
TALK
DATA
TO ME

eDiscovery Unraveled

Uncovering Process Realities and Technological Integration

Dr. Pragyansmita Nayak
Chief Data Scientist
Hitachi Vantara Federal

Jonathan Ferguson
Digital Solutions Lead
Hitachi Vantara Federal

TALK DATA TO ME

October 25th

About Carahsoft

HITACHI
Inspire the Next

We are a trusted government IT solutions provider, delivering software and support solutions to Federal, State, Local, and Education customers. Carahsoft maintains dedicated teams to support sales and marketing for all its vendors.

TALK DATA TO ME

Meet Pragyansmita

HITACHI
Inspire the Next

Job	Chief Data Scientist
Background	• 20 years of Federal Government experience • Federal Financial Management, Malware Detection, Waste Management, Records Management, Network Forensics and more.
Certifications	Ph.D. in Computational Sciences and Informatics
I love to	Learn and appreciate the “Art to the Science” process of data science.
Email	Pragyan.Nayak@HitachiVantaraFederal.com
LinkedIn	https://www.linkedin.com/in/pragyansmita/

TALK DATA TO ME

Meet Jonathan

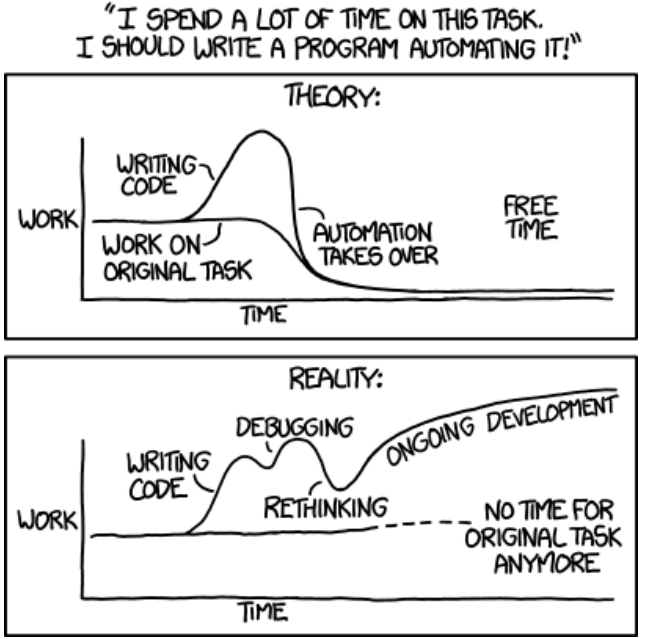
HITACHI
Inspire the Next

Job	Digital Solutions Lead
Background	19 years in Information Technology 5 years serving DoD/USAF 5 months with Hitachi Vantara Federal
I love to	Coach lacrosse, cycle and spend time with family!
Email	Jonathan.Ferguson@HitachiVantaraFederal.com
LinkedIn	https://www.linkedin.com/in/jfergusondc/

TALK DATA TO ME

Agenda

- eDiscovery
- FOIA and Analytics
- Multimodal Analytics
- Success stories and challenges
- Questions and Polls
- Next Steps



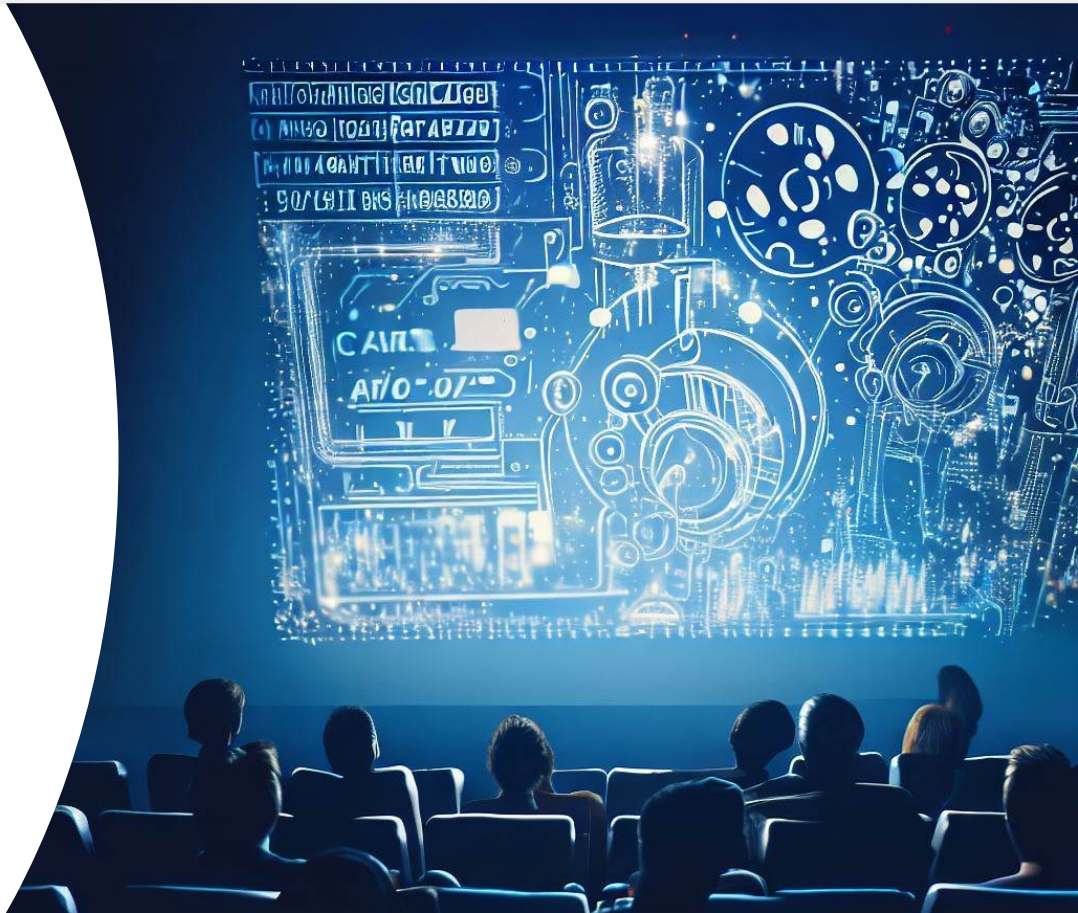
https://xkcd.com/1319/

TALK DATA TO ME

AUDIENCE POLL

What does eDiscovery mean to you?

Note: We will ask this yet again at the end of this session



eDISCOVERY, FOIA AND ANALYTICS

eDiscovery

- Short for electronic discovery
- Process of identifying, collecting, and producing electronically stored information (ESI) for legal proceedings or investigations
- Use of technology to search, review, and analyze vast amounts of digital data (in contrast to manual approach)
- Used as evidence in litigation, regulatory compliance, data breach impact analysis or internal investigations

Connection to FOIA and Analytics

HITACHI
Inspire the Next

eDiscovery and the Freedom of Information Act (FOIA) are connected through their shared objective of accessing and retrieving electronic information.

eDiscovery

- ✓ Guided by specific rules and procedures established by the legal system
- ✓ Process is typically governed by court orders
- ✓ Parties involved in the legal matter have obligations to preserve and produce relevant electronic evidence

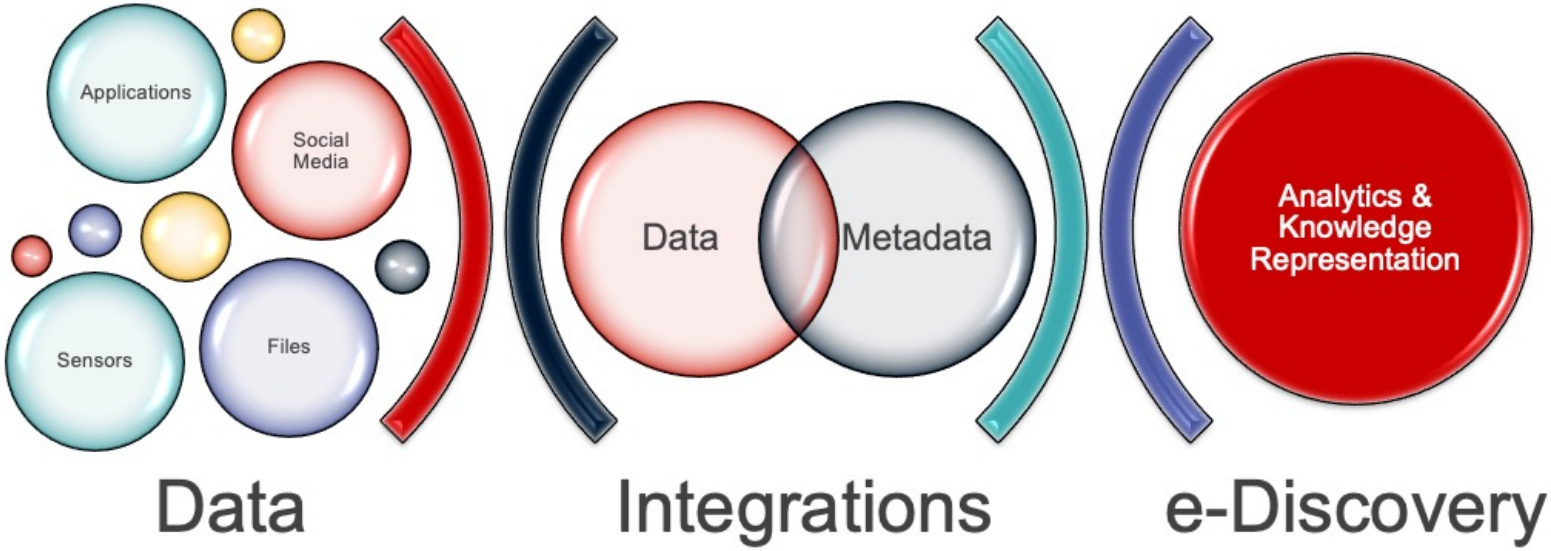
FOIA

- ✓ Set of procedures and timelines for submitting requests to government agencies
- ✓ Agencies review the requested records and determine if any exemptions apply, such as for national security, personal privacy, or confidential business information
- ✓ Provide the requested information or justify any denials

MULTIMODAL eDISCOVERY

Data, Metadata and Knowledge Representation

HITACHI
Inspire the Next



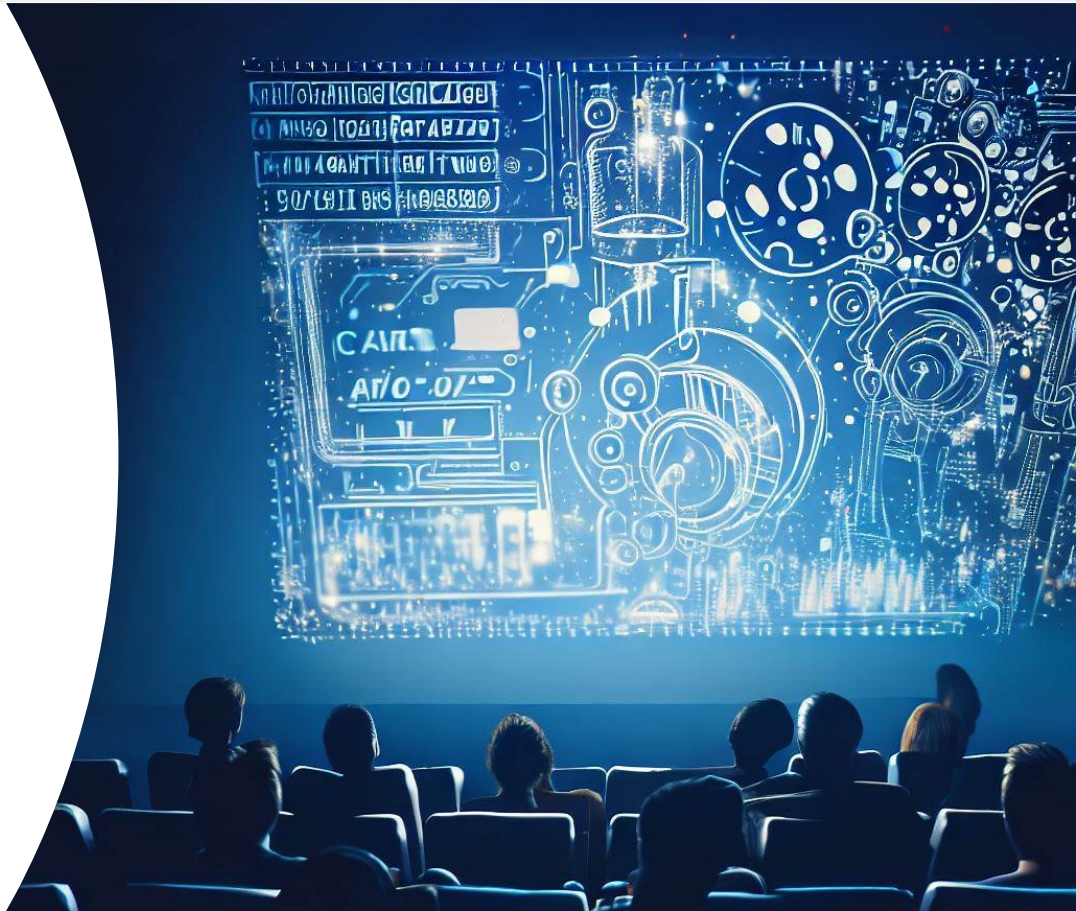
Multiple modalities or types of data analysis combining various methods, such as text analytics, visual analytics, audio/video/image analytics, metadata analysis, machine learning and more.

TALK DATA TO ME

AUDIENCE POLL

If you could use eDiscovery to investigate and uncover more about any historical figure or celebrity, who would it be and why?

No, you can't say your own name! ☺

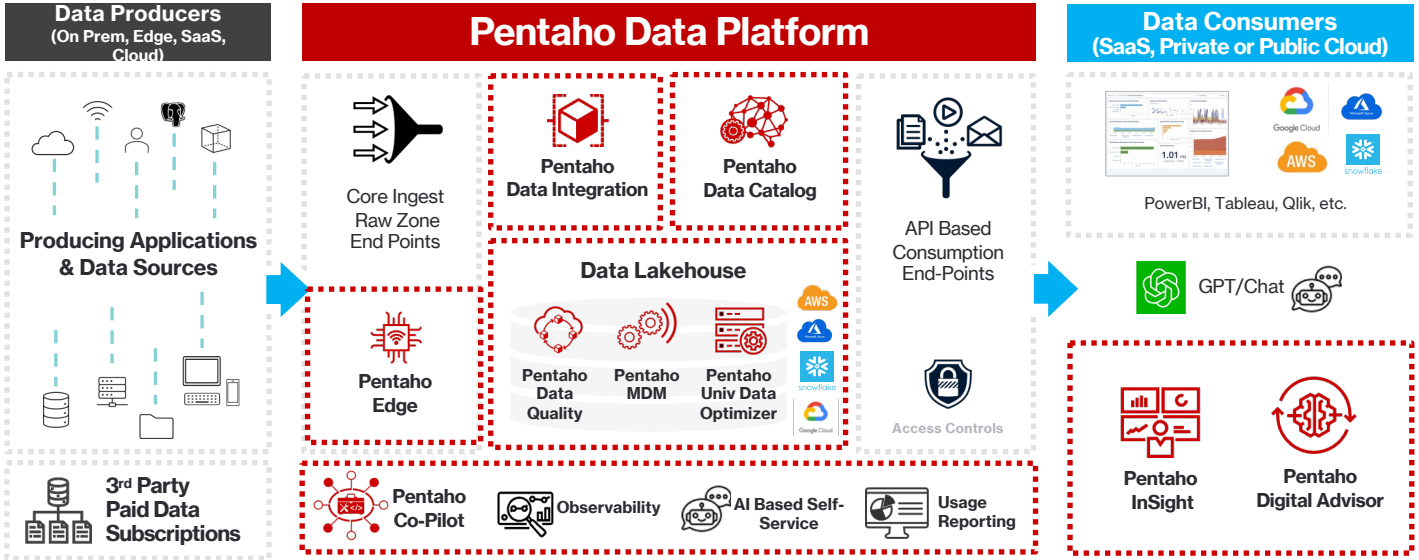


eDISCOVERY SOLUTIONS

Pentaho Data Platform | Data Driven for Business Outcomes

HITACHI
Inspire the Next

Fully functional (ingest, transform, process, store, enrich, search, access & analyze) **Pentaho Data Platform** works with a target cloud account with consistency and a significantly reduced turnaround time of days instead of months. **Pentaho Data Platform** connects data to cloud technologies that allows enterprises to quickly set up a data platform in a composable, modular, scalable, flexible, and extensible manner.



ROI & Time to Value

ROI will occur upon each use. Value realization will become faster and easier. Qualified projects will ramp quicker, more consistently, and with higher data quality.

3rd Party Connectivity

IBM Mainframe, DB2, SQL Server, Posgres, Sybase, NetApp, Dell/EMC, HPE, Windows Files (CIFS), Hadoop, SAP HANA, Teradata, Vertica, Oracle, Tableau, Ab Initio, AWS Hadoop, AWS Glue, AWS Redshift, AWS RDS, AWS S3, Azure PowerBI, Cloudera, Snowflake, MongoDB

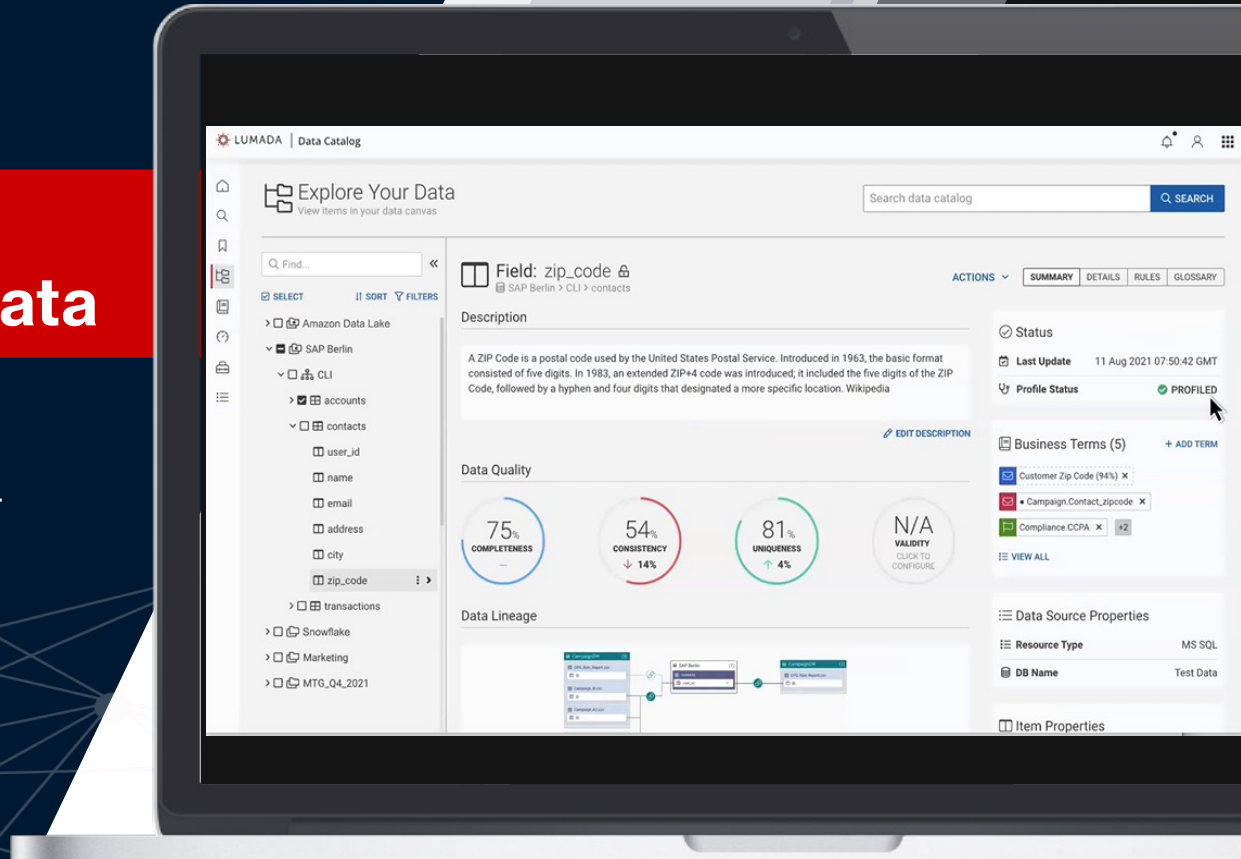
Outcome

The **Pentaho Data Platform** can be used to rapidly manifest a working **Data Platform** which can be enhanced and/or modified as required for private and public cloud. Typical user base are line of business teams who want immediate functionality but are not yet "**Data Driven**."

Data Catalog

Powerful Decisions Based on Trusted Data

- Automated data discovery for analytics & governance
- Accessing quality and lineage for trust
- Empower standardization with business glossary
- Data compliance for risk reduction
- Data lineage for governance



Search To Your Needs

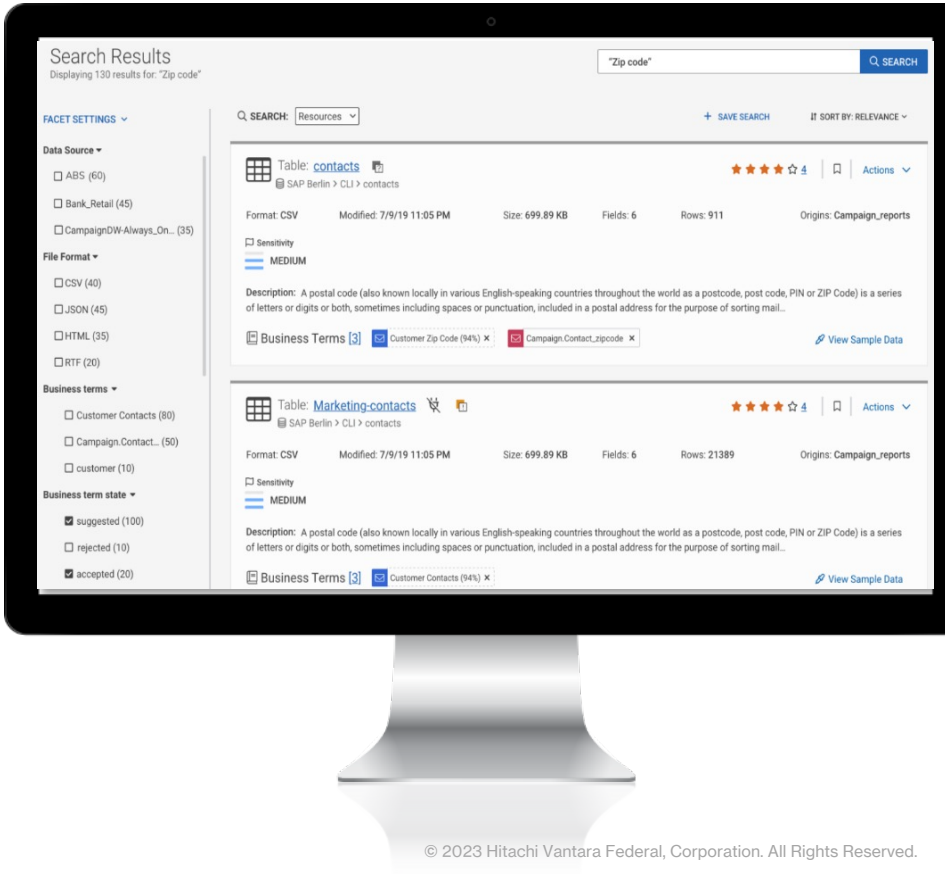
HITACHI
Inspire the Next

Amazon-like search experience

- Familiar experience for finding data, no training needed
- Accelerate finding all matching content
- At-a-glance views provide quick assessment

Domain-aware search

- Searches are performed using domain-specific glossaries of your teams

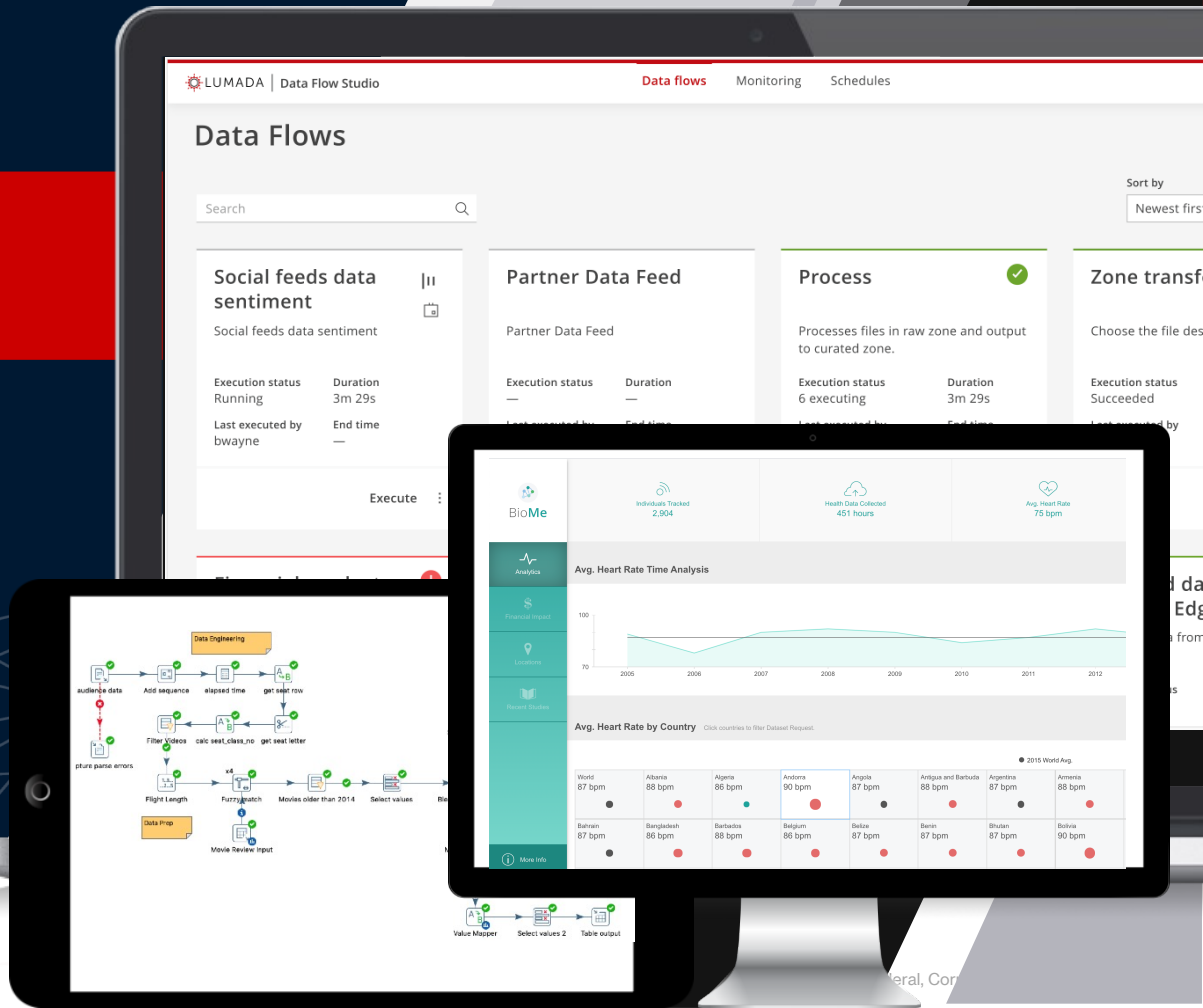


TALK DATA TO ME

Pentaho

Data Integration & Analytics

- Data pipelines & analytics
- No-code environment
- Any data, any system
- Enterprise scale and security
- Open-Source foundation
- Micro-Services Architecture



Data Platform

HITACHI
Inspire the Next

A unified foundation for an organization's data-driven initiatives, promoting efficient data management, analysis, and collaboration across various teams and departments.

Enables organizations to harness the power of data to make informed decisions, drive innovation, and achieve business objectives.

TALK DATA TO ME

eDISCOVERY SUCCESS STORIES AND CHALLENGES

eDiscovery Use-cases

HITACHI
Inspire the Next

Anomaly Detection

Identify outliers or abnormal patterns within multimodal data. Anomalies may indicate potential data breaches, unusual behavior, or fraudulent activities, enabling organizations to promptly address such incidents.

Document Classification

Machine learning algorithms, such as Support Vector Machines (SVM), Naive Bayes, or Random Forest, can be employed to classify documents based on specific criteria. Learn from labeled training data and prioritize document review, reduce manual effort and thereby, accelerate the identification of relevant documents.

Image and Video Analytics

Real-time data lake-based analytic platform with NLP for data extraction with a centralized repository for structured and unstructured files. Machine learning algorithms to evaluate time series data. Digitize and automate market regulation, structure and transparency functions and improve time-to-market for its core business operations.

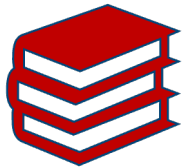
Recommender Systems

Recommender systems leverage machine learning algorithms, such as collaborative filtering or content-based filtering, to suggest relevant documents, similar cases, or potential connections based on the user's interactions or data characteristics. This assists in knowledge discovery, improving efficiency and aiding decision-making.

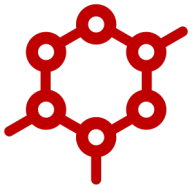
TALK DATA TO ME

Text Analytics eDiscovery

HITACHI
Inspire the Next



Natural Language Processing (NLP): Named Entity Recognition (NER), sentiment analysis, topic modeling, and language models (e.g., word embeddings, Transformers) can help uncover key entities, sentiments, and themes within documents, aiding in faster analysis and understanding of textual data.

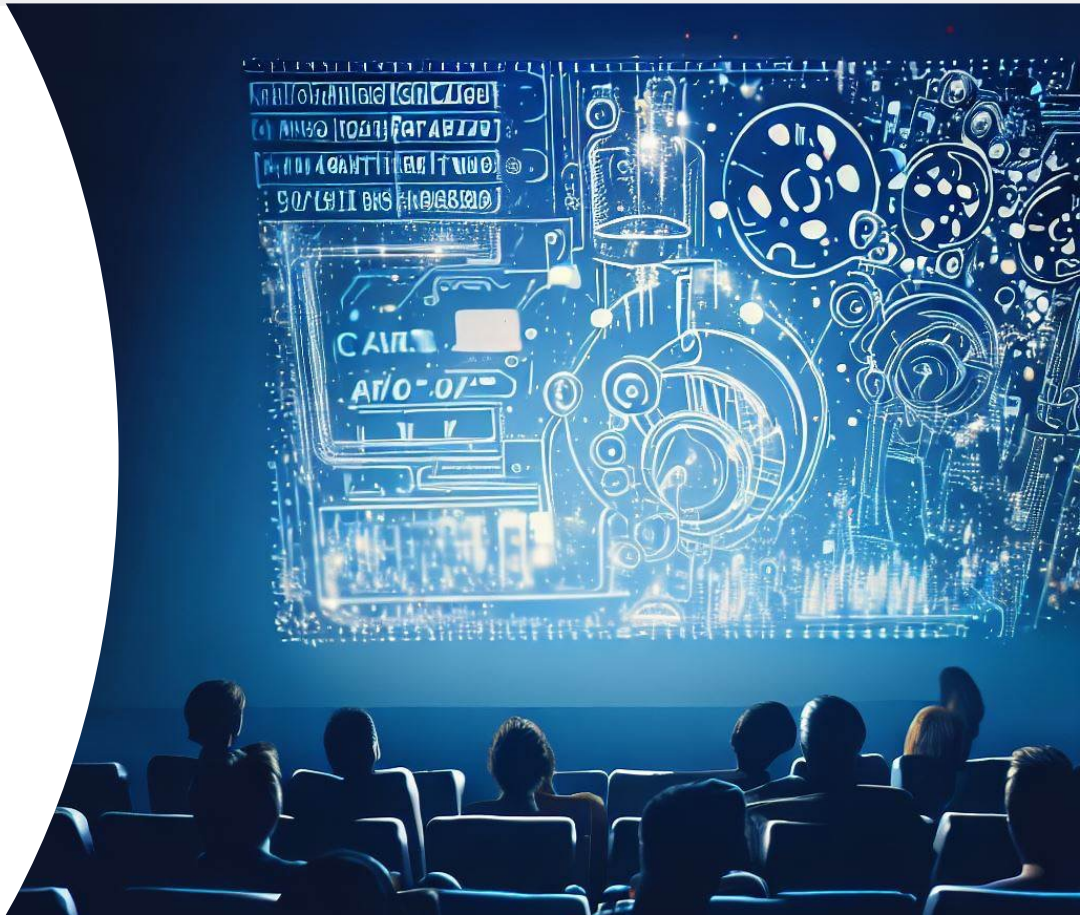


Clustering: Unsupervised learning algorithms like k-means clustering or Latent Dirichlet Allocation (LDA) can group similar documents together based on their content. Clustering assists in organizing large datasets, identifying document clusters, and discovering hidden patterns or topics within the data.

AUDIENCE POLL

What does eDiscovery mean to you?

Note: We will ask this yet again at the end of this session



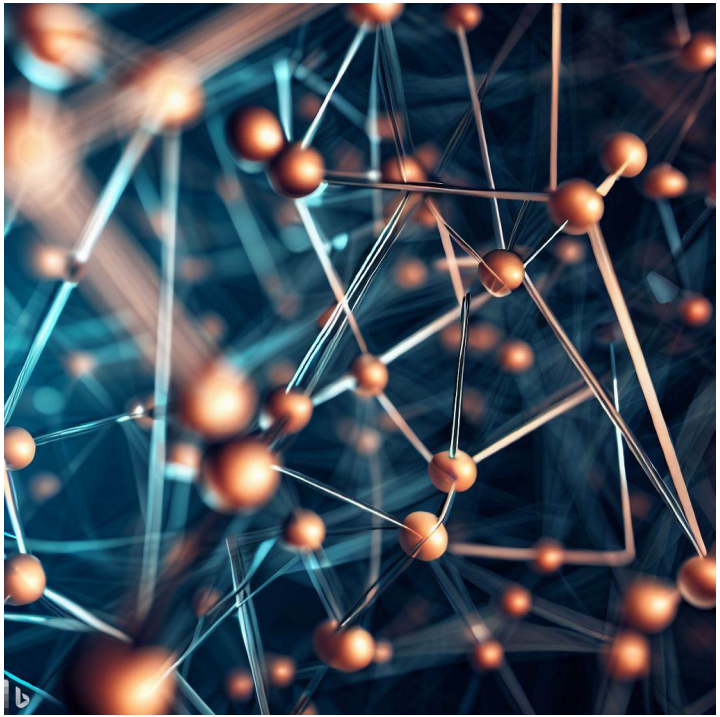
AUDIENCE POLL

Did eDiscovery become more relatable to your use-cases post this webinar?

HITACHI
Inspire the Next

- Yes
- No

Please share the use-cases in the chat window or reach out to us to discuss and learn more.



TALK DATA TO ME

Q&A

TAKE ACTION

T-Shirt Raffle

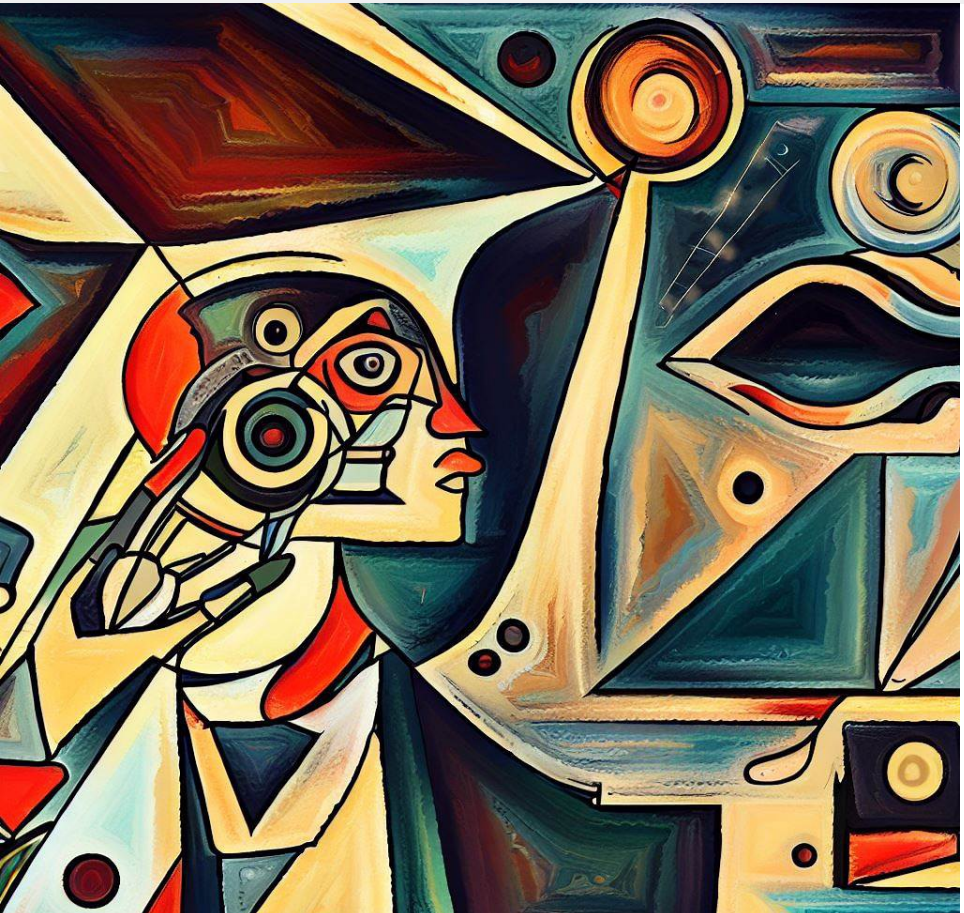
HITACHI
Inspire the Next

- 10 attendees will win a “Talk Data to Me” T-shirt
- Drawing takes place from today’s attendees
- Winners will be notified via email within 24 hours that they have won
- We will collect winners’ address and size to ship out T shirts

TALK DATA TO ME

Next Steps

HITACHI
Inspire the Next



1. Continue Learning:
2. Schedule a Discussion:
3. See a Demo:
4. Attend a future Talk Data to Me Webinar
5. Provide us feedback on your topics of interest

A virtual series from Hitachi Vantara Federal

Talk Data to Me

ON DEMAND | DataOps: Unleashing the Power of Data for Agency Innovation

ON DEMAND | Navigating the Data Storage Maze: Conquering Security, Complexity, and Efficiency Challenges

Stream Previous Sessions & Register for Future Sessions

HITACHI
Inspire the Next

TALK DATA TO ME


TALK
DATA
TO ME

Keep in Touch

HITACHI
Inspire the Next

Jonathan Ferguson

jonathan.ferguson@hitachivantarafederal.com

LinkedIn: jfergusondc

Dr. Pragyan Nayak

pragyan.nayak@hitachivantarafederal.com

LinkedIn: pragyansmita