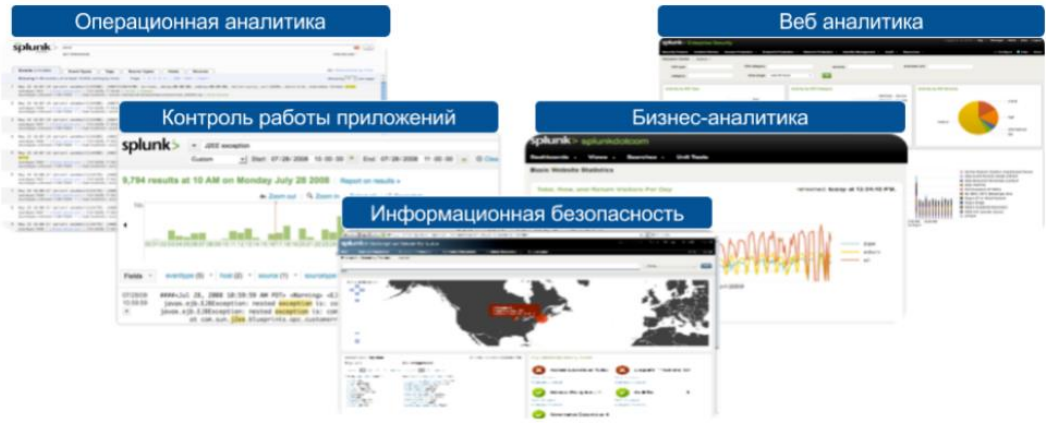
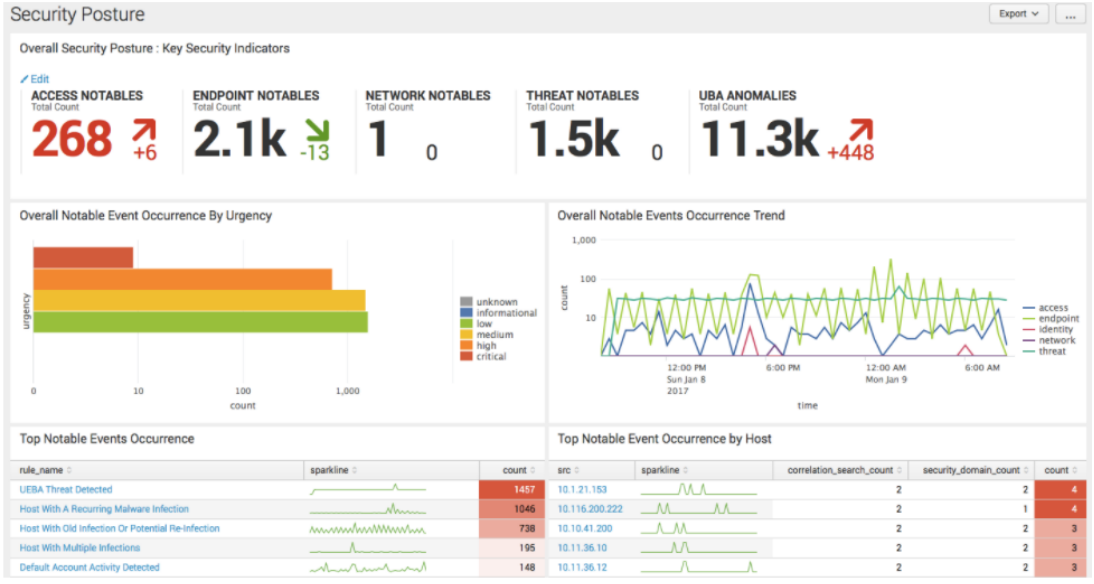
C310. Splunk — общее описание платформы, базовые особенности установки и архитектуры

Splunk это платформа для сбора, хранения, обработки и анализа машинных данных, то есть логов. На сегодняшний день является крайне популярной в США и в Европе и постепенно выходит на другие рынки, включая Россию. Одной из главных особенностей платформы является то, что она может работать с данными практически из любых источников, и поэтому список возможных применений системы очень широк.



Splunk, в большинстве случаев, (автоматически или с помощью аддонов) разбирает входные данные на поля и значения и в последствии обрабатывает их. Обработка происходит посредством SPL запросов (специальный язык от Splunk), с помощью которого можно строить различные выборки и таблицы, сортировать, фильтровать, агрегировать, строить отчеты, создавать вычисляемые поля, обращаться как к внутренним, так и внешним справочникам, создавать дашборды, с широким спектром визуализации и делать алерты (например по результату выполнения запроса отправлять тикеты в Service Desk). Все это можно упаковать в свое персональное приложение.



Основные отличия или сильные стороны Splunk

Real Time Architecture: Splunk осуществляет сбор, поиск, мониторинг и анализ по различным и достаточно большим (сотни Тб данных в день) объемам данных в режиме реального времени и все это — одна система.

Почему это важно? Потому что Splunk может обеспечить сбор данных в реальном времени из тысяч разнородных источников — и это может быть как физический или виртуальный хост, так и облако. Также Splunk поддерживает поиск не только в реальном времени, но и по всему временному промежутку, данные за который были собраны. То есть мы можем осуществлять поиск, мониторинг, оповещения, отчетность и анализ за любое время (исторический данные и данные в реальном времени в одном решении). И наконец, Splunk обеспечивает быстрый результат и высокую интерактивность поисковых запросов на чрезвычайно больших объемах данных.
Universal Machine Data Platform: Splunk является универсальной платформой для машинных данных, которая обеспечивает комплексный сбор данных, их обработку и анализ. Таким образом мы можем индексировать любые машинные данные с отметкой о времени независимо от структуры и формата. Splunk способен объединить в себе машинные данные + бизнес данные + пользовательские данные, что делает его крайне универсальным.
Schema on the Fly: Splunk осуществляет поиски по времени, то есть вам не нужно заранее знать структуру данных, чтобы сформировать запрос. Вы можете выбрать промежуток во времени, ввести пару ключевых слов и быстро ознакомиться с данными. Нет никаких жестких ограничений на столбцы, таблицы и прочее. Это сильно повышает гибкость системы. Также любой запрос можно остановить, поставить на паузу или показать промежуточные результаты.
Agile Reporting & Analytics: Splunk предоставляет широкие возможности по построению аналитики, отчетов и их визуализации. Помимо целевых данных, система также может обращаться ко внешним справочникам, например в SQL БД. Также хотелось бы сказать, что Splunk достаточно открытая система и вы всегда можете дописать свой модуль, хотя возможности визуализации ну очень разнообразны.
Scales from Desktop to Enterprise: Splunk использует технологию MapReduce, что обеспечивает распределение нагрузок и горизонтальную масштабируемость системы, то есть мы можем начать с одного сервера для Splunk, а при увеличении данных — быстро добавить пару новых серверов и распределить нагрузку. Также благодаря технологии MapReduce Splunk может быстро перерабатывать реально большие объемы данных, не требуя выдающегося железа.
Fast Time to Value: Splunk позволяет быстро получить результат от использования. Внедрение занимает часы или дни, а не недели и месяцы. Тоже самое с масштабированием и эксплуатацией.
Passionate & Vibrant Community: У Splunk есть очень качественное, а главное бесплатное комьюнити, которое включает:
Splunk Base — портал, содержащий всевозможные приложения и аддоны, 99% из которых бесплатно
Splunk Answers — форум с большим числом вопросов/ответов и живых участников
Splunk Dev — портал для разработчиков
Splunk Dock — полная база знаний продукта
Бесплатная версия Splunk с ограничением на индексацию в 500 Мб в день доступна на официальном сайте компании, единственное что вам нужно сделать это пройти регистрацию.

Системные требования

Platform	Recommended hardware capacity/configuration
Non-Windows platforms	2x six-core, 2+ GHz CPU, 12GB RAM, Redundant Array of Independent Disks (RAID) 0 or 1+0, with a 64 bit OS installed.
Windows platforms	2x six-core, 2+ GHz CPU, 12GB RAM, RAID 0 or 1+0, with a 64-bit OS installed.

Splunk поддерживает как 32-bit, так и 64-bit разрядную архитектуру. Ниже представлены таблицы с доступными платформами для Splunk отдельно для Unix и Microsoft. В последнем столбце таблицы находится информация о Splunk Universal Forwarder. Это отдельный дистрибутив и отдельная роль в платформе Splunk, которая выступает в качестве агента и отвечает исключительно за сбор логов и пересылку их на сервер.

Unix

Operating system	Architecture	Enterprise	Free	Trial	Universal Forwarder
Solaris 10 and 11	x86 (64-bit)	D	D	D	✔
	SPARC				✔
Linux, kernel version 2.6 and later	x86 (64-bit)	✔	✔	✔	✔
	x86 (32-bit)				D
Linux, kernel version 3.x and later	x86 (64-bit)	✔	✔	✔	✔
	x86 (32-bit)				D
PowerLinux, kernel version 2.6 and later	PowerPC				✔
zLinux, 2.6 and later	s390x				✔
FreeBSD 9	x86 (64-bit)				✔
FreeBSD 10	x86 (64-bit)				✔
Mac OS X 10.10 and 10.11	Intel		✔	✔	✔
AIX 7.1 and 7.2	PowerPC				✔
AIX 6.1	PowerPC				D
HP/UX† 11i v3	Itanium				✔
ARM Linux	ARM				A

A – версия доступна для скачивания, но не имеет официальной поддержки
D – версия в данный момент поддерживается, но в будущих релизах компания может снять ее с официальной поддержки

Windows

Operating system	Architecture	Enterprise	Free	Trial	Universal Forwarder
Windows Server 2008 R2	x86 (64-bit)	D	D	D	D
Windows Server 2012 and Server 2012 R2	x86 (64-bit)	✔	✔	✔	✔
Windows 8, 8.1, and 10	x86 (64-bit)		✔	✔	✔
	x86 (32-bit)		***	***	✔

D – версия в данный момент поддерживается, но в будущих релизах компания может снять ее с официальной поддержки
… — версия поддерживается, но Splunk не рекомендует использовать данную архитектуру

Установка

После того как вы скачали установочный файл просто запускайте установку и по умолчанию система встанет в базовой конфигурации. Подробная пошаговая инструкция по установке на Windows здесь, на Unix системы здесь.

После установки Splunk должен быть доступен через веб интерфейс порт 8000: localhost:8000 и после смены пароля и входа вы увидите следующий интерфейс.