Splunk Spark Integration

Gang Tao

About Me

- Software Engineer with 15+ Years experience
- Now architect working on Data acquisition and Cloud App
- Used to be working on BI, ERP and other Enterprise application development
- Like data science and open source

Splunk Company Overview

Company

- Global HQs:
  - San Francisco
  - London
  - Hong Kong
- 1,800+ employees globally
- Annual Revenue: $450.9M (YoY +49%)
- NASDAQ: SPLK

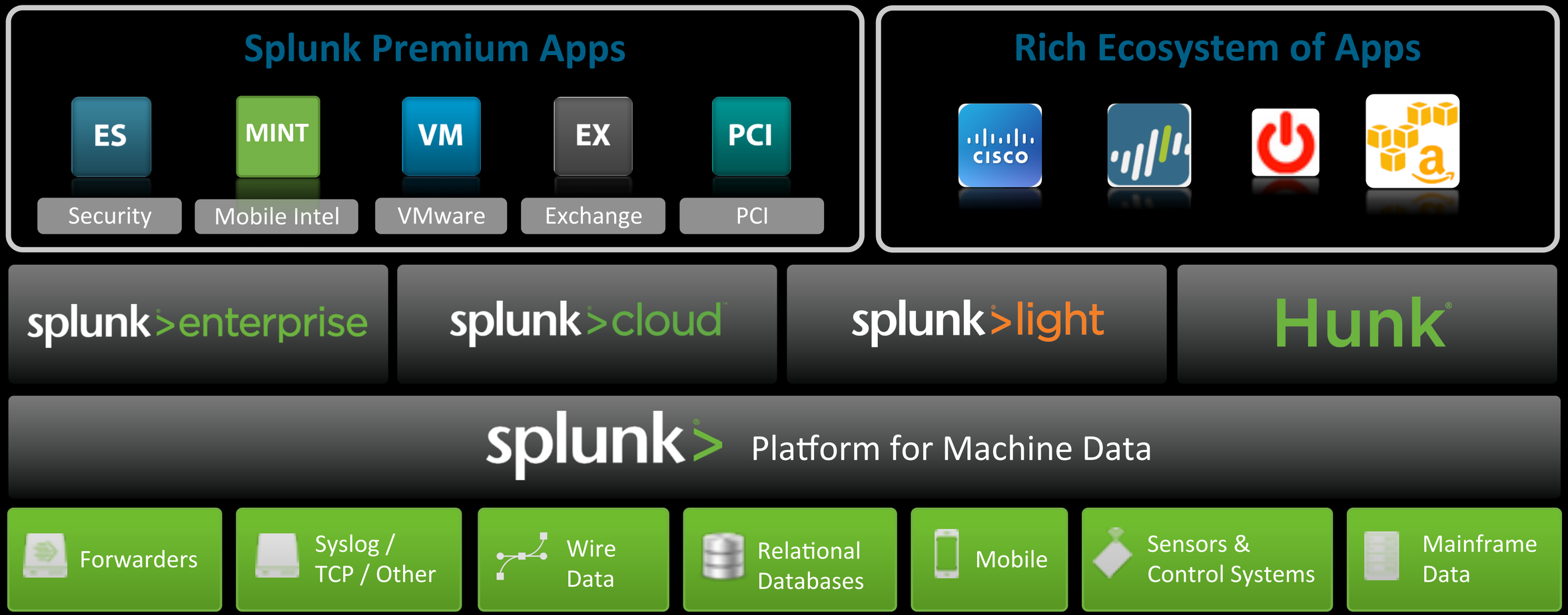
Products

- Free trial to massive scale
- Splunk products:
  - Splunk Enterprise
  - Splunk Cloud
  - Hunk
  - Splunk Light
  - Splunk MINT
  - Premium Solutions

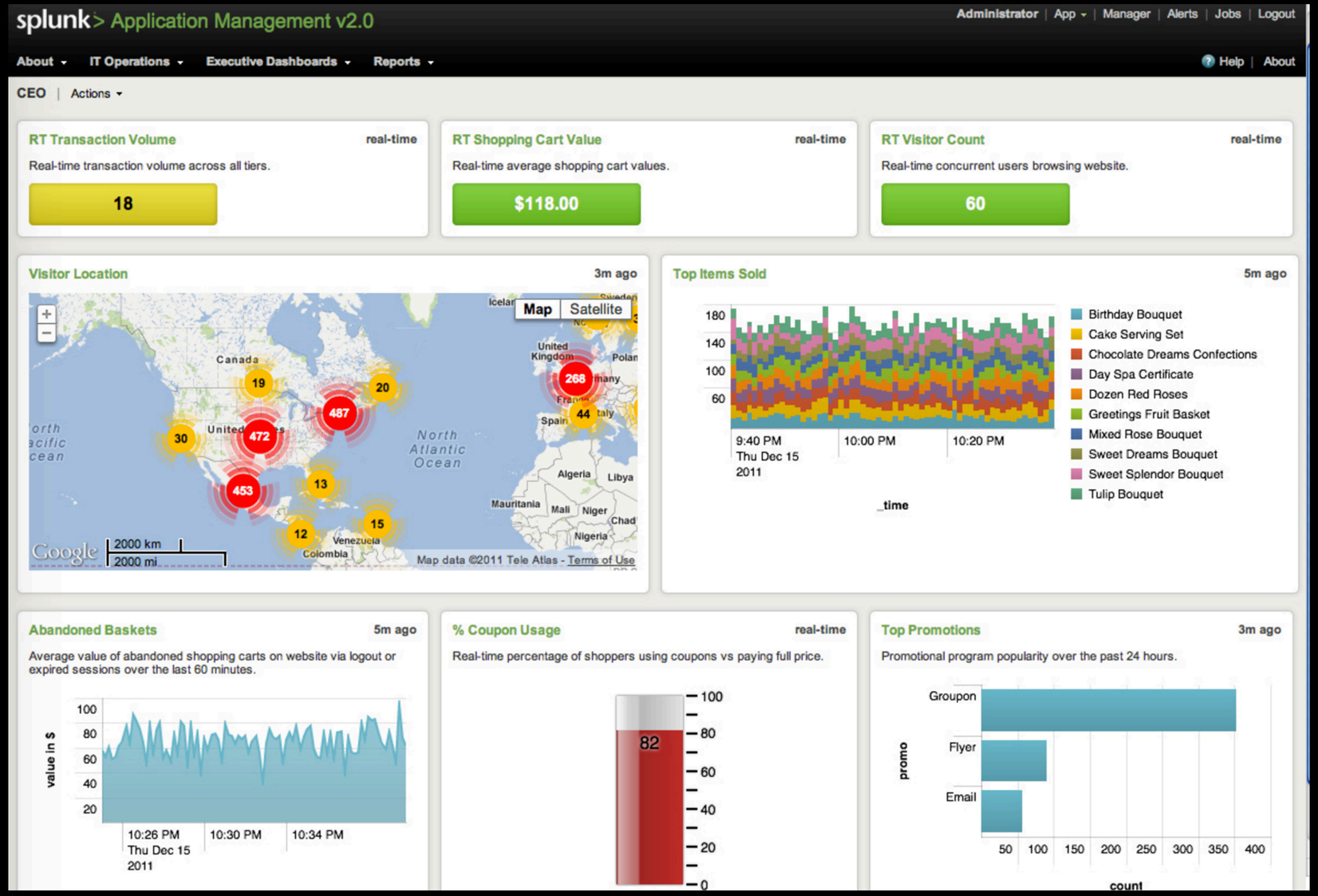
Customers

- 10,000+ customers
- Across 100 countries
- Small to large organizations
- More than 80 of the Fortune 100
- Largest license:
  - 400+ Terabytes/day

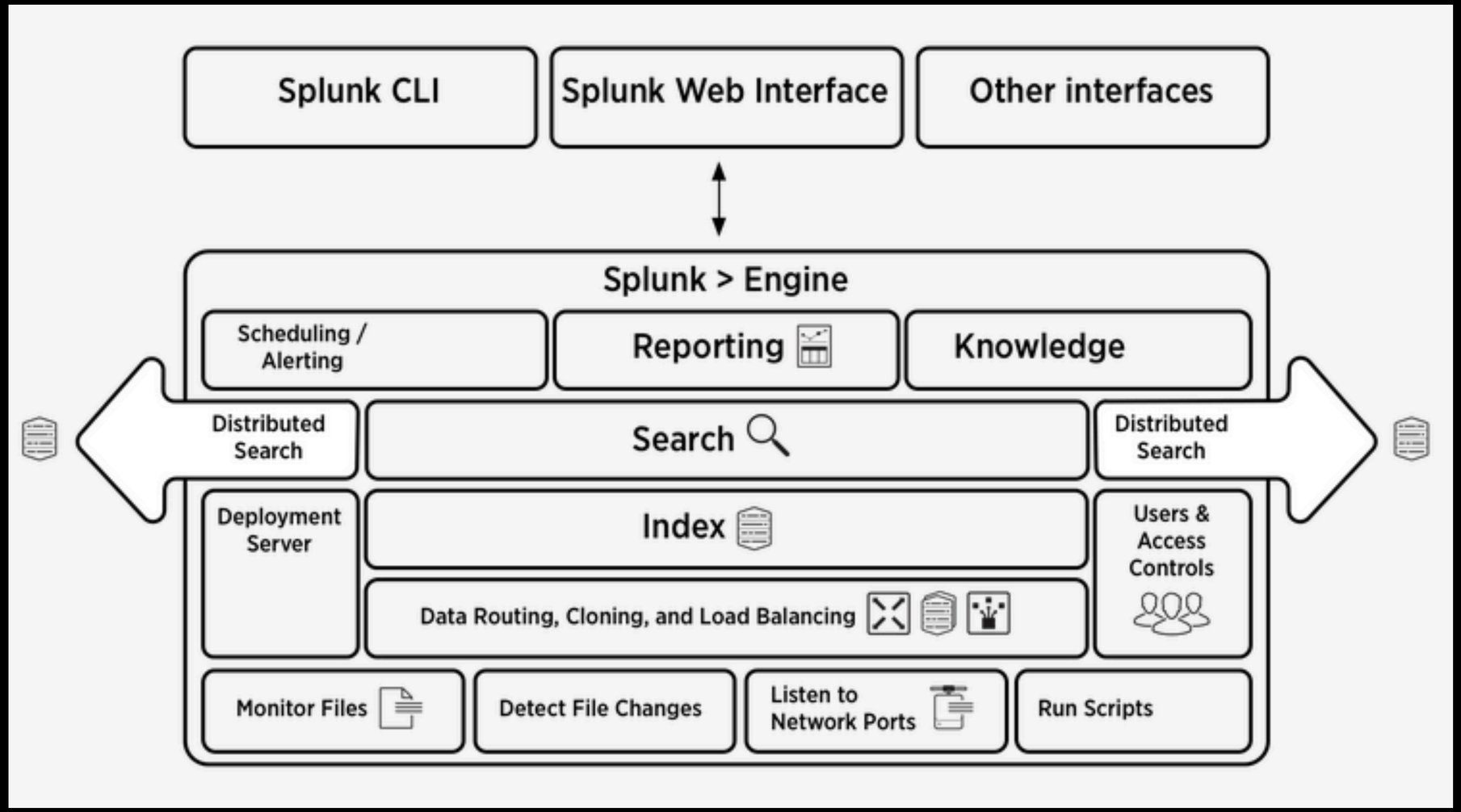
Splunk – a Machine Data Platform



Demo


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
United States
North Atlantic Ocean
Algeria
Libya
Mauritania
Mali
Niger
Chad
Nigeria
Venezuela
Colombia
19
20
487
30
472
268
44
453
13
15
12
Google
2000 km
2000 mi
Map data ©2011 Tele Atlas - Terms of Use
Top Items Sold
5m ago
180
140
100
60
9:40 PM
Thu Dec 15
2011
10:00 PM
10:20 PM
_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.
value in $
100
80
60
40
20
10:26 PM
Thu Dec 15
2011
10:30 PM
10:34 PM
% Coupon Usage
real-time
Real-time percentage of shoppers using coupons vs paying full price.
82
100
80
60
40
20
0
Top Promotions
3m ago
Promotional program popularity over the past 24 hours.
promo
Groupon
Flyer
Email
50
100
150
200
250
300
350
400
count

Splunk Technical Stack



Presenting

Processing

Store

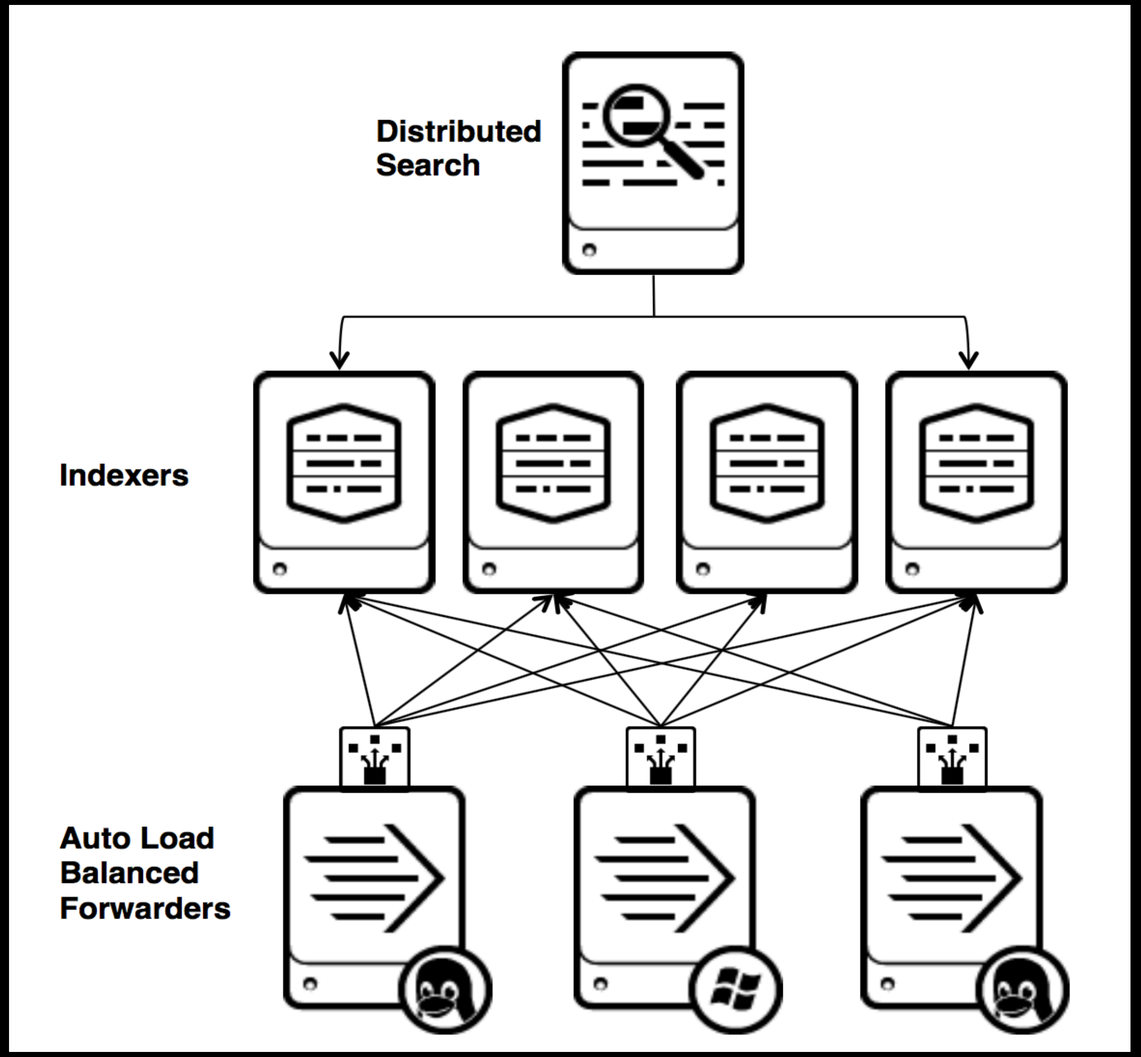
Acquisition

Splunk Deployment Architecture

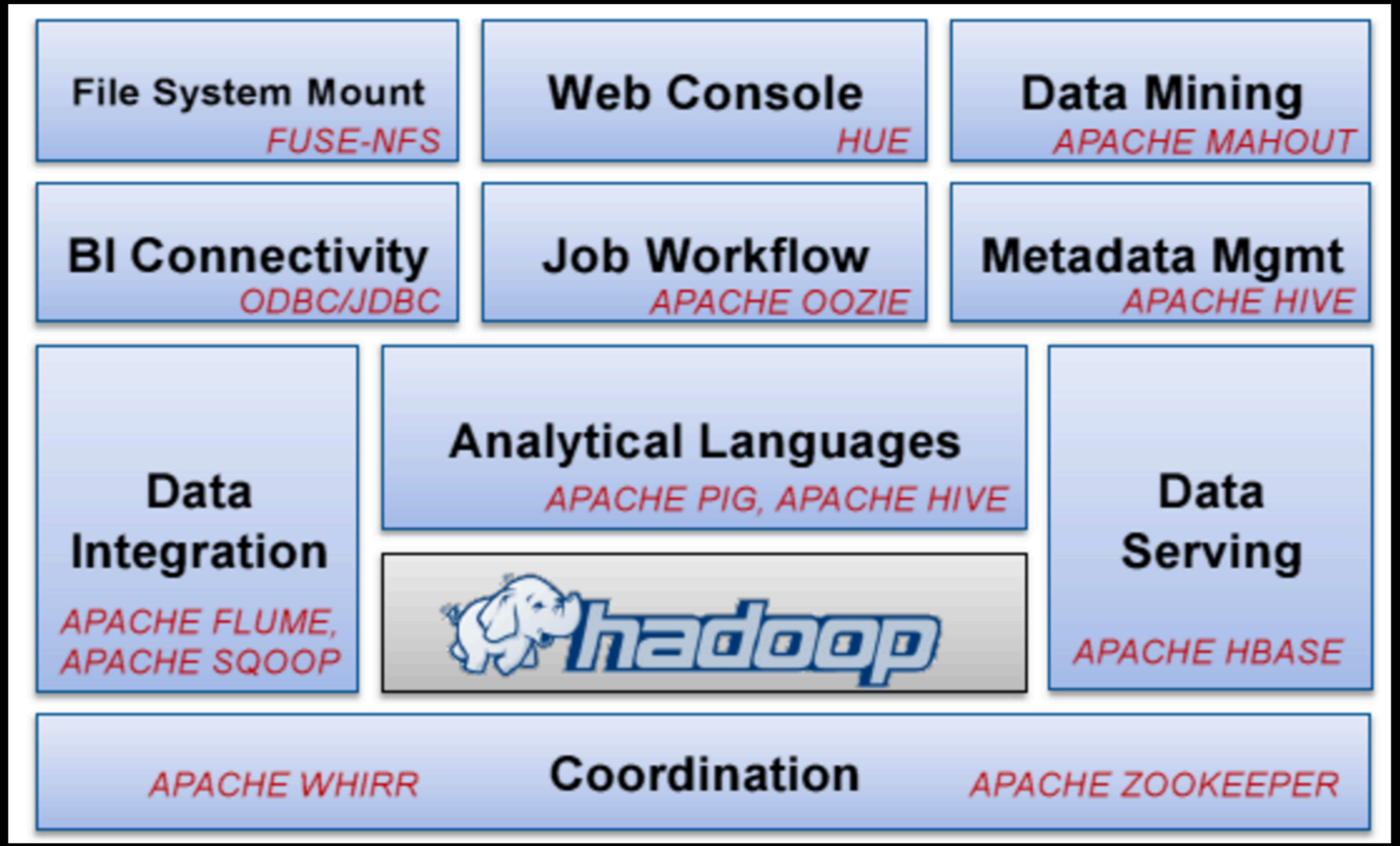
Indexer
store data, transform row data into events and searches the indexed data in response to search requests.

Search Head
directs search requests to a set of indexers, merges the results and presents them to the user

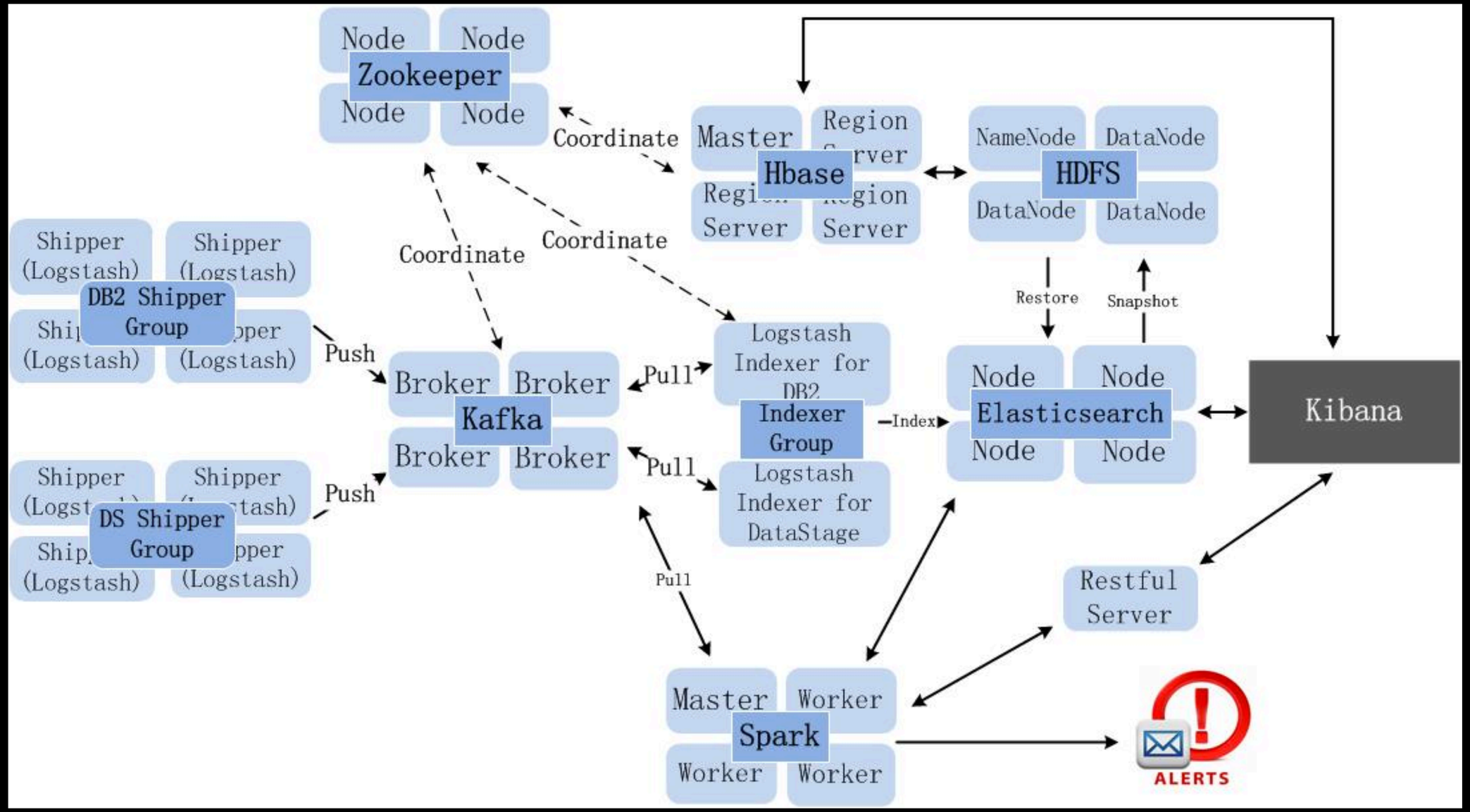
Forwarder
get data into indexers



Splunk VS Open Source



Splunk VS Open Source



SQL of Machine Data - SPL

SPL – Splunk Processing Language

- SQL
- *nix Pipe
- Google Search

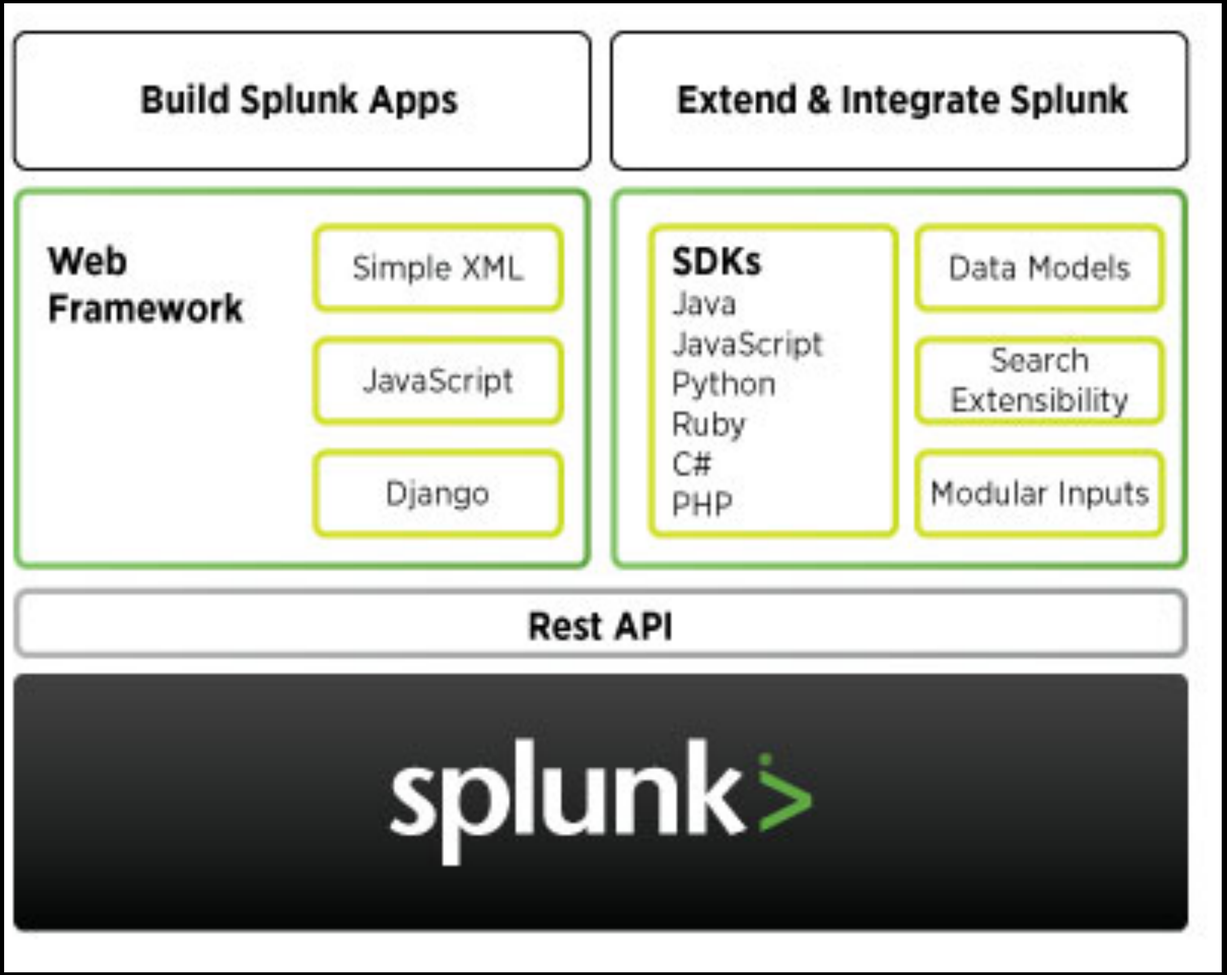
```
SELECT *
FROM mytable

WHERE (mycolumn1="true" OR
mycolumn2="red") AND mycolumn3="blue"
```

```
source=mytable
AND (mycolumn1="true" OR
mycolumn2="red")

AND mycolumn3="blue"
```

Extensibility - Splunk App



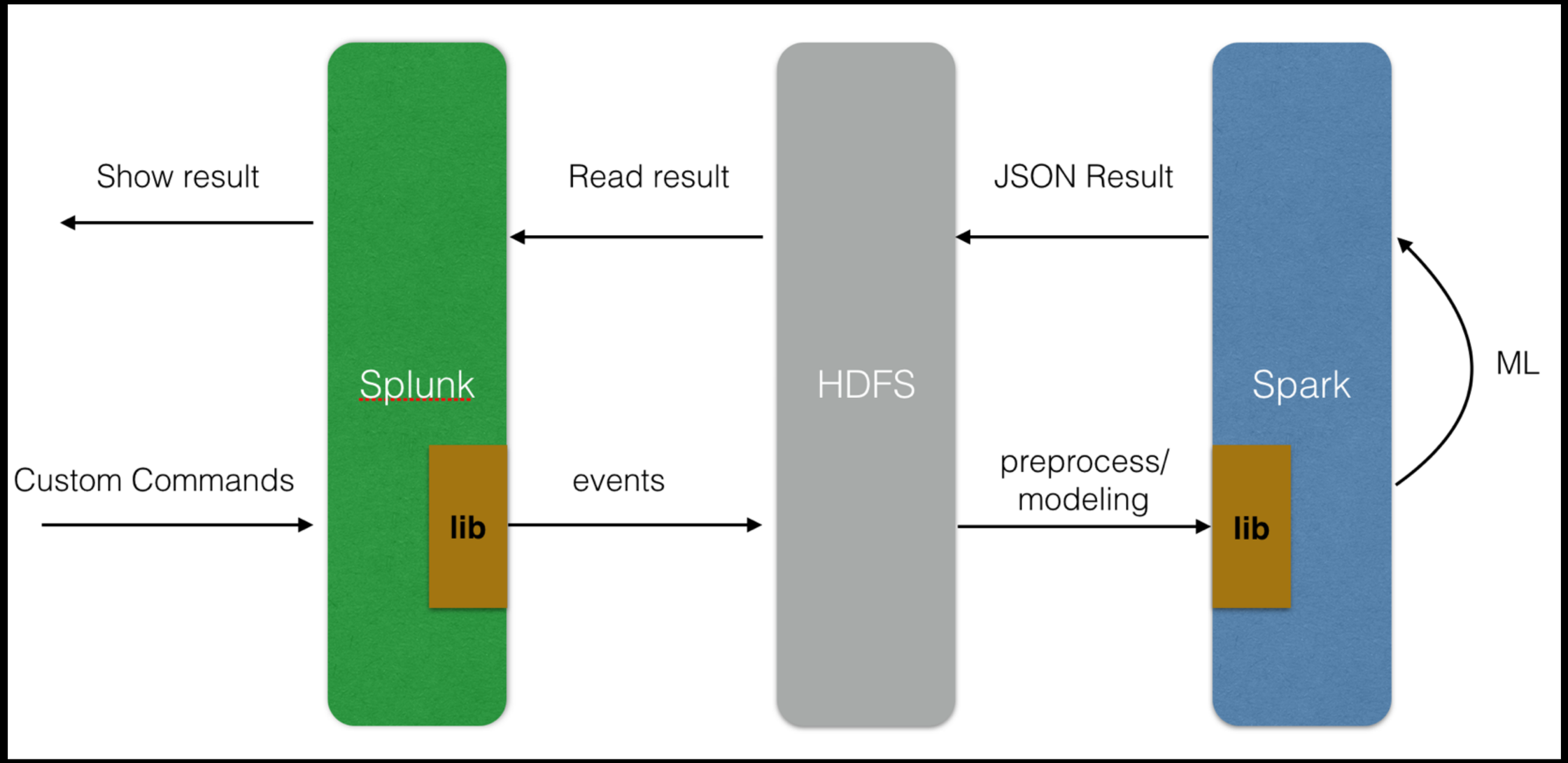
http://apps.splunk.com/

- ★ Enterprise Security
- ★ ITSI
- ★ DB Connect
- ★ Technology Add-ons

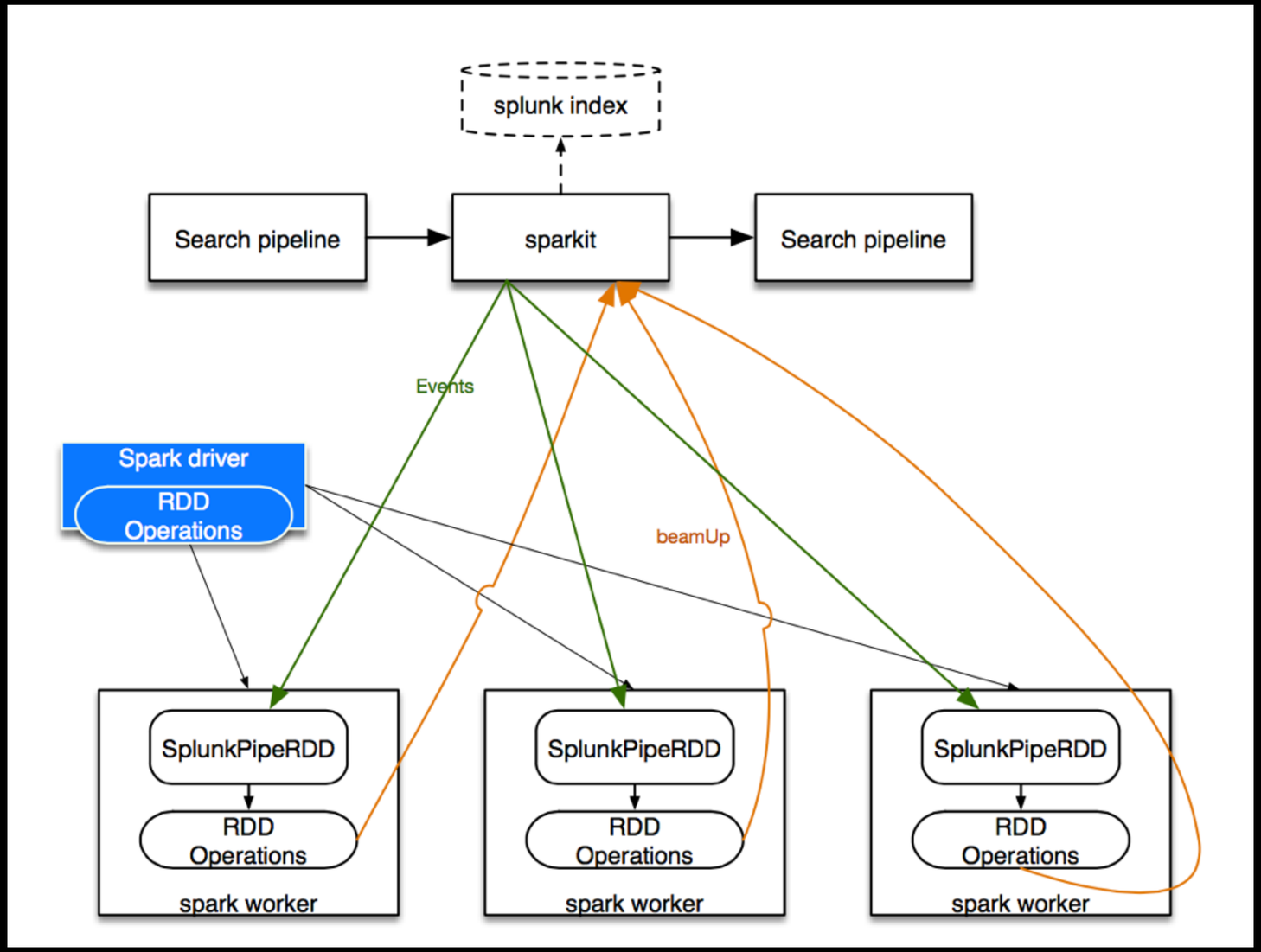
Why Integration?

- Splunk to Spark
  - Data Ingestion
  - Unstructure/Semi Structure data Indexing
  - Data processing with Splunk search
  - Data Presenting
- Spark to Splunk
  - Powerful computing capability
  - Machine Learning
  - Open Source community

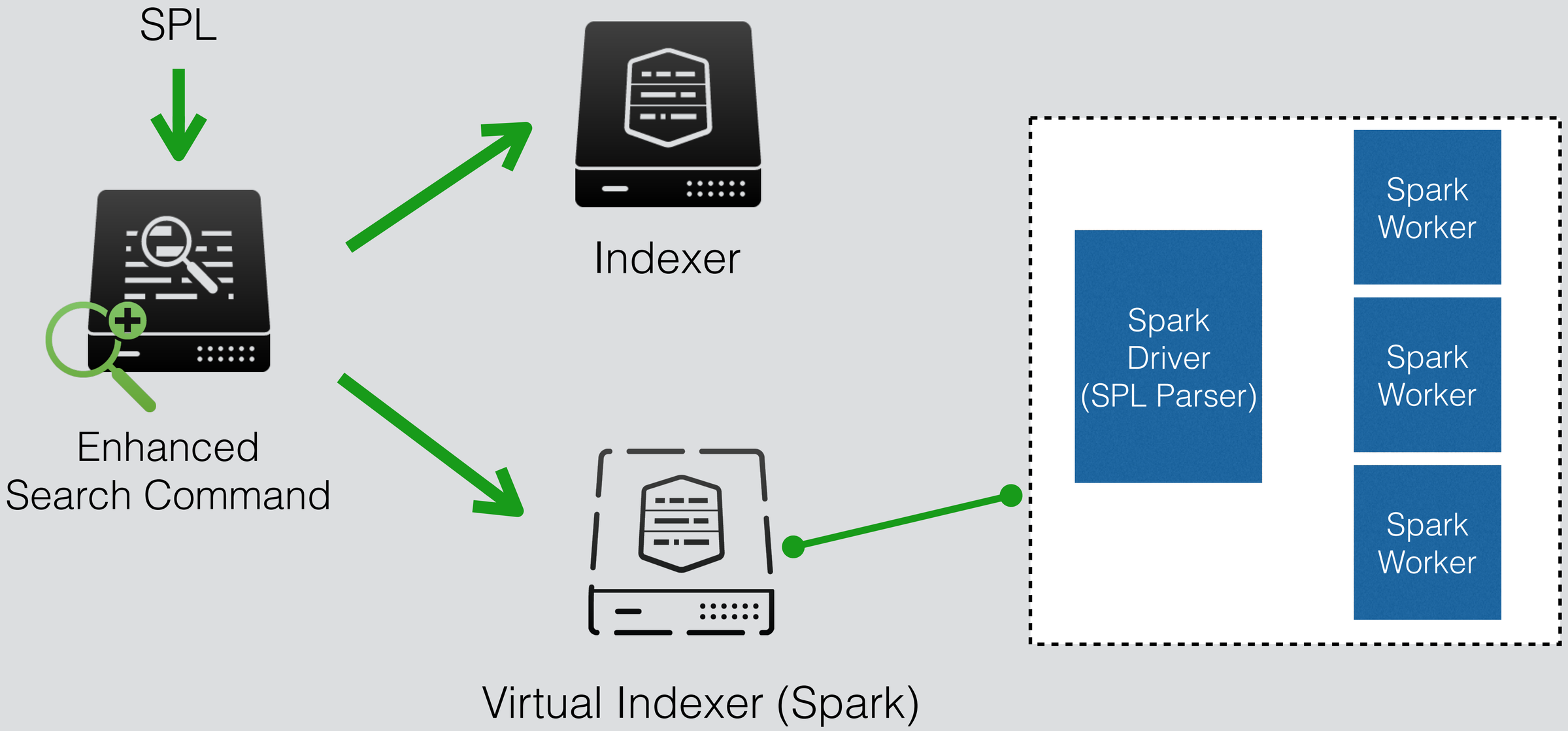
Solution A


Show result
Read result
JSON Result
Splunk
HDFS
Spark
ML
Custom Commands
events
preprocess/
modeling
lib
lib

Solution B


splunk index
Search pipeline
sparkit
Search pipeline
Events
Spark driver
RDD Operations
beamUp
SplunkPipeRDD
RDD Operations
spark worker
SplunkPipeRDD
RDD Operations
spark worker
SplunkPipeRDD
RDD Operations
spark worker

Solution C


SPL
Indexer
Enhanced
Search Command
Virtual Indexer (Spark)
Spark
Driver
(SPL Parser)
Spark
Worker
Spark
Worker
Spark
Worker

Challenges

- Avoid big data movement
- keep good user experience
- Adapt to SPL concept


Q
A


Thanks!