splunk> turn data into doing™

안녕하세요. 스플렁크 코리아입니다.

1월 17일 토크웨비나 생방송 중 진행된 Q&A 답변을 정리하여 드립니다.

[DSP, DFS 관련]

Q. dfsjob 명령어를 사용하니깐 기능이 비활성화라고 나오는데요, 활성화는 어디서 하나요?

A. DFS는 스플렁크의 UI/SPL는 그대로 사용하면서 외부의 DFS노드(Spark Cluster)의 리소스를 활용해서 대량 데이터 분석/조인 등의 작업을 가속화하는 기능입니다. 따라서 별도의 dfs 노드를 설치/구성하고 연결을 정의해야 합니다. 설치 등의 내용은 하기의 링크를 참조하시기 바랍니다. https://docs.splunk.com/Documentation/DFS/1.1.0/DFS/Overview

Q. DSP, DFS 흥미롭네요..사전 준비사항이나 limit확인은 어디가 할까요.?

A. 문서 사이트를 참고하실 수 있습니다.
https://docs.splunk.com/Documentation/DSP

https://docs.splunk.com/Documentation/DFS/1.1.0/DFS/Overview

Q. dfs 나 dsp를 쓰려면 기존 라이센스와는 별도로 구매가 필요한 것인가요?

A. 네, 다른 프리미엄 앱처럼, DFS,DSP 각각 별도의 가격 체계가 있습니다.
DFS: https://www.splunk.com/en_us/software/pricing/faqs.html#splunk-dfs
DSP: https://www.splunk.com/en_us/software/pricing/faqs.html#splunk-dsp

---

[스플렁크 플랫폼]

Q. 수집시 압축기능및 로그파일이 있을텐데 별도 에디터등으로 쉽게 분석이 가능한가요?

A. 네. 스플렁크는 수집시 원본데이터(로그데이터등 텍스트데이터)를 압축 저장하면서 원본. 문자열별로 reverse index를 생성합니다. 이는 스플렁크의 UI를 통해SPL(Search Process Language)라는 명령어를 이용해서 간단히 할 수 있습니다. SPL의 활용 방법등은 하기의 링크를 참조해 주세요.

https://conf.splunk.com/files/2017/slides/power-of-spl.pdf

Q. 실시간으로 방대한양으로 수집되는 빅데이터도 자유자재로 분석이 가능할지요?

A. 네. 많은 저희 고객들이 Splunk로 분석하고 계십니다. 해외의 대형 고객 중 한 곳은 하루 Splunk로 수집되는 데이터량이 8페타바이트를 넘는 고객도 있습니다.

Q. API 연동 등을 통해서 외부 프로그램에서 데이터 연동해서 받을 수 있나요?

A. 네, 가능합니다. https://dev.splunk.com/enterprise/ 를 참조하시면 API Reference와 각 언어별 SDK 를 사용하시는 가이드가 잘 정리되어 있습니다.

Q. 스플렁크 또한 크버네티스와 도커를 이용하여 구축 가능한가요? 구축시 쉬운가요? / 관련 샘플 매누얼 받아볼수 있을까요?

A. 네 가능합니다. 도커환경의 스플렁크도 서포트되며, Kubernetes Operator 를 사용해서 쿠버네티스 환경에 손쉽게 배포하고 사용하실 수 있습니다. 하기의 링크를 참조하시면 좋을 것 같습니다.

https://github.com/splunk/splunk-operator
https://github.com/splunk/docker-splunk
https://github.com/splunk/splunk-connect-for-kubernetes

Q. 여러 종류의 비정형로그데이터를 어떠한 정제 및 가공작업없이 바로 불러들여서 사용이 가능한건가요?

A. 네 맞습니다. Predefine된 데이터타입에 대한 자동 필드 추출 외에도 필요한 필드를 UI에서 정의하거나, regex를 이용해서 정의하고 분석하실 수 있습니다.

Q. 스플렁크로 쌓인 데이터들을 분석하여 원하는 결과를 실시간으로 RDBMS로 전송하거나 Front단에서 해당 결과를 보여줄 수 있는지요?

A. Splunk 쿼리 실행 결과를 DB에 넣어 줄 수 있습니다. "Front단"이 어떤 의미이신지 모르겠습니다만, WebUI나 Mobile app등을 이용해서 Splunk쿼리 결과를 볼 수 있습니다.

A. db connect를 이용해서 스플렁크의 데이터를 Database로 전송/적재하거나 대시보드를 이용하면 될 것 같습니다.

Q. 데이터 소스들 어떤 것들로부터 받을 수 있나요? 예를 들어 구글 아날리킥도 가능할까요?

A. 기본적으로 텍스트로 읽을 수 있는 모든 종류의 데이터에 대한 수집,분석이 가능합니다. File Tailing, Script Input, Syslog, SNMP, HTTP Event Collect, Modular Input등의 다양한 수집 방법을 지원하며, 각 소스 별로 미리 만들어진 수집기(Addon) 를 splunkbase.splunk.com 에서 다운받아서 활용하실 수 있습니다. 문의하신 Google Analytic도 app이 나와있습니다. (https://splunk-base.splunk.com/app/4174/)

Q. ETL 쪽에 많이 쓰이는 추세인가요?

A. 데이터에 대한 정제 및 정규화 작업이 수집 전 설계과정에서 필요 없는 스플렁크의 Schema on the Fly 기능때문에 어떤 고객은 Splunk를 ETL처럼 사용하는 고객도 있습니다만, 제대로 사용하는 고객은 수집/저장/분석/대시보드/Alert 등 다양한 전체 기능을 활용하고 계십니다.

Q. IT조직의 급속한 발전 속에 AI와 머신러닝을 이용한 기업의 협업및 자동화를 효과적으로 지원하기위한 개발/운영주기 통합 모델링 구현방안 궁금합니다.

A. 스플렁크는 지난 4~5년간 보안분야에서 내부자 통제를 위한 AI 솔루션인 UBA, 머신러닝을 IT운영분야,보안관제 분야에 쉽게 적용하도록 도움을 주는 ITSI(IT Service Intelligence), ES(Enterprise Security) 그리고 운영 협업을 위한 VictorOps와 자동화분야의 Phantom등의 다양한 솔루션들을 개발 및 인수하여 고객의 업무환경에 대한 효과적인 지원을 하고 있습니다.

Q. 웹 서비스나 동영상 스트리밍 서비스에 지연요소분석이나 트래픽 급등 시 즉시 알람 및 급등 콘텐츠분 석하는데도 활용 할 수 있는지요?

A. 네 가능합니다. conf.splunk.com 사이트에 보시면 저희 글로벌 고객들의 많은 usecase들을 보실 수 있습니다.

[스플렁크 특장점]

Q. 데이터수집할경우 하둡과의 차이점이 있다면 어떤것이 있을까요? 예를들어 인덱싱, 필터링하여 찾는 방식 등등

A. 데이터 형태에 관계없이 하나의 형태로 저장/인덱싱을 수행한다는 점이 매우 다릅니다. 하둡의 경우 Flume/sqoop등의 수집기를 통해 하둡의 저장공간 HDFS에 저장후 ETL,M/R

프로세스 과정을 통해 데이터를 정제한 후 SQL on Hadoop등의 추가적인 분석엔진을 통해 분석하고 이를 다른 시각화 툴등에서 활용하는 데이터 파이프라인을 일일이 구축해 나가는 플랫폼이라면, 스플렁크는 정제등의 과정없이 스플렁크에 인덱싱한 후 인덱싱 데이터에 대한 검색/분석/시각화/조건에 따른 경보 생성등 전체 프로세스를 단일 제품내에서 지원하는 플랫폼이라고 보시면 될 것 같습니다. 또한 워크로드 측면에서도 하둡이 배치성 작업 중심이라면, 스플렁크는 검색 중심의 워크로드 처리에 강점을 보인다고 보시면 좋을 것 같습니다.

Q. 클라우데라와 비교하면 어떤 장점이 있나요?

A. 클라우데라는 하둡 배포본 업체이므로, 하둡과 Splunk을 차이점을 말씀하시는 것으로. 생각됩니다. 위의 질문 중 하둡과의 차이 질문에 대한 답을 참조 부탁드립니다. 많은 장점이 있습니다만, 가장 큰 차이는 "생산성"이라고 말씀드릴 수 있습니다.

Q. AI를 위해서는 데이터가 중요하며 빅데이터 처리를 위한 솔루션이 매우 핫할 것으로 생각합니다. 최근 빅데이터 수집, 처리할 수 있는 ELK Stack이 핫 한데요. Splunk에서는 ELK 스택과 기능, 성능적인 면에서 어떤 차별점이 있는지요?

A. 기능, 성능 측면에서 많은 차이가 있습니다.기능 측면에서는 확장성과 사용성, 대시보드, 드릴다운 지원 등 많은 차이가 있습니다. 하나 예를 들자면, lookup과 같은 기능이 있어서 데이터를 enrich할 수 있는 기능은 ELK사용하시던 분이 Splunk를 쓰게되면 매우 좋아하시는 기능 중 하나입니다.
성능 역시 두 제품의 특성이 다릅니다. ELK를 적은양의 데이터를 간단하게 조회하는 상황에서 장점이 있으나, 노드당 저장할 수 있는 데이터 크기에 한계가 많고, 대용량 처리에 문제가 생깁니다. 수집시에도 json으로 포멧을 바꿔야 하는 것도 힘든 점 중 하나 입니다.

Q. Elastic Search와 비교를 해주실 수 있을까요?

A. 서로 출발점이 다른 제품이라(Elasticsearch는 검색엔진, 스플렁크는 시계열 인덱싱) 1:1로. 비교하기엔 무리가 있지만 관련된 내용에 대한 하기 영어 세션을 참조하시면 좋을 것 같습니다.
https://conf.splunk.com/watch/conf-online.html?search=FN1455#/

Q. Splunk는 그럼 데이터 저장은 하둡과 다른 컨셉으로 진행되는건가요? 전혀 다른 빅데이터의 새로운 플렛폼으로 이해하면 되는지 궁금합니다.

A. 네. Splunk는 특허를 보유한 자체 데이터 저장 포맷 (타임시리즈 인덱스 구조)을 가지고 있습니다. HDFS 파일시스템을 이용하는 하둡과는 다르게 일반적인 파일시스템 스토리지에 설치가 가능합니다. 설치 가능 OS/filesystem등은 하기의 링크를 참조해 주세요.
https://docs.splunk.com/Documentation/Splunk/8.0.1/Installation/Systemrequirements

Q. 스플렁크의 대안, 대체(?) 를 위한 로그분석 툴(?, Elastic Stack, 로그신, 로그자이오, 큐박스 등등)이 언급되고 있습니다. Splunk 만의 특화된 점이 어떤 것이 있을까요?

A. 어떻게 보면 스플렁크는 이런 로그 분석 솔루션의 할아버지격? 으로 이 시장을 만든 솔루션이라고 말씀드릴 수 있을 것 같구요, 하나하나 나열하기 힘들 것 같습니다만, 가장 중요한 점은 조직에 적용했을 때 나오는 "생산성"이 아닐까 싶습니다. 또한 계속 스플렁크에서 나가는 방향이 로그분석 과 이에 연결되는 다양한 작업들(예: 협업, 자동화, 모니터링) 에 대한 생산성 증가를 위한 제품들, 그리고 단순한 로그를 검색 하는데서 나아가서 이를 IT운영 분석 모니터링이나 보안 관제에 쉽게 활용하기 위한 ITSI나 ES등의 솔루션들은 타 경쟁업체가 아직 가지지 못한 기능들이라고 말씀드릴 수 있을 것 같습니다.

[대시보드 관련]

Q. Splunk에서 대시보드 성능 개선을 위해서savedsearch를 이용해 대시보드 내 검색문을 줄이는 방법이 특별히 좋은 규칙이나 장점이 뭘까요?

A. 대시보드 성능 개선은 다양한 측면이 있습니다. datamodel 가속화나 report 가속화, summary 등 다양한 검색 가속화 기능을 적용하거나, base search를 적용해서 한번 수행한 검색 결과를 재활용하는 것도 방법입니다.
DB에서 SQL tuning이 매우 효과적인 튜닝 방법인 것처럼, Splunk에서 SPL튜닝 또한 매우 효과적일 수 있습니다.

Q. 상당히 괜찮아 보입니다. 대쉬보드는 커스터마이징을 어디까지 할 수있을까요?

A. 기본 UI에서 편집하는 것으로도 대부분의 IT운영, 보안관제에 필요한 대시보드는 만들 수 있습니다. 기본 UI framework인 SimpleXML 은 일반 웹 기반 어플리케이션으로 javascript나 css를 통해 커스터마이징 할 수도 있으며, Splunk Dashboard App https://splunkbase.splunk.com/app/4710/ 을 사용하면 보다 정교한 완전 커스터마이징을 Drag-and-

---

Drop형태로 구현할 수도 있습니다. 고급 대시보드 작성을 위해서는 Splunk 웹UI가 가지고 있는 웹프레임워크를 이용할 수 도 있습니다. 자세한내용은 https://dev.splunk.com을 참고하시면 됩니다.

[기타 질문]

Q. SPL 을 이용한 튜닝을 잘하는 파트너가 있나요? 혹은 스플렁크에서 지원을 하나여?

A. 국내 23개 스플렁크 파트너사의 100여명 이상의 엔지니어 풀을 가지고 다양한 고객 지원을. 하고 있으며, 저희 Professional Service를 계약하시면 보다 고품질의 서비스를 제공받으실 수 있습니다.

Q. 다양한 데이터 수집 및 분석에 있어서 GDPR 과 같은 개인정보보호에 대한 대응방안이나 솔루션을 제공하는지요?

A. 아래 링크에 있는 저희 고객 사례를 참고해보시면 좋을 것 같습니다.
https://conf.splunk.com/watch/conf-online.html?search=GDPR#/

Q. 기업의 다양한 보안, 문서, DB 등 Indexing data를 hashing 기술로 자동화 할 수 있도록. 포함 되었나요?

A. 인덱싱된 데이터의 보호로 이해하면 될까요? 스플렁크는 다양한 권한 관리 기능으로 적재된 데이터에 대한 접근제어 기능과 더불어 위 변조 방지, 민감데이터 마스킹등의 기능을 제공하며 다양한 보안 규약을 준수하는 제품입니다.

Q. 최근 성공적인 레퍼런스 적용 시 어떤 효과가 있었는지요?

A. conf.splunk.com 사이트를 참고하시면 저희 글로벌 고객들의 적용 사례와 평가를 보실 수 있습니다.

Q. Splunk AR에서 분석한 내용에 대해 해당 디바이스에 AR로 매핑하기 위해서는 데이터 이외 형상 정보도 관리가 되어야 할 것 같은데요. Splunk AR을 위해 디바이스 형상을 어떤 식으로 관리를 하시는지요?

A. 네, 말씀하신데로 각각의 디바이스에 대한 tag정보는 Splunk Cloud Gateway 를 통해. 등록하고 관리됩니다. 하기의 문서를 참조하시면 좋을 것 같습니다.
https://docs.splunk.com/Documentation/AR/2.0.0/UseSplunkAR/GetStartedWithAR

Q. 스플렁크의 장점은 쉽게 원하는 View를 스스로 만들 수 있다는 것인데, 배울 수있는 커리큘럼이 많지는 않은 것 같습니다. 추천할 만한 커리큘럼이 있으신지요?

A. Splunk Fundamental 1 무료 코스를 먼저 들어보시면 어떨까요? 저희 교육 웹사이트에 접속해서 수강하실 수 있습니다. https://www.splunk.com/en_us/training.html