

Title: Reverse CDN

**A Pavilion Data Systems, Inc. study report enabling attributable market insights.**

Author: **Mediumroast Barrista Robot**

Creation Date: **2022-01-16 23:38**

Findings

# Introduction

The mediumroast.io will automatically generate key themes, text summaries, and detect relevant snippets from the interactions associated to a study. It is then up to you to read, digest, and find meaningful points from these interaction data and document them here. In that spirit this paragraph should be rewritten based upon your research into these data.

# Opportunity

Opportunies document the real outcomes from your research into the interactions associated to the study. It is recommended that you write a quick summary related to core things you have found.

* Transparency - Product or service plans that are completely transparent, to all stakeholders and constitutents, are easy to follow and clear in direction. Further inherent transparency enables all involved parties to build community and rally behind your target audience. The mediumroast.io enables the construction of transparent plans.
* Evidence Based - Too often product, engineering and marketing teams believe in their own vision instead of the vision of their users. This misalignment between internal and external visions is evident when plans, for products and services, lack attributable evidence. That is because it is impossible to show where key elements, leading to the vision, originated from. Easily constructing evidence based plans is a key value of the mediumroast.io.
* Community - When different groups have differing opinions on what the future holds for the product or service, the potential for community is dashed. Using both Transparency and Evidence Based plans the mediumroast.io can be the kernel needed to build up a community that is passionate about user and market needs instead of internal opinion.

# Actions

To improve completeness of this study the following actions and next steps are documented with appropriate statuses.

1. Augment your study with relevant information, add interactions and companies, and more generally have fun.|Status: In progress
2. Activate your mediumroast.io organization and create your first study.|Status: Done

Key Themes by Sub-Study

These are the themes for every sub-study that you elected to include in the report. The format for each theme section covering summary and detailed/discrete themes includes the theme identifier and description/definition, a fun theme fortune, the themes tags, and a listing of quotes. Note these data are pulled from the system, and if you'd like to modify them you would need to modify them in the system prior to generating this report. Additionally, if you make changes to this report they aren't reflected in the system.

# Sub-Study Identifier: 2 — Using competitive white papers look for competitive insights.

## Summary Theme

This is the summary theme for the entire sub-study. Using natural language processing mediumroast.io has detected the associated tags and quotes matched to the tags. It is a generalization of the sub-study and accompanies the discrete detailed themes.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Business owners rethink the flow of data in the cloud with data management and analytics. [system generated]

**Tags:** *business data work | business owners | data data | data management | data security | edge cloud | enterprise data | flow management | public cloud | rethink data | stream processing analytics*

### Theme Quotes

* RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—FROM EDGE TO CLOUD | 35 | CHAPTER SEVEN Data Security and Data Management In addition to DataOps a unique part of data management that deserves a special examination is data security.management and analytics were RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—FROM EDGE TO CLOUD | 34 | SEAGATE POV DataOps: Getting to Customer Satisfaction and Profit Through Data This survey established that a there they need to interrogate it will flow between environments smooth-functioning DataOps is key the data at their disposal.RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—FROM EDGE TO CLOUD | 19 | The public cloud benefits that extend to the multicloud ecosystem include: Fast growth and scale.management and analytics solutions.
* 3 Data-in-Motion Philosophy SOLUTION BRIEF Stream Processing and Analytics With Loads of Control and Options The Stream Processing and Analytics The third tenet of capabilities is stream processing and analytics.STREAM PROCESSING FLOW MANAGEMENT STREAMS MESSAGING AND ANALYTICS DATA PIPELINES ’s data-in-motion philosophy is rooted in the complementary powers that are brought to the table by Apache NiFi for flow management Apache Kafka for streams messaging and Apache Flink for stream processing and analytics.Those are best addressed by a combination of flow management and stream processing and analytics capabilities in your streaming architecture.Data capture and flow management at the edge; to We list the stages below that highlights how Provisioning that data directly to/from your Kafka messaging backbone; and/or best-in-class compute engines became Stream processing and analytics our data-in-motion vision.
* Scientists Business Conclusion Analysts and Building a Platform that makes the most of Kafka by enabling true stream other IT and data processing and analytics is not easy.data from How does one ensure all the pieces work together and are transactional enterprise grade?It may be used as the underpinning of stream processing and analytics or as a data distribution hub.enterprise targets.

## Detailed Themes

Here are the discrete themes for the sub-study. Since we need enough unique tags to check the individual interactions in the sub-study for relevant text snippets and quotes you may find discrete themes without quotes. We do not strike them from the report because we believe it is important to see what the mediumroast.io detected for transparency purposes.

### Detailed Theme Identifier: 0

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Apj of business on the edge of a cloud in europe [system generated]

**Tags:** *apj | business | china | cloud | data | edge | europe | kafka | management*

#### Theme Quotes by Interaction

##### 202007010000-Reverse CDN-Seagate

* infrastructure operations and business units 23% 20% 16% 15% 18% 15% 15% 15% 15% 11% 13% 11% 13% 12% 11% 9% 9% 10% 6% 7% Europe NA China APJ Europe NA China APJ Europe NA China APJ Europe NA China APJ Europe NA China APJ Meeting increases in data Improving ease of use Improving availability and Integrating third-party capacity requirements increasing uptime services to use alongside modern infrastructure 10% 12% 11% 13% 9% 8% 7% 9% 7% 8% 7% 9% 7% 6% 5% 6% Europe NA China APJ Europe NA China APJ Europe NA China APJ Europe NA China APJ 1 The EMEA Datasphere IDC 2019 RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—FROM EDGE TO CLOUD | 51 | CHAPTER FOUR North America (NA) The survey queried enterprises in the United States and Canada.
* infrastructure operations and business units 23% 20% 15% 16% 18% 15% 15% 15% 15% 13% 11% 13% 11% 12% 11% 9% 9% 10% 6% 7% NA China Europe APJ NA China Europe APJ NA China Europe APJ NA China Europe APJ NA China Europe APJ Meeting increases in data Improving ease of use Improving availability and Integrating third-party capacity requirements increasing uptime services to use alongside modern infrastructure 10% 12% 11% 13% 8% 9% 9% 7% 7% 8% 7% 9% 7% 6% 5% 6% NA China Europe APJ NA China Europe APJ NA China Europe APJ NA China Europe APJ SEAGATE POV A Data Growth Leader In the next two years enterprise astonishing 49% of expected increased use of SaaS applications.
* Proportion of Respondents Reporting Their Organizations FIGURE 1 Have a Single Enterprise-Wide App for Policy Management Back-up/recovery Container orchestration Policy management Data discovery Data classification 45% 40% 39% 38% 36% 36% 36% 33% 36% 35% 36% 38% 33% 32% 29% 31% 34% 30% 31% 27% APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe Metadata management Recovery orchestration Data security for data Data security for data Data migration tiering being stored on the move or placement 38% 39% 38% 41% 34% 36% 34% 35% 37% 29% 35% 34% 35% 34% 30% 28% 28% 28% 30% 27% APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe DATA BY IDC To take a region like APJ and technologically developed countries.
* FIGURE 11 Main Challenges in Exploiting Data’s Potential Making collected Managing the storage Ensuring that needed Ensuring the security of Making the different silos of data usable of collected data data is collected collected data collected data available 41% 39% 39% 39% 39% 38% 39% 41% 37% 38% 35% 35% 37% 38% 35% 33% 32% 31% 24% 33% China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ Getting the required Having the technology in Establishing data Building the people resources Connecting the curated resources to manage place to analyze data management governance needed to analyze data data with data users collected data and processes 37% 30% 27% 27% 29% 29% 29% 26% 24% 23% 24% 23% 22% 8% 22% 19% 21% 6% 6% 11% China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ China NA Europe APJ RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—FROM EDGE TO CLOUD | 47 | CHAPTER THREE Europe The survey queried enterprises from four European countries: the United Kingdom France Germany and Russia.

**Frequency:** *895*

##### 202104140000-Reverse CDN-Cloudera

* delivers an enterprise data cloud for any data We deliver on that vision by providing an integrated platform of best-in-class data streaming anywhere from the Edge to AI.
* Data capture and flow management at the edge; to We list the stages below that highlights how Provisioning that data directly to/from your Kafka messaging backbone; and/or best-in-class compute engines became Stream processing and analytics our data-in-motion vision.
* Kafka enabled the business to scale that volume across multiple Stream processing and analytics is how you generate real-time analytical insights from on-premises and cloud environments.
* cured “Kafka Blindness” see the white paper “Manage Monitor and Replicate Apache Kafka Across the Enterprise and Cloud.” Full Visibility of Streams Messaging The streams messaging tenet requires high scalability and stability and the best-in-class compute engine in that space is Apache Kafka.

**Frequency:** *895*

##### 201802010000-Reverse CDN-Striim

* Kafka data.
* end-to-end How does one deliver data from Kafka to targets like Hadoop databases or cloud storage?
* The platform enables Data Scientists Business Analysts and other IT and data professionals to get the most value out of Kafka without having to learn and code to APIs.
* Real-time data ingestion into Kafka from many Enterprise data sources 2 GETTING DATA INTO KAFKA When writing to When users are considering how to get data into Kafka they need to Kafka in determine how to collect source data in a streaming fashion and how to “massage” and transform that data into the required format on Kafka.

**Frequency:** *895*

### Detailed Theme Identifier: 1

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** A teddy bear in a dark room [system generated]

**Tags:**

#### Theme Quotes by Interaction

##### 202007010000-Reverse CDN-Seagate

* mediumroast.io was unable to find a relevant quote or text snippet for this theme.

**Frequency:** *None*

##### 202104140000-Reverse CDN-Cloudera

* mediumroast.io was unable to find a relevant quote or text snippet for this theme.

**Frequency:** *None*

##### 201802010000-Reverse CDN-Striim

* mediumroast.io was unable to find a relevant quote or text snippet for this theme.

**Frequency:** *None*

### Detailed Theme Identifier: 2

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** I think i need to stop thinking about them. [system generated]

**Tags:** *stop think them*

#### Theme Quotes by Interaction

##### 202007010000-Reverse CDN-Seagate

* mediumroast.io was unable to find a relevant quote or text snippet for this theme.

**Frequency:** *None*

##### 202104140000-Reverse CDN-Cloudera

* mediumroast.io was unable to find a relevant quote or text snippet for this theme.

**Frequency:** *None*

##### 201802010000-Reverse CDN-Striim

* mediumroast.io was unable to find a relevant quote or text snippet for this theme.

**Frequency:** *None*

# Sub-Study Identifier: default — Gain a clear understanding of the problem space to build a proper product plan for streaming storage.

## Summary Theme

This is the summary theme for the entire sub-study. Using natural language processing mediumroast.io has detected the associated tags and quotes matched to the tags. It is a generalization of the sub-study and accompanies the discrete detailed themes.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Aggregation of different kinds of data in real time for different types of business. [system generated]

**Tags:** *credit risk | data aggregation | data credit | data data | data lineage | data management | different kinds | high performance | lines business | little bit | move data | open source | real time | small dollar | use case*

### Theme Quotes

* So basically we have real time use cases for example customer are our website and we need to real time so they fill in some forms.Some real like can be any type in real business scenario right?So it really depends on also previously depends on the use case priorities we first move and move the data used by some critical use use case for example our customers information and call logs you just mentioned claim information for example some other finance finance data you live data they have their own life they are based still actually not not yet.And second one I think about but I just mentioned the data lineage and most of more data management I think data lineage and also the like kind of automatic data quality checking because there are like emerging patterns or regardless of legacy problems.
* That is data lineage.That's not the case.So we had to look at it as well we had like several things were driving this one was that like we didn't want to go stand up a data center in all those different countries the time the you know the expense of that was was high.And I but I also you know rather than have the lock in if I you open source with a service backing behind it.
* For our applications performance in various lines of business as well that's our use case for Splunk.Yeah so our use case behind Kafka is it's a little bit different than what would have been Spark is the so we use that for for basically real time data streaming between our applications.So I'm also thinking about the real time credit checks that we need for small dollar like like basically instant approval on credit cards small dollar credit cards trying to do real time like KYC checks and fraud fraud data passage so like you know monitoring transactions in real time sending you a you know a text message to verify a transaction in real time that's that's all continuing to to grow I would say we're probably we are just moving from medium to large we're moving from tier one to tier bank tier one bank status.We have we have Informatica on the insurance side but not for real time data transfer that is more data aggregation side for merging data between various insurance lines of business.
* So I think that that's a use case right?But then there's an operational data store that is like near real time.I think that technology that you have because it's highly you know like high performance right.So I thought that would be a good use case.
* So we have time ingesting in real time.Those are my SQL is really what we use I believe so a little bit of MongoDB as well but my okay.So data room very rapidly every blood sample that we process generates about 800 gigabytes of so of data we also receive in the industry data sharing the half as a compliant fashion be identified we nobody cares about the identification of the data with the identify the data we share with pharmaceutical companies so they can use that to come up with medications for cancers quicker and then they share their data data with us as well to make our solution be more effective on all different cancers right.Yeah so the DVS stuff is like first licensed but you concerned about that a little bit?
* Or is it small business and large customers?And then finding a way to get to move to real time faster.Just give me some details that way to help you out a little bit please.VR Satish So let's let's talk a little bit more about ETL.
* And some people use SSD for very high performance use case like Bloomberg of the world they have this real time terminals running.I won't call high performance high capacity storage you're talking about 12 terabyte drives that time.So I said look I mean in addition to going into a business case so we said look I'm not going to give it away we're open open source it we're going to make a proprietary product with a certain price pricing.So that includes basically we sell the open source distribution that goes to use by data scientist analyst security admins and etc.

## Detailed Themes

Here are the discrete themes for the sub-study. Since we need enough unique tags to check the individual interactions in the sub-study for relevant text snippets and quotes you may find discrete themes without quotes. We do not strike them from the report because we believe it is important to see what the mediumroast.io detected for transparency purposes.

### Detailed Theme Identifier: 1

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** I know a lot of people who call me a bad guy but i think i really will be good at it. [system generated]

**Tags:** *are | bad | call | great | know | me | one | really | say | will*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* So it really depends on also previously depends on the use case priorities we first move and move the data used by some critical use use case for example our customers information and call logs you just mentioned claim information for example some other finance finance data you live data they have their own life they are based still actually not not yet.
* So from what I see so the train is really moving to cloud.
* So basically we have real time use cases for example customer are our website and we need to real time so they fill in some forms.
* Because different company had the choice that's in the financial service across different companies.

**Frequency:** *178*

##### 202201041932-Reverse CDN-Meditronics

* Well for us most of our data right now is in was in sequel service moving to H base and I'll probably move from there as well.
* That's what mostly a database does.
* So you know a lot of time thinking about you know integrating with databases storage snapshots application integration all kinds of stuff like that.
* So we had to look at it as well we had like several things were driving this one was that like we didn't want to go stand up a data center in all those different countries the time the you know the expense of that was was high.

**Frequency:** *178*

##### 202201060719-Reverse CDN-Guidepoint

* We are having to find ways to stream and move that data query data sources for publicly available information.
* And so a lot of those requirements became active by the different regulators in 2011 and 12.
* So I'm also thinking about the real time credit checks that we need for small dollar like like basically instant approval on credit cards small dollar credit cards trying to do real time like KYC checks and fraud fraud data passage so like you know monitoring transactions in real time sending you a you know a text message to verify a transaction in real time that's that's all continuing to to grow I would say we're probably we are just moving from medium to large we're moving from tier one to tier bank tier one bank status.
* And like another example is sometimes not not necessarily between applications which is what we use Splunk for which is applications traffic monitoring like basically the the go between traffic between various you know applications on our side while it's transparent to the customer we have a lot of systems in the background that are supporting different lines of business within the same application.

**Frequency:** *178*

##### 202110061800-Reverse CDN-GuidePoint

* So it starts with you know I mean let's say moving from for example a lot of you know shooting but we're moving from Caterpillar to other scenario right?
* Because workload may mean different things to different people.
* And process a lot of time you know sometimes you know once one other time was already kind of transformation.
* So I mean as the different and like I mentioned I mean I cater to the different verticals.

**Frequency:** *178*

##### 202110061100-Reverse CDN-GuidePoint

* So we have time ingesting in real time.
* Sorry we definitely use a database.
* So so what a lot of there's a lot of SQL for sure.
* We will do a lot I've been doing a lot of testing with it.

**Frequency:** *178*

##### 202110071330-Reverse CDN-GuidePoint

* What are the different stages different processes they do in a typical pipeline?
* But I'll give you an example of what I ran into Time Warner Cable I was the lead architect that helped them move from a batch orientation to a real time compute model.
* And that information you know customer information you know product detail all that stuff?
* That's why we had a lot of customers right now moving to more of a real time environment you know please update on the fly.

**Frequency:** *178*

##### 202111080900-Reverse CDN-GuidePoint

* So we we spend a lot of time billing.
* And then how do you without moving data?
* It's it's very different.
* And then on top moving from yarn to Kubernetes.

**Frequency:** *178*

### Detailed Theme Identifier: 2

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Dna example showing a lot of different information moving around a database. [system generated]

**Tags:** *database | different | dna | example | information | lot | move | moving | time*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* Because different company had the choice that's in the financial service across different companies.
* So we will say.
* So for the for the large company ITT the also some would say these are not.
* I think because because I was hearing probably most I mean more from the series in the financial service companies to get I think that can be probably not an issue in our app or like you know impact companies.

**Frequency:** *128*

##### 202201041932-Reverse CDN-Meditronics

* And I'd say that that's that's huge.
* Especially when there's you know some medical space or financial space whatever.
* Now I'm not saying you have to be compatible with that I'd say go for it.
* There is a bunch of them out there on I've been watching this for a long time.

**Frequency:** *128*

##### 202201060719-Reverse CDN-Guidepoint

* There's there's underwriting and credit risk.
* And every time those decisions or actions are made by customers it changes a little bit our our risk picture and our risk posture depending on what it is that somebody is doing.
* And some of the mortgage crisis items and other things that were popping up during that time I moved moves over into risk risk management compliance and then i They say it risk management at that point.
* We have we have Informatica on the insurance side but not for real time data transfer that is more data aggregation side for merging data between various insurance lines of business.

**Frequency:** *128*

##### 202110061800-Reverse CDN-GuidePoint

* So I would say at this time?
* Alright so then what you would need to do you know I mean because I'm a smaller companies you know if you want let's say if you want to focus on financial services you know what I'll do focus focus on like credit unions for example right?
* SUMMARY KEYWORDS data cloud moving run lambda application product storage hear pipeline aws financial services provide customers business problem high performance storage focus people petabytes Sure.
* Time you know not as much as America but still so yeah so it's hard to get into you know large company when you're new your company yourself?

**Frequency:** *128*

##### 202110061100-Reverse CDN-GuidePoint

* So we have time ingesting in real time.
* By the two companies.
* So essentially data is just the bread and butter of the company.
* I would say it's not I would say it's not as sort of a compliance requirements.

**Frequency:** *128*

##### 202110071330-Reverse CDN-GuidePoint

* We also have customers in financial.
* It's a it's a big problem for all the automotive companies right now.
* Or is it small business and large customers?
* I wrote your company name down.

**Frequency:** *128*

##### 202111080900-Reverse CDN-GuidePoint

* And it's hard time getting the solution to market because how do you get the customers?
* So how many customers are going to have that problem?
* So they had a hard time.
* I've seen revenue as obviously the challenges and I advise companies on that.

**Frequency:** *128*

### Detailed Theme Identifier: 3

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** A company says that customers are at risk from time to time due to poor credit. [system generated]

**Tags:** *business | companies | company | credit | customers | financial | insurance | risk | say | time*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* Everything is for me it's more like if like you mentioned if your storage really high performance right.
* So in this part it's really high high volumes.
* So we will say.
* So they do have live performance issues.

**Frequency:** *85*

##### 202201041932-Reverse CDN-Meditronics

* And I'd say that that's that's huge.
* So that definitely like it is not treated as like individual components.
* So we had to look at it as well we had like several things were driving this one was that like we didn't want to go stand up a data center in all those different countries the time the you know the expense of that was was high.
* Well and when you hear things or when you see kind of what's been what it does and like the New York Times example that I gave you know one of things I saw from the storage space my time of various ice was storage vendors were very eager to put interfaces on top of their spindles to be able to or their you know SSDs to be able to land bytes on them.

**Frequency:** *85*

##### 202201060719-Reverse CDN-Guidepoint

* let's say that again which one?
* For our applications performance in various lines of business as well that's our use case for Splunk.
* So we were looking at.
* We were using data dog.

**Frequency:** *85*

##### 202110061800-Reverse CDN-GuidePoint

* SUMMARY KEYWORDS data cloud moving run lambda application product storage hear pipeline aws financial services provide customers business problem high performance storage focus people petabytes Sure.
* I think that technology that you have because it's highly you know like high performance right.
* And that requires you know performance right.
* And you know from my team you know we were horizontal right?

**Frequency:** *85*

##### 202110061100-Reverse CDN-GuidePoint

* And so it's been kind of I would say storage has been definitely one area of challenge with smoke as well as the indexers Wait so that's sort of Performance is what we do is very important way.
* So I will say you know there's definitely like definitely looking at objects that are currently using object storage but we're looking at it.
* So we are basically you know the storage for slug with we returned like so we have actually one petabyte right now or roughly one petabyte of storage?
* That's that's definitely appealing.

**Frequency:** *85*

##### 202110071330-Reverse CDN-GuidePoint

* Customers struggle with what to put on faster drives versus what not to put on faster drive they put all that crap out there.
* old data drives.
* When I say fast it can mean for example you know running functions inside storage as an exam.
* Yeah but in fact I may be right when I say we have one of the world's largest data analytics running on our system.

**Frequency:** *85*

##### 202111080900-Reverse CDN-GuidePoint

* I won't call high performance high capacity storage you're talking about 12 terabyte drives that time.
* Yeah definitely.
* Okay so I mean be sigley You have a high performance storage you have.
* Definitely we at Scara we are primarily doing very high capacity storage.

**Frequency:** *85*

### Detailed Theme Identifier: 4

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** If i could put a ssd into a drive i would definitely put it in a storage. [system generated]

**Tags:** *definitely | drives | high | one | performance | put | say | ssd | storage | were*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* So it's either hybrid or a private cloud.
* So this year either Microsoft or Google will collaborate with either do the lights either hybrid or cloud our private cloud?
* And I use Cloud most of the time it's not like public cloud.
* So especially based on cloud so we don't really work with one provider like like you is probably like the provider the storage class.

**Frequency:** *67*

##### 202201041932-Reverse CDN-Meditronics

* SUMMARY KEYWORDS data kafka cloud run storage application architecture prem steve database transformation postgres terms metadata s3 aws amazon lambda functions patient device So let me go.
* That is cloud agnostic.
* So we are now in cloud to?
* And you can use you can use essentially you can develop your application and with the primitives that I expect in the cloud so stuff like serverless.

**Frequency:** *67*

##### 202201060719-Reverse CDN-Guidepoint

* Yeah we are trying to move more and more to cloud now.
* And so we need to expand the resiliency and security around our cloud architecture over the next 12 to 18 months.
* Now granted through the VPN we were able to use VPN that still drew on processing power from the on prem strictly on prem you know architecture that we had we those that that bandwidth and that that data flow we had to move more of that to some distributed cloud capabilities.
* There is some regulatory requirements in terms of backup systems and business continuity right now where and this is getting a little bit outside my wheelhouse on the total cloud architecture.

**Frequency:** *67*

##### 202110061800-Reverse CDN-GuidePoint

* Or you know cloud?
* And again you know if you look at it from you know cloud point of view right I mean cloud majority of the cloud you know is today at least is you know I mean data warehousing cloud like redshift.
* So I mean so the cloud either they will become your customer or those providers.
* So actually you know if we do go to cloud we use a lot of encryption as much as we can.

**Frequency:** *67*

##### 202110061100-Reverse CDN-GuidePoint

* You know in the England England entity we had to actually do a private kind of a colocation kind of a private cloud if you will because of the fact that we could not use AWS.
* So what are the situations like that even if you're you know the the side of AWS is going to be you know cloud provider of choice you may run into a country where you cannot use AWS and what solution are you offering?
* And then from there disrupt the here of the cloud to AWS.
* This is a this is validated architecture right?

**Frequency:** *67*

##### 202110071330-Reverse CDN-GuidePoint

* We stood up a private cloud.
* Well there has to be the interconnect between GMs environment and your cloud provider.
* Given the mandate that you have to move to the cloud?
* But the world in general is trying to move to the cloud.

**Frequency:** *67*

##### 202111080900-Reverse CDN-GuidePoint

* And once they do that or maybe it's a cloud that layer can be cloud or on prem did admin there.
* So it's kind of heads against the cloud guys to not just take it and not contribute back.
* Otherwise today you have to go to AWS s3 you have to go to Teradata you have to go to you know some open old RDBMS move the data undocumented engineers had to move the data out to a central location.
* And all the use cases?

**Frequency:** *67*

### Detailed Theme Identifier: 5

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Microsoft uses aws to move its cloud to the private sector. [system generated]

**Tags:** *architecture | aws | either | go cloud | microsoft | move cloud | private | provider | use*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* And with a B test.
* One is more for traditional machine learning ones.
* And also with a little bit of machine learning flows with ml ops with AV test functionality I think that's really the last mile to succeed for any data driven project right.
* The you have a training environment where you train your models in 3d models or how do you do?

**Frequency:** *31*

##### 202201041932-Reverse CDN-Meditronics

* So if somebody is making a decision like this is really important to them and you should you should test this right.
* And then basically like we are like looking into this focus area do you have any like suggestions or reference architecture mean again?
* And one things I've seen is I've done you know I've been learning myself and kind of looked at what other people have been doing as well is you know Kafka is just become massively important.
* And you know so in that spirit coffee just seems to be a really natural play in that area to say Okay what what could be done to make it so that you know if I have my application and I've got you know such a such a dependency on Kafka is plays such a key part of my architecture you know that that just speaks to opportunity to be able to go and innovate in the storage space to say how do I make that faster?

**Frequency:** *31*

##### 202201060719-Reverse CDN-Guidepoint

* What what you know we kind of need to plan for that starting in the next about 18 months.
* And so that requires kind of those late live data feeds that like Apache Kafka has you know has perfected Yeah so we have a we have a pretty large relationship at the MDM level with IBM and Oracle.
* So there was like a 10% shock there between basically March of two of 20.
* Now granted through the VPN we were able to use VPN that still drew on processing power from the on prem strictly on prem you know architecture that we had we those that that bandwidth and that that data flow we had to move more of that to some distributed cloud capabilities.

**Frequency:** *31*

##### 202110061800-Reverse CDN-GuidePoint

* So I mentioned because I think artificial intelligence and machine learning is power hungry?
* So artificial intelligence machine learning would be a good use case for what you hear.
* So should you just mentioned inferencing now so what are you doing is actually you're moving the complexity to inferencing them right.
* So but but like I said I would pick an area and you asked me about Wells Fargo?

**Frequency:** *31*

##### 202110061100-Reverse CDN-GuidePoint

* We're you know if you have a Data Challenge right they're always trying to figure out how do we have machine machine learning AI in a fast storage that can quickly you know index and quickly figure out you know what subset of the data is really relevant to this specific patient right as opposed to just scanning it across the whole you know storage and having to keep everything on expensive storage.
* Can be natively supported because they were using HVR for computer simulate CR.
* Like that we will fall into another area where we have to call you guys for support to open tickets right.
* And so it's been kind of I would say storage has been definitely one area of challenge with smoke as well as the indexers Wait so that's sort of Performance is what we do is very important way.

**Frequency:** *31*

##### 202110071330-Reverse CDN-GuidePoint

* I have responsibility for the computer aided test area.
* And I had responsibility for AI in the Computer Aided test area along with big data problems as well.
* And what we were trying to do is figure out the issues early on in the Computer Aided test area doing testing the design specifications before a vehicle even got produced.
* Well for inferencing purposes yes.

**Frequency:** *31*

##### 202111080900-Reverse CDN-GuidePoint

* Are you talking about inference of AI inference?
* Again these are these are things I'm learning.
* Just like you know I will say I think Iceland does it sometimes I think they have different models.
* Okay that's the fourth part I think inference.

**Frequency:** *31*

### Detailed Theme Identifier: 6

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** A computer aided test of inference in the area of interest. [system generated]

**Tags:** *computer aided test | inference | inferencing | machine learning | models | test area*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* Was the problems and problems.
* So there are a lot of legacy issues.
* So they do have live performance issues.
* The data moms I mean it's not huge like big tech companies.

**Frequency:** *30*

##### 202201041932-Reverse CDN-Meditronics

* And there's a lot of problems in that space to be solved for sure.
* That's really really big.
* They wanted to essentially is down who said there's no problem can be just as they can't be solved by another level of abstraction.
* They had a big announcement event.

**Frequency:** *30*

##### 202201060719-Reverse CDN-Guidepoint

* For retention we have issues.
* We're having issues syncing that up.
* SUMMARY KEYWORDS data kafka transactions years moving requirements prem applications product business environment query bit real customer storage people batched talk bank no problem.
* And so we were having accuracy issues in terms of posting those transactions based on time that you know timezone requirements of what was really qualifying as end of day.

**Frequency:** *30*

##### 202110061800-Reverse CDN-GuidePoint

* You know and then you're looking for a problem that could be solved given that solution.
* issue or challenge.
* We'll solve the problem.
* But the problem is is process.

**Frequency:** *30*

##### 202110061100-Reverse CDN-GuidePoint

* So that's one challenge that we have in DPFs.
* 300 gigs so it's a big lever pushing a lot to it.
* So now we got IBM ndd in an environment which is a bit of a challenge so and then.
* I would say yes there's there's times we still experience you know when we're When our volumes increase we do experience some storage performance issues we've tried to add more compute to it.

**Frequency:** *30*

##### 202110071330-Reverse CDN-GuidePoint

* It's a big problem.
* Two big issues?
* It's a big problem to this day.
* So you're saying the loading is problem solved by an issue?

**Frequency:** *30*

##### 202111080900-Reverse CDN-GuidePoint

* So it's a challenge.
* It is a big problem.
* I mean when I came in one of the big product was Dr was a big problem.
* I mean density shall the way I'll frame the problems they basically it's a cost problem not NVMe material procuring the solution to that?

**Frequency:** *30*

### Detailed Theme Identifier: 7

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** A solution to a big problem that has been solved. [system generated]

**Tags:** *big problem | challenge | fix | issues | problems | solution | solved*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* So basically we have real time use cases for example customer are our website and we need to real time so they fill in some forms.
* So I can be real time can be bypassed.
* But we don't have I mean we don't have that time critical and large amount of data that need to use Splunk yet then at least kept up class Spark streaming.
* I mean with with the smart AI real time and Splunk.

**Frequency:** *29*

##### 202201041932-Reverse CDN-Meditronics

* And you can use you can use essentially you can develop your application and with the primitives that I expect in the cloud so stuff like serverless.
* So right now we're using Apache Atlas for some of that stuff.
* So the for the larger data that we have today like Kafka?
* It is not the Kafka topic Kafka traffic not handle 70 terabytes.

**Frequency:** *29*

##### 202201060719-Reverse CDN-Guidepoint

* That is real time data streaming that we use Kafka for managing that Traffic where you basically have to have real or near real time data.
* Yeah so our use case behind Kafka is it's a little bit different than what would have been Spark is the so we use that for for basically real time data streaming between our applications.
* And then what we have data we have well we have data dog we have a few others but we use Kafka in our environment which is also an Apache product.
* So the data streaming is is what we use we try to keep it as real time as possible between basically our our customer facing mobile application and our mobile banking like PC side we have a third version and all of those get aggregated.

**Frequency:** *29*

##### 202110061800-Reverse CDN-GuidePoint

* But then there's an operational data store that is like near real time.
* So we use you know all three Carnegie mentioned we use data bricks we use you know redshift we use BigQuery.
* And then you know time to market.
* And process a lot of time you know sometimes you know once one other time was already kind of transformation.

**Frequency:** *29*

##### 202110061100-Reverse CDN-GuidePoint

* So we have time ingesting in real time.
* We do have to do with a transport loop to transforming data all the time as we receive data.
* So data room very rapidly every blood sample that we process generates about 800 gigabytes of so of data we also receive in the industry data sharing the half as a compliant fashion be identified we nobody cares about the identification of the data with the identify the data we share with pharmaceutical companies so they can use that to come up with medications for cancers quicker and then they share their data data with us as well to make our solution be more effective on all different cancers right.
* Sorry we definitely use a database.

**Frequency:** *29*

##### 202110071330-Reverse CDN-GuidePoint

* you know this this real time thing is I mean a lot of people are using Apache Kafka and all that you know messaging and all that stuff trying to get into real time.
* Don't turn me to Apache Kafka is even going to be viable or not.
* And then finding a way to get to move to real time faster.
* And so what I ended up doing was I brought in a few vendors that had some real time capabilities.

**Frequency:** *29*

##### 202111080900-Reverse CDN-GuidePoint

* And some people use SSD for very high performance use case like Bloomberg of the world they have this real time terminals running.
* But what you're saying is more like a tearing real time tearing under the hood?
* And all the use cases?
* Data Plane is a pivot from a quote unquote typical Apache license which is very open to AGPL.

**Frequency:** *29*

### Detailed Theme Identifier: 8

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Using kafka to stream data in real time [system generated]

**Tags:** *apache | kafka | real time data | streaming | use*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* And the table can eat already like 3040 years.
* So we previously have on premise one but actually it has been several years.
* So I think that's probably one.
* So if you are not really probably IT guy like from like with 20 years if yours is really hard for someone to understand all this or has a holistic view about the data lineage or how the data model connects between them.

**Frequency:** *28*

##### 202201041932-Reverse CDN-Meditronics

* But there's a company called muta that I started looking at probably five years ago they came out of CA they have some data masking capabilities it's really quite interesting.
* And I'm guessing that our cluster size is probably about eight or 10.
* Even if you put 10 million boxes.
* Then you need is probably where they're going to go.

**Frequency:** *28*

##### 202201060719-Reverse CDN-Guidepoint

* We by regulation have to retain that data for up to 10 years.
* Yeah and you know it's interesting one thing that has not been published yet is and is the information requirements under a retention requirements under GDPR.
* And then we'll we'll do gradual migration probably I would say over the next two to three years get to 5060.
* And we have to be able to access that for 10 years for regulatory queries and reasons.

**Frequency:** *28*

##### 202110061800-Reverse CDN-GuidePoint

* But Informatica has been around for many years.
* So so I've been in the industry for about 32 years.
* And you know from my team you know we were horizontal right?
* And then I'm getting a neighbor who find 10 code I mean you should do this.

**Frequency:** *28*

##### 202110061100-Reverse CDN-GuidePoint

* So that's not really relevant to us you know GDPR like everybody else.
* I mean we're talking you know this magnitude of I would say somewhere between five to 10 petabytes of data.
* And so DNA sequences so like when I joined forces you know we were very good for pretty much lung cancer and breast cancer was second but everything else we were not very good at.
* I would say it's not I would say it's not as sort of a compliance requirements.

**Frequency:** *28*

##### 202110071330-Reverse CDN-GuidePoint

* But I spent quite a few years there with them.
* So three and a half models are probably the best for something like that.
* I want to see here the requirements.
* And then I was at Walmart for four years as the lead architect with their big data problem they had.

**Frequency:** *28*

##### 202111080900-Reverse CDN-GuidePoint

* And we are burning the probably an year and a half I can go and look at a number more but that's roughly lasting three years or five years.
* It's is it going to become big problem in 10 years?
* So I think is it's I'm almost seeing you in my shoes like three years ago.
* And we need to guarantee the product for five years.

**Frequency:** *28*

### Detailed Theme Identifier: 9

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** The gdpr requirements for retention were three years. [system generated]

**Tags:** *10 years | gdpr | probably | retention requirements | three years | were*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* I think overall company wise they say that for data users I don't see a huge cat as we expect okay.
* So basically we have real time use cases for example customer are our website and we need to real time so they fill in some forms.
* So I can be real time can be bypassed.
* And I use Cloud most of the time it's not like public cloud.

**Frequency:** *26*

##### 202201041932-Reverse CDN-Meditronics

* And we was really you know highly extremely impatient.
* And that take the network it?
* So you know I take a look at that.
* All right and then running away?

**Frequency:** *26*

##### 202201060719-Reverse CDN-Guidepoint

* And I think that's going to be a I think it's gonna be a very slow shift.
* And we are running into that decision.
* That was a 10% shift there that was prompted by COVID.
* You have to be able to process those requests in real time or for instance on a life insurance annuity if they want to take a distribution and And they want to link that distribution to a bank account.

**Frequency:** *26*

##### 202110061800-Reverse CDN-GuidePoint

* So no movement is extremely costly.
* So because you have to take you for your words right?
* But two things are needs are changing quickly.
* So the challenge that you're running into masscue how old is the company?

**Frequency:** *26*

##### 202110061100-Reverse CDN-GuidePoint

* So we have time ingesting in real time.
* Are you also like extremely likely to get from IBM as well?
* So so that that was sort of anything can expect to be like had I would?
* And especially we are running sort of workloads on AWS.

**Frequency:** *26*

##### 202110071330-Reverse CDN-GuidePoint

* On we moved everything to Pivotal Pivotal platform.
* Take care.
* Remember pivotal?
* But ab initio is extremely fast.

**Frequency:** *26*

##### 202111080900-Reverse CDN-GuidePoint

* So their means pieces was Spark is slow which is slow.
* Because the moment you expect them it's this is going to take forever.
* How do you run a query so that you can get the time inside quickly.
* If you don't expect them to make the change.

**Frequency:** *26*

### Detailed Theme Identifier: 10

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** A man is slowing down to watch a pivotal moment of his running, which is extremely fast. [system generated]

**Tags:** *expect | extremely | pivotal | quickly | running | shift | slow | take | time*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* GPU need.
* So you need to you can't really just do POC you need to deploy code that is more efficient.
* So that's why we need to reinvent.
* So you need to work closely with IT team.

**Frequency:** *25*

##### 202201041932-Reverse CDN-Meditronics

* And they need the help to be able to get there.
* Stuff like lambda functions or some kind of you know Functions as a Service give me the scaling give me give me allow me to work with those primitives that I'm used to and then give me the benefit of either performance or you know a cost or something like that.
* Then you need is probably where they're going to go.
* Yeah I mean I think you need to do more than ingestion.

**Frequency:** *25*

##### 202201060719-Reverse CDN-Guidepoint

* It's throttled by regulation and therefore in some cases you need that batch processing or you need an Informatica to hold that for you.
* And we were also having issues in terms of handling what I would call your in some cases you you would need to be able to what was it called there was a terminology where you need to part of the account or that product would need to be processed real time and part of it might need to be aggregated at the end of the day.
* Splunk is the one that would help us monitor that that action those actions and and so the security part is what we would monitor access management patterns.
* How long do we need to keep that?

**Frequency:** *25*

##### 202110061800-Reverse CDN-GuidePoint

* Yeah interact and kind of help you indirectly I can't help it.
* How can you You help them?
* I can give you all the data but it's huge.
* Well we use lambda you know for obviously you know when we have new use cases instead of you know I mean spending a lot of money trying to figure out what you want to learn but initially we use it actually a lot of time is to you know prove a concept actually you know because lambda is fine but it won't give us the performance that we really need.

**Frequency:** *25*

##### 202110061100-Reverse CDN-GuidePoint

* Give an example like you know we started a entity in London England actually.
* So when we we need to make sure that the data has no PHA in it right?
* And so like I'll give an example like garden health you know they do cancer pre screening by blood draw.
* Because there's two places you know so I guess on the AWS side I guess lambda is kind of addressing those pain points but our premise will be kind of like really just if a storage can actually process that internally it really would help drastically.

**Frequency:** *25*

##### 202110071330-Reverse CDN-GuidePoint

* It's relational.
* This is not relational?
* This is not relational.
* Didn't need one.

**Frequency:** *25*

##### 202111080900-Reverse CDN-GuidePoint

* Yeah you need to learn from those experiments.
* Right So then I'll give you one example right there on my experiments equal is successful or failed whatever.
* So I can give you my feedback.
* So I think I would say learn from some of the guys like what big stream is doing.

**Frequency:** *25*

### Detailed Theme Identifier: 11

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Cant help but learn the kinks of relational relationships. [system generated]

**Tags:** *cant | give | help | kinks interact | need learn experiments | relational*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* So we basically use had to and also spark all this kind of stuff.
* Basically.
* But we don't have I mean we don't have that time critical and large amount of data that need to use Splunk yet then at least kept up class Spark streaming.
* And then this on top of this we have three layers.

**Frequency:** *19*

##### 202201041932-Reverse CDN-Meditronics

* You can also look at you know stuff like Spark and that as well.
* But this wildly dominant Spark Kafka.
* So right now we're using Apache Atlas for some of that stuff.
* But being able to have you know they have a really good story for you know the tools of you know the tools that are so dominant like Spark is another one.

**Frequency:** *19*

##### 202201060719-Reverse CDN-Guidepoint

* And then you mentioned Apache Spark we trialed that but we did not implement it which we can talk about but I'm at the enterprise level.
* And then did we change our pricing based on that and treat the customer fairly right?
* Yeah so our use case behind Kafka is it's a little bit different than what would have been Spark is the so we use that for for basically real time data streaming between our applications.
* Yeah so we we went and did a trial on Spark and hold on pulling up some notes on this.

**Frequency:** *19*

##### 202110061800-Reverse CDN-GuidePoint

* So initially we would you know use lambda then we use Python and go you know whatever works for us spark.
* Go to the conferences right.
* And maybe go and listen in.
* You know I guess of course you know like anything else you know you have encryption and all kinds of stuff.

**Frequency:** *19*

##### 202110061100-Reverse CDN-GuidePoint

* So we have the on prem version of Spark right so we put on prem stuff and then we use VMware to run it on top of VMware basically.
* Right and then some cost will go down to it.
* So that's that's basically the the paint we're having.
* Go into support if it's not supportive but it will be an issue right.

**Frequency:** *19*

##### 202110071330-Reverse CDN-GuidePoint

* This is rules change you got to go back into the ETL code and change it.
* stuff.
* And it's really important if you guys can tackle this problem inside your environment would be how familiar are you with Spark?
* So you could go to Fry's and buy the stuff.

**Frequency:** *19*

##### 202111080900-Reverse CDN-GuidePoint

* So then go up upstream then like in all the cases the commonality I'm seeing is you're trying to make small changes I would say go make a change in Spark and provide a spark.
* And if you can say Look I'm going to change the spark for you.
* And then on top we used to have spark Apache Spark which is what Databricks does often similar to confluent Kafka hive which is basically another very performant high scale sequel analytics and theater stuff on top which is basically mostly on workflow management.
* I mean spark in memory but Spark is slow.

**Frequency:** *19*

### Detailed Theme Identifier: 12

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** A spark goes around the top of an apache to change the shape of the fire. [system generated]

**Tags:** *apache spark | basically | change spark | changes | go change | stuff | top*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* GPU need.
* So we and now the trend is everything must be better VR to API or micro services.
* So basically so we have we have two very powerful servers over basically.
* So GPU sed are using that for us?

**Frequency:** *19*

##### 202201041932-Reverse CDN-Meditronics

* And I'm guessing right now the cluster size.
* Because what I'm thinking about is just what are you competing against in environment.
* I mean I can tell you right now when we did lift and shift it was 70 roughly 70 virtual machines.
* And I'm guessing that our cluster size is probably about eight or 10.

**Frequency:** *19*

##### 202201060719-Reverse CDN-Guidepoint

* And so in that instance you you have to aggregate and you batch process or cluster process those transactions right.
* And I would tell you the more aged it gets the more that we push it from environment to environment.
* And so I cut across all lines of business and then I also validate a lot of our infrastructure and data privacy and workflows against you know FF IEC like IT handbook requirements.
* And that that is that is an environment that would be you know basically cue seed and cleanse before it lands and the retention environment.

**Frequency:** *19*

##### 202110061800-Reverse CDN-GuidePoint

* So yeah so so what happens?
* So what happens is you know you kind of create data right which is sitting somewhere.
* So what happens is each of these companies they have their own fiefdom right there so they have their own way of doing business.
* We're back.

**Frequency:** *19*

##### 202110061100-Reverse CDN-GuidePoint

* Question servers there's so what we do that on premise happens not so much computing infrastructure because all physical is virtual.
* We have a lot of within the VMware environment also there's a lot of like some of the ETL in the redo that happens in the VMware environment Splunk and VMware.
* So that's we have dedicated 256 servers in a cluster.
* That's we are VMware shop the functionalist perspective so that when a DNA sequence that happens that's all in a hydraulic computing physical infrastructure or metal once once that workload is just the reports that are generated we don't store that in a high performance computing environments that gets pushed to the VMware infrastructure on prem.

**Frequency:** *19*

##### 202110071330-Reverse CDN-GuidePoint

* Is a cluster cluster.
* VR Satish Yes.
* So your environment?
* And then the transformation happens from the data warehouse.

**Frequency:** *19*

##### 202111080900-Reverse CDN-GuidePoint

* So Somebody we do go by VR AR VR se like how do That is okay.
* I'm talking about sort of use case we create about one terabyte data in our servers roughly a petabyte in a month.
* This industry is as you go into the edge computing you are completely different paradigm where the box is somewhere out there you cannot manage it.
* But we are completely in the pre package AI all the way from Quality Inspection honors I mean what's your kind of do right edge computing with in situ processing inference from storage.

**Frequency:** *19*

### Detailed Theme Identifier: 13

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Vmware is a virtual machine that runs on a cluster of servers. [system generated]

**Tags:** *cluster | computing | gpu | happens | infrastructure | servers | virtual | vmware environment | vr*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* For ETL part actually there are different data layer.
* Normally we have three layers and then on the second layer is more than ETL.
* And then on top of that there is a vacation layer are more like data use usage layer from the third layer actually our team my team who will be able to do that because after you have ETL is America.
* I mean tools are platform I use right.

**Frequency:** *18*

##### 202201041932-Reverse CDN-Meditronics

* But operating on that and like any tools like snowflake or any ETL tools it becomes very expensive.
* But being able to have you know they have a really good story for you know the tools of you know the tools that are so dominant like Spark is another one.
* that's gonna depend on a lot of things a lot of it will be dependent upon more of the legal agreements that are in place.
* And you know maybe we'll just do a lot of stuff to it.

**Frequency:** *18*

##### 202201060719-Reverse CDN-Guidepoint

* So we were looking at.
* We were using data dog.
* And this this all goes back to customer experience CX capabilities you know the tolerance for wait periods is almost nothing right?
* Now granted through the VPN we were able to use VPN that still drew on processing power from the on prem strictly on prem you know architecture that we had we those that that bandwidth and that that data flow we had to move more of that to some distributed cloud capabilities.

**Frequency:** *18*

##### 202110061800-Reverse CDN-GuidePoint

* And today you know we used to be ETL.
* So we have to understand you know I mean what's our current ETL?
* Okay no matter what you do today the biggest cost is ETL.
* So we do that and then you know we have ETL so the You know and when we do that you know they're the Pebble Time right?

**Frequency:** *18*

##### 202110061100-Reverse CDN-GuidePoint

* So ETL you want at all?
* We have a lot of within the VMware environment also there's a lot of like some of the ETL in the redo that happens in the VMware environment Splunk and VMware.
* Because right now we're like so we're at a point that you don't see any more growth in our ETL jobs and we're gonna be we're gonna have to do something because it's not going to be able to complete.
* So so what a lot of there's a lot of SQL for sure.

**Frequency:** *18*

##### 202110071330-Reverse CDN-GuidePoint

* ETL tools?
* So how is ETL window?
* You know there is a lot of movement for from ETL.
* But most of the ETL capabilities that I've seen even now a lot of the business rules are inside the coded.

**Frequency:** *18*

##### 202111080900-Reverse CDN-GuidePoint

* They have a lot of problem compaction and convection meaning like you end up a lot of small files a lot of and then following with that now you have to maintain that file you end up with a lot of metadata on the file.
* But I haven't seen a lot.
* No that makes a lot of sense.
* So we we spend a lot of time billing.

**Frequency:** *18*

### Detailed Theme Identifier: 14

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Etl lets you use a lot of the tools you already have. [system generated]

**Tags:** *etl capabilities | etl etl | etl tools | lets | lot | rules | were*

#### Theme Quotes by Interaction

##### 202110191700-Reverse CDN-GuidePoint

* So for example most of like courses me t plus t plus two.
* And this is more cost effective.
* So they do have live performance issues.
* So that definitely will be an add on value.

**Frequency:** *17*

##### 202201041932-Reverse CDN-Meditronics

* Yeah so that's our costs are non trivial.
* Stuff like lambda functions or some kind of you know Functions as a Service give me the scaling give me give me allow me to work with those primitives that I'm used to and then give me the benefit of either performance or you know a cost or something like that.
* So that definitely like it is not treated as like individual components.
* Because after like basically like it helps in E ELT in replay and other things as a storage for a couple of weeks or one week or something like that.

**Frequency:** *17*

##### 202201060719-Reverse CDN-Guidepoint

* But we are having lower operating costs.
* For our applications performance in various lines of business as well that's our use case for Splunk.
* So there was like a 10% shock there between basically March of two of 20.
* Now granted through the VPN we were able to use VPN that still drew on processing power from the on prem strictly on prem you know architecture that we had we those that that bandwidth and that that data flow we had to move more of that to some distributed cloud capabilities.

**Frequency:** *17*

##### 202110061800-Reverse CDN-GuidePoint

* But their costs you know for performance and all that.
* So Madeira 2 million is pretty good you know a simple product and you know good cost and performance.
* Now as ELT right.
* And that requires you know performance right.

**Frequency:** *17*

##### 202110061100-Reverse CDN-GuidePoint

* And so but yeah that's mainly the two is one and everything you mentioned plus plus.
* So yeah that's that's definitely you know something that's important and depending on you know the cost and then also the performance you know and you know and also I'm assuming it would have a lower operational overhead flow right?
* Right and then some cost will go down to it.
* That's that's definitely appealing.

**Frequency:** *17*

##### 202110071330-Reverse CDN-GuidePoint

* And put a hybrid model in place that addresses you know cost benefit.
* The traditional to ELT?
* But they'll cut down the cost enormously.
* Okay you're not you're not really concerned about performance just get it out of there.

**Frequency:** *17*

##### 202111080900-Reverse CDN-GuidePoint

* Yeah definitely.
* I mean if it's possible right again that's something you'd look at cost benefit.
* And then internally will do all the work but you get a cost benefit.
* I mean we saw about teknicks benefit of query performance by being the middle layer where it's the view.

**Frequency:** *17*

### Detailed Theme Identifier: 15

Each theme, if quotes/snippets are detected, has appropriate data pulled from interactions listed below. Additionally, the frequency of each theme is reported to provide the user with a sense of how strong the theme is relative to the other themes in the sub-study.

**Definition:** Description generated by the Mediumroast Caffeine machine intelligence service, please replace the definition with your own.

**Fortune:** Elts benefit from a lower cost of entry and performance [system generated]

**Tags:** *cost benefit | costs | definitely | elt | lower | madeira | performance | plus*

#### Theme Quotes by Interaction

References

## 202101310000-Reverse CDN-Cloudera

Date: 2021-01-31 00:00 | Sub-Study Identifier: 1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ FORM -K \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (Mark One) ☒ ANNUAL REPORT PURSUANT TO SECTION OR (d) OF THE SECURITIES EXCHANGE ACT OF For the fiscal year ended January or ☐ TRANSITION REPORT PURSUANT TO SECTION OR (d) OF THE SECURITIES EXCHANGE ACT OF For the transition period from\_\_\_\_\_\_\_\_to\_\_\_\_\_\_\_\_ Commission File Number: 001- CLOUDERA INC. ( Exact name of registrant as specified in its charter) Delaware -2 (State or other jurisdiction ...

Interaction Resource: [202101310000-Reverse CDN-Cloudera](http://mr-02:9000/pavilion/202101310000-AMER-US-CALIFORNIA-PALO ALTO-ICT-Reverse CDN-Cloudera-SEC.pdf)

## 202106160000-Reverse CDN-Confluent

Date: 2021-06-16 00:00 | Sub-Study Identifier: 1

As filed with the Securities and Exchange Commission on June . We have two classes of authorized common Class A common stock and Class B common stock. Each share of Class A common stock is entitled to one vote. Each share of Class B common stock is entitled to votes and is convertible at any time into one share of Class A common stock. Investing in our Class A common stock involves risks. Any representation to the contrary is a criminal offense. million for the years ended December and and $. Th...

Interaction Resource: [202106160000-Reverse CDN-Confluent](http://mr-02:9000/pavilion/202106160000-AMER-US-CALIFORNIA-MOUNTAIN VIEW-ICT-Reverse CDN-Confluent-SEC.pdf)

## 201802010000-Reverse CDN-Striim

Date: 2018-02-01 00:00 | Sub-Study Identifier: 2

Get More Value from Apache Kafka with Enterprise-Grade Streaming Integration and SQL-Based Stream Processing Striim White Paper February 2018 Get More Value from Apache Kafka Table of Contents Executive Summary ................................................ 1 Make Kafka Easy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1 Ingestion into Kafka . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ....

Interaction Resource: [201802010000-Reverse CDN-Striim](http://mr-02:9000/pavilion/201802010000-AMER-US-CALIFORNIA-PALO ALTO-ICT-Reverse CDN-Striim-Whitepaper.pdf)

## 202007010000-Reverse CDN-Seagate

Date: 2020-07-01 00:00 | Sub-Study Identifier: 2

Seagate authored the Kingdom France Germany Russia following report to share its forecasts Australia Japan India South Korea and opinions based on the survey Taiwan and China. RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—FROM EDGE TO CLOUD | 3 | Summary of Findings Today’s unprecedented growth to businesses goes untapped. is DataOps—the discipline Only 32% of data available to connecting data creators with enterprises is put to work. RETHINK DATA: PUT MORE OF YOUR BUSINESS DATA TO WORK—F...

Interaction Resource: [202007010000-Reverse CDN-Seagate](http://mr-02:9000/pavilion/202007010000-AMER-US-CALIFORNIA-CUPERTINO-ICT-Reverse CDN-Seagate-Whitepaper.pdf)

## 202104140000-Reverse CDN-Cloudera

Date: 2021-04-14 00:00 | Sub-Study Identifier: 2

SOLUTION BRIEF DATA-IN-MOTION PHILOSOPHY A Blueprint for Enterprise-wide Streaming Data Architecture Don’t Let Your End-to-End Streaming Architecture Fall Short Real Life Data-in-Motion It’s not enough to have the best messaging solution at the heart of your end-to-end streaming Below describes how the data-in-motion architecture. successfully modernized their messaging architecture to support a new line of This solution brief describes Cloudera’s data-in-motion philosophy and is meant as a blue...

Interaction Resource: [202104140000-Reverse CDN-Cloudera](http://mr-02:9000/pavilion/202104140000-AMER-US-CALIFORNIA-PALO ALTO-ICT-Reverse CDN-Cloudera-Whitepaper.pdf)

## 202110061100-Reverse CDN-GuidePoint

Date: 2021-10-06 11:00 | Sub-Study Identifier: default

SUMMARY KEYWORDS data prem storage aws cloud pipelines run lambda functions petabytes blood sample performance nvme hardware platform software functions servers warehousing file workloads Well you know the list of challenges that you know I can just think of you know a lot of times I mean you know AWS is great but sometimes you really want to do multi multi cloud right or you may have a hybrid environment. I think I'm beginning to think that may I should tackle that issue. And so like I'll give ...

Interaction Resource: [202110061100-Reverse CDN-GuidePoint](http://mr-02:9000/pavilion/202110061100-AMER-US-CALIFORNIA-SANTA CLARA-Expert Networking Services-Reverse CDN-GuidePoint-Transcript.pdf)

## 202110061800-Reverse CDN-GuidePoint

Date: 2021-10-06 18:00 | Sub-Study Identifier: default

SUMMARY KEYWORDS data cloud moving run lambda application product storage hear pipeline aws financial services provide customers business problem high performance storage focus people petabytes Sure. So you know I've started you know dealt with the main trends and sparks system and all that in the next hour. And everybody's doing using the cloud right? I think that technology that you have because it's highly you know like high performance right. So I mentioned because I think artificial intelli...

Interaction Resource: [202110061800-Reverse CDN-GuidePoint](http://mr-02:9000/pavilion/202110061800-AMER-US-CALIFORNIA-SANTA CLARA-Expert Networking Services-Reverse CDN-GuidePoint-Transcript.pdf)

## 202110071330-Reverse CDN-GuidePoint

Date: 2021-10-07 13:30 | Sub-Study Identifier: default

UMMARY KEYWORDS data storage customers file run informatica system people business database big environment move models important transformation etl fast gm write Hi What are we on the call MCA? And then he left HP and he went to work as CIO for General Motors. Yeah but in fact I may be right when I say we have one of the world's largest data analytics running on our system. So yeah So we do have some really large customers who really are dealing with a lot of data right? Be temporary storage be...

Interaction Resource: [202110071330-Reverse CDN-GuidePoint](http://mr-02:9000/pavilion/202110071330-AMER-US-CALIFORNIA-SANTA CLARA-Expert Networking Services-Reverse CDN-GuidePoint-Transcript.pdf)

## 202110191700-Reverse CDN-GuidePoint

Date: 2021-10-19 17:00 | Sub-Study Identifier: default

SUMMARY KEYWORDS data storage cloud customer layer model platform server applications gpu problems file power bi real move deploy system api transform pipelines Yeah so I think so just any recap. But the major actually computer science and my bio I mean working experience our background is more on a plea Amin application building applications in the financial service. So from what I see so the train is really moving to cloud. Because I know if I say that data savvy on the fire so first is the in...

Interaction Resource: [202110191700-Reverse CDN-GuidePoint](http://mr-02:9000/pavilion/202110191700-AMER-US-CALIFORNIA-SANTA CLARA-Expert Networking Services-Reverse CDN-GuidePoint-Transcript.pdf)

## 202111080900-Reverse CDN-GuidePoint

Date: 2021-11-08 09:00 | Sub-Study Identifier: default

SUMMARY KEYWORDS data storage run tearing customer inference layer solution spark case product nvme ssd talking built big problem provide query company Good morning this is Symmetra here. So it's kind of heads against the cloud guys to not just take it and not contribute back. And both discuss products be combined became Cloudera data platform a CDP as we know in the new company yep. I won't call high performance high capacity storage you're talking about 12 terabyte drives that time. I mean I m...

Interaction Resource: [202111080900-Reverse CDN-GuidePoint](http://mr-02:9000/pavilion/202111080900-AMER-US-CALIFORNIA-SANTA CLARA-Expert Networking Services-Reverse CDN-GuidePoint-Transcript.pdf)

## 202201041932-Reverse CDN-Meditronics

Date: 2022-01-04 19:32 | Sub-Study Identifier: default

SUMMARY KEYWORDS data kafka cloud run storage application architecture prem steve database transformation postgres terms metadata s3 aws amazon lambda functions patient device So let me go. I mean you know the vast majority of architectures you will see when you look at any architectural presentation is going to have Kafka in it in a pretty important role. And what I then look at you know and that's today with current architectures and that when you look at where things are going actually I was ...

Interaction Resource: [202201041932-Reverse CDN-Meditronics](http://mr-02:9000/pavilion/202201041932-AMER-US-CALIFORNIA-SANTA CLARA-Healthcare-Reverse CDN-Meditronics-Transcript.pdf)

## 202201060719-Reverse CDN-Guidepoint

Date: 2022-01-06 07:19 | Sub-Study Identifier: default

SUMMARY KEYWORDS data kafka transactions years moving requirements prem applications product business environment query bit real customer storage people batched talk bank no problem. All of that data gets aggregated at the enterprise level to monitor our portfolio risk levels in multiple lines of business. So Kafka is where we use to stream data from customer facing to essentially back end monitoring systems. And to seed in the evenings not live necessarily. But when you have multiple lines of b...

Interaction Resource: [202201060719-Reverse CDN-Guidepoint](http://mr-02:9000/pavilion/202201060719-AMER-US-CALIFORNIA-SANTA CLARA-Finance-Reverse CDN-Guidepoint-Transcript.pdf)