 every year.

USABILITY

Datagres' PerfAccel provides administrators with full command and control of the data in the grid through a single console. A single pane combines analytics and insight through performance dashboards, as well as a simple command line, for running commands i.e. creation of cache/source, deletion of cache/source, adjusting sizes and so on. Its flexible interface lets users configure their own policies of persistent cache, pre-fetching, predictive cache, real-time cache size configuration and auto-caching hundreds of NFS mount points.

PerfAccel Commands are easy to use and an administrator could learn them in a few minutes. One of the big advantages of PerfAccel solution is to enable hardware agnostic cache devices. It gives system administrators a choice to cautiously upgrade the grid to faster local storage as budgets allow. PerfAccel provides flexibility in configuring data management options and seamless working with the cache devices independent of their type and location.

Datagres Technologies Inc

2600 EL CAMINO REAL, Palo Alto, CA 94306 Phone: 510-402-4365

www.datagres.com



All of the documentation provided in this document, is copyright Datagres Technologies Inc. Datagres PerfAccel is a patent pending technology from Datagres Technologies Inc. Information in this document is provided in connection with Datagres products. No license, express or implied, by estoppel or otherwise, to any Datagres intellectual property rights is granted by this document. Except as provided in Datagres's Terms and Conditions of Sale for such products.

Datagres and PerfAccel are trademarks or registered trademarks of Datagres Technologies Inc or its subsidiaries in the United States and other countries. Copyright © 2015, Datagres Technologies Inc. All Rights Reserved. Datagres may make changes to specifications and product descriptions at any time, without notice.