

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

Welcome to FYI, the four-year innovation podcast. This show offers an intellectual discussion on technologically-enabled disruption because investing in innovation starts with understanding it. To learn more, visit arc-invest.com.

Arc Invest is a registered investment advisor focused on investing in disruptive innovation. This podcast is for informational purposes only and should not be relied upon as a basis for investment decisions. It does not constitute either explicitly or implicitly any provision of services or products by arc. All statements may regarding companies or securities are strictly beliefs and points of view held by arc or podcast guests and are not endorsements or recommendations by arc to buy, sell, or hold any security. Clients of arc investment management may maintain positions in the securities discussed in this podcast.

Hello and welcome to FYI, Arc's four-year innovation podcast. Today, we're extremely happy to welcome David Flynn on the show. David is the CEO and founder of Hammerspace, a data orchestration company that's doing some really interesting things. David, thanks for joining us today. Frank, really glad to be here. Thanks for having me on.

To start us off and kind of set the stage, could you give us a bit about your background and your relationship with data over your career? Oh gee, well, I'll be dating myself, but I taught myself programming when I was in high school on the Commodore 64. I had to confess that to Steve Wozniak when he was working with me at Fusion I.O. that I grew up on the poor side of town and couldn't afford an Apple computer and learned on the Commodore 64. I guess I've been lucky to have an interesting career. I was part of Larry Ellison's attempt to disrupt the Wintel duopoly with the network computer, a smart terminal, and then later interactive TV that was Liberate Technologies and the company Network Computer, Inc. before that.

I had the privilege of being chief architect at an outfit that built some of the world's largest supercomputers, Linux networks back in the early 2000s, where built some of the first Infiniband-based supercomputers, the first 100 node, the first 1000 node, the first 10,000 node Linux cluster that used Infiniband. We built those mostly for the U.S. Department of Energy, the likes of Los Alamos and more and sliver more. It's all got a taste for high performance computing.

I grew up back east in Huntsville, Alabama, where Warner Von Braun settled in the U.S. and did a lot of design work on Back to the Saturn V and the Space Shuttle and so forth.

I had my father was in the Strategic Defense Initiative, the Star Wars program. It gave me some exposure to that sort of thing. In high school, I ended up with a job working at Computer Science Corporation and built a flight simulator for an Army missile system. My first start was building three-dimensional graphics stuff while in high school. The largest claim to fame to date, because Hammer Space is definitely going to be a much bigger deal, was the company Fusion I.O., where we introduced high performance solid state for enterprise in mass. That was what ultimately became NVMe, form factor flash. In reality, it was the first high performance SSDs period. We disrupted the storage array business by showing that a single device could have more performance than an entire multi-million dollar storage array. Once you put those directly into each server, you can get crazy, crazy performance levels. I'm pleased to say it's nearly a quarter trillion dollar a year industry NVMe flash alone, much less the entire enterprise SSD space. Fusion I.O. was very successful, debuted on the New York Stock Exchange back in 2012. It's now part of Western

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

Digital. This is a really interesting disruption story. I would say just solid state drives and flash storage, overtaking, hard spinning disk drives. How did you end up being early into that disruption? What did it look like inside a company in a data center that was being penetrated by a new type of storage that was more performant? Well, you know, it was purely luck for where I had been in my career. I had gone from doing embedded systems, Larry Ellison's network computer

Inc. building these smart terminals and then ultimately TV set-top boxes that had browser technology on them. I'd gone from doing that to building very high-end supercomputers and this had given me exposure to NAND flash technology as you might embed in consumer electronic devices and then the architectures and design for massive parallelism and see these tiny embedded systems and very large supercomputers actually are more similar than you would think

because on both ends of that spectrum, you have to eke out every last ounce of horsepower of the hardware. It's in the middle that you get fat and lazy. It's in traditional corporate IT that you're kind of sloppy with all of that, but if you're building a mobile device or a cell phone or something, especially back in that era, you didn't have any kind of leeway to give and the supercomputing as well. The interesting part here, Frank, is that consumer electronics and very high volume is ultimately what drove the demand and volume for NAND flash and it was in taking that technology that had become cheap, high density, it had become viable. We took it into the enterprise space because corporate, especially corporate IT, had been, like I said, a little bit lazy in the middle there and hadn't really learned to exploit all of the performance of hardware.

Yeah, it's interesting. It's a common thing we see. You could look at battery adoption in electric vehicles similarly where a technology comes out that's better on some metrics, but maybe more expensive when it first comes out. You need something to drive demand to increase unit production

to bring costs down over time that actually expands the market opportunities where the first Teslas were very expensive, but over time costs have come down to the point where now the total cost of ownership of a Tesla Model 3 is lower than a Toyota Camry. I think the same thing sounds like happened in the flash space. Yeah, you could say that consumer electronics drove demand for lithium-ion batteries and then you make the leap to where you have even potentially much

higher volume in vehicles and now you have a foundation for it that didn't exist before.

That was the exact same thing with NAND flash and it had a major impact. We ended up selling over \$1 billion of that product to Facebook starting when Facebook was smaller than MySpace and it was their early adoption of flash technology that allowed them to scale their service in a way that others couldn't. It was a huge differentiator. At one point we disrupted ourselves. They were able to serve from two data centers the entire US when they had planned on putting in three. When we expected a large purchase shortly after they had gone public, suddenly they found that they could be much more efficient and operate from just two data centers which took and left something like \$100 million hole in our revenue forecast for the year.

Well, maybe that's a good launching point in thinking of how Facebook, which is now meta, has gone from smaller than MySpace to multiple applications with over a billion users and many data centers all over the world. How has the evolution of data centers changed

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

an organization's relationship with data and how does that really tie into your move from the hardware space to the software space with Hammerspace?

Well, Hammerspace is really an outgrowth of the experience with Fusion I.O. and the realization that managing decentralized data, having data be physically distributed across flash inside of servers in the case of Fusion I.O. and Solid State, or across separate third-party storage systems and cloud services, or across whole data centers. These are three different scalings of the same localization problem. Data today is a highly localized asset. The computing has to happen in close proximity to the data and the data is very immobile. We still live in a world where data, the main mode of using data is store and copy. Storage is the entire multibillion-dollar data storage industry and copy is in one form or another the foundation of all data management and the entire multibillion-dollar data management industry. It is our contention that that entire methodology is broken and that we ultimately have to move to where data is presumed to always be in motion and simply accessible from anywhere while the data is moved from behind the data presentation layer, behind the file system. Having it so that data is continuously accessible from a file system capable of spanning all of this infrastructure and putting the data movement function behind this global parallel file system, that's what is data orchestration.

It's not just data management by another fancy name because data management presumes that you're

copying stuff from without. This is where the movement is happening from the inside and that means you're actually solving the seeming paradox of how do you have data local to all of the places where you might need it, local to different data centers, local on different storage systems, even local on the different servers with their embedded flash. How do you have data local in all of those places without ever having copied it? It seems paradoxical to say you've never copied it. Well, how can it be local there? Well, in our world, those aren't copies. The file system, the global file system has simply moved the data to be on the right system where you're going to need it. That's data orchestration where the movement happens transparently from within, not as a copy from without, and where you're able to have data be local to everywhere you need it simultaneously. Yeah, I think that this difference between data management and data orchestration is really important to understand what Hammer Space is doing. Let me give my example of how I've explained it from my past life as a software developer and see if that checks out with you, especially this concept of data management being at a different layer is on top of your data versus behind the scenes with your data. Let's say working on a cloud EC2 instance on Amazon, a compute server, and I want to do some manipulations on data that's stored on an Amazon S3 bucket

or some storage in my data center. I basically, in my Python script, I'm going to write, okay, get the data from the S2 bucket, copy it to the local file system of the server I'm working on, I'm going to make some updates, and then I'm going to send that file back and update it or create a new iteration of it in some kind. That's kind of the data management world where I have compute and I'm copying data over to where my server is. And with Hammer Space, the data is just, seems like once the server is connected to the global file system, the data exists locally because I can just reference a file on that global file system directly and manipulate it directly without doing this kind of copy down and then push out. Without the get input to an object store. Well, you've hit on something else that is actually extremely germane to this discussion,

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

and that is that the big public cloud vendors starting with Amazon introduced a simplified API for access of data, super simple storage, S3, SSS. And it's basically a dumped down file system that's not a real file system. And it puts the burden on every application vendor or user to use get inputs from user space, rest interfaces. It's not a real file system. The operating system doesn't know how to attach to it, doesn't know how to load data objects from it. You can't put even your binaries, your program, your executables on it and have it run from it. It's not a true POSIX compliant file system, not standards based file system. Not only that, but there are now a gazillion different dialects of S3. You've got Azure Blob, you've got S3, you've got all of the different on-prem vendors. And those, by the way, are not all compatible with each other. They're different interfaces entirely. And all of this was so you could get to scale so that the cloud vendors could host multiple tenants, lots of different companies at a global scale. And that was kind of throwing the baby out with the bathwater. And then they convinced the whole industry, you need to rewrite your applications to do your own data management, to do your own copy of data, copying it down out of the object store onto something local, maybe into memory, maybe onto disk, copy it back up. Well, that is a cop out and was never going to be able to scale. And what we're seeing that is so amazing here is that AI workloads need real file systems and need to be able to move quickly because researchers don't have time to go and recode everything to use object gets inputs. So even very, very large scale, these hyperscale companies that are at the forefront of the competition with AI, they suddenly have found religion and need real file systems, shared file systems, not local file systems, but a global file system. So it has really been, you could say a perfect storm for the need for hammer space, because now you need to collect data from many different sources and feed it, not just to one, because it's not just one AI, it's many different models. And you need to feed it into each of those. And these are overlays. This is separate from the application that was using that data in the first place. So while we used to have a one-to-one relationship, this application put its date on this storage, this application put its date on this storage, maybe they shared the same storage system, maybe a different bucket or something, or maybe they weren't, but you really had, you know, a one-to-one. Now you have a many to many, we have to collect up data from many different data sets and feed it into many different models. Oh, and there's a third many, because these are all cross products and it's a three-dimensional cross products, many applications, many different models that they need to be fed to, and you need to run it and in different distinct data centers where you have different hardware, because this requires in GPUs and it's very bursty. So the last thing you can afford to do is spend that kind of money, actually spend the time waiting on getting GPU shipped, you can't do that. So even some of the largest of the automakers who are competing in the world for autonomous vehicles, they can't get their hands on GPUs and are using not just one, but all of the clouds as ways to get access to that shared resource. In other words, we have to move to a least model, a time-shared model to be able to do that, but then the challenge is, well, how do you get the data there? And that's exactly what we solve. So Hammerspace is solving the issue of decentralizing data so that you can solve this many to many to many problems in the AI world, and it's solving how to feed data at super high performance for read workloads for the ingest of data, for write workloads for the checkpointing to capture a snapshot of the state

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

of things so that you can restart when something fails, because when you're doing, you know, thousands of GPUs, something's going to fail and you don't want to throw away the entire job. So it really is, and I had a gentleman actually, Los Alamos National Labs, where they are also, in addition to their research around the nuclear arsenal and so forth, they are doing AI work now, and it's a whole new kind of workload that needs different types of the ability to feed performance. And Hammerspace really solves those three things, the data orchestration, the read performance, and write performance at kind of a new level that the world of enterprise NAS has never really done before. And that's because the architecture is a true parallel file system. It's more akin to the file systems that you would see in the supercomputing world. Yeah, I think a good way to maybe tie together those kind of performance traits with what Hammerspace is doing. You describe it as a global parallel file system. And I think all of those words are probably important. So global, I think, is kind of obvious. Data is globally distributed, and our users or the applications that are accessing that data are globally distributed, and we need to have a way to orchestrate the connection of those two. Describe parallel and what that actually means. And then you were talking about file systems and comparing them to object storage. So understanding are file systems always better than object storage? Or when would you use one or the other? And what does that look like? Both of those words would be great to go more into. So parallel means that the data path where you're accessing the data can be done in an end to end fashion, many to many fashion. So you can have many different storage systems

or nodes, and you can have many different clients. And the data is diffuse across them without any choke points. So you get perfectly linear scalability. If you double the number of storage nodes, you double the number of clients, you get double the overall performance. Interestingly enough, that is not the case in any enterprise NAS system today, whether you're talking NetApp, Vast, Isilon, the Dell stuff, Cumulo, those systems are called scale out NAS, and they keel over. You stop getting additional performance at some point because metadata becomes the choke point. The magical design difference with a parallel file system is that the control plane and the metadata is separated off to the side, and the data path is direct and parallel. It's the separation of metadata from data from the control plane from the data path, which allows you to have scalable data path in traditional enterprise NAS, even the most advanced new generation from the likes of Vast or Cumulo or Pure Storage, those systems, metadata and data are combined. And that means it becomes a choke point when it comes to the distribution of data across the storage systems. So parallel is not just lip service, it is representative of very fundamental architectural difference, which for the first time is available as a standards-based NFS as an enterprise NAS protocol. And that's because folks like at Los Alamos 15 years ago conceived of this and started an academic effort in that direction and Hammerspace picked that up and used it not just to solve performance, which was what it was designed for, but we used it to solve orchestration because fundamentally you need to separate metadata from data to be able to allow the data to move freely and to be orchestrated like we're talking about. So we use this architecture to solve performance and orchestration, sort of either ends of that spectrum. And that's one of the things that makes Hammerspace so utterly unique is we're introducing a new class of enterprise NAS, one that's true parallel NAS, not just scale out or scale up like NetApp in the original.

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

Yeah, it's super interesting. And then file systems. So you were comparing some of the benefits of file systems to object storage. Should we totally get rid of object storage or do we still need that too? Well, no, object storage is great because it's cheap because it allows the vendors to build these things without having to build file systems. So we should be thankful because what we have built is a file system that can sit atop different storage systems with their own internal file systems. And Hammerspace, think of it as an overlay file system with this parallel access gives you aggregate performance. Well, an object store is a simplified file system that's easier for folks to build. That's why it's fundamentally cheaper as a technology and it puts more of the burden on the folks using it. So there's definitely a place for it as a cheap and deep for archival tier for mass storage tier. And Hammerspace can orchestrate data onto object storage just like we can orchestrate it onto file storage and we can orchestrate it onto flash inside of server storage. So when we talk about being a global parallel file system, global doesn't just mean geographically something that can span different parts of the globe, different data centers. It also means it can sit atop of any form of storage infrastructure, file, block or object storage infrastructure from any vendor. You could say universal in that sense in that it's able to sit on those different forms, not just limited to say NVMe over fabrics, like say vast technologies limited to certain types of things. So global meaning universal and only then can you truly claim to be an orchestration platform because you have to be able to orchestrate across all different storage infrastructure. So yeah, every single one of those words global parallel and file system means something. File system also means that the operating system knows how to attach to it natively and consume data from it through traditional read and write interfaces so that applications don't have to be rewritten. And that's because many of these AI applications, most if not all of these AI applications speak, you know, POSIX file IO depend upon the OS to do that thing. I mean, operating systems have been built over the past 30, 40 years and they have APIs for doing this. It didn't do anybody, any service to say let's throw all that out and create these alternative interfaces to, you know, rest based user space. And it was really just a cop out to make it simpler to build the storage systems because it was starting to be a handicap because they're so hard to build. Yeah, that last part's the difference in my coding example, but in your Python script between doing read this file from the file system, which is a very straightforward thing you learn in Python 101. And get this file from an Amazon S3 bucket, which is a different set of code you need to write versus if you're getting it from Azure Blob Storage versus any other type of kind of what looks like external storage. And that's how Hammerspace simplifies things. It doesn't matter where the data is actually stored or on what type of storage you can just read it in very simply. Another way to say it is we've implemented the most universal interface to data. And that's the interface that OS is know how to do. You can always do the other ones. You can interact with it as an object store if you want to Hammerspace. But there's not much need to do that when you can interact with it as a true file system. And it's not just the APIs, it's the semantics, the fact that you can do random reads and writes that you can go and read a tiny little section, that you can rewrite a tiny little section of it, that you get very low latency that you're getting, local access performance. But the example I like to use is just loading a program off of it. An OS doesn't know how to load a program from an object store. It doesn't exist to the OS. So you can't navigate around and do an LS. You can't list the directory. You can't move the files

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

around. You can't rename things. So the whole slew of how you interact in a home directory moving files and directors around, none of that works. And that's why people have recreated sort of these web interfaces. You go through a browser and that's a cop out for what you really need to do. And AI researchers need to be able to work with petabyte scale data sets and have the most native simplified interface. And that is a file system.

Yeah, I think that paints the picture, the value proposition really clearly. I like the most universal interface as an analogy. What kind of companies need this? You've mentioned AI a few times. Makes sense, the demands there for highly performant access to data.

Highly distributed data. Highly distributed data, right?

On a file system, yeah. From many different, from infrastructure that is itself distributed, right? The data is distributed, the infrastructure is distributed, and having a universal interface so that folks aren't wasting their time. Because the innovation cycles in AI are so fast, you can't waste the time worrying about this, right? The interfaces, you've got to be able to use it off the shelf. The other industries that really make a difference are the traditional ones that use large amounts of data, especially unstructured data. So everything from making movies and video

games, to designing microchips, to designing drugs, bio and life sciences, energy and energy exploration, building spaceships, airplanes, rocket ships, or cars, industrial design.

What all of these have in common is, and what unstructured data tends to be, is some form of imaging, whether it's images that we see with our eyes, images that you print on microchips, the images of genomes and proteins, or the three-dimensional models of things, designing of parts. And that is the largest portion of data. If you take the universe of all data, 90% of it is unstructured. Only the minority, the 10% is structured, meaning it has fields, like databases, like your bank account balance is one field in a database. Those are very small. The thing about AI is, AI has opened up the ability for machine perception for computers to actually be able to perceive what's in images, perceive images to perceive, let's call it higher dimensional, much more nuanced things. And so now all of a sudden, unstructured data has suddenly

come to the forefront. It's become part of the game. And it's not just the recent data, it's also old data. So now you have something that suddenly surfaces, instead of the 10th percentile of data needing to be on a performance system to feed your data warehouse or your transactional databases,

now we have this bulk data that needs to be there to be used as training sets.

Yeah, so it sounds like, or would it be correct to say that the explosion of interest in generative AI has accelerated the demand for a global parallel file system? Absolutely, absolutely.

It's accelerated the demand for access to unstructured data from many different sources and fed into many different regions and from a more universal interface. So it really is, like I say, the perfect storm for what I set out nearly 10 years ago to build when we set out to build this because building file systems is not easy. It's the hardest part of the operating system or operating environment. And we're talking here about the file system for the super cloud or software defined cloud. This is really about the data platform that allows you to use infrastructure anywhere as a service. So the data platform has really been the key to all operating systems operating environments. And so timing couldn't have been better for this to be

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

true in the future. And we're seeing that now with companies like Jeff Bezos Blue Origin is using Hammerspace for their design of rocket ships and rocket engines as a way to make data available in the different facilities where they design them, where they manufacture, where they test fire, where they launch them, where they monitor these rockets. All of that is brought together with a data platform that a data orchestration platform with Hammerspace that also gives them the ability to tap into the cloud. So let that sink in for a minute. Jeff Bezos own company Blue Origin is using Hammerspace as the way to do hybrid cloud. Yeah, it's pretty incredible. I mean, there's obviously it's, and it seems like the way that the demands of our data are growing from the applications and the different types of storage and even environments when you're talking about space. Is there any that you're like over the next five to 10 years that you're most particularly excited about Hammerspace's opportunity or just things going on in the industry related to data? Well, clearly the thing that has the most buzz today is around AI and machine learning. And that, you know, I really think is going to transform so many different industries. And there's major investments appropriately so going into those systems and the architecture and design for it, due to these new drivers like, you know, multiple data sources, sending it to multiple places to run models on it, needing a universal interface. These things are driving that demand. To see that being done at scale, we have now one of the companies that I have history with, I can't state names, sold product to them in the past at other companies, is at the forefront of the AI arms race, because they have a platform where they can directly monetize this with their very large user base. And they have now adopted Hammerspace as the data platform for that AI work.

And so some of the language models that are out there now are being trained using Hammerspace as the data ingest, as Checkpoint is, you know, the AI researchers scratch pad. I'm super stoked about that because I had been, you know, tracking what was going on with language models and so forth and just it's jaw droppingly impressive to see what's going on there. And now to be able to have a part in that is very rewarding. And the interesting part about this, Frank, is that it's recursive. Hammerspace abstracts data from the infrastructure so that we can allow the movement of data to happen transparently, not disrupt how you're accessing it. Well, that allows us for the first time to truly automate data movement. And what we have done as a company is

we've used a machine learning technique to decide where to place data and when to move it using a financial arbitrage simulation. It is more of a machine learning technique. It's the wisdom of the markets. It's using a resource auction model to determine what data down to the individual files should live on which storage systems across a, you know, a global environment. And it's interesting that if you think about it, what we're talking about here is enabling AI to be used to accelerate the rate at which we can do the AI research and AI innovation. And that's why you get this exponential speed up is when you start using these techniques to make it faster to be able to train the models to iterate on them and move even quicker with it.

Yeah, that's super cool. I love that example. I was going to ask if you're using AI at Hammerspace, and that makes a lot of sense. Basically, all the companies that are looking how to optimize their cloud storage budgets, Hammerspace knows from your orchestration layer how much you're accessing data, who is accessing it. Okay, let's move it to a low cost region. That's a lower

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

tier of performance if people aren't using it. As soon as demand speaks up, let's upgrade it, et cetera, driving that logic based on machine learning. And in the future, what this looks like is, you know, five years from now, 10 years from now, people are going to look back and go, how did we live? How did we survive in a world where we had to copy shit around between storage? Because once data becomes orchestrated, it actually, and here's the bizarre thing, data is more permanent in an orchestration platform. It's not bound to the lifetime of any one storage system. It simply lives in the ether. It lives in an alternate dimension. That's why we called the company Hammerspace. That represents the alternate dimension within which you can have an infinite amount of storage and access it instantly. What Hyperspace is for travel and sci-fi, Hammerspace is for the storage of things. It's that bubble universe where you can stick things, right? And that is, you know, the interesting part is, you know, going to that whole concept of disruptive innovation and volume markets, pioneering technologies that work their way back into the more premium markets, right? So, consumer electronics, driving flash technology, and then that goes back into the data center, or driving battery technology that goes back into vehicles and into the power grid. Well, we have already, as consumers, moved to a data orchestrated world. When you go and access data on your phone, you don't even think about it. If you lose your phone and you get another one, all of your data is there. When you go from your phone to a tablet to your laptop, your data is just there. We don't think about it anymore. Every last text message, email, photo, video, you don't, as a user, you don't accept an application that doesn't give you that data orchestrated model. And yet, in the corporate IT world, we haven't gotten there yet. So really, I'm just doing the same thing again, which is reapplying these principles that we have grown accustomed to, that your data simply exists independent of any of the physical systems that might be holding it, and applying that to the foundational petabyte and exabyte scale data that makes these very data-intensive businesses, so it's no longer through storing it that it's permanent, it's through having that data live in an orchestrated platform. Yeah, that's super interesting. I like the concept of how you've taken your past life in the hardware industry and helped take a technology pioneer maybe in consumer electronics and bring it to the data center and the hardware side, and now you're doing the same thing on the software side. Correct, correct. I am a software engineer by trade, and doing Fusion.io and NVMe Flash was really a means to an end. This is the logical conclusion of that, is then how do you put a file system across that flash that's scattered throughout servers? How do you put a file system across the different storage systems and services on-prem and in the cloud, and then how do you put that file system in a way where you can have the same file system be serving data from within multiple data centers, active, active, read and write, and it's the same data set? And here's another interesting thing. This doesn't just transform how businesses work with their data, it transforms how businesses work together to collaborate on data. It will change how data as a product is supplied between companies working together in a layered market to build data into a more refined product. Think of it as the supply chain of data, the data supply chain. To date, the way that is done is by copy and merger between the storage systems at the different companies. So if you're a movie studio and you contract out certain scenes to be done by a contract studio, you're copying stuff back and forth. With Hammerspace,

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

and we have customers doing this now, they can simply possess the same file system in both organizations. And what I write over here, he's able to read over there, and what he writes over there, he can read over here. In other words, you're sharing data by reference to the same live environment as opposed to sharing data by copy and merger. Yeah, that's really interesting. I hadn't thought about that angle of helping not just one organization access disparate data, but helping organizations collaborate together. If you think about that, this is really what's kind of missing from internet. Internet had a model of being stateless, right? It's completely stateless.

Well, how do you have now durable state, and especially valuable and data that needs to be secured? How can you have that spread at a global scale? And data orchestration sort of provides that notion of durable persistence of data that can even go across organizations. And the beauty is, it's always in the same file system, secured by the same access controls, the same audit trail. By eliminating data copy as your principal mechanism for moving data, now you can keep Pandora's box

closed. Because as soon as you copy something, you've scattered it to the wind. There's no way you can collect it back. You don't know where all it goes. But within Hammerspace, if I change the permissions on this file, it's changed globally. If I set up audit trail, I've got an audit trail globally. And I can tell whoever accesses it wherever they access it. I would argue that compliance and security have largely been theater until you can move away from copying data as your main tool of data management. Because copying violates any of that. Yeah, it sounds like simultaneously a massive unlocking capability and also a simplification of an organization or multiple organizations interaction with data.

That's right. And the ability to simplify even processes that in the past, they were in conflict. To get the agility, you would copy stuff around, which would violate security and compliance. So those things were fundamentally in direct opposition to each other. But when you move to an orchestrated world where data movement is happening from behind the file system, now your security and compliance is no longer in conflict with performance access and agile and universal access to data. So it fundamentally changed the game by different factoring. And that factoring that makes it possible is the separation of the control plane from the data path, the metadata from the data. And that is, by the way, the same thing which enabled software-defined networking back in the day. Super interesting. So it's a common design principle. It's just not been applied to data as a way to decouple data from the storage it's stored on, right?

Yeah, I like that. Maybe one last question to close this out. How long do you think this transition for this technology to really take hold in the average enterprise or data center or even cloud? How long is this going to take? What's the adoption cycle going to look like?

Oh, I wish I could say it was going to be faster. But we saw solid state. And by the way, adoption of flash was much more one-dimensional. Performance, performance, performance is the main driver for it. Hammer space is much more intimate on many different dimensions. And fundamentally changes how you interact with data. So it has a lot more inertia to fight against and will have a lot more momentum once you get there. So I expect five years from now it will be a well-accepted principle, sort of like server virtualization. That's the closest analogy.

We used to live in a world where you had to feed CD-ROMs into server chassis. You had to rack and stack the servers in the first place. This is how you spun up a new server. Now you spin up a new server by a few clicks of the mouse and it's in the virtual. So server virtualization was the

[Transcript] FYI - For Your Innovation / Decentralized Data and AI Workloads with Hammerspace CEO David Flynn

decoupling of the compute from the computer running it. And data orchestration is the decoupling of the data from the storage system storing it. And just like server virtualization took a while, but reached a tipping point where it was accepted that you needed to go there, I think it's going to be about five years before it's accepted that you simply have to go orchestrated. And it's a different paradigm. To go to a data orchestrated paradigm, your data gains its permanence by living in perpetual motion in an orchestration system. And that's just counter intuitive versus data being permanent by storing it, which seems so natural. If you want to be permanent, store it. You want to put it somewhere else, copy it. Well, that will change, but it'll take at least five years. But then we'll see probably 10 years before it's penetrated the market and really transformed it. But who knows? All that could change more abruptly with AI. Right. Moving the industry so, so quickly. All games could be off if we hit singularity. Right. Yeah. Exciting times. And it seems like there's a lot of opportunity ahead. So I think that's a good place to end it. Thanks so much for coming on the show, David. My pleasure. Thanks, Frank.

ARC believes that the information presented is accurate and was obtained from sources that ARC believes to be reliable. However, ARC does not guarantee the accuracy or completeness of any information. And such information may be subject to change without notice from ARC. Historical results are not indications of future results. Certain of the statements contained in this podcast may be statements of future expectations and other forward looking statements that are based on ARC's current views and assumptions and involve known unknown risks and uncertainties that could cause actual results, performance or events that differ materially from those expressed or implied in such statements.