VALORA technologies

PowerHouse™ AutoClassification to Manage Content from 30,000 Employees Worldwide:
Legal Hold, Retention & Classification of Terminated Employee Files

SITUATION
This technology and AI leader has a large and diverse workforce, with many new hires and terminations each year. In keeping with their corporate policies around content management, each departing employee's hard drive, shared and personal file shares, and email stores are forensically imaged and preserved for the long term.

Unfortunately, this approach has two big drawbacks:
1. With a workforce of more than 30,000 worldwide, the forensic images soon stacked up to outrageous data storage proportions exceeding 30 TB.
2. As time went on, no one really knew what those forensic images contained.

CHALLENGE
With increasing litigation, regulation, compliance and data stewardship challenges, this corporation needed a solution that wouldn't break the bank or any policies for data management. That's when they turned to Valora.

SOLUTION
Throughout this project, PowerHouse processed 100s of GB of forensically imaged data every day in order to identify and tag the content across a wide array of metadata attributes, such as Document Type, Subject, and Topic Areas, Custodians, Authors, Recipient & Copyees (CC/BCC), PII and sensitive data, Business Unit, Office Location, and Employee Role

SOLUTIONS APPLIED:
- AutoClassification
- Document Analytics
- Electronic File Processing
- OCR & Text Extraction
- Analytics & Data Mining

PRODUCTS USED:
- PowerHouse
- BlackCat

After the initial data extraction and analysis, the PowerHouse Rules Engine made multiple management dispositions for each file, such as: ROT, Relevance & Privilege, Retention & Legal Hold, Data Privacy & Security.

Once attributes and disposition are assessed, preview images, searchable text and the myriad attribute and rules tags are sent for rich data display in Valora's BlackCat data visualization add-on module.

RESULTS
The entire approach exemplifies Valora's Data Under Management™ technique that keeps content up to date, properly tagged and managed on a persistent, evergreen basis.