
Listen to your data™
splunk>



Splunking Big Data

Stephen Sorkin

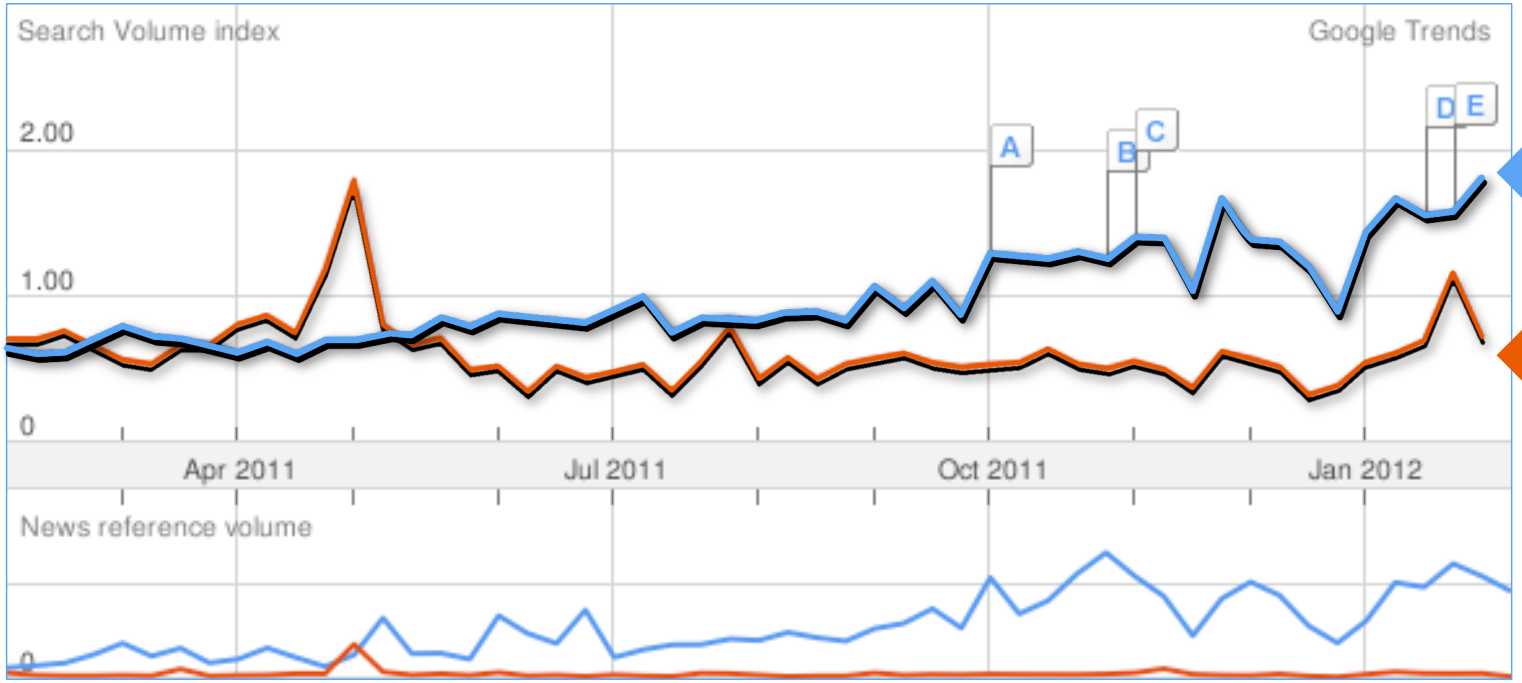
VP Engineering

February 23, 2012

splunk>

Big Data vs. Barack Obama

Search Volume Last 12 Months



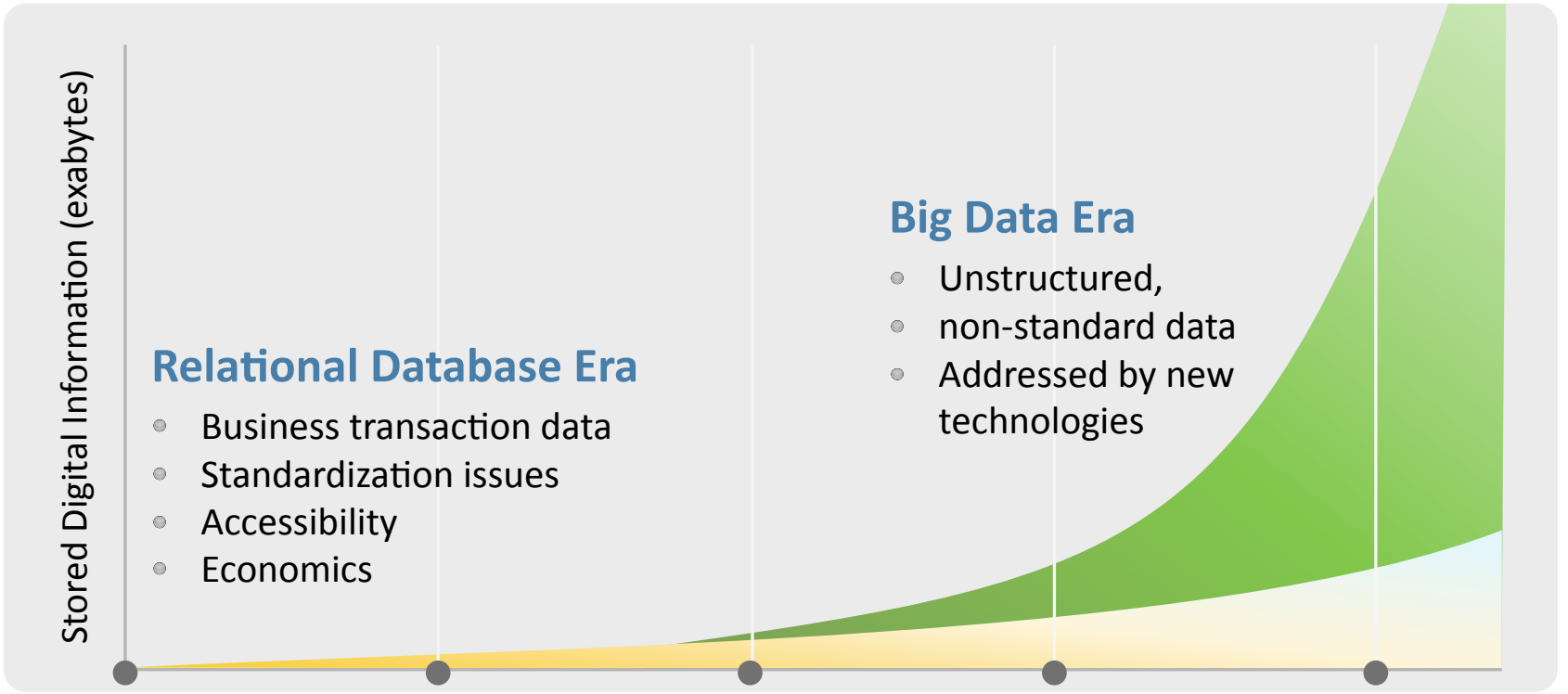
Big Data

Barack Obama

Copyright © 2011, Splunk Inc.

splunk> Listen to your data.

Traditional Approaches Solve a Different Problem



Copyright © 2011, Splunk Inc.

splunk> Listen to your data.

Splunk is a Big Data Solution

Ready when you are

- Easy to download and deploy
- Pre-integrated, end-to-end functionality
- All the enterprise-grade features you expect!

Powerful all-in-one solution

- Collects data from tens of thousands of sources
- Advanced real-time and historical analysis of data
- Fast, custom visualizations for IT and business users

Performance at scale

- Proven at multi-terabyte scale per day
- Upwards of PB under management
- 3,300+ customers

Copyright © 2011, Splunk Inc.

Splunk Collects and Indexes Any Machine Data



Customer Facing Data

- Click-stream data
- Shopping cart data
- Online transaction data

Outside the Datacenter

- Manufacturing, logistics…
- CDRs & IPDRs
- Power consumption
- RFID data
- GPS data

Windows

- Registry
- Event logs
- File system
- sysinternals

Linux/Unix

- Configurations
- syslog
- File system
- ps, iostat, top

Virtualization & Cloud

- Hypervisor
- Guest OS, Apps
- Cloud

Applications

- Web logs
- Log4J, JMS, JMX
- .NET events
- Code and scripts

Databases

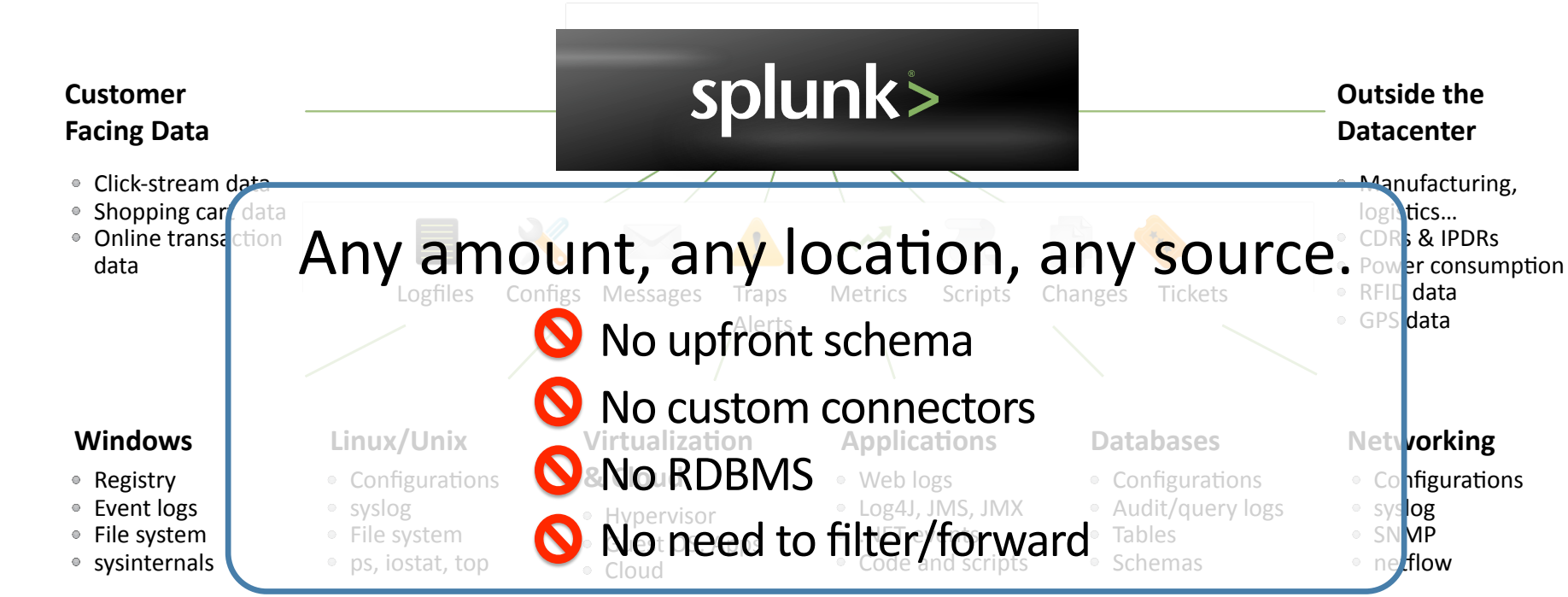
- Configurations
- Audit/query logs
- Tables
- Schemas

Networking

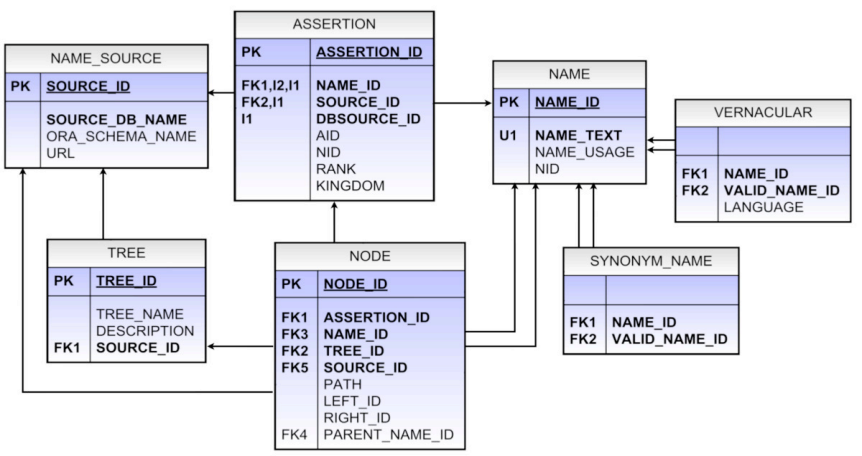
- Configurations
- syslog
- SNMP
- netflow

Copyright © 2011, Splunk Inc.

Splunk Collects and Indexes Any Machine Data



RDBMS/SQL – Early-structure Binding



SELECT customers.* FROM customers
WHERE customers.customer_id NOT IN(SELECT customer_id FROM orders WHERE year(orders.order_date) = 2004)

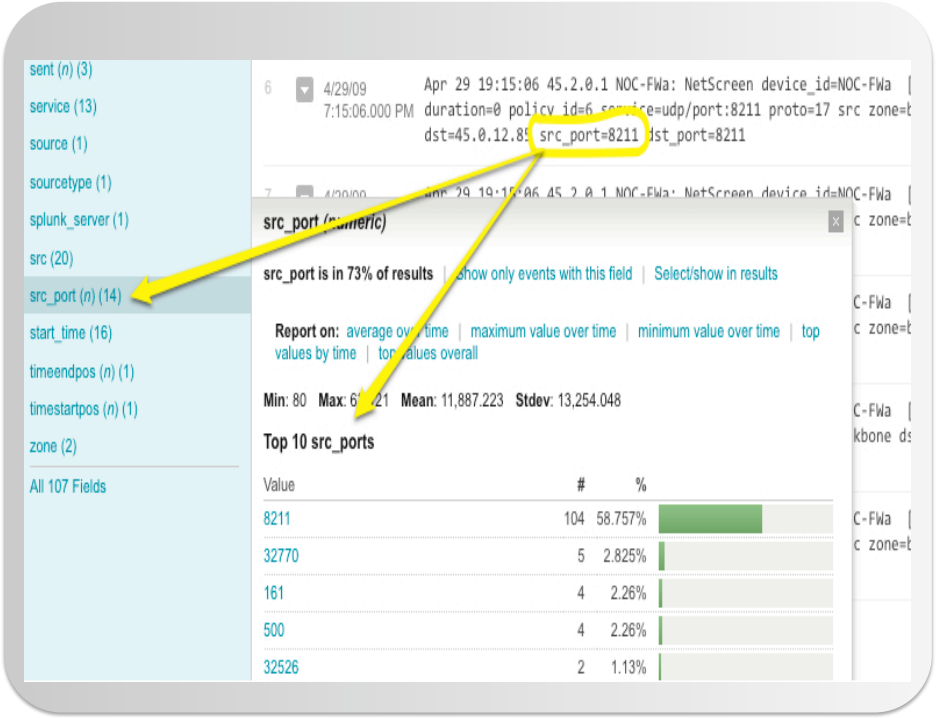
Structure

- Schema – created at *design time*
- Queries – understood at *design time* for maximum performance

Data

- Homogeneous – must fit into tables or be converted to fit into tables
- Must exactly match constraints

splunk> Late-structure Binding



Structure

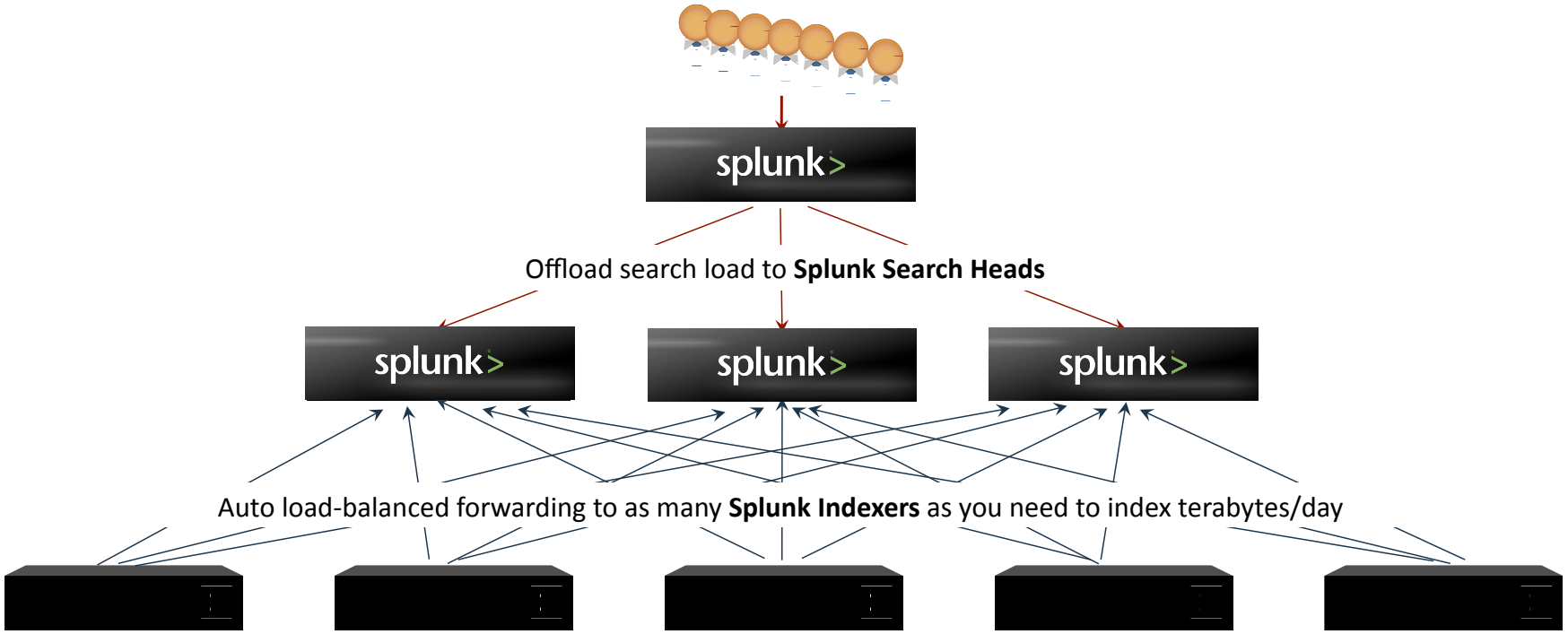
- Schema-less
- Created at *search time*
- Queries/Searches can be *ad*-hoc

Data

- Heterogeneous – can come from any text-based source
- Constantly changing
- No conversion required, no constraints

Copyright © 2011, Splunk Inc.

Massive Linear Scalability to Tens of TBs/Day



Send data from 1000s of servers using combination of **Splunk Forwarders**, syslog, WMI, message queues, or other remote protocols

Copyright © 2011, Splunk Inc.


Listen to your data™
splunk>