Big Data Analytics

Splunk

Lecture #1

What is Splunk?

Tool for searching, monitoring, and analyzing machine-generated big data

Captures, indexes and correlates real-time data in a searchable repository from which it can generate graphs, reports, alerts, dashboards and visualizations

splunk>

Log Files · Configurations · Wire Data · Alerts · Metrics · Scripts · Changes · Tickets

Sensors · Security · Custom Applications · Networks · Databases · Servers · Smartphones and Devices · Virtual Machines · Web Services

What is Splunk?

Machine Data

      accessible across organization

      identify patterns

      diagnose potential problems

      provide intelligence
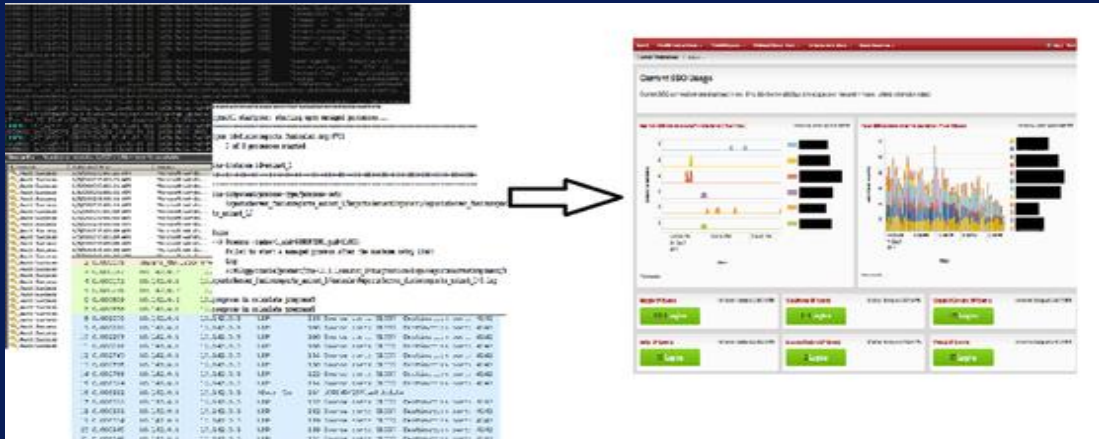
Why Splunk?

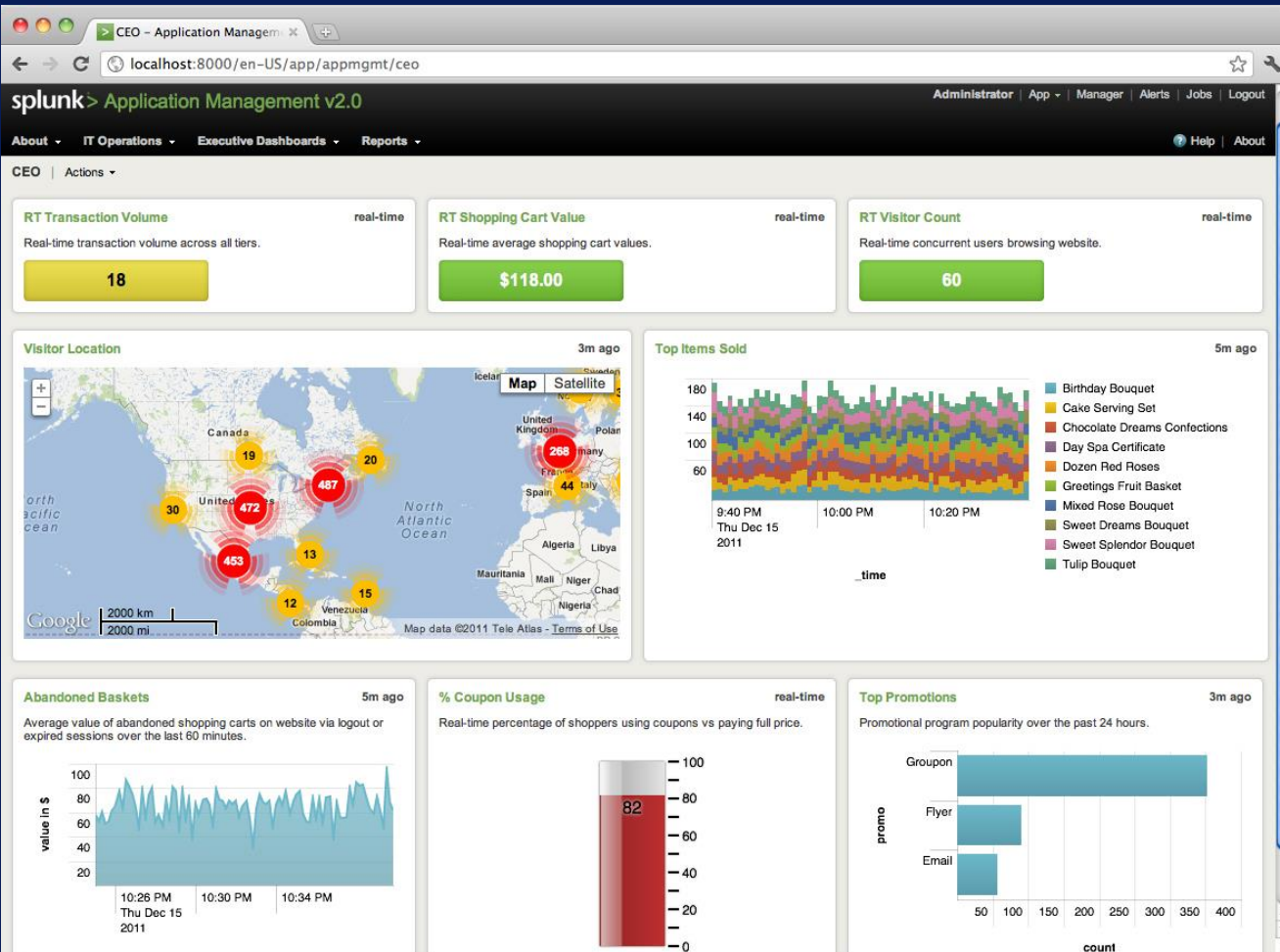Machines produce great volumes of data

From technology to business value

What can Splunk do?

Google-like tool for machine data

- Search

- Investigate

- Troubleshoot

- Monitor

- Visualize

- Alert

splunk> Application Management v2.0

Administrator | App ▾ | Manager | Alerts | Jobs | Logout

About ▾   IT Operations ▾   Executive Dashboards ▾   Reports ▾

? Help | About

CEO | Actions ▾

RT Transaction Volume   real-time

Real-time transaction volume across all tiers.

18

RT Shopping Cart Value   real-time

Real-time average shopping cart values.

$118.00

RT Visitor Count   real-time

Real-time concurrent users browsing website.

60

Visitor Location   3m ago

Map | Satellite

Canada

19   20

487

30   United   472
453
13
12   Venezuela   15
Colombia

Google 2000 km
2000 mi

Iceland

United
Kingdom
268

France   44   Italy
Spain

Algeria Libya

Mauritania Mali Niger Chad
Nigeria

North
Atlantic
Ocean

North
Pacific
Ocean

Map data ©2011 Tele Atlas - Terms of Use

Top Items Sold   5m ago

180
140
100
60

9:40 PM   10:00 PM   10:20 PM
Thu Dec 15
2011

_time

■ Birthday Bouquet
■ Cake Serving Set
■ Chocolate Dreams Confections
■ Day Spa Certificate
■ Dozen Red Roses
■ Greetings Fruit Basket
■ Mixed Rose Bouquet
■ Sweet Dreams Bouquet
■ Sweet Splendor Bouquet
■ Tulip Bouquet

Abandoned Baskets   5m ago

Average value of abandoned shopping carts on website via logout or
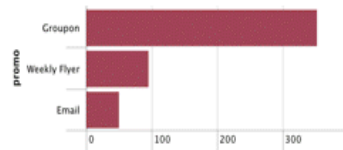expired sessions over the last 60 minutes.

100
80
60
40
20

value in $

10:26 PM   10:30 PM   10:34 PM
Thu Dec 15
2011

% Coupon Usage   real-time

Real-time percentage of shoppers using coupons vs paying full price.

100
80
82
60
40
20
0

Top Promotions   3m ago

Promotional program popularity over the past 24 hours.

Groupon

Flyer

Email

promo

50 100 150 200 250 300 350 400

count

Splunk Features

Central Repository

Data Access for Analytics

Structure and meaning of data

Visualization

Applications

App Dev and App Mgmt. | IT Operations | Security and Compliance | Digital Intelligence | Business Analytics | Industrial Data and Internet of Things

Developer Platform (REST API, SDKs)

splunk>

Splunk Architecture

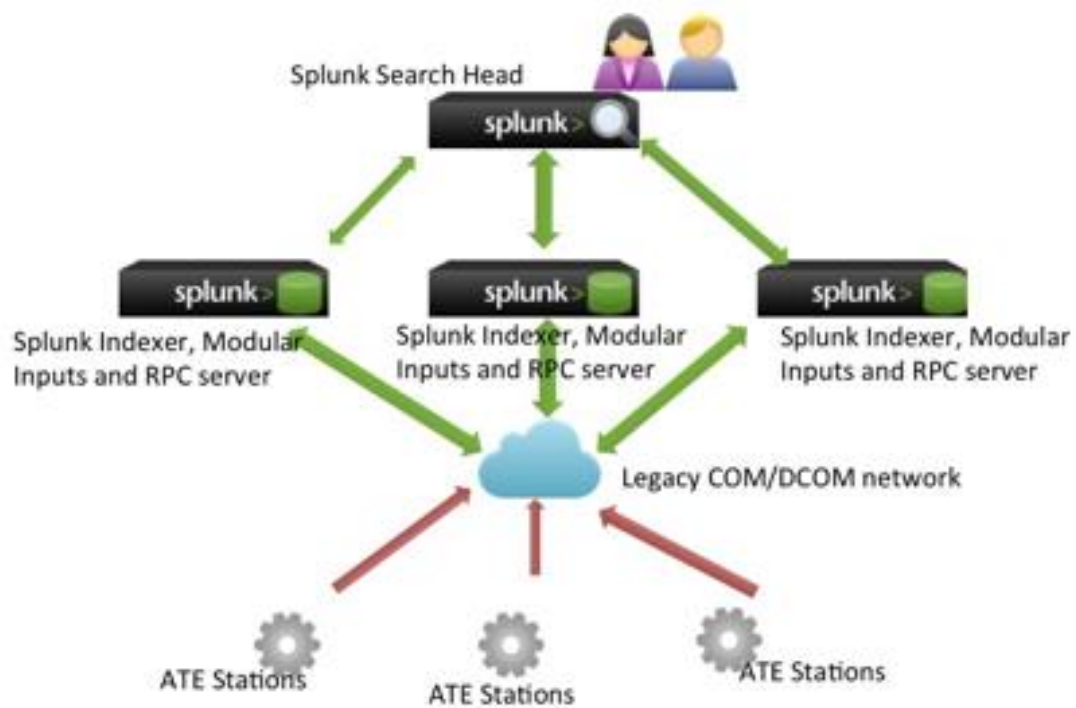4 Major components

     Search Head

     Forwader

     Indexer

     Deployment Server

Splunk Architecture
Search Head

GUI for Splunk search, analysis and reporting

Splunk Architecture Forwader

Universal forwarder (UF)
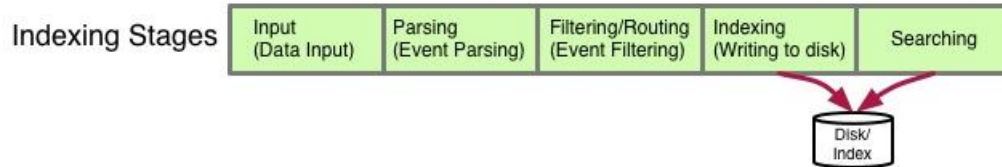
Heavy weight forwarder (HWF)
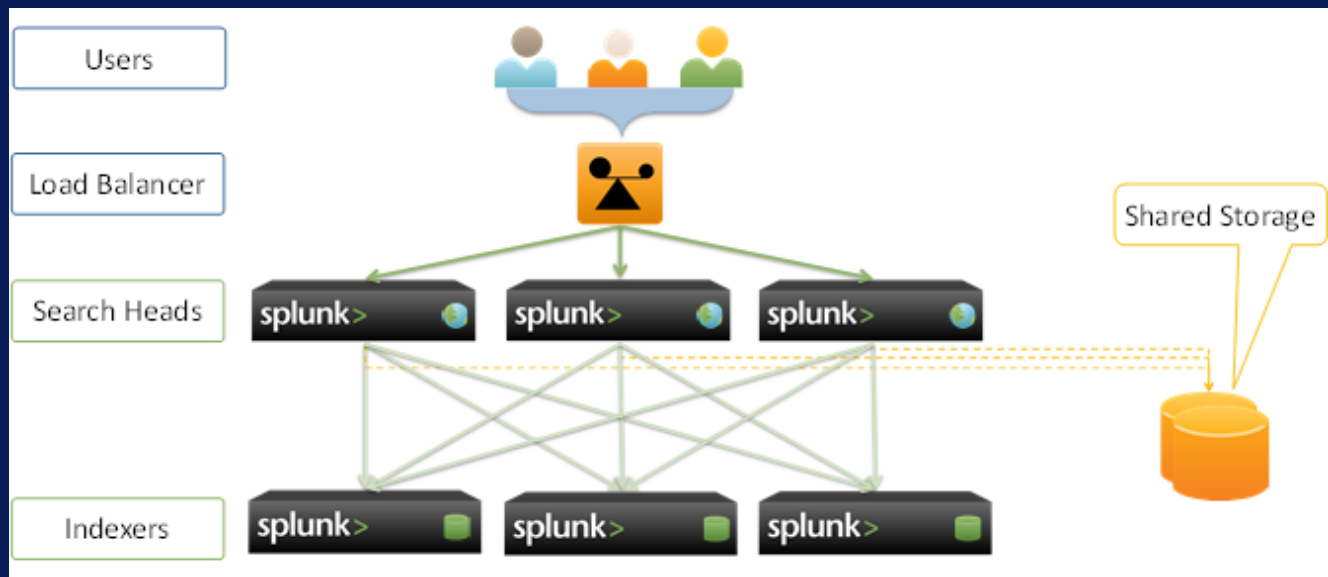
Splunk Architecture Indexer

**Indexing incoming data
Searching the indexed data**

Splunk Architecture Indexer

Indexing Stages

| | Input (Data Input) | Parsing (Event Parsing) | Filtering/Routing (Event Filtering) | Indexing (Writing to disk) | Searching |

Disk/ Index
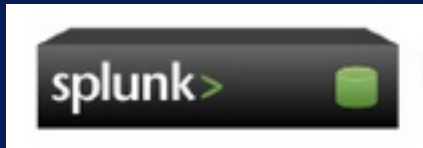
Splunk Architecture Deployment Server

Host and deploy apps to different components

Deploy technology add-ons to forwarders and indexers for index-time knowledge

What are some Splunk Use Cases?

Use Cases

IT and operations

  Index data from firewalls

  Intrusion detection system

Use Cases

Web Analytics

       Web site performance metrics

       Efficacy of the on line promotions

       Web traffic and stream downloads

Use Cases

Internet of Things

   Metropolitan data
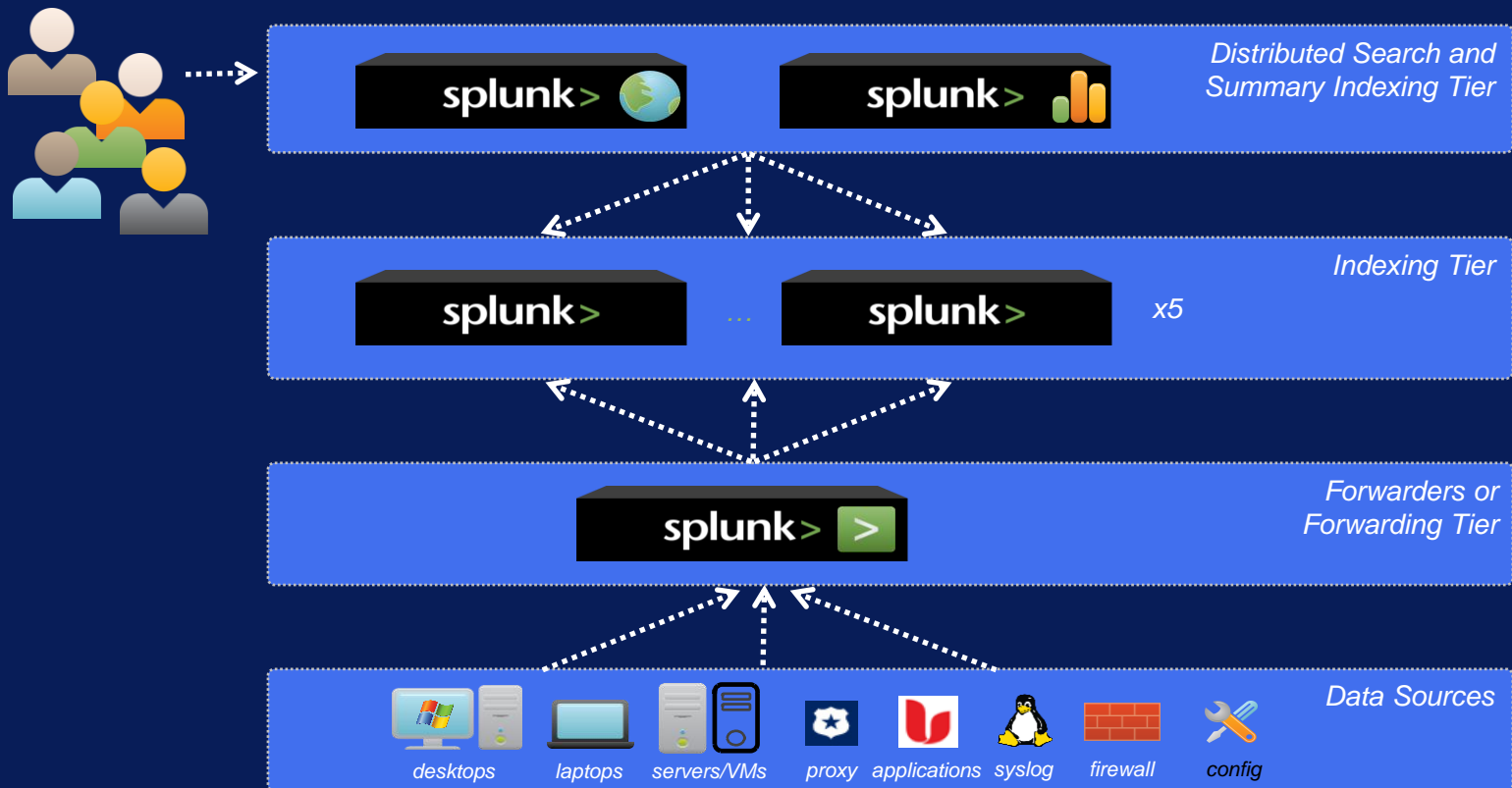
   Wi-Fi enabled Nest

   Elevator usage

Use Cases

Security

    Security log data

    Network log data

Compliance

    Auditing standards

Splunk Architecture

Splunk and Big Data

Data Sources → Splunk for Analytics → Hadoop for ETL