Big Data Analytics

Splunk

Lecture #1

What is Splunk?

By Wikipedia

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

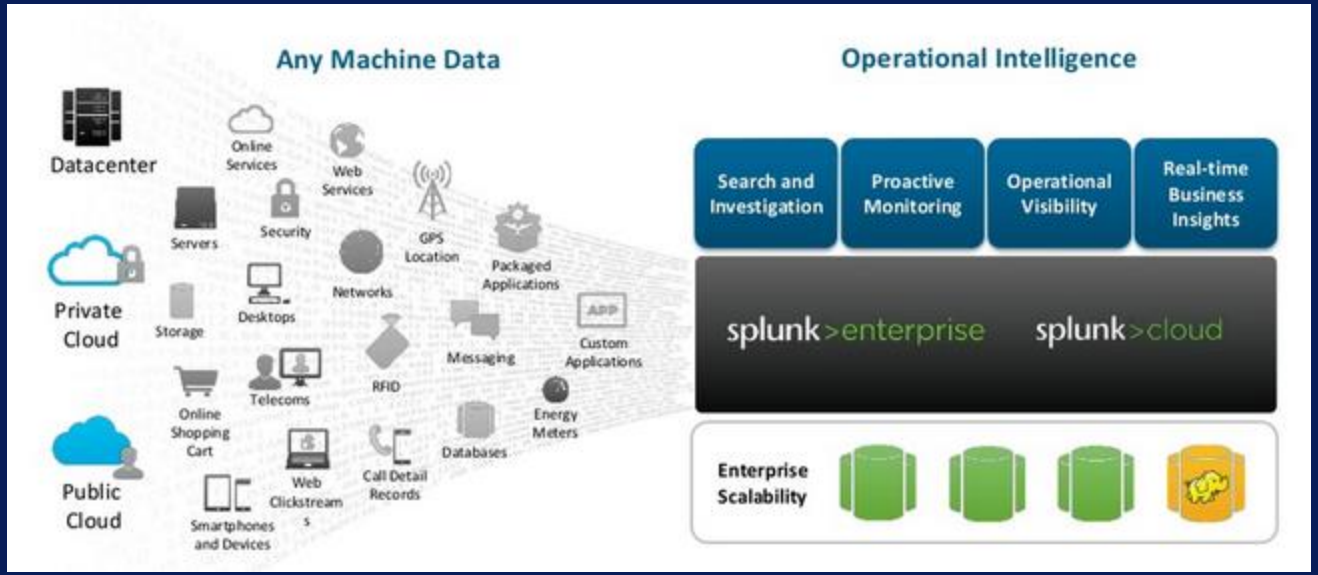

splunk>
Log Files
Configurations
Wire Data
Alerts
Metrics
Scripts
Changes
Tickets
Sensors
Security
APP
Custom Applications
Networks
Databases
Servers
Smartphones and Devices
VM
Virtual Machines
Web Services

What is Splunk?

By Wikipedia

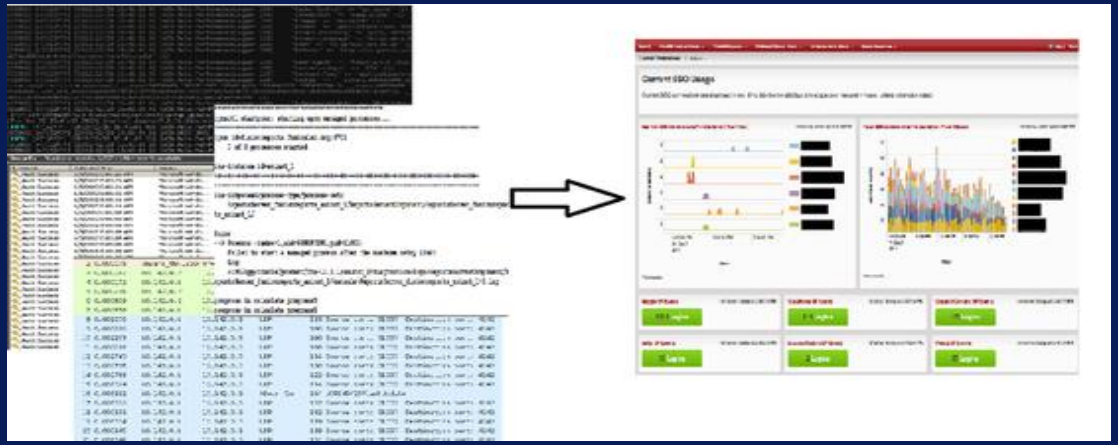
Machine Data

- accessible across organization
- identify patterns
- diagnose potential problems
- provide intelligence


Any Machine Data
Datacenter
Private Cloud
Public Cloud
Online Services
Web Services
Servers
Security
GPS Location
Packaged Applications
Networks
Desktops
Storage
APP
Custom Applications
Messaging
RFID
Telecoms
Online Shopping Cart
Energy Meters
Databases
Web Clickstreams
Call Detail Records
Smartphones and Devices
Operational Intelligence
Search and Investigation
Proactive Monitoring
Operational Visibility
Real-time Business Insights
splunk>enterprise
splunk>cloud
Enterprise Scalability

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

CEO – Application Managem

localhost:8000/en-US/app/appmgmt/ceo

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

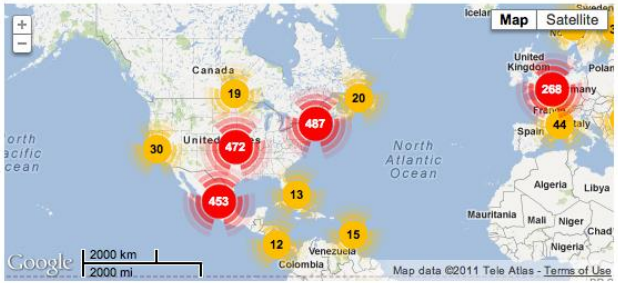
real-time

Real-time concurrent users browsing website.

60

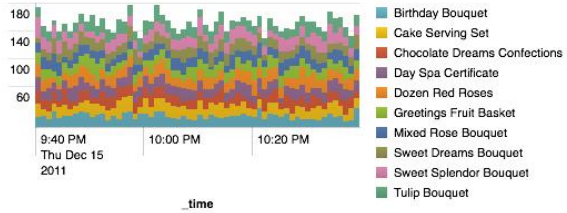
Visitor Location

3m ago



Top Items Sold

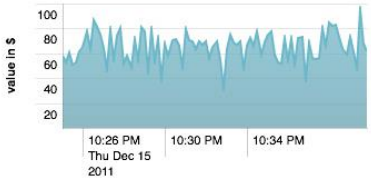
5m ago



Abandoned Baskets

5m ago

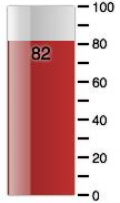
Average value of abandoned shopping carts on website via logout or expired sessions over the last 60 minutes.



% Coupon Usage

real-time

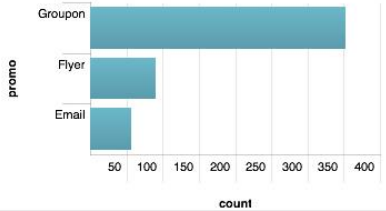
Real-time percentage of shoppers using coupons vs paying full price.



Top Promotions

3m ago

Promotional program popularity over the past 24 hours.



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.

19

$151.00
average purchase

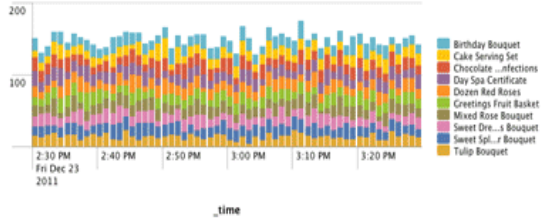
RT Visitor Count — real-time

Real-time concurrent users browsing website.

62

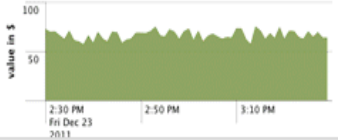
Visitor Location — < 1m ago

Top Items Sold — 5m ago



Abandoned Baskets — 5m ago

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

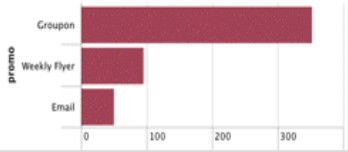


% Coupon Usage — real-time

Real-time percentage of shoppers using coupons at checkout.

Top Promotions — 1s ago

Promotional program popularity over the past 24 hours.



Splunk Features

Central Repository

Data Access for Analytics

Structure and meaning of data

Visualization

Applications

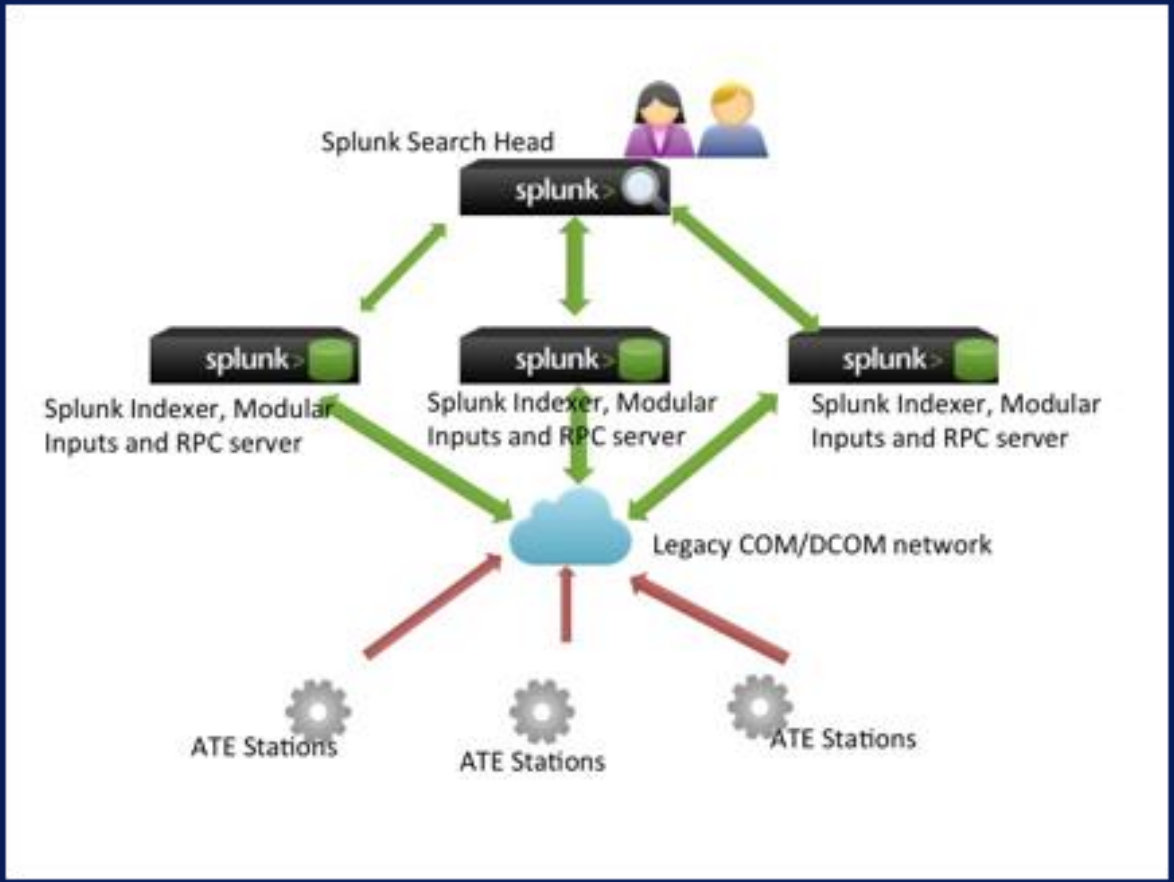
Splunk Architecture

4 Major components

- Search Head
- Forwader
- Indexer
- Deployment Server

Splunk Architecture
Search Head

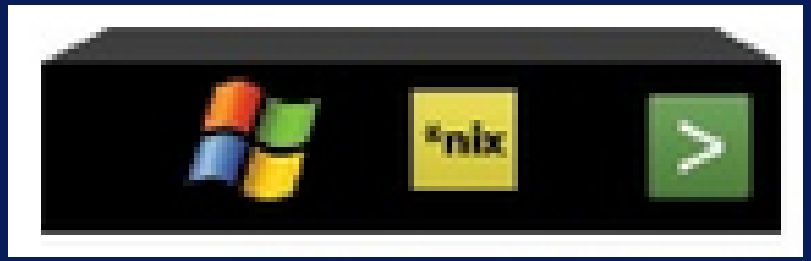
GUI for Splunk search, analysis and reporting

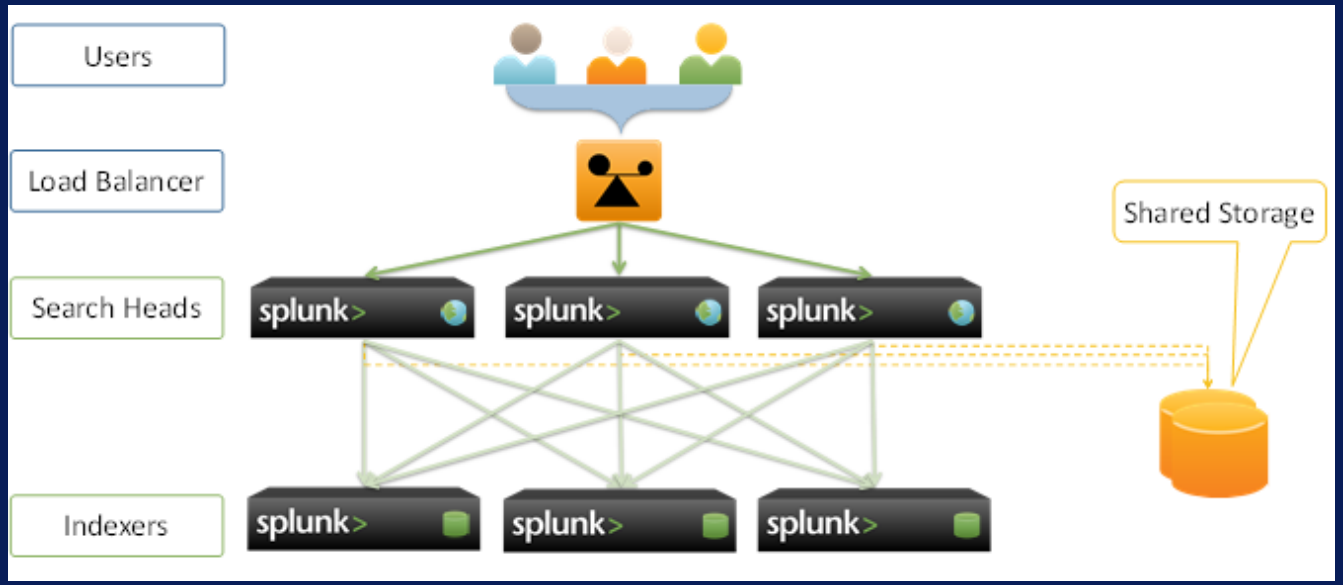

Splunk Search Head
splunk>
splunk>
splunk>
splunk>
Splunk Indexer, Modular Inputs and RPC server
Splunk Indexer, Modular Inputs and RPC server
Splunk Indexer, Modular Inputs and RPC server
Legacy COM/DCOM network
ATE Stations
ATE Stations
ATE Stations

Splunk Architecture Forwader

Universal forwarder (UF)

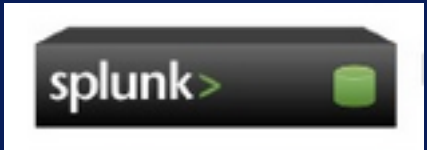
Heavy weight forwarder (HWF)




Users
Load Balancer
Search Heads
Indexers
Shared Storage
splunk>
splunk>
splunk>
splunk>
splunk>
splunk>

Splunk Architecture
Indexer

Indexing incoming data
Searching the indexed data



Splunk Architecture
Indexer

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

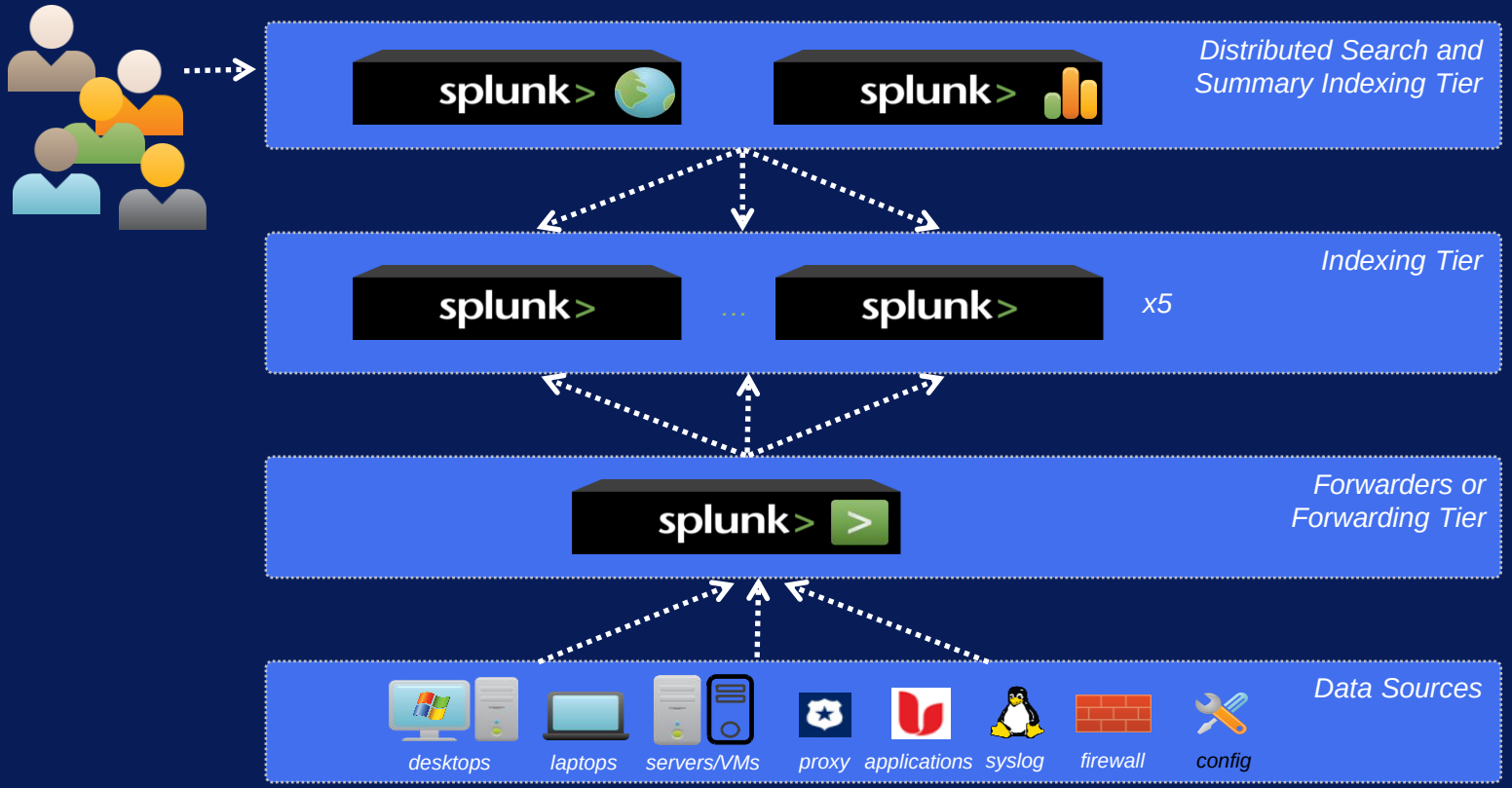
Use Cases

- Internet of Things
  - Metropolitan data
  - Wi-Fi enabled Nest
  - Elevator usage

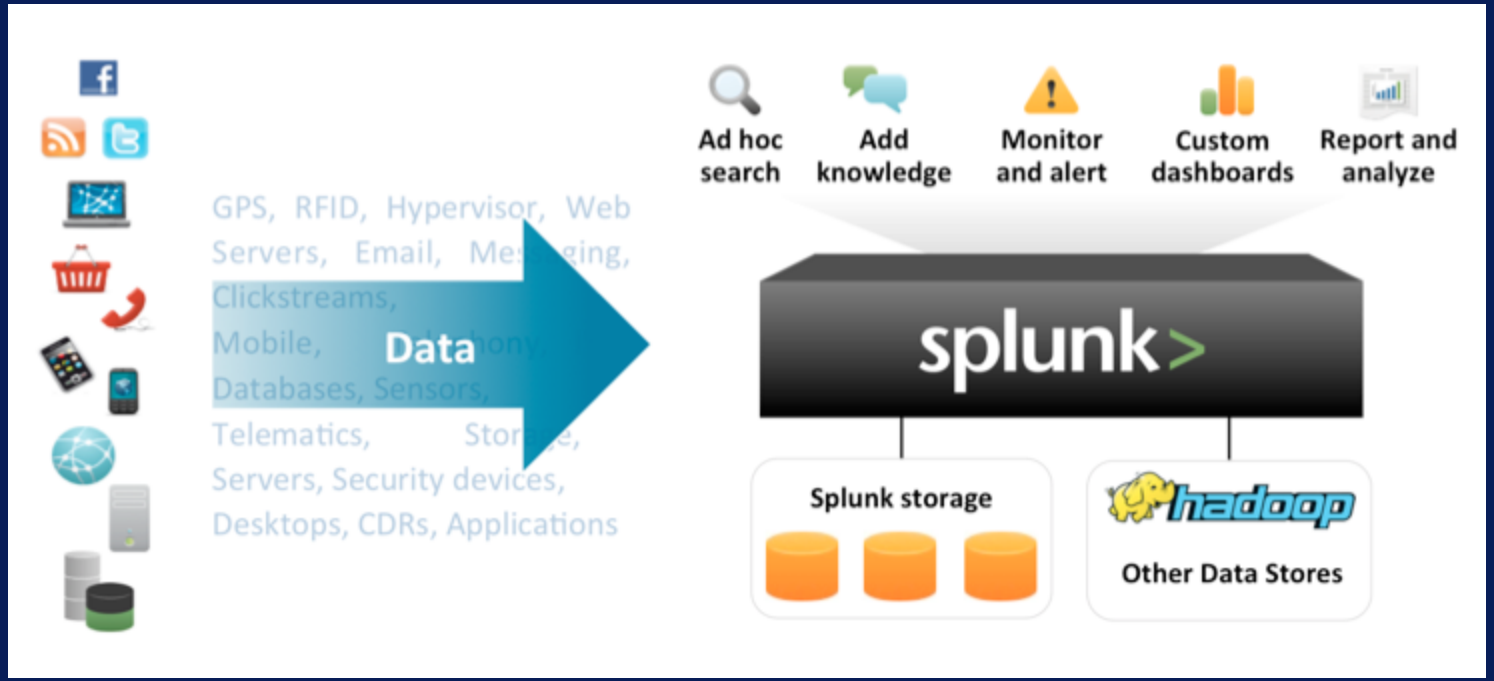
Use Cases

- Security
  - Security log data
  - Network log data
- Compliance
  - Auditing standards

Splunk Architecture



Splunk and Big Data



Data Sources

Splunk for Analytics

Hadoop for ETL