

서버 內 개인정보검색과 검색관리

주식회사 프렌트리
장 창영 대표



개인정보 어디에 있나요?

개인정보는 암호화, 가명/익명화/삭제 등 보호조치를 해야 합니다.

그러려면 먼저 개인정보를 식별해야 합니다.

지금 사용하고 있는 PC에 개인정보가 있겠죠?

어떤 폴더, 어떤 파일안에 개인정보가 있을까요?

그러면 더 복잡하고 다양한 IT 환경에서는 어떨까요?

개인정보를 먼저 찾아야 → 보호조치가 가능합니다.



Enterprise 저장소 - 서버, 클라우드, 데이터베이스, 빅데이터



User Storage - Workstations, Cloud, Email



서버 內 개인정보 검색을 위한 기반기술

- ✓ 검색엔진
- ✓ 검색관리를 위한 워크플로우

검색엔진

다양한 종류의 파일들 내에 있는 개인정보를 잘 읽어 낼 수 있느냐?

- MS오피스,HWP
- Open Source - Star Office / Open Office / Libre Office
- Open standards - PDF, RTF, HTML, XML, CSV, TXT
- File Databases - Access, DBase, SQLite, MSSQL MDF & LDF
- Backup - Microsoft Binary / BKF
- Images - JPG, GIF, BMP, PNG, TIFF, PDF (embedded), FAX**
- Compressed TAR, ZIP (all types), Gzip (all types), Bzip2, ALZ/EGG
- Generic binary and ASCII text file support

→ Native format decoding

서버 內 개인정보 검색을 위한 기반기술

- ✓ 검색엔진
- ✓ 검색관리를 위한 워크플로우

검색엔진

운영중인 서버에 부하를 주지 않고 검색 작업을 잘 수행할 수 있느냐?

- 개인정보검색보다 업무 프로세스가 우선합니다.
- CPU/MEM를 최소한으로 사용하여 검색작업이 수행되어야 합니다.
- 다중코어 CPU의 경우 1코어만 사용하여 검색할 수 있도록, 엔진성능이 높아야 합니다.
- 개인정보검색과 업무프로세스 사이의 우선순위를 설정할 수 있어야 합니다.

→ C/C++ 등 기계 친화적인 언어로 개발된 엔진

서버 內 개인정보 검색을 위한 기반기술

- ✓ 검색엔진
- ✓ 검색관리를 위한 워크플로우

검색엔진

대용량, 다중 압축파일에 대한 검색을 잘 수행할 수 있어야 합니다.

- 가장 위험한 파일은 다중 압축하여 보관중인 백업파일들 입니다.
- 다중 압축된 파일을 잘 읽어내기 위해서는 네이티브 포맷 디코딩이 필수 입니다.
- 큰 파일을 메모리로 직접 불러 올리면, 서버에 장애를 발생 시킬 수 있습니다.
- 파일을 작은 블록으로 나누어서 읽을 수 있는 기술이 필요합니다.

→ File chunk, Native format decoding

서버 內 개인정보 검색을 위한 기반기술

- ✓ 검색엔진
- ✓ 검색관리를 위한 워크플로우

검색엔진

다양한 인프라 환경을 지원할 수 있어야 합니다.

- Windows, Linux, UNIX등 모든 OS를 지원할 수 있어야 합니다.
- AWS,Azure,GCP,Openshift,K8S등 Public/private cloud도 모두 지원해야 합니다.
- 특히, PaaS환경의 경우에는 검색결과의 위치를 특정할 수 있는 기술 또한 필요합니다.
- Oracle, MSSQL등 상용되는 모든 DBMS를 지원해야 합니다.

→ 제조사의 역량과 지속가능성

서버 內 개인정보 검색을 위한 기반기술

- ✓ 검색엔진
- ✓ 검색관리를 위한 워크플로우

검색관리를 위한 워크플로우

개인정보보호 담당자와 현업부서와의 소통이 무엇보다 중요합니다.

- 수백/수천대의 서버들을 대상으로 개인정보를 검색하는 업무는 쉽지가 않습니다.
- 조직의 규모에 따라 정책설정,검색수행과 결과확인,조치 등 업무의 분장이 필요합니다.
- 이를 뒷받침하는 워크플로우(UI)를 제공해야만 원활한 소통이 가능합니다.

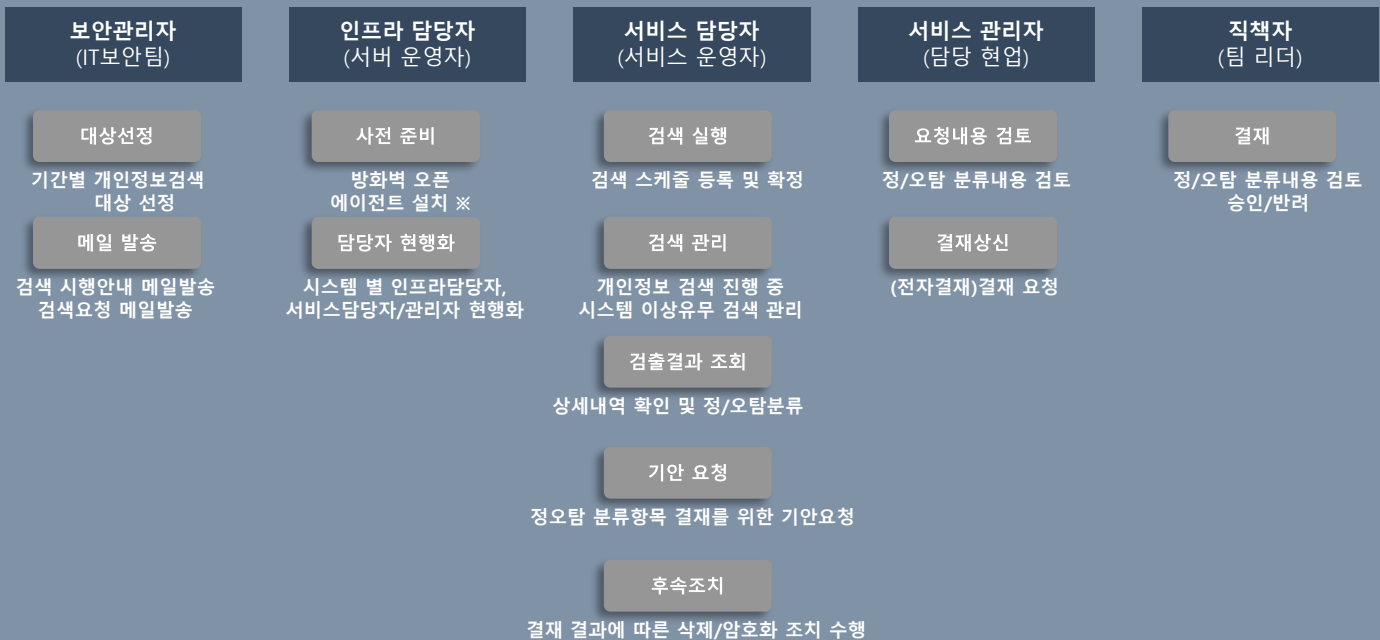
➔ 성공적인 개인정보검색은 소통이 중요

서버 內 개인정보 검색을 위한 기반기술

- ✓ 검색엔진
- ✓ 검색관리를 위한 워크플로우

검색관리를 위한 워크플로우

다양한 조직규모와 업무분장을 지원할 수 있는 프로세스를 지원해야 하며, 인사DB, SSO등 유관 시스템들과의 연동을 지원해야 합니다.



맺음말

개인정보보호는 개인정보가 인프라 內 어디에 있는지를 찾아내는 것이 우선합니다.

이를 위해서는,

안정적인 검색 엔진을 기초로 하여, 검색결과에 대한 사후 관리가 가능한 워크플로우가 필수적입니다.

감사합니다