

## IT MANAGERS REGION 9 SEMINAR

## Day 1 - Session 1

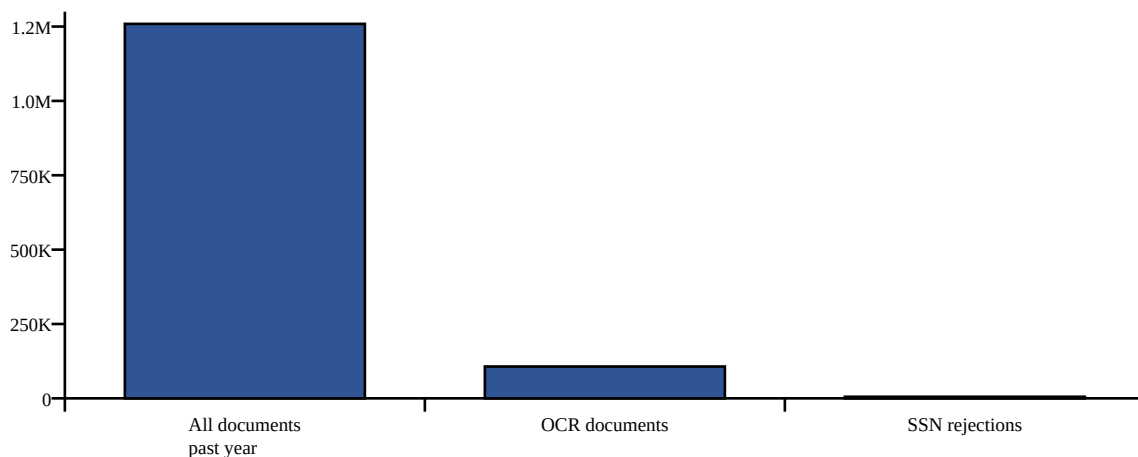
# Bankruptcy Documents - BKDocs.us

**Presenter: Justin A. Crone****Independence Software, LLC****Tuesday, September 22, 2026**

|                       |                                                                                                                                                                                                 |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Session focus</b>  | BKDocs usage statistics, infrastructure resilience, database and server architecture, client modernization, document handling, encryption performance and planned health-monitoring automation. |
| <b>Document style</b> | Detailed organized seminar notes. Discussion is consolidated by subject rather than reproduced as a running transcript.                                                                         |

**Key Takeaways**

- The session combined operational statistics with a detailed look at how BKDocs is engineered and maintained.
- The full recording adds important material that was missing from the earlier 30-minute capture, particularly OCR architecture, hybrid encryption and automated health checks.
- Resilience was a recurring theme: limited public exposure, ISP DDoS mitigation, RAID, replication, backups, encryption and validation were described as separate layers rather than a single control.



## 1. System Usage, Office Comparisons and Document Volume

Justin opened with office-by-office usage figures intended to give the Region 9 group a sense of how heavily each trustee office uses BKDocs. The comparison separated cumulative documents since an office joined the service from the most recent 12-month activity. He also showed the information graphically, using one series for total historical volume and another for the prior year.

The broader system processed approximately **1,259,000 documents** during the prior year. The discussion made clear that office policy affects volume: offices that require nearly every document to be submitted through the service naturally produce substantially higher counts than offices that still accept paper or use other channels.

## OCR and Social Security Number screening

Approximately **107,000 documents** went through OCR processing, and about **5,300** were rejected because Social Security numbers were detected. That is roughly five percent of the OCR population. When a document is rejected, it is deleted and the submitting user receives an email notification.

Justin said the average OCR processing time for a typical document of roughly 15 pages had been reduced to about 30 seconds. Larger uploads, such as complete tax-return packages that include instruction pages, naturally take longer.

## 2. DDoS Event and ISP Traffic Scrubbing

Justin described a June distributed denial-of-service event directed at a backup database server. The targeted server was not a normal public application server; it primarily held a backup database and also performed some OCR work. Its exposed access was tightly limited, including SSH on a nonstandard port restricted to a single IP address.

The ISP detected the abnormal inbound traffic and moved it into its traffic-scrubbing environment. Legitimate traffic could be passed through, but because the machine did not host public services, essentially none of the attack traffic represented legitimate use. The two incidents together lasted roughly 12 minutes and were contained without reported damage to the server.

Justin did not claim to know why that specific IP address was targeted. He noted that ISP addresses can be recycled and speculated that the address may have had some prior history before the current server had it. The practical point was that the service provider's upstream mitigation and the server's limited exposure kept the event from becoming a larger outage.

## 3. July Database Drive Failure and Recovery

Over the July 4 weekend, a hard drive failed in the primary database server. The database storage was configured as RAID 1, so the data was mirrored, but Justin intentionally took the service offline for maintenance rather than continuing normal operation while the array was degraded.

Before hardware replacement, he made an additional database backup while the database was still accessible and compared it with the replicated backup database. The primary and backup servers were expected to be one-to-one copies, and that comparison provided another integrity check before proceeding.

The hosting provider replaced the failed drive under its hardware service commitment. After the server returned, Justin rebuilt the RAID array, verified database integrity, and checked synchronization against the last replication point before bringing the service back online. Total downtime was described as roughly five to six hours.

The disk volumes are encrypted at the operating-system level. The failed physical disk therefore contained encrypted data when removed, and the provider's process destroys/recycles the failed media. The incident was used to illustrate why mirrored storage, replication, backups, encryption and documented recovery checks all matter together.

## 4. Database Platform: MariaDB Today, PostgreSQL Under Evaluation

BKDocs moved away from MySQL years ago and currently uses MariaDB. Justin discussed the supported life of the production and backup database versions and the plan to keep the environment on supported releases.

PostgreSQL is being explored rather than presented as an immediate migration. Justin cited security characteristics, backend performance and PostgreSQL's mature JSON capabilities as reasons it is attractive. API-driven services often return delivery status and other results as JSON, so native querying of JSON data is useful.

He also discussed prior experience with PostGIS and geospatial workloads as an example of PostgreSQL's depth. At the same time, he emphasized that a conversion would not be trivial. PostgreSQL is stricter about some datatypes, so existing data would have to be carefully mapped, converted and repeatedly validated. The likely benefits are primarily in backend processing rather than an obvious change users would see on the front end.

## 5. Server Farm, Role Separation and Private Backend Communication

The BKDocs environment has multiple servers with distinct responsibilities, including application, API, database, backup/storage and supporting processing roles. Two newer servers were added as replacements for aging hardware rather than simply to increase the server count.

The API server was separated from the public application server after usage grew and high-volume download periods began producing latency. Separating those workloads reduced competition between interactive public activity and office download/API activity.

The servers are hosted with the same provider but are not all in the same physical location. They communicate with one another over a private virtual network rather than exposing database traffic directly to the public Internet.

### **Document flow through the environment**

When a debtor, attorney or other user uploads a PDF, it first reaches the application layer. The system performs its checks, including malware scanning and Social Security number screening, encrypts the document, and then transfers it across the private backend network to storage.

When a trustee office requests a document through the API/client, the request is handled through the API layer and the document is retrieved from backend storage. Database queries likewise remain on the private backend network; the database itself is not directly exposed to the Internet.

## **6. PHP and Java Modernization**

The web application had remained on PHP 7.4 longer than desired because of older libraries and dependencies. Some libraries had to be replaced, updated or rewritten. Ubuntu's backported security fixes provided some protection during the transition, but Justin described the eventual move to PHP 8.5 as a significant modernization effort.

The trustee client is Java-based. The publicly available client at the time of the presentation was described as version 2.14 using Java 21. A coming client release was expected to move to Java 25 using the Eclipse Temurin runtime, in part because of licensing considerations.

Justin noted that future client releases may occur more frequently even when application code has not materially changed, because the bundled runtime itself may need to be updated.

## **7. Malformed PDFs and a Safer Failure Mode**

Some PDFs display normally in Acrobat or another viewer even though their internal structure is malformed. Utility-company bills were mentioned as a recurring historical example. The problem becomes visible when BKDocs attempts to add its case/date stamp layer during download.

Rather than repeatedly trying to repair or alter the original PDF simply so the stamp can be applied, the planned behavior is to fail more gracefully: record the condition, skip the stamp when necessary, and continue delivering the underlying document. This avoids turning a stamping problem into a document-delivery failure.

## **8. Chapter 12 / Multiple Database Support**

The discussion covered offices that also handle Chapter 12 matters and therefore need more than one trustee/database configuration. The client can support a second database and separate scheduled processing.

A prior change involving names/security keys created a condition in which Chapter 12 exports could fall back to Chapter 13 keys. A separate installation/directory was used as a practical workaround while the multi-database behavior was corrected. The larger point was that the client must keep database identity, keys and scheduled jobs isolated when one office operates multiple trustee datasets.

## **9. File Deletion Issue**

Justin described a minor client issue in which files were not being deleted as expected because a line of code had inadvertently been commented out. Logging helped identify the cause. The fix restores the intended deletion behavior.

Until corrected behavior is deployed everywhere, offices seeing retained files should verify the deletion status and use manual cleanup where needed rather than assuming the normal aging/deletion process has completed.

## **10. OCR Engine Optimization - Detail from the Continuation**

The later portion of the presentation explained why OCR performance improved. The older design started a new OCR-engine instance for each document, so every job paid the startup cost and threads could compete inefficiently for work.

The revised design keeps multiple OCR engines available and uses a weighted scheduler to distribute documents across them. That avoids repeatedly initializing the engine and reduces situations where one worker receives disproportionate work while others sit idle.

Justin connected this architecture change directly to the improvement from roughly 50-second processing toward approximately 30 seconds for an average document.

## **11. Hybrid Encryption to Improve Large Download Performance**

Justin also discussed changing the encryption design to improve high-volume download performance. The concept combines fast symmetric AES encryption for the document itself with asymmetric encryption used to protect the AES key.

The key can be rotated across configurable batches instead of performing the more computationally expensive asymmetric operation against every complete file. For offices downloading large batches, particularly on older hardware, he estimated that work taking roughly 10-20 minutes under the older approach could potentially fall below five minutes.

The security goal is not to remove encryption; it is to use the appropriate cryptographic method for the bulk data and reserve the expensive asymmetric operation for protecting the key.

## **12. Automated Office Health Checks / PK2s Concept**

A future project discussed with Debbie Smith would automate recurring checks that office administrators may currently perform manually. Examples included monitoring free disk space, identifying missing database indexes and flagging case/data inconsistencies.

The intended model is exception-based monitoring: the system repeatedly evaluates known conditions and tells the office when something requires attention. That would reduce dependence on an administrator remembering to run a checklist and could surface problems before they become user-visible failures.