splunk> .conf2017

Bring Context To Your Machine Data With Hadoop, RDBMS & Splunk

Raanan Dagan and Rohit Pujari

September 25, 2017 | Washington, DC

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Agenda

- Splunk Big Data Architecture
- Alternative Open Source Approach
- Real-World Customer Architecture
- End-to-end Demonstration

Big Data Technologies

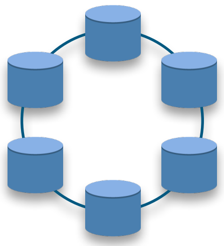
Relational Database Structured

- Schema at Write
- SQL
- ETL

RDBMS
Oracle, MySQL, IBM DB2, Teradata

NoSQL Semi-Structured

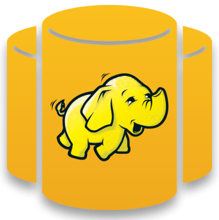
- Schema at Read
- Key-Value, Column, Document & Other Stores



Cassandra, HBase, MongoDB

Hadoop Semi-Structured

- Schema at Read
- MapReduce
- HDFS Storage

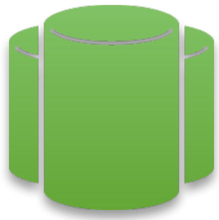


MapReduce

Distributed File System

Splunk

- Schema at Read
- Search
- Real-Time Indexing



Time-Series, Unstructured, Heterogeneous

Splunk: Open And Extensible

Databases = Splunk DB Connect (Hive, Impala, Oracle)

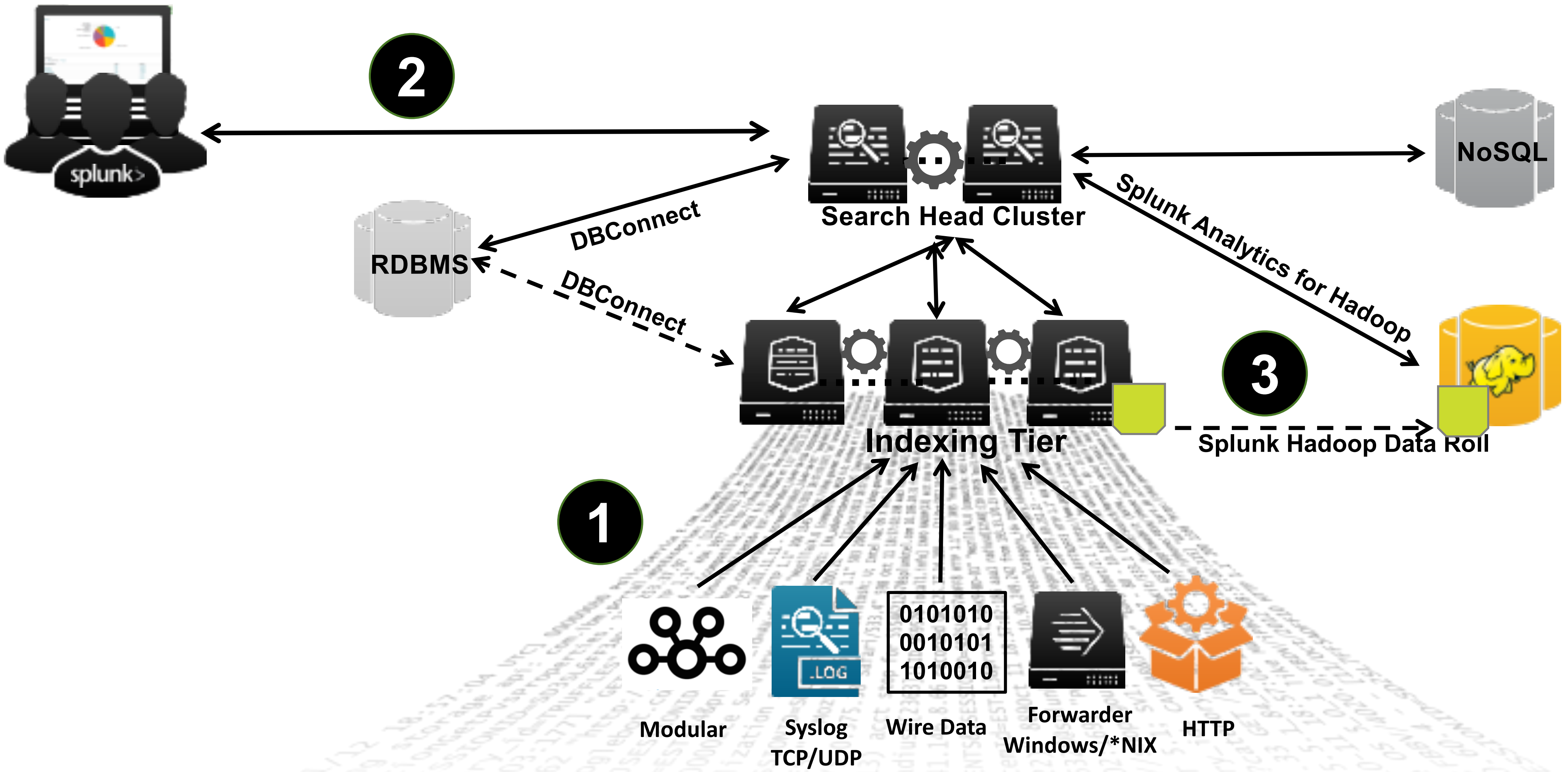
Hadoop = Analytics for Hadoop, Hadoop Data Roll, Connect

Kafka = Splunk Kafka Add-On, Kafka with HEC

Spark = Spark SQL

NoSQL = MongoDB, Hbase, Cassandra apps

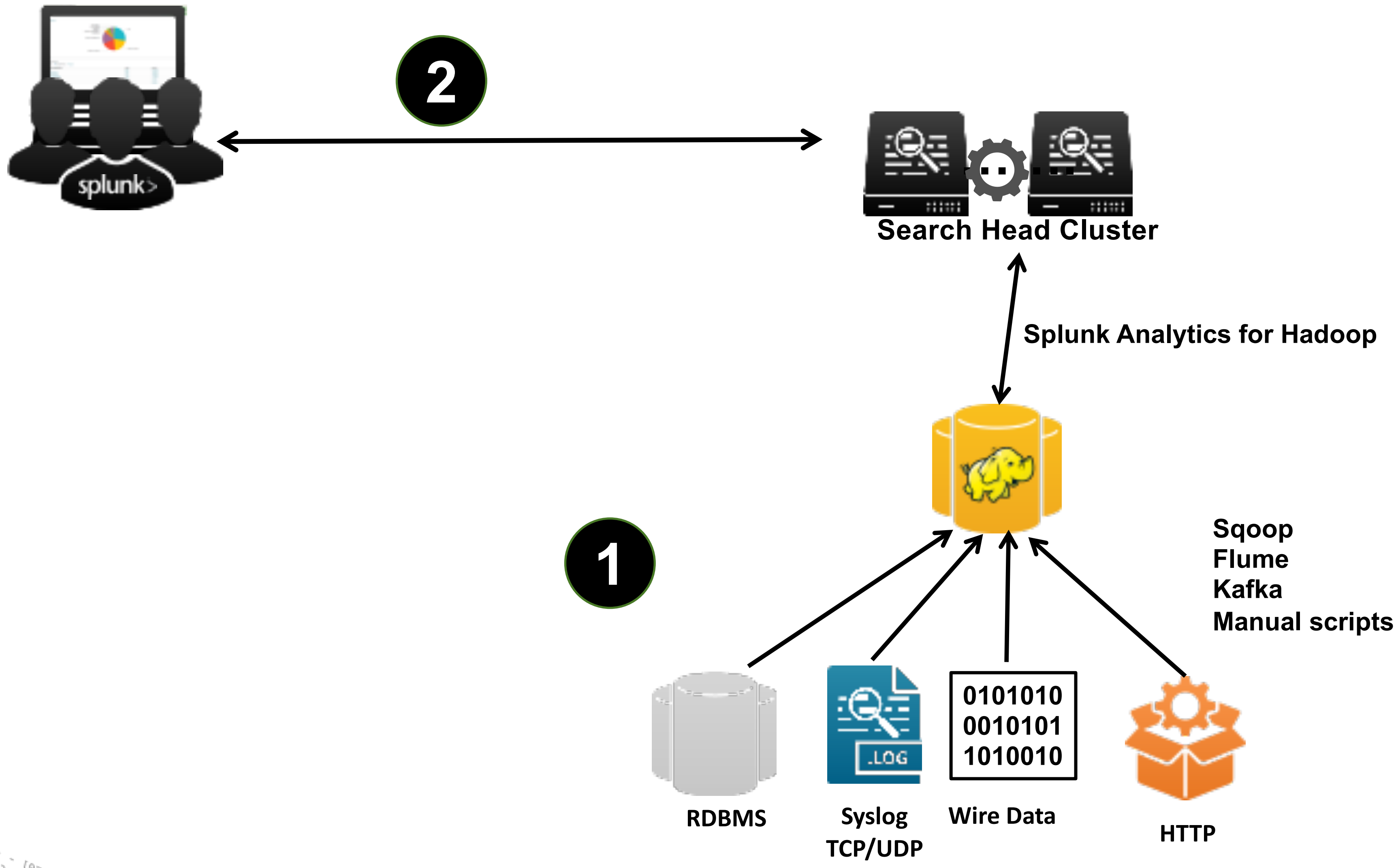
Splunk Enterprise Architecture



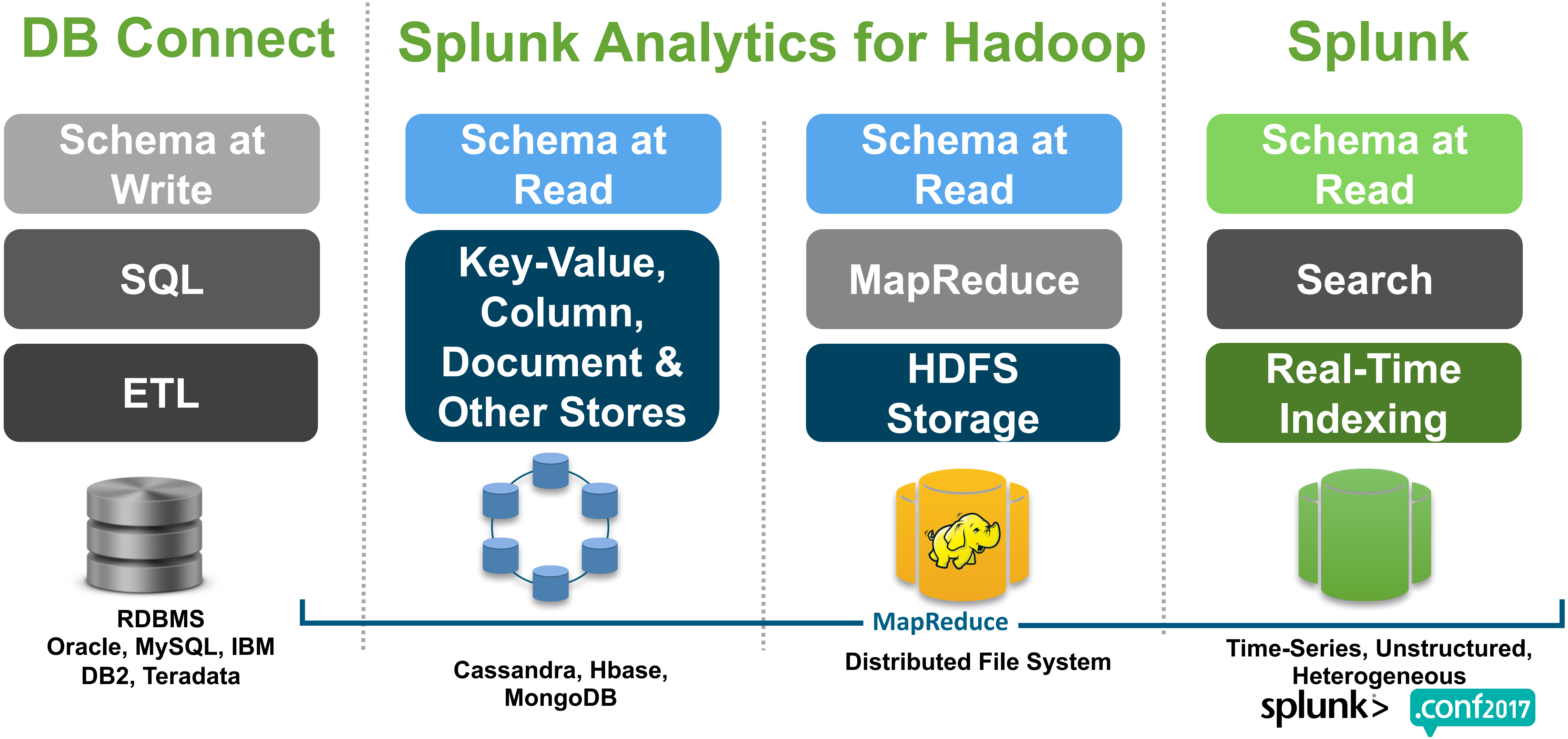
Splunk And Hadoop - Products

- **Splunk Analytics for Hadoop:**
  - Analyze Hadoop Data using Hadoop MapReduce Processing
- **Splunk Hadoop Connect:**
  - Export data from Splunk to Hadoop
- **Hadoop Data Roll**
  - Archive Splunk indexers to Hadoop
- **Splunk Monitor Hadoop:**
  - Monitor Hadoop

Splunk & Hadoop Architecture



Splunk Big Data Technologies



DB Connect

Schema at Write

SQL

ETL

RDBMS
Oracle, MySQL, IBM DB2, Teradata

Splunk Analytics for Hadoop

Schema at Read

Key-Value, Column, Document & Other Stores

Cassandra, Hbase, MongoDB

Schema at Read

MapReduce

HDFS Storage

Distributed File System

Splunk

Schema at Read

Search

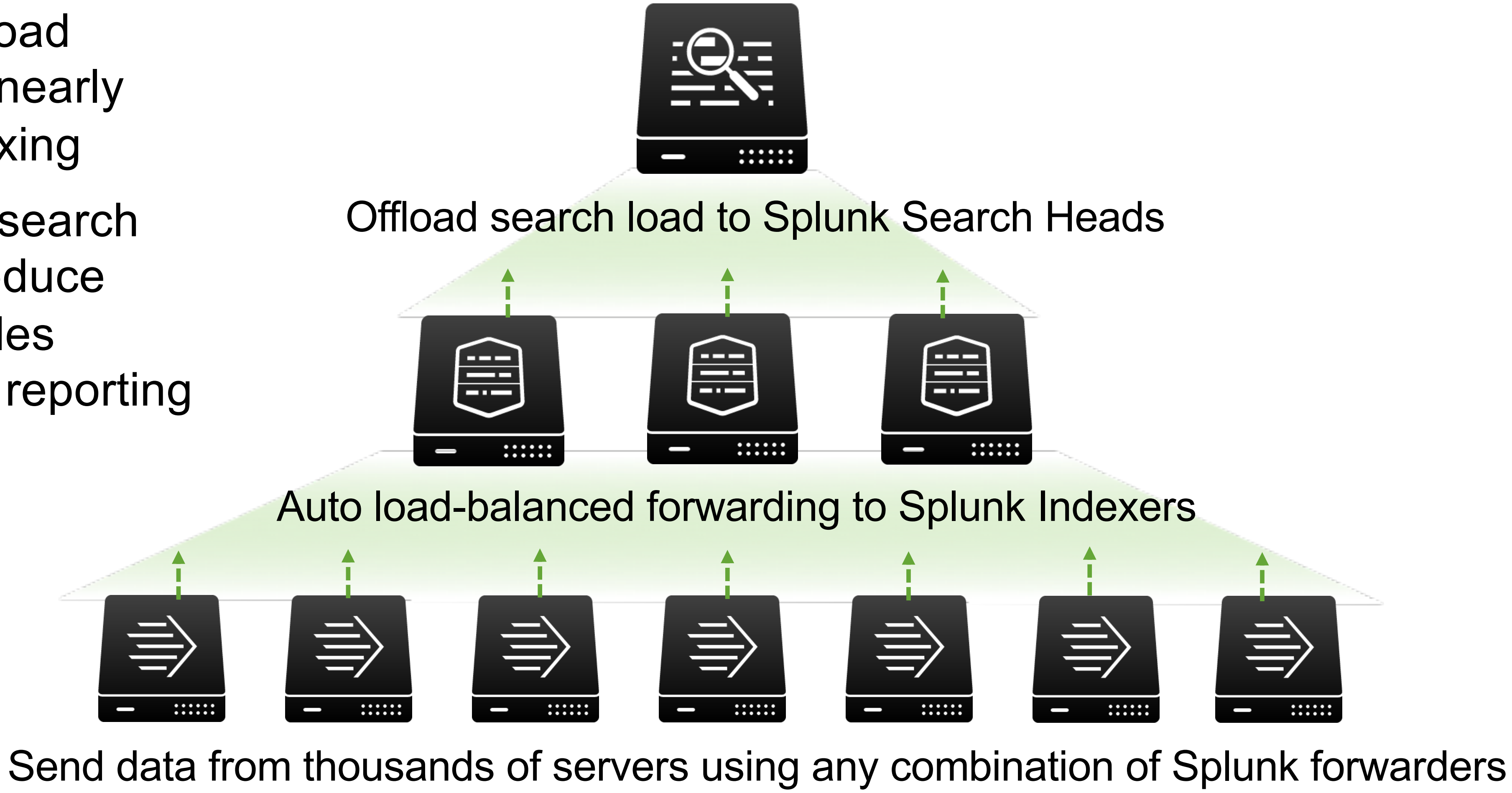
Real-Time Indexing

Time-Series, Unstructured, Heterogeneous

Splunk Scalability

Enterprise-class Availability and Scale

- Automatic load balancing linearly scales indexing
- Distributed search and MapReduce linearly scales search and reporting

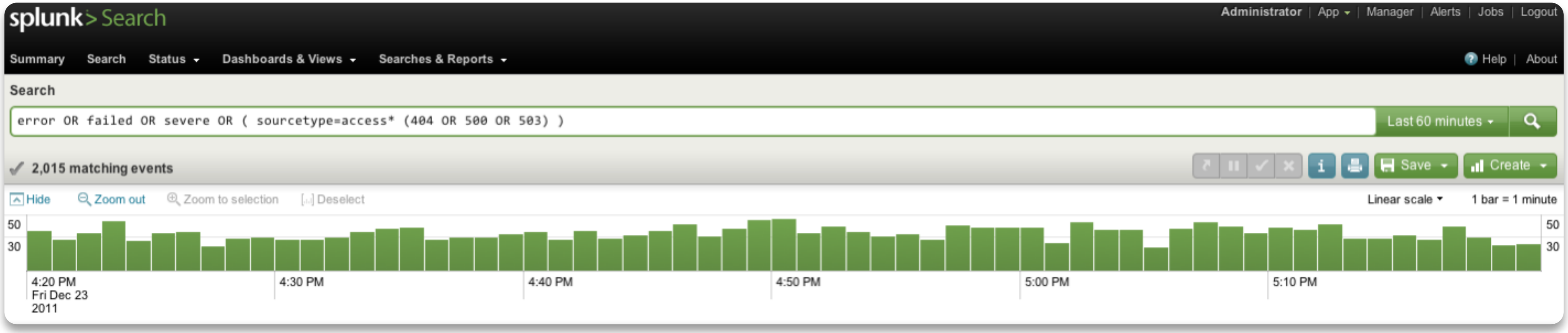


Offload search load to Splunk Search Heads

Auto load-balanced forwarding to Splunk Indexers

Send data from thousands of servers using any combination of Splunk forwarders

Splunk Real-Time Analytics

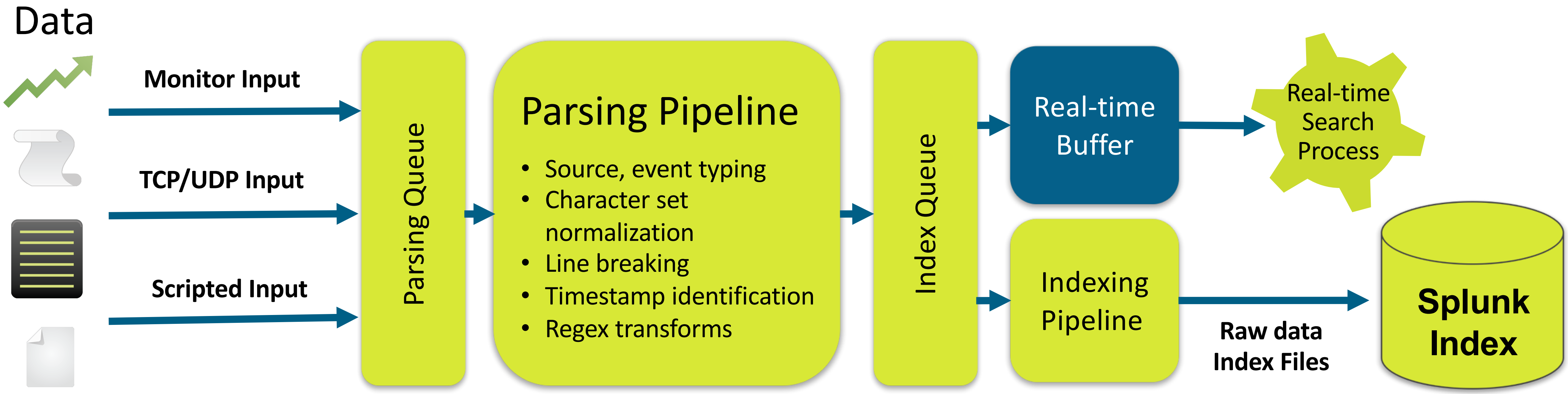


Search

```
error OR failed OR severe OR ( sourcetype=access* (404 OR 500 OR 503) )
```

Last 60 minutes

2,015 matching events



Parsing Pipeline

- Source, event typing
- Character set normalization
- Line breaking
- Timestamp identification
- Regex transforms

Splunk With Hadoop - Unique Features

Virtual Index

- Enables seamless use of the Splunk technology stack on data wherever it rests
- Natively handles MapReduce

Schema-on-the-fly

- Structure applied at search time
- No brittle schema
- Automatically find patterns and trends

Flexibility and Fast Time to Value

- Interactive search
- Preview results while MapReduce jobs run
- Drag-and-drop analytics

Security: Access Control, Pass Through Authentication, Kerberos

What About Structured Data?

Customer profile

Product attributes

Employee details

Pricing and Rate plans

Asset info

Use Cases For Structured Data In Splunk

Index machine data from databases, such as logs or sales records

Enrich machine data with high-level data, such as customer records

Update structured databases with Splunk info, such as risk scores

Interactively browse structured and unstructured data from Splunk reports

Machine Data Delivers Real-time Insights

Media server logs (machine data)

Phone Number

IP Address

Track ID

```
Mar 01 19:18:50:000 aaa2 radiusd[12548]:[ID 959576 local1.info] INFO RADOP(13) acct start for
2172618992@splunktel.com 10.164.232.181 from 12.130.60.5 recorded OK.
2013-03-01 19:18:50:150 10.2.1.34 GET /sync/addtolibrary/01011207201000005652000000000053 - 80 -
10.164.232.181 "Mozilla/5.0 (iPhone; CPU iPhone OS 5_0_1 like Mac OS X) AppleWebKit/534.46 (KHTML,
like Gecko) Version/5.1 Mobile/9A405 Safari/7534.48.3" 503 0 0 825 1680
Mar 01 19:18:50:163 aaa2 radiusd[12548]:[ID 959576 local1.info] INFO RADOP(13) acct stop for
2172618992@splunktel.com 10.164.232.181 from 12.130.60.5 recorded OK.
```

Structured Data Contains Business Context

Phone number

IP address

Track ID

Media server logs (machine data)

```
Mar 01 19:18:50:000 aaa2 radiusd[12548]:[ID 959576 local1.info] IN             start for
2172618992@splunktel.com 10.164.232.181 from 12.130.60.5 recorded OK.
2013-03-01 19:18:50:150 10.2.1.34 GET /sync/addtolibrary/01011207201000005652000000000053 - 80 -
10.164.232.181 "Mozilla/5.0 (iPhone; CPU iPhone OS 5_0_1 like Mac OS X) AppleWebKit/534.46 (KHTML,
like Gecko) Version/5.1 Mobile/9A405 Safari/7534.48.3" 503 0 0 825 1680
Mar 01 19:18:50:163 aaa2 radiusd[12548]:[ID 959576 local1.info] INFO RADOP(13) acct stop for
2172618992@splunktel.com 10.164.232.181 from 12.130.60.5 recorded OK.
```

Customer, product databases

Track ID	Artist	Title	Format ID	Run time
01011207201000005652000000000053	Maroon 5	Moves like Jagger	MP3	4:30

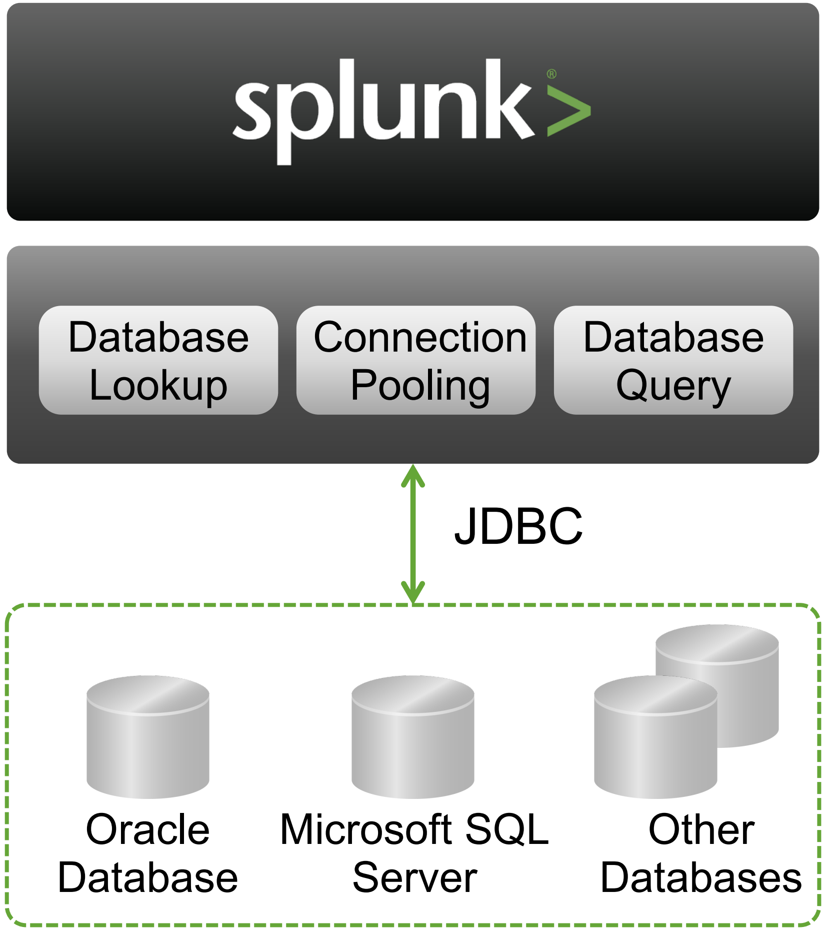
Phone #	**Subscriber ID**
2172618992	53546

Subscriber ID	First Name	Last Name	Age	State	Customer Score
53546	Jim	Morrison	25	CA	93

Splunk DB Connect

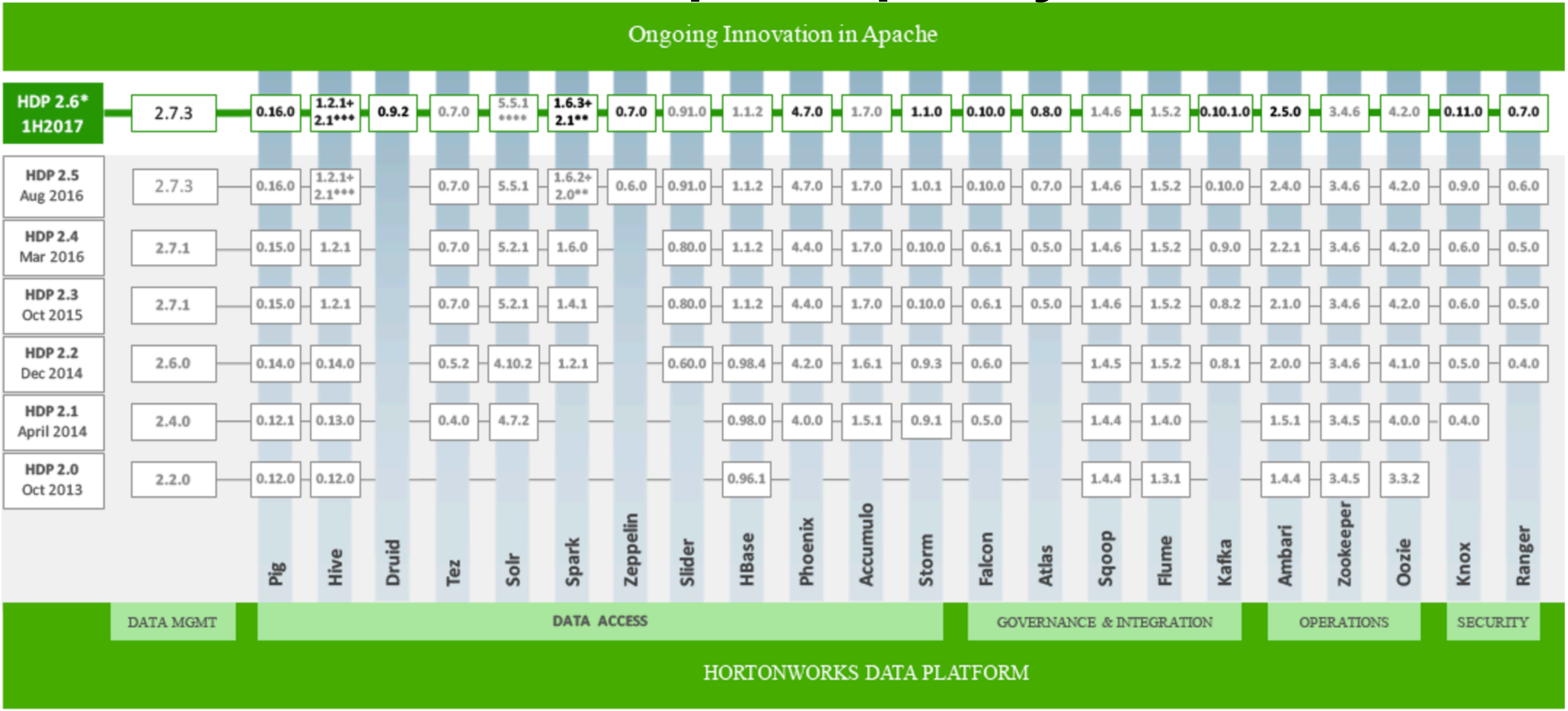
Reliable, scalable, real-time integration between Splunk and traditional relational databases

- Enrich search results with additional business context
- Easily import data into Splunk for deeper analysis
- Ingest, transform machine data in Splunk and export it to relational databases
- Integrate multiple DBs concurrently
- Simple set-up, non-evasive and secure



Open Source Alternatives

Hadoop Complexity



Ongoing Innovation in Apache

	DATA MGMT	DATA ACCESS												GOVERNANCE & INTEGRATION					OPERATIONS			SECURITY	
		Pig	Hive	Druid	Tez	Solr	Spark	Zeppelin	Slider	HBase	Phoenix	Accumulo	Storm	Falcon	Atlas	Sqoop	Flume	Kafka	Ambari	Zookeeper	Oozie	Knox	Ranger
HDP 2.6* 1H2017	2.7.3	0.16.0	1.2.1+ 2.1***	0.9.2	0.7.0	5.5.1 ****	1.6.3+ 2.1**	0.7.0	0.91.0	1.1.2	4.7.0	1.7.0	1.1.0	0.10.0	0.8.0	1.4.6	1.5.2	0.10.1.0	2.5.0	3.4.6	4.2.0	0.11.0	0.7.0
HDP 2.5 Aug 2016	2.7.3	0.16.0	1.2.1+ 2.1***		0.7.0	5.5.1	1.6.2+ 2.0**	0.6.0	0.91.0	1.1.2	4.7.0	1.7.0	1.0.1	0.10.0	0.7.0	1.4.6	1.5.2	0.10.0	2.4.0	3.4.6	4.2.0	0.9.0	0.6.0
HDP 2.4 Mar 2016	2.7.1	0.15.0	1.2.1		0.7.0	5.2.1	1.6.0		0.80.0	1.1.2	4.4.0	1.7.0	0.10.0	0.6.1	0.5.0	1.4.6	1.5.2	0.9.0	2.2.1	3.4.6	4.2.0	0.6.0	0.5.0
HDP 2.3 Oct 2015	2.7.1	0.15.0	1.2.1		0.7.0	5.2.1	1.4.1		0.80.0	1.1.2	4.4.0	1.7.0	0.10.0	0.6.1	0.5.0	1.4.6	1.5.2	0.8.2	2.1.0	3.4.6	4.2.0	0.6.0	0.5.0
HDP 2.2 Dec 2014	2.6.0	0.14.0	0.14.0		0.5.2	4.10.2	1.2.1		0.60.0	0.98.4	4.2.0	1.6.1	0.9.3	0.6.0		1.4.5	1.5.2	0.8.1	2.0.0	3.4.6	4.1.0	0.5.0	0.4.0
HDP 2.1 April 2014	2.4.0	0.12.1	0.13.0		0.4.0	4.7.2				0.98.0	4.0.0	1.5.1	0.9.1	0.5.0		1.4.4	1.4.0		1.5.1	3.4.5	4.0.0	0.4.0	
HDP 2.0 Oct 2013	2.2.0	0.12.0	0.12.0							0.96.1						1.4.4	1.3.1		1.4.4	3.4.5	3.3.2		

HORTONWORKS DATA PLATFORM

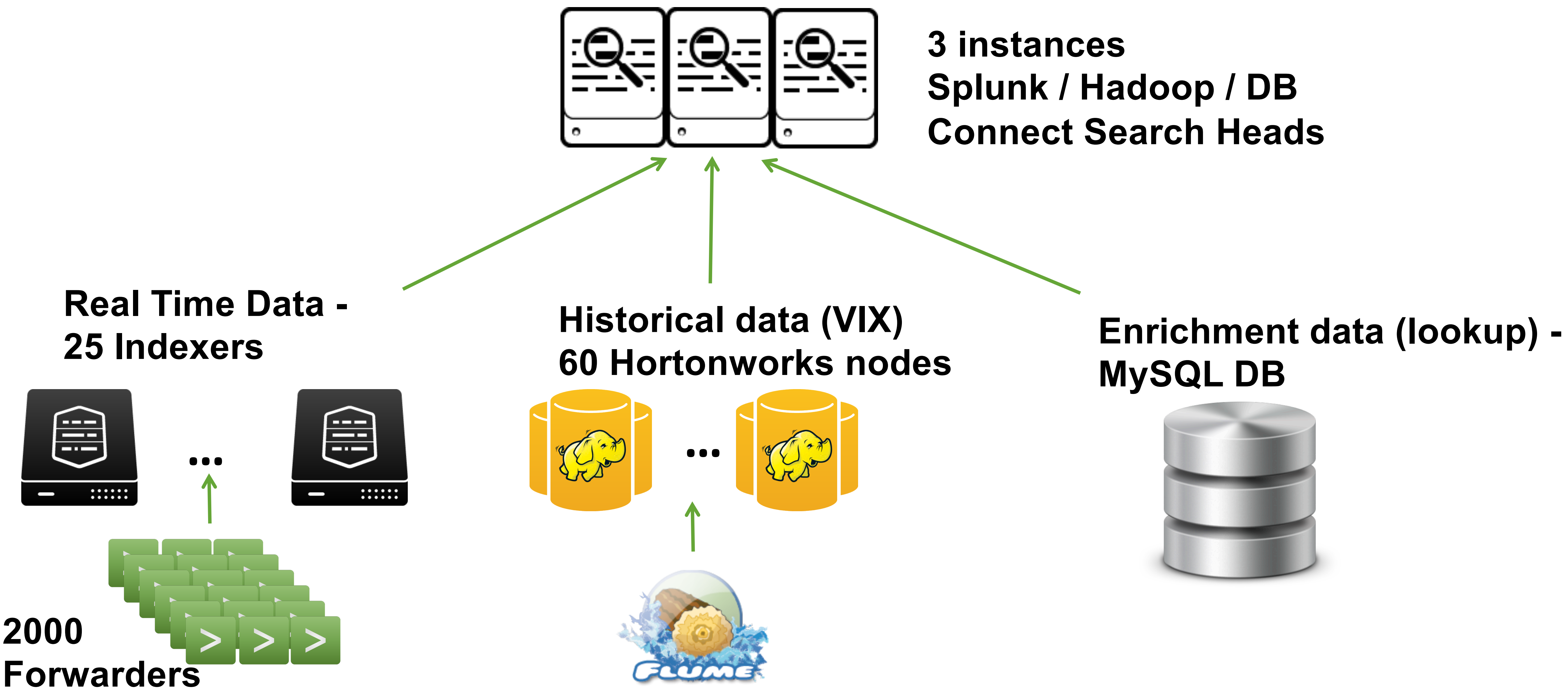
splunk> .conf2017

Hadoop Use Cases

Splunk use cases	Application Delivery	IT Operations	Security and Compliance
Splunk or Hadoop use cases	Business Analytics	IoT	
Hadoop use cases	ETL for RDBMS		

Customer Architecture

Summary Architecture



3 instances
Splunk / Hadoop / DB
Connect Search Heads

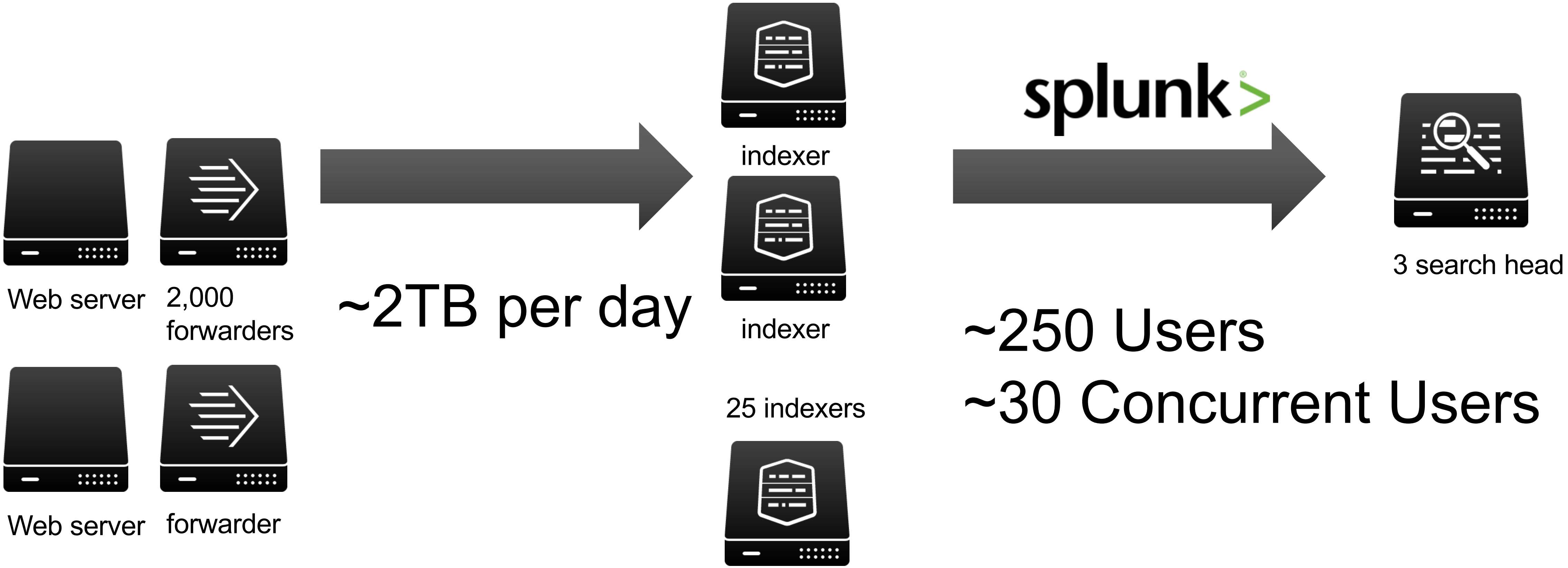
Real Time Data -
25 Indexers

Historical data (VIX)
60 Hortonworks nodes

Enrichment data (lookup) -
MySQL DB

2000
Forwarders

Splunk Deployment Architecture



splunk>

Web server

2,000 forwarders

Web server

forwarder

~2TB per day

indexer

indexer

25 indexers

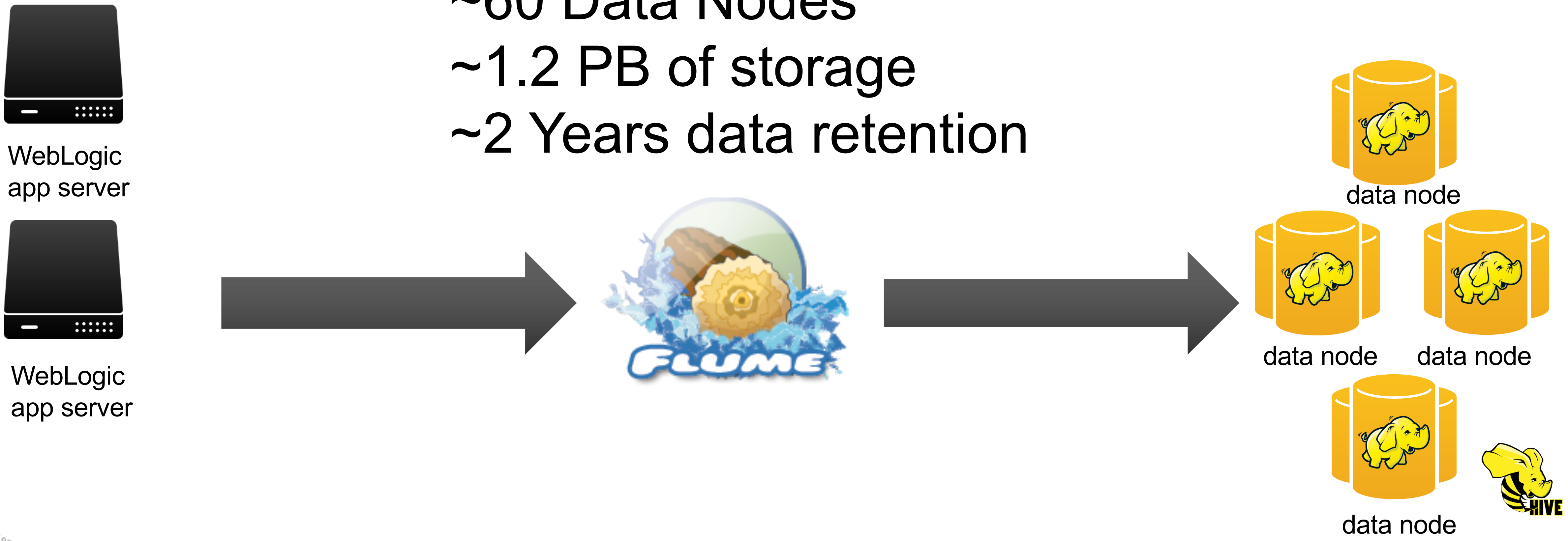
3 search head

~250 Users

~30 Concurrent Users

Hadoop Architecture

~30 Flume Agents
~60 Data Nodes
~1.2 PB of storage
~2 Years data retention



WebLogic app server

WebLogic app server

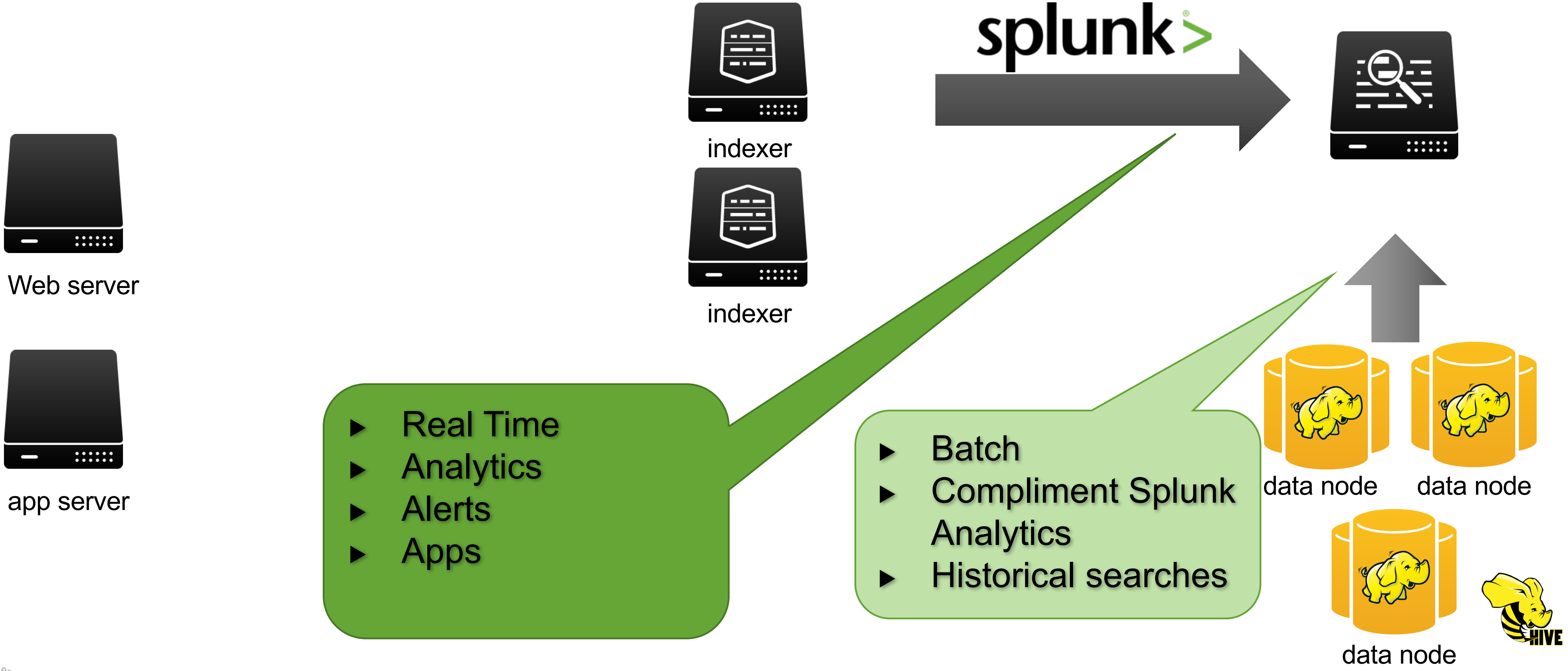
data node

data node

data node

data node

Splunk + Hadoop = All The Data

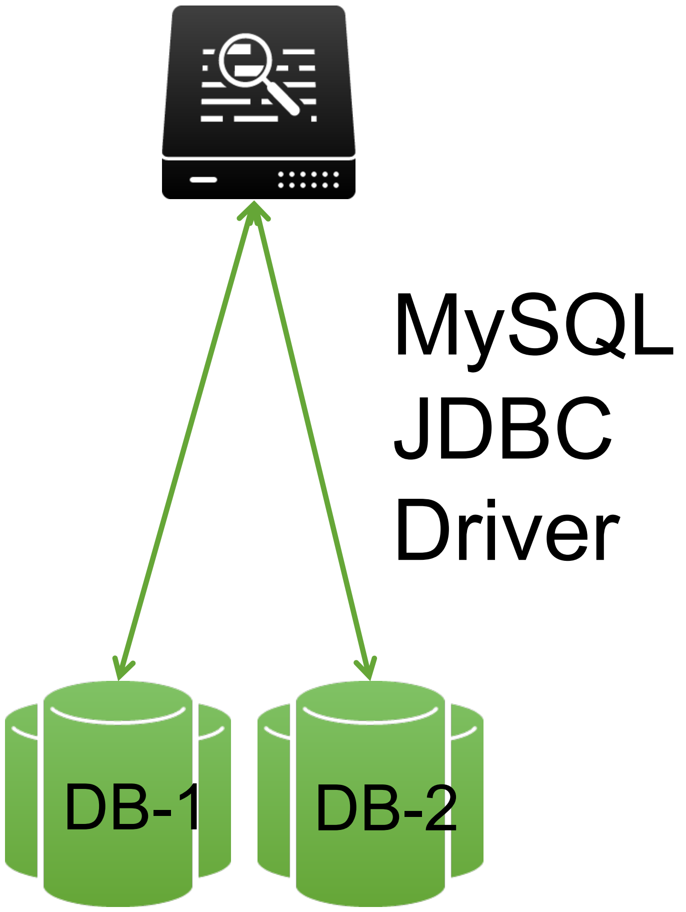


- Real Time
- Analytics
- Alerts
- Apps

- Batch
- Compliment Splunk Analytics
- Historical searches

DB Connect Architecture

- Install DB Connect on a Search Head
- Use DB Connect for Lookup
- Several Lookups coming from two different MySQL Databases
- Lookup Enrich log data with business insight



MySQL JDBC Driver

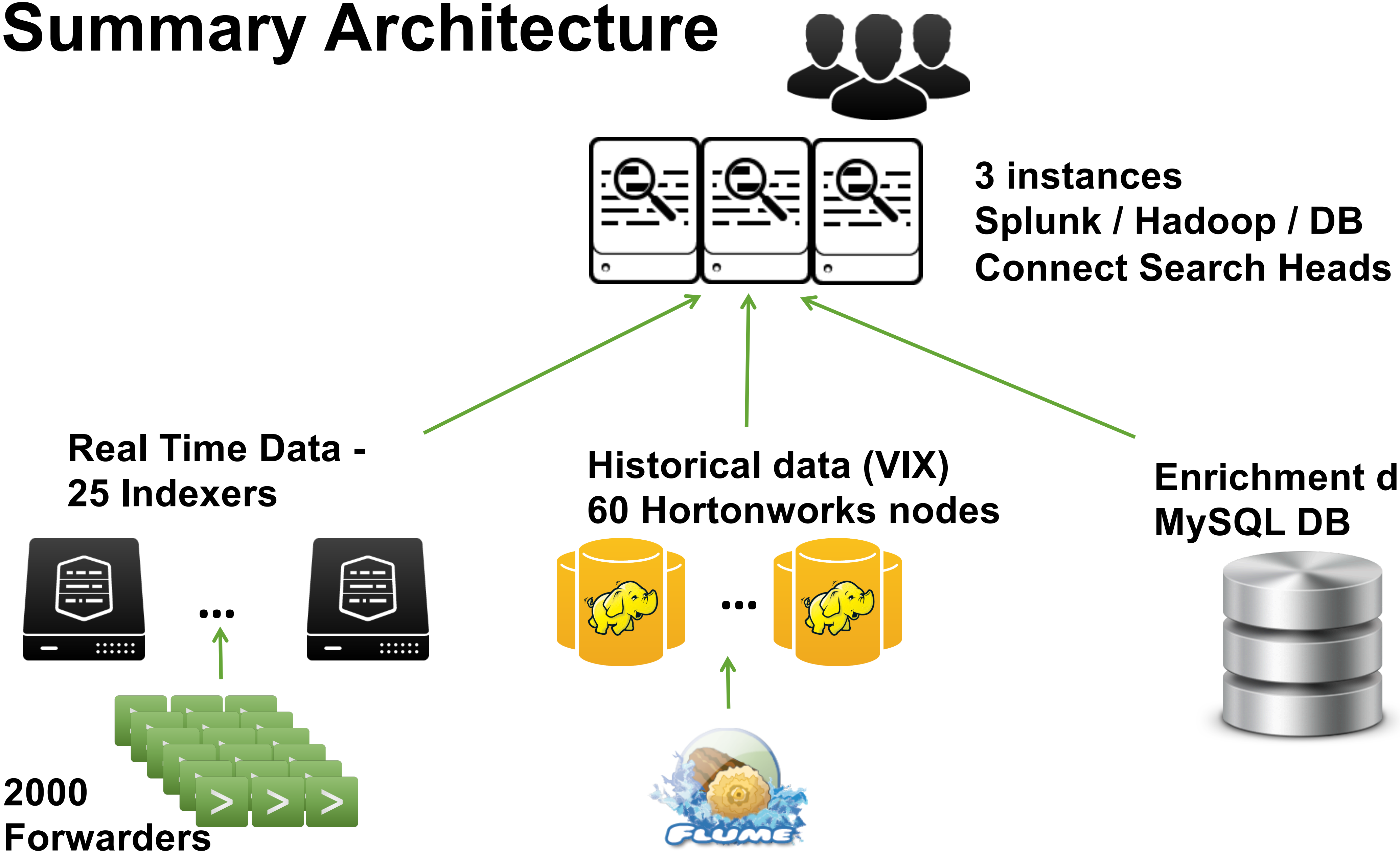
DB-1

DB-2

DB - Architecture Performance Impact

Command	Connection	Architecture
Indexing		
Inputs - dbmon-tail ** **Recommended**	Medium number of connections (Small amount of data - only delta)	DB to Index (connection pooling)
Inputs – dbmon-dump	Small amount of connections (Lots of data per connection)	DB to Index (connection pooling)
Outputs	Lots of DB Connections (Small amount of data)	Search Head to DB (connection pooling)
Not Indexing		
Search – DBXQuery	Lots of DB Connections	DB to Search Head
Lookups ** **Selected this option**	Lots of DB Connections	DB to Search Head

Summary Architecture



3 instances
Splunk / Hadoop / DB
Connect Search Heads

Real Time Data -
25 Indexers

Historical data (VIX)
60 Hortonworks nodes

Enrichment data (lookup) -
MySQL DB

2000
Forwarders

Customer Architecture Demo

Summary

- Splunk is open and extensible
- Splunk enables you to combine data from multiple sources for enriched insights
- Splunk can complement and fill the gaps in open source technologies

Thank You

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017