Webcast Presentation

The Master's Conference™

Handling Non-Traditional Data Sources in eDiscovery

HAYSTACK

July 2020

Michael D. Sarlo, EnCE, CBE, RCA, CCLO, CCPA

Partner and Senior EVP of eDiscovery and Digital Forensics

Michael is a Partner and Senior Executive Vice President of eDiscovery and Digital Forensics at HaystackID.

In this role, Michael facilitates all operations related to electronic discovery, digital forensics, and litigation strategy both in the US and abroad while working on highly complex forensic and e-Discovery projects.

Michael is fluent in a wide variety of digital forensic, ECA, e-Discovery processing and hosting technologies in addition to the IT infrastructure associated with deploying and administering such tools. He is also a frequent speaker on digital forensics best practices, GDPR issues, ESI treatment methodologies, and cost containment strategies for working with massive volumes of data.

Michael is an expert witness and holds certifications around several forensic tools including Cellebrite, EnCase, and BlackLight. He is also a Relativity Certified Administrator (RCA).

HaystackID | msarlo@haystackid.com | HaystackID.com | +1.617.943.2960

John Wilson, ACE, AME, CBE

CISO and President of Forensics, HaystackID

John provides expertise and expert witness services to help companies address various matters related to digital forensics and electronic discovery (eDiscovery), including leading investigations, ensuring proper preservation of evidence items and chain of custody. He develops processes, creates workflows, leads implementation projects as well as GDPR data mapping services for clients including major financial institutions, Fortune 100 companies, Am Law 100 law firms as well as many other organizations small and large. In addition, he provides expert witness services and consulting in matters of all sizes. His work spans some of the largest litigations and matters on record in the United States and many of the 39 countries where has worked on cases.

John is a certified forensic examiner, licensed private investigator and information technology veteran. He has over two decades of experience working with the US Government and both public and private companies. He serves clients in a variety of industries and is an advisor to outside counsel, general counsel and in-house executives on best practices.

HaystackID | jwilson@haystackid.com | HaystackID.com | +1.678.673.3463

Handling Non-Traditional Data Sources in eDiscovery

Agenda

- Non-Traditional ESI Fundamentals and Considerations
- Audio and Video Discovery
- Cloud Data, Collaboration Suites
- Short Message Collaboration
- The Expanding Internet of Things (IoT)

Non-Traditional Data Presents Challenges Across the EDRM

- **Identification** - How to define the scope of potentially discoverable data
- **Control** - Unclear in some cases who has custody & control
- **Preservation** - Sources often provide no means of preservation – legal hold?
- **Collection** - Likewise, obtaining the data may require specialized procedures/tools
- **Processing** - eDiscovery tools are not often not able to handle nontraditional data
- **Analysis** - Many data sources are incompatible with routinely-used analytical functions
- **Review** - Volume, unitization, format and context require specialized review protocols
- **Production** - Traditional forms of production may not apply

Other Issues

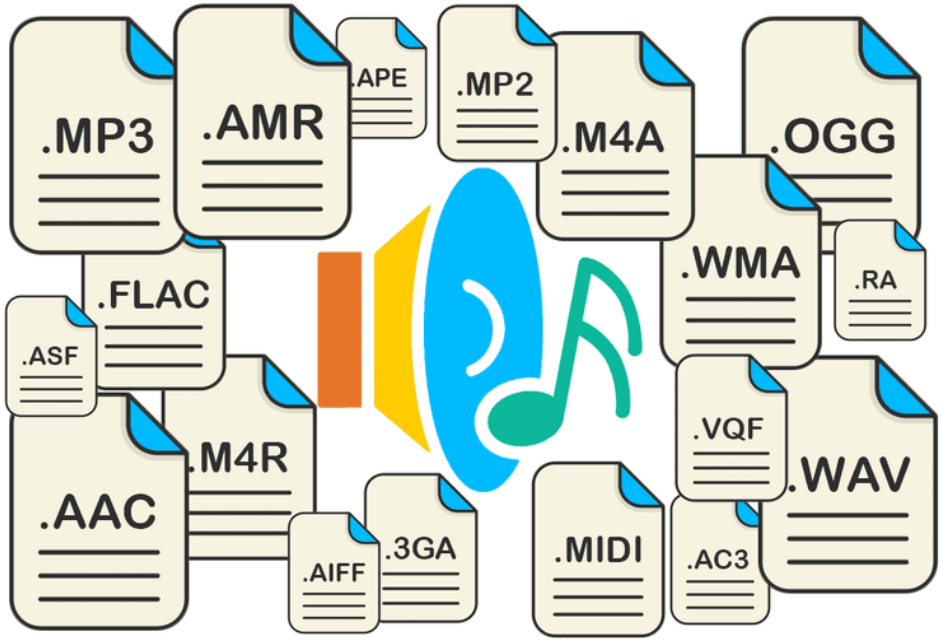
- **Awareness** - Are companies aware of its existence? Location? Method of access?
- **Volume** - The potential enormity can be daunting
- **Meaning** - Context is often required in order to understand individual data points
- **Volatility** - Data sources are often ephemeral by nature, and systems may not have a mechanism for preservation, or means to preserve without freezing system
- **Complexity** - Data is frequently stored in complicated data structures and formats

Audio-Video Discovery

Audio Discovery

- **Volume, Preservation and Collection Challenges**
  - 50% of corporation customer interaction tracking exists in audio
  - Volumes can be massive, and organizations may only save a certain time period of data before older recordings are overwritten.
  - Metadata often tracked in databases that also may require custom collection methods and/or may not easily lend to legal hold

- **Diversity of Formats**
  - Many different CODECs (hardware and software)
  - Recording quality tends to be poor
  - Proprietary formats may require reverse engineering
  - Audio often needs to be converted and/or standardized to format supported by eDiscovery tools that handle audio



Audio Discovery (cont.)

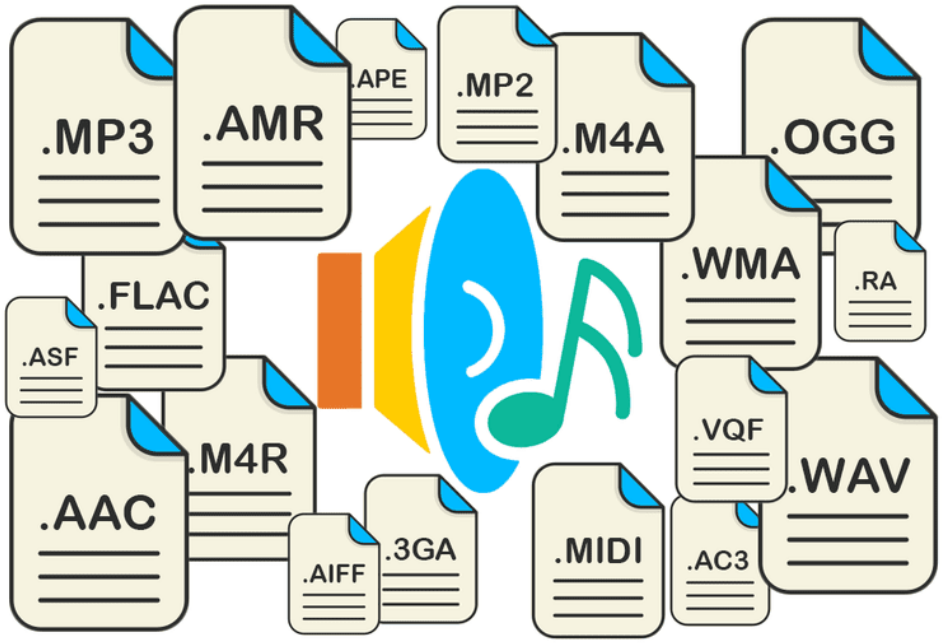
- **State of Voice-to-Text Transcription Technology**

Authenticity.ai

VERITONE

nexidia

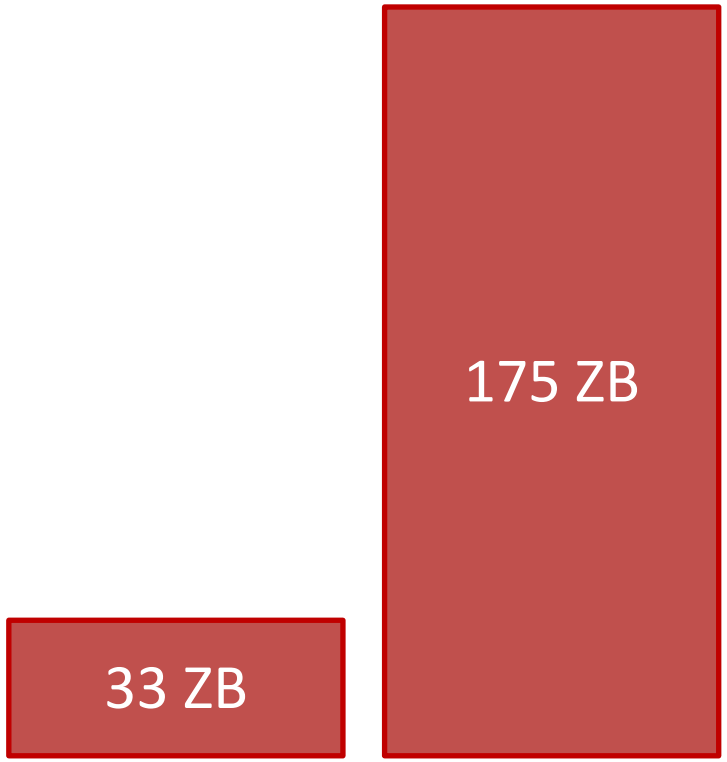
- **Review and Production Considerations**
  - Sampling can help gauge initial transcription quality
  - Transcription may be iterative with different providers
  - Leverage fuzzy search to isolate targets for manual review
  - Establish open dialogue with opposing parties re burden



Data in the Cloud

Data Volume: Accelerating Growth

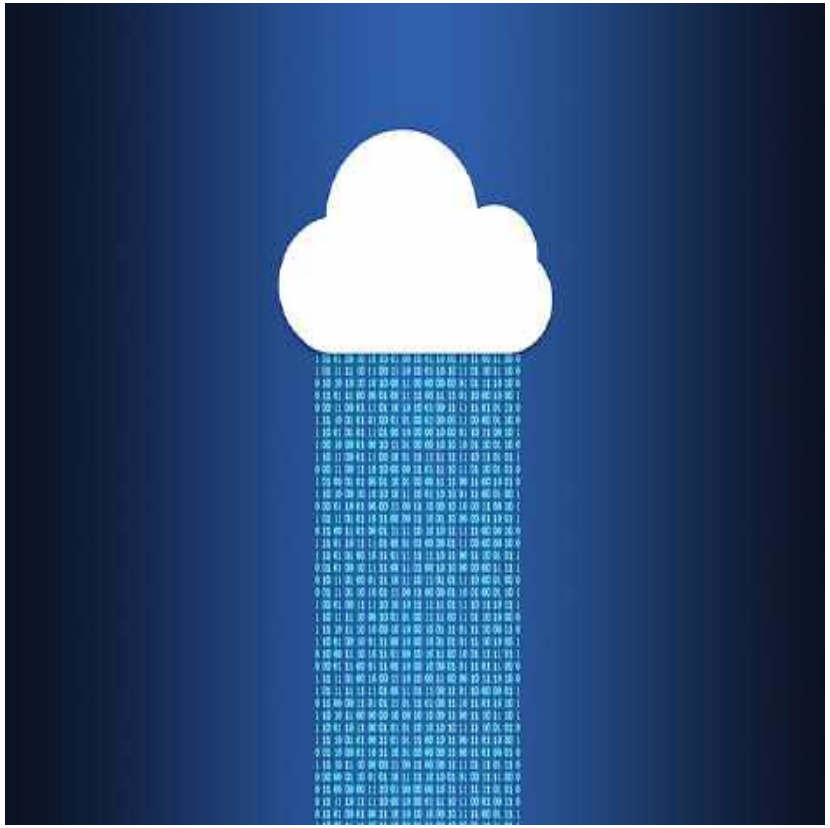
- The world's data is predicted to grow from 33 ZBs (zettabytes) in 2018 to a 175 ZBs by 2025.
- 90% of the world's data has been created in the last two years alone.
- Over 8 million people use voice control every month.
- Each day 95 million photos and videos are shared on Instagram.
- Almost 1 billion text messages are sent every hour.
- More than 156 million emails are sent out every minute.
- Every second there are 1,724,125 spam emails sent.



Data in the Cloud

- **eDiscovery Challenges:**
  - Where does the data actually reside?
  - Who has custody and control (*Gordon Partners v. Blumenthal*)?
  - Does the provider support preservation and collection?
  - Native format?

- **Considerations**
  - Ethical obligations re: competence extend to cloud data sources
  - Tools and API's can help in collection help but have limits
  - Consultants can provide valuable assistance
  - Understand the potential pitfalls of self preservation



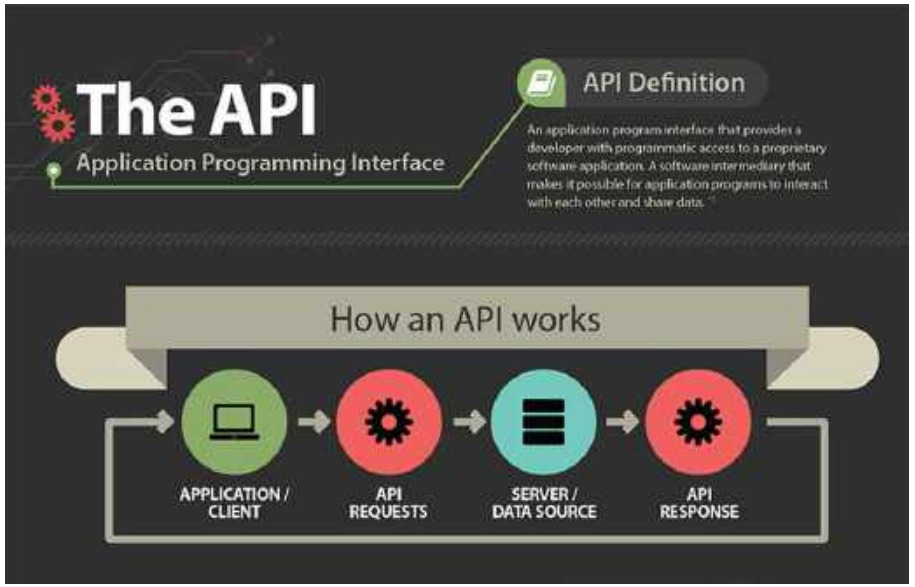
Cloud-Based Productivity Suites

- More and more organizations are moving to the cloud, and different products lend themselves to search, preservation, and extraction better than others. Post extraction, significant transformation may be required for review.
- Web-based applications may interact with other web-based applications or versions of the same application.
- Many cloud service suites have robust API hooks that allow third party applications to interact with core applications that may be able to create, modify, or store data that appears resident in one system.

Connecting with The Cloud

What Is An API?

- An application programming interface (API) is an interface or communication protocol between a client and a server intended to simplify the building of client-side software. It has been described as a “contract” between the client and the server, such that if the client makes a request in a specific format, it will always get a response in a specific format or initiate a defined action.
- An API may be for a web-based system, operating system, database system, computer hardware, or software library.

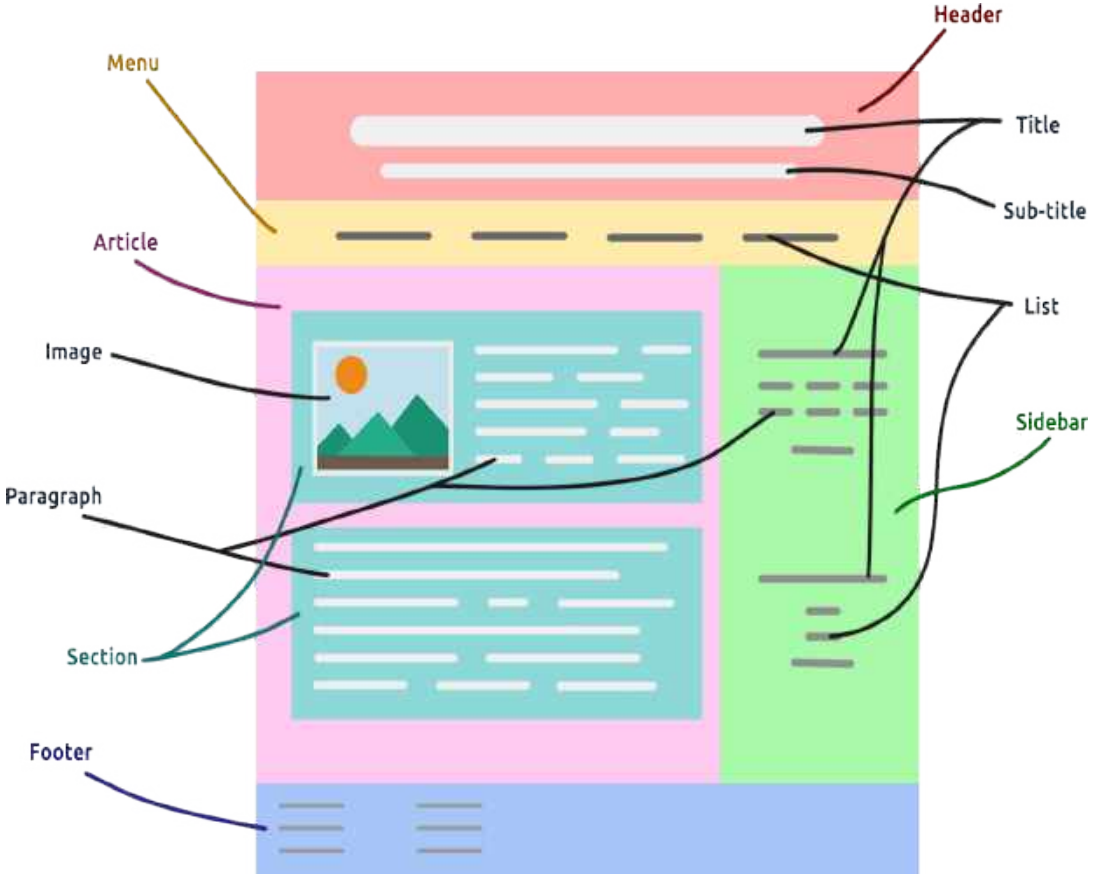


- An API specification can take many forms, but often includes specifications for routines, data structures, object classes, variables, or remote calls. POSIX, Windows API and ASPI are examples of different forms of APIs. Documentation for the API usually is provided to facilitate usage and implementation.

API DOCUMENTATION CAN OFTEN BE OUT OF DATE OR SIMPLY INCORRECT
TESTS AROUND REPEATABILITY AND EXPECTED OUTCOMES MUST BE PERFORMED BEFORE GOING ALL IN

APIs Explained

- **Most systems with a Web API can be collected.**
- **Authentication and access vary by system**
- **Two major methods of collection:**
  - Data + Metadata – Think Box.com, DropBox, Confluence
  - Web Page Imaging – Think Web Capture
- **Most APIs have a hierarchical model**
  - We request all top-level objects in the system
  - We request all sub-objects of each of those top-level objects
  - We recurse down through the structure until we've collected all pertinent records
  - This data can be formatted as needed.



Chat Data Types

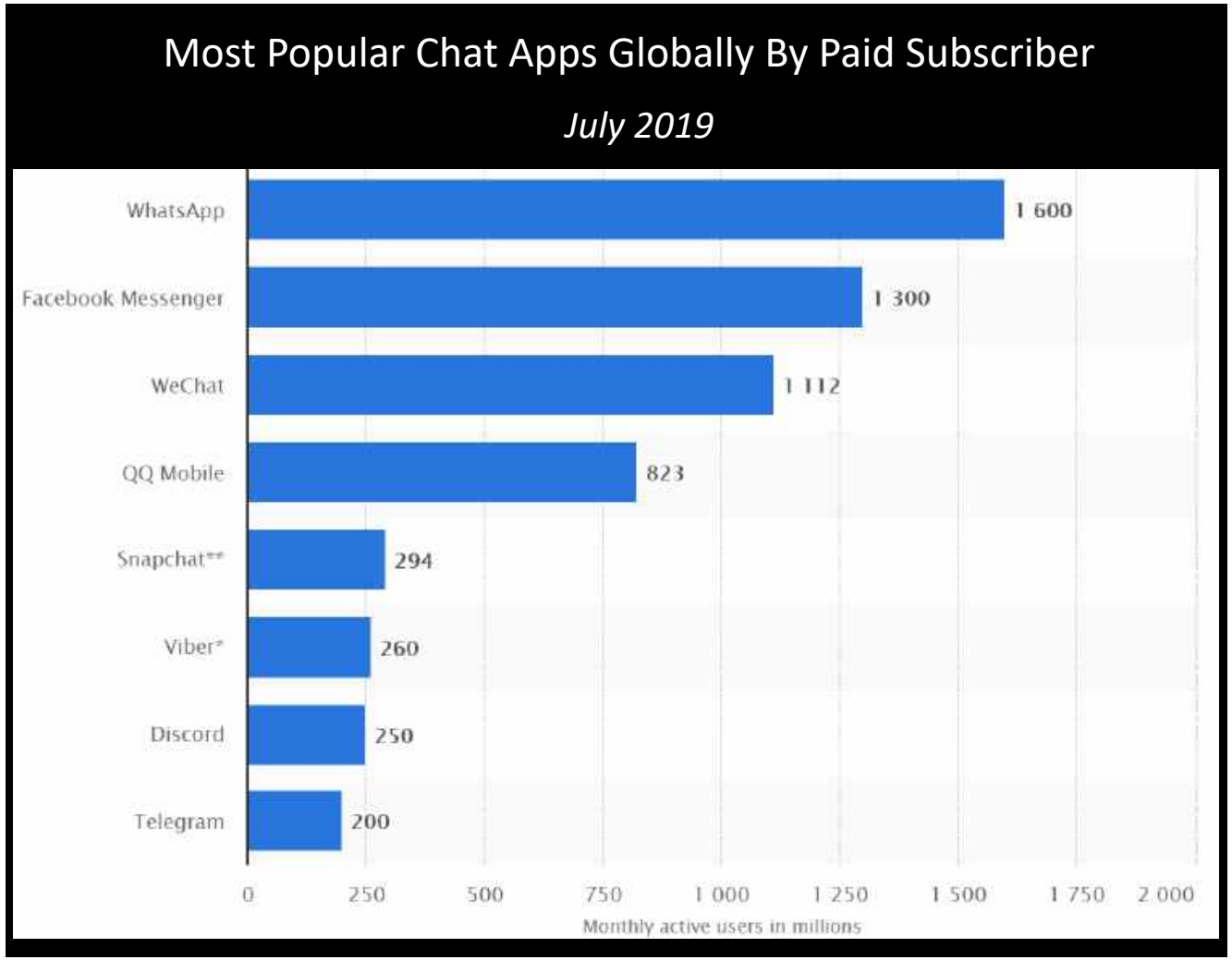
Considerations and Challenges

Standard Chats

Skype - Available in Office 365 Depending On Logging Settings as an Exchange Object. For on-premise configurations, chat logging may need to be pulled from desktops or exchange server if enterprise logging is enabled with Skype for Business Server.

Bloomberg- Many variations of Bloomberg implementations, and changes to the underlying structure of XML files that contain content - must check every time to understand versioning differences. Additionally, attachments are within a separate GZIP file for both Bloomberg Mail AND Chats.

Domino / Lotus Sametime - In almost all cases, Domino enterprise level SameTime logging is not enabled due to massive performance issues when the service is turned on. Workstations needs to be scanned for local SameTime logs, which may also be encrypted like local NSFs with compendium ID Files which may be inaccessible or managed at the server level.



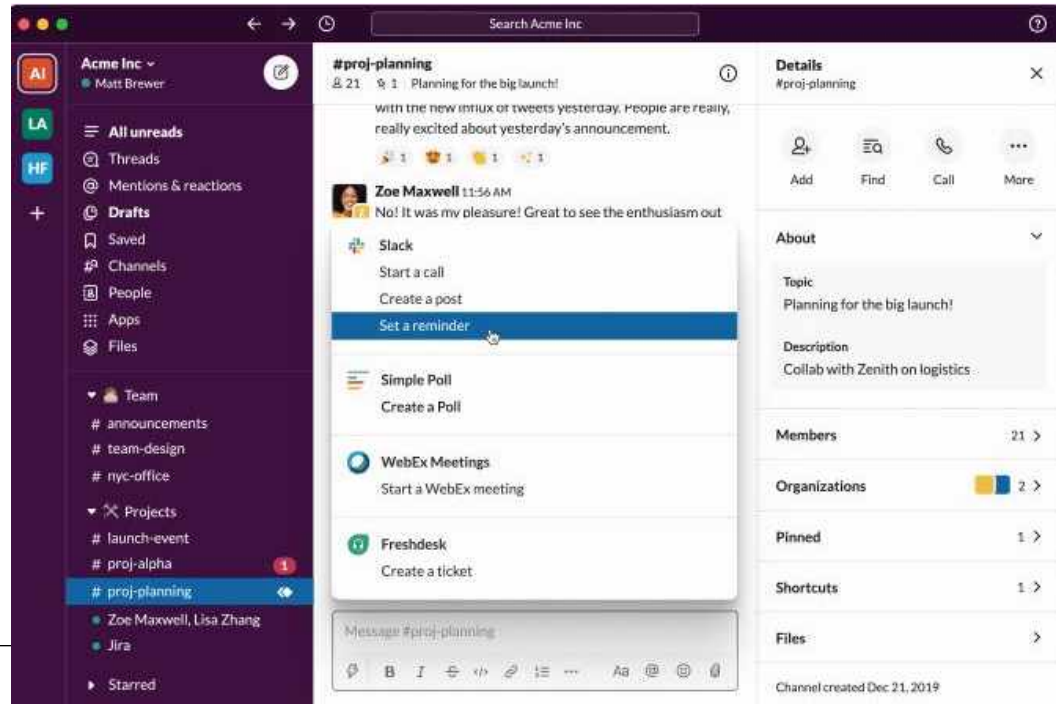
Next Generation Chat Platforms & Application Discovery

Considerations, Challenges, and Methods

Jay Stacks

Slack Compliance Export Example


```
[
    {
        "client_msg_id": “d234d122-ad98-41ae-94ee-0f673e2fa38f3",
        "type": "message",
        "text": “@U34F54D2, I deleted the data as per @U56RED123’s instructions",
        "user": "UR56TE344",
        "ts": "1549562926.000000",
        "team": “T64RF231“
        “reply_count”: 1,
        “replies: [
            “user”: “@U34F54D2”
            “ts”:  "1549562973.000100"
    }
```

HAYSTACK

Slack

Built in Exports

Export Pros:

- Collects entire organization (corporate)
- Contains historical edits and deletions not available via OAuth
- All non-attachment data delivered at once

Export Cons:

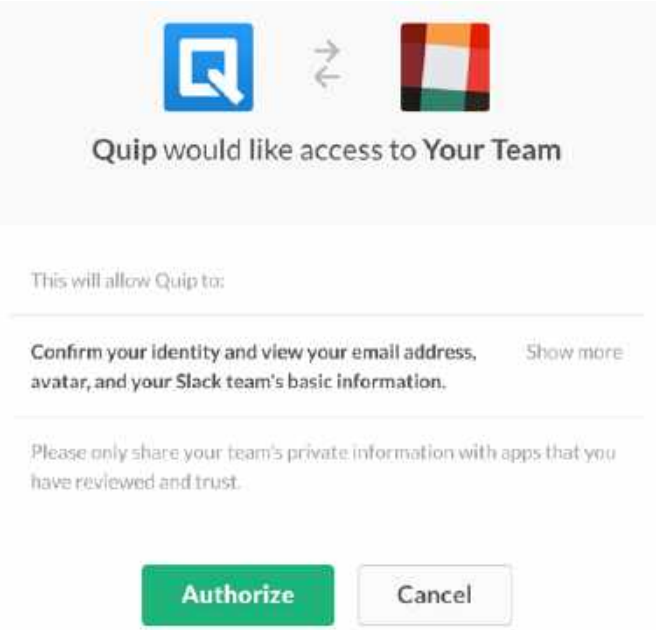
- Scope of exports varies by plan
- Collects entire organization indiscriminately
- Large instances can have extremely unwieldy export sizes
- Attachments must be downloaded separately from main export

ATTACHMENTS MUST BE DOWNLOADED AS SEPARATE OBJECTS, FROM A PUBLICLY AVAILABLE RANDOMIZED LINK – THIS CAN POSE ACCESS AND/OR SECURITY ISSUES.

Slack

Targeted Custodian OAuth Collection Method

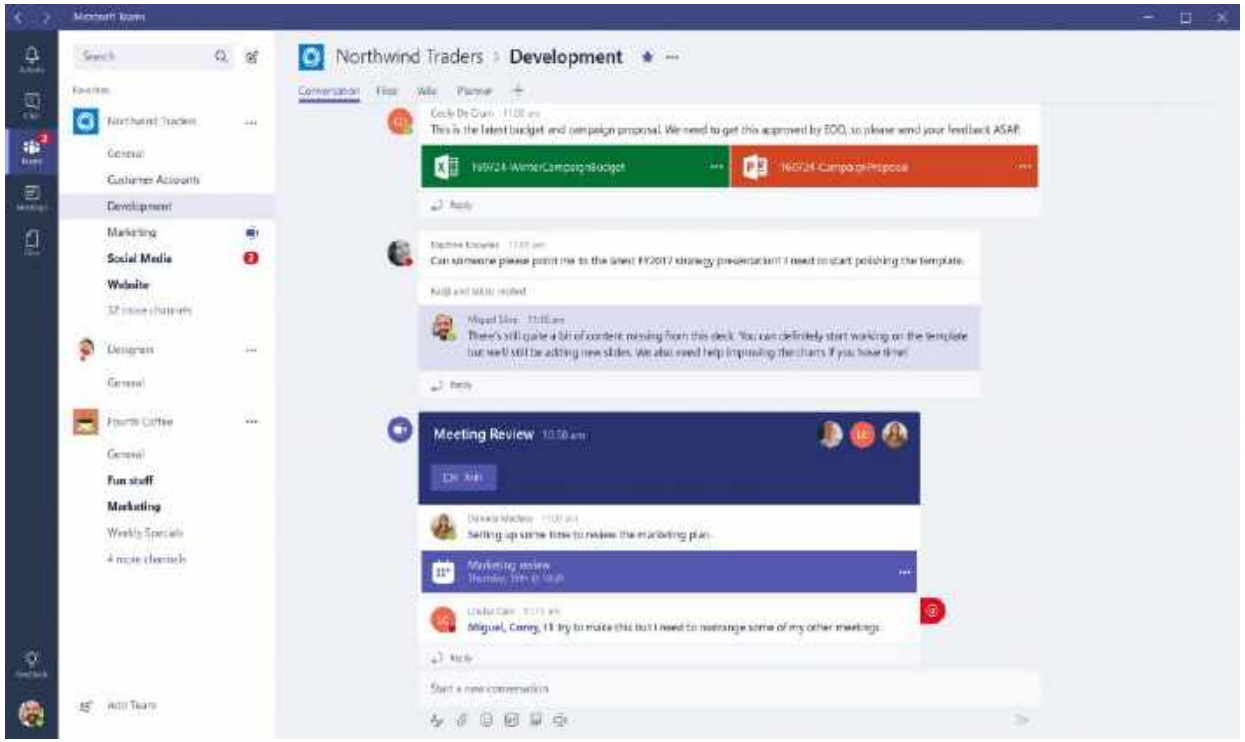
- OAuth Collection Pros
  - Allows limited collection without unnecessarily invading user privacy
  - Reduces total collection size
  - Is a more targeted and precise method of collection
- OAuth Collection Cons
  - Collects data “as-is”
  - Requires authentication by each custodian being collected
  - Alerts custodian to collection process
  - Subject to rate limitations by Slack



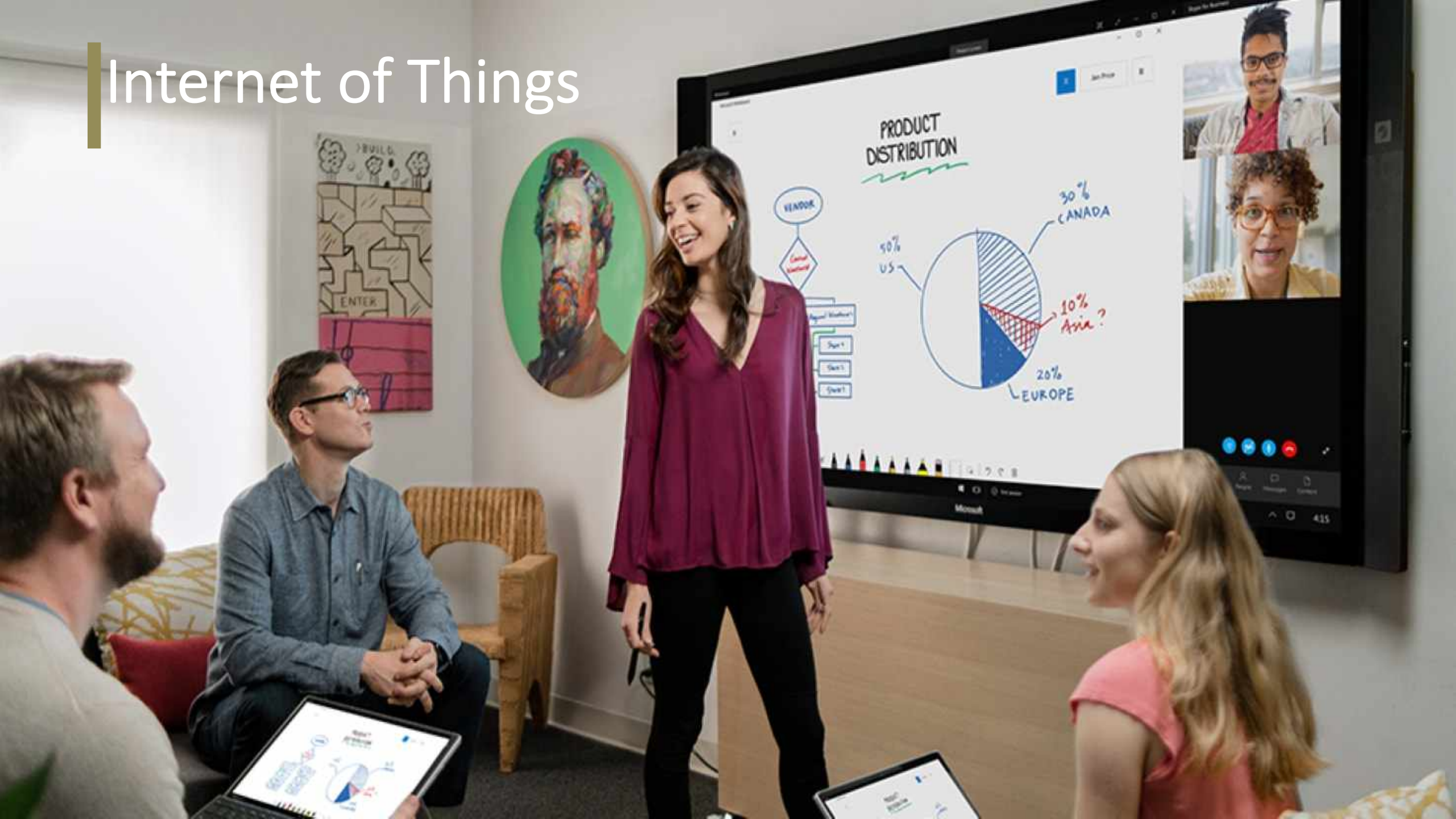
IF A USER IS DELETED FROM THE SYSTEM OR DELETES THEIR MESSAGES, THEN ASSOCIATED MESSAGES WILL NOT BE ACCESSIBLE USING AN OAUTH METHOD.

Microsoft Teams

- **Deeply Integrated in Windows**
- **Cannot Disable in Office 365**
- **Skype Replacement**
- **Robust Third-Party App Hub**
- **Compliance Center export capability limited – difficult to extract look and feel of Teams**



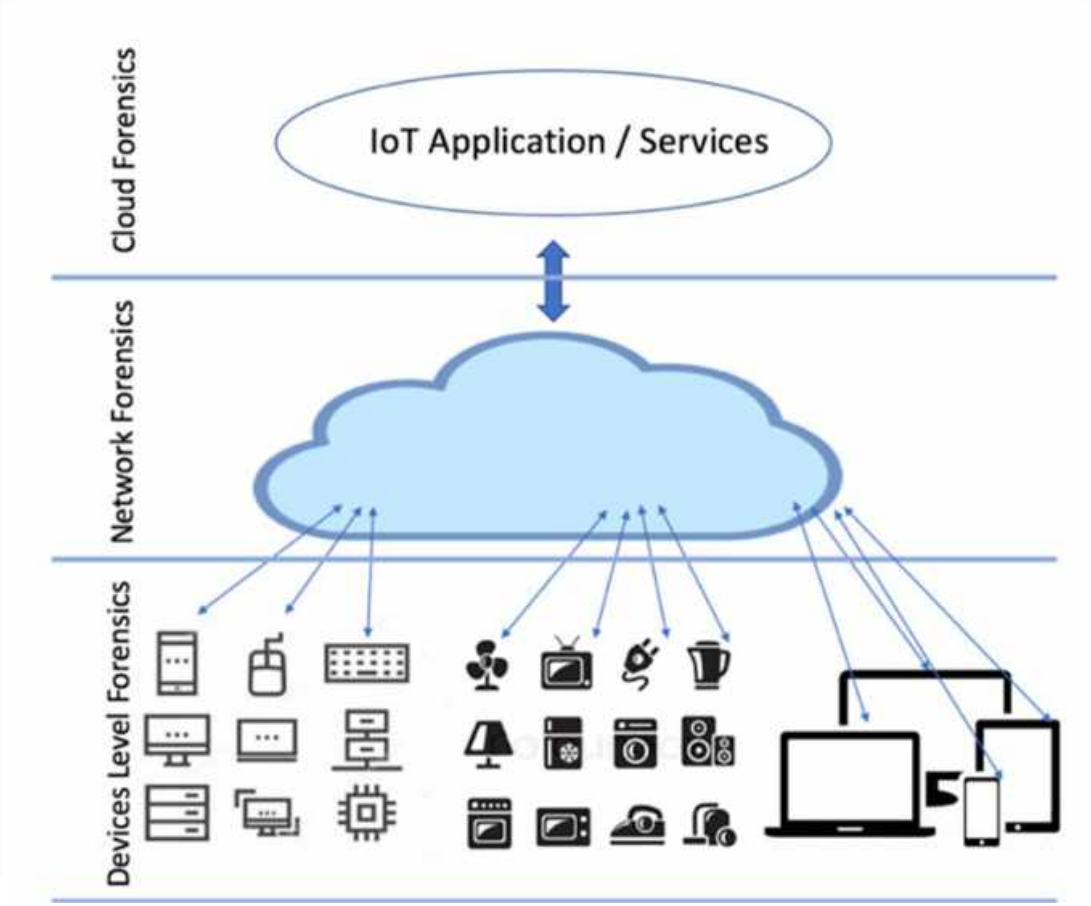
Internet of Things



Internet of Things: Challenges and Framework

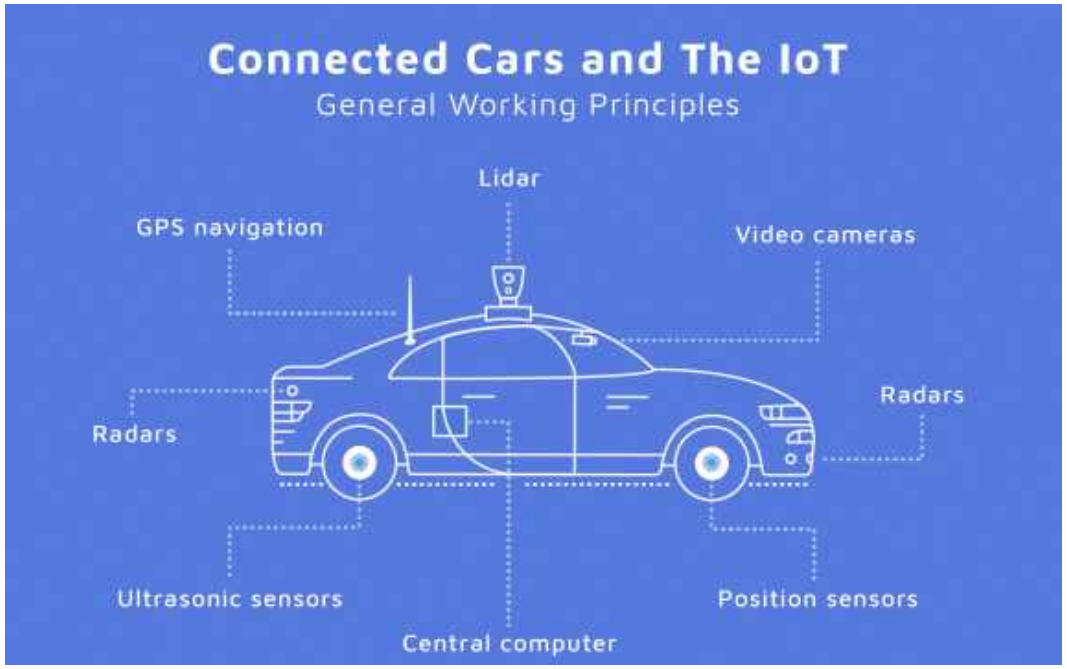
- **IoT eDiscovery Challenges**
  - Data is not readily accessible
  - Data is extremely volatile
  - May require cross correlation amongst many data points:
    - On a Physical Device (Alexa, Google Home, Etc.)
    - Within an App (iOS Health Data)
    - Within the Cloud (Alexa Voice Recordings)
    - Networking equipment

- **Extraction and Analysis is A Science Project**
  - Devices evolve daily with continuous updates
  - Devices are often an operating frameworks for individual applications with fluid update cycles
  - Repeatable methods for extraction are difficult, and results must continuously be retested with each eDiscovery exerc to ensure defensibility



Internet of Things: Many Different Data Points

- **Video cameras** (forward, rear, side) — detect traffic lights and signals of other cars; read road signs; detect pedestrians, cyclists and obstacles; provide 360 degrees of visibility around the car;
- **Radars** (forward, rear) — detect objects and their velocities at long distances; are able to see the car ahead through heavy rain, fog and dust; monitor the position of other vehicles nearby; measure the distance to objects and help parking;
- **Lidar** — a rotatable radar on the roof that scans the environment up to 100 meters and creates a 3D map of the location
- **Ultrasonic sensors** — complement the vision, detecting hard and soft objects, and measuring the position of objects close to the vehicle
- **Position sensors** — may be built in the wheels to sense their movements and detect the car position on the map
- **GPS navigation** — catches signals from satellites to provide more accurate positioning
- **Central computer** — provides analytics of the gathered data and influences decision-making



IoT: Expanding Data Points Beyond the Cloud

From edge sensors to the centralized cloud

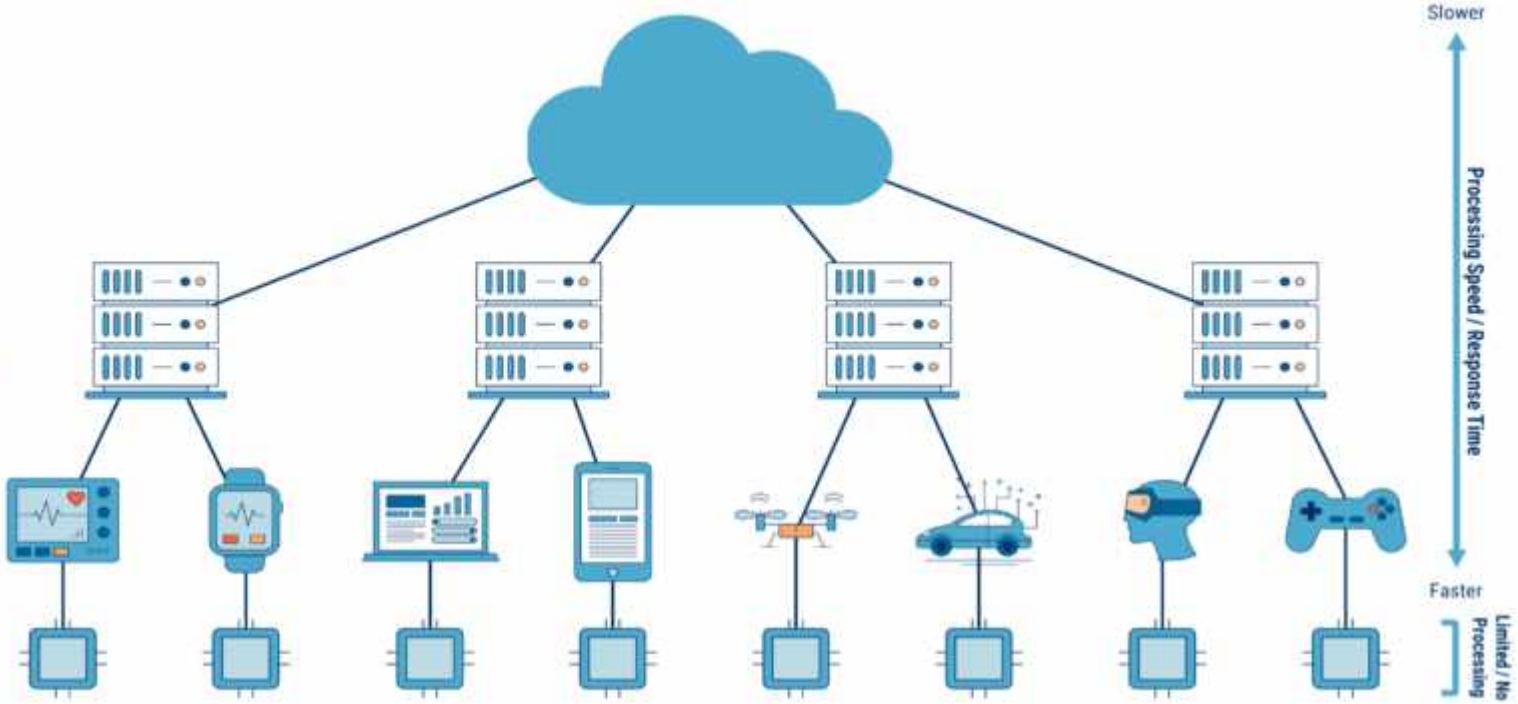
The edge computing ecosystem is comprised of four primary areas

Centralized Cloud
Centralized data centers are farthest from the network edge. However, they offer a much greater density of compute, storage, and networking resources.

Edge Infrastructure
Small, distributed data centers that provide a resource-dense midpoint between edge devices and the centralized cloud. Low roundtrip latencies of 5 – 10ms.

Edge Devices
Real-time data processing within devices based on application needs. Processing limitations present.

Edge Sensors & Chips
Data collection & origination.



	Fog Nodes Closest to IoT Devices	Fog Aggregation Nodes	Cloud
Response time	Milliseconds to subsecond	Seconds to minutes	Minutes, days, weeks
Application examples	M2M communication Haptics[2], including telemedicine and training	Visualization Simple analytics	Big data analytics Graphical dashboards
How long IoT data is stored	Transient	Short duration: perhaps hours, days, or weeks	Months or years
Geographic coverage	Very local: for example, one city block	Wider	Global

HAYSTACK

About HaystackID

Why HaystackID

HaystackID is a **specialized eDiscovery services firm** that helps corporations and law firms find, listen to, and learn from data when they face complex, data-intensive investigations and litigation.

Focus	Enablers	Markets	Workstreams	Models
Forensics	Experts	Law Firms	Audits	Pay for Use
Early Case Insight	Technology	Corporations	Investigations	Subscriptions
Legal Review	Integration	Governments	Litigation	MSA*

HaystackID serves more than 500 of the world's leading corporations and law firms from North American and European locations. HaystackID's combination of expertise and technical excellence coupled with a culture of white glove customer service make it the **alternative legal services provider** that is *big enough to matter but small enough to care.*

* Managed Services Agreement

Questions and Answers

The Master's Conference™

HAYSTACK