# PLA1735A – Getting to Know Splunk's Data Streaming Technology


**Thor Taylor**

Director of Product | Splunk

**Poornima Devaraj**

Senior Product Manager | Splunk

# Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans 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 may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

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 functionalities described or to include any such feature or functionality in a future release.

# Thor Taylor

Director of Product

# Poornima Devaraj

Senior Product Manager


splunk> .conf20

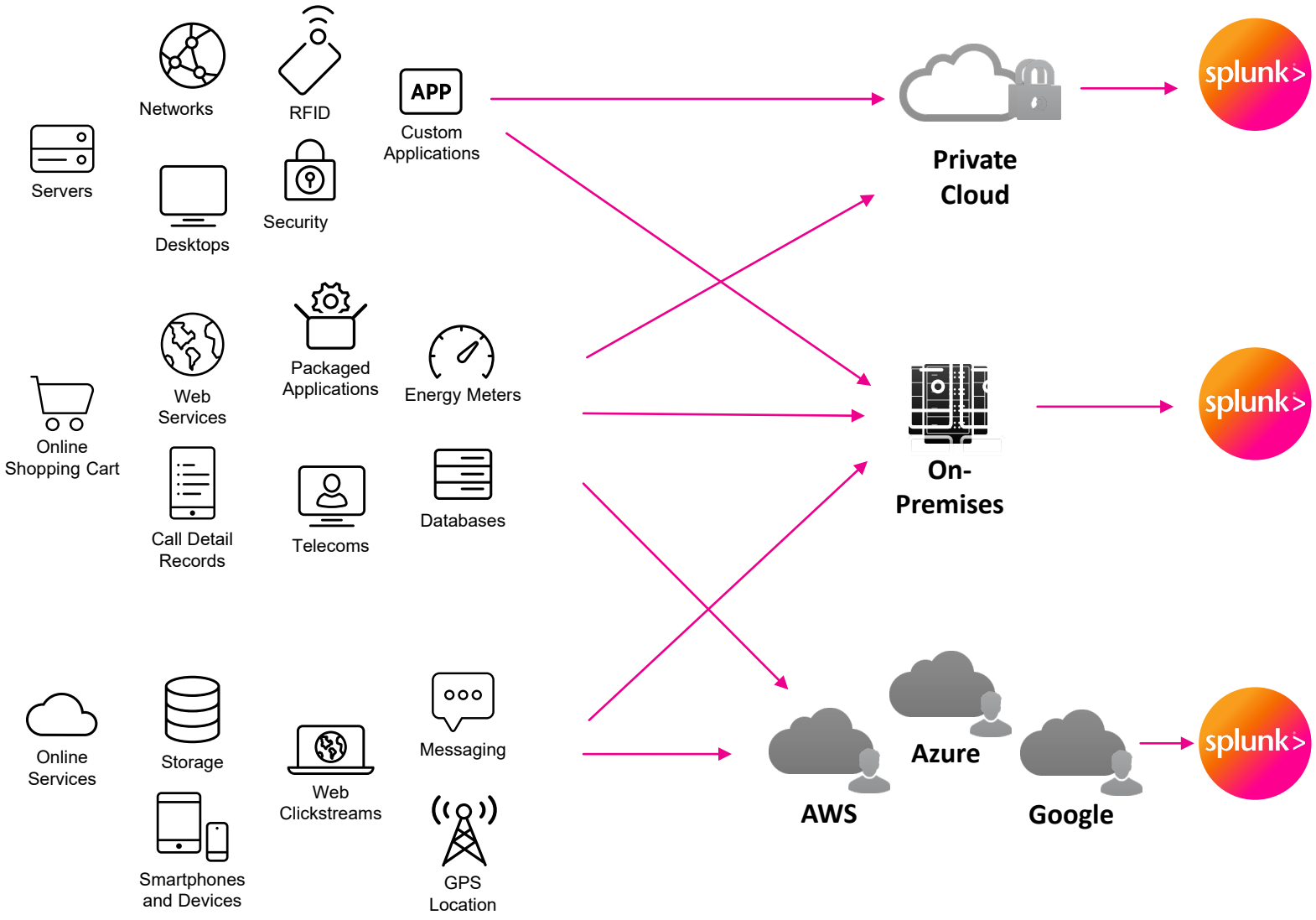
# New Names, Same Great Technologies

VictorOps → is now **Splunk On-Call**

SignalFx Infrastructure Monitoring → is now **Splunk Infrastructure Monitoring**

SignalFx Microservices APM → is now **Splunk APM**

# The Data Fabric




1. How important is your data in driving business decisions?
2. What happens if data is lost because of failures in transport?
3. Do you know when data stops flowing or if data drift occurs?
4. Do you have the ability to isolate high value data from low value, noisy data?
5. Can you enrich your data, before it gets downstream using lookups or ML?
6. Overall how can you make the handling and transportation of data more predictable and visible?

# Turning Real-time Data Into Action is Hard

## Data Generated

**1.7MB**
every second
**2,500PB**
every day

## Hybrid Environment

**Cloud** adoption grows but…
**65%** of Enterprise workloads are on premises

## Business Actions

**Insights to Data**
Customer purchases, supply chains, online purchasing

# Customer Challenges

## Control

Massive amounts of data make it hard to collect, protect and deliver the right data to the right users and systems

## Visibility

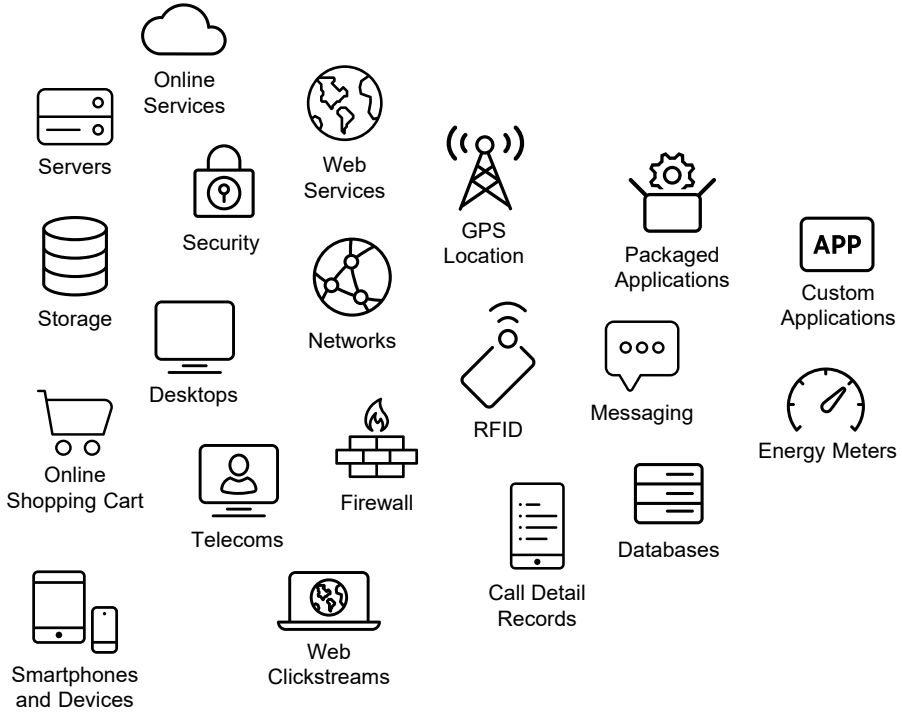
Data driven decision making is challenged with multiple instances, subsidiaries, on-premise + cloud/multi-cloud

## Insights

Generate business-critical insights faster to remain competitive in data-driven environment

# Splunk Data Stream Processor

## A real-time stream processing solution that collects, processes and delivers data to Splunk and other destinations in milliseconds



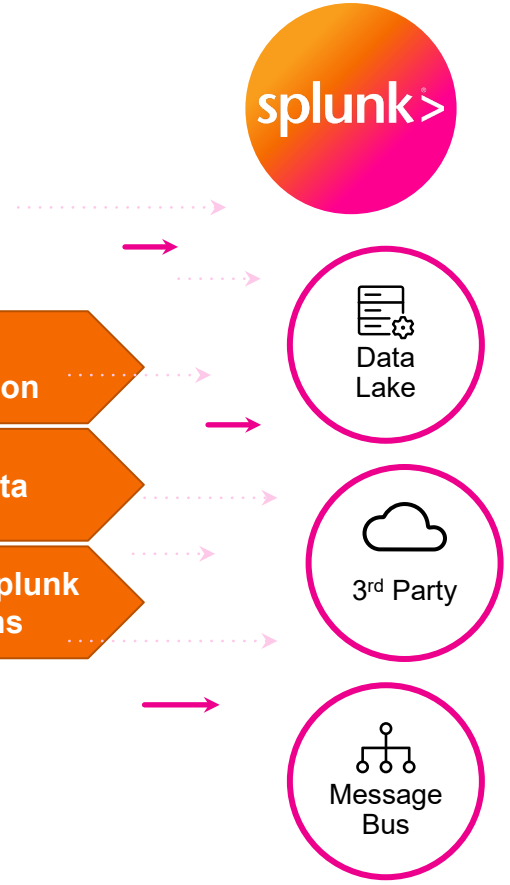

**Splunk Data Stream Processor**

| | |
|---|---|
| Filter | Enrich |
| Normalize | Transform |
| Aggregate | Learning |
| Mask Sensitive Data | |
| Detect Data Patterns or Conditions | |

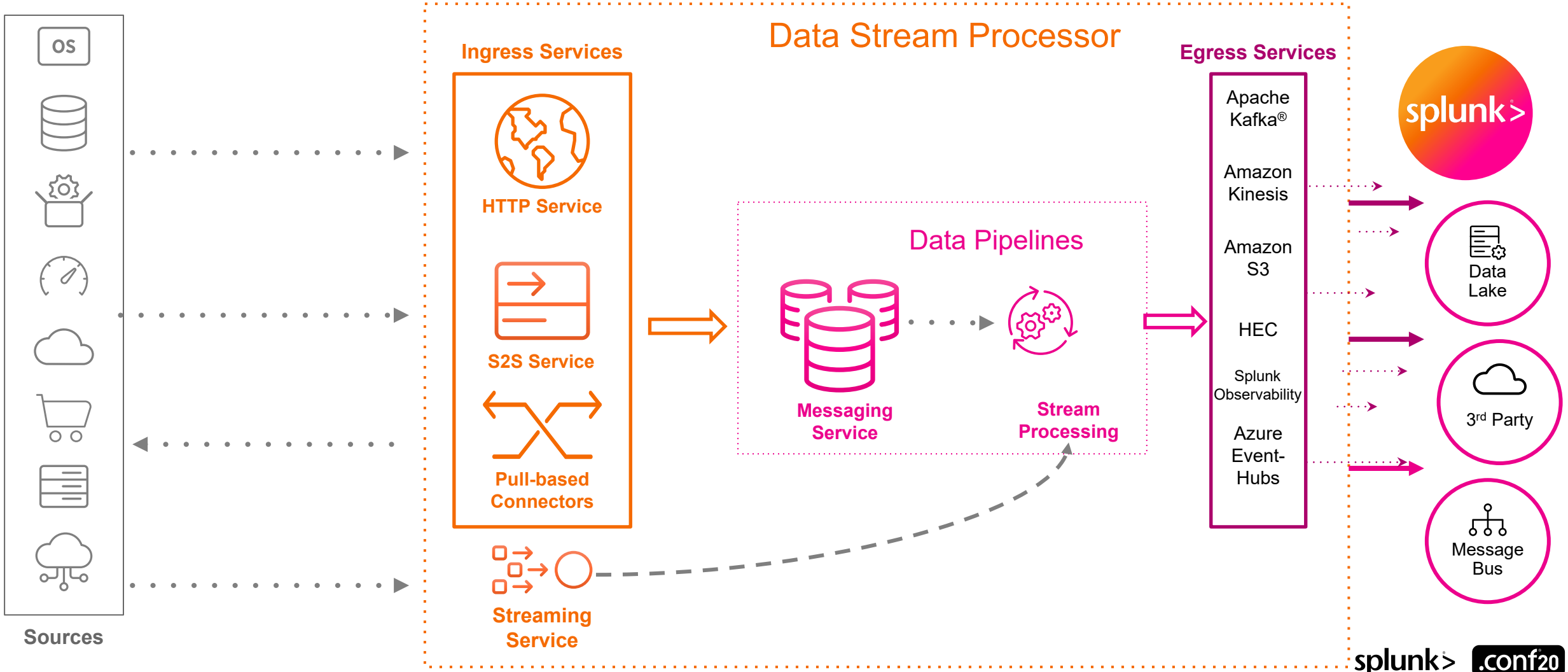
**Turn Raw Data Into High-value Information**

**Protect Sensitive Data**

**Distribute Data To Splunk Or Other Destinations**



# DSP Architecture



# Refresher – DSP 1.1

Additional sources and sinks enable enterprise data collection and distribution:

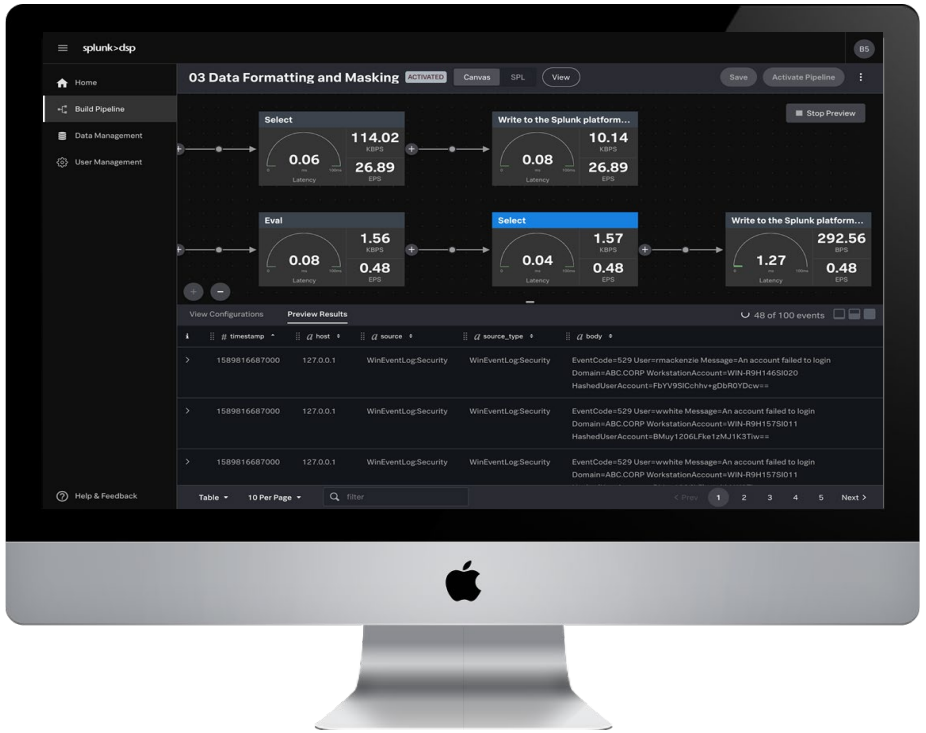
- AWS S3, GCP Monitoring Metrics, Microsoft Office 365, Splunk Observability + Splunk Enterprise – ingest and route data to other teams and systems

Updated Infrastructure

- Flink update improving performance and support for new Sinks
- Language updated to SPL2 simplifying learning DSP
- Move from Kafka to Pulsar to improve scale and resiliency

Data Compliance

- Uncover sensitive information in-stream, then mask before routing downstream



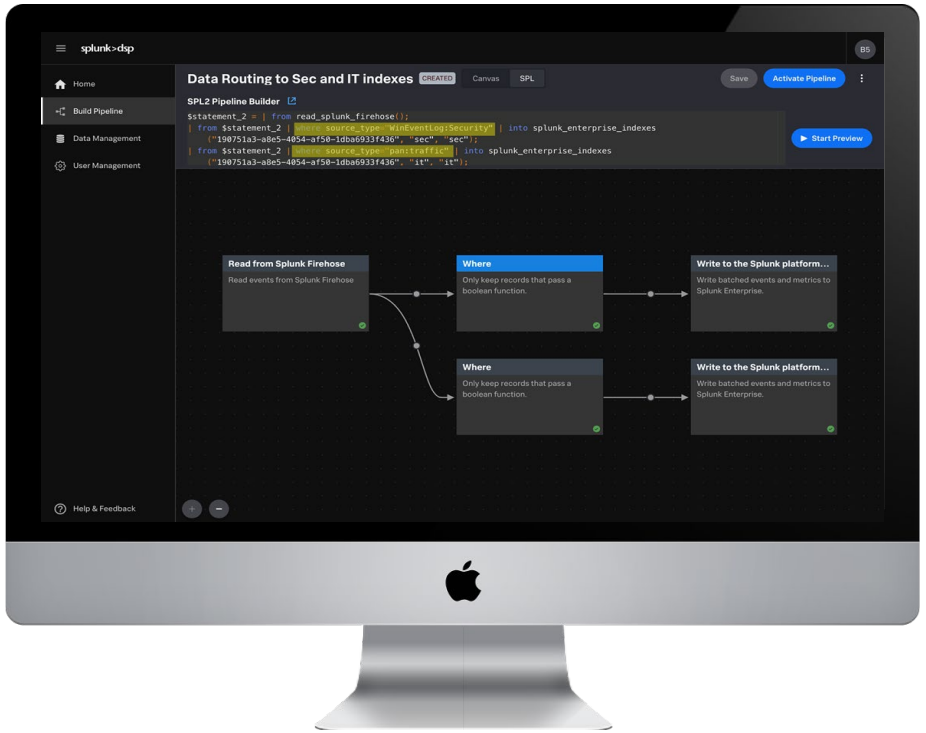
# Coming Soon DSP 1.2

## Data access + support for multi-cloud environments:

- GCP Pub/Sub, Azure Event Hub – ingest and route data to other teams and systems + manage cloud infrastructure sprawl
- Convert logs to metrics and traces on the stream – route to Splunk Observability

## Accelerate insights w/enrichment

- Lookups provide in-stream data enrichment at scale, making downstream search more relevant and accurate
- Unbounded ML functionality on the stream w/dedicated GUI for easy access
- New content updates supporting timestamp and linebreak automation



# Beta Signup Today! (SaaS)

## Supported Data Sources*:

Apache Pulsar, Amazon Kinesis, Splunk (Universal Forwarder, Heavy Weight Forwarder), Azure Event Hubs, Syslog (Splunk connect for Syslog), HEC, Google Pub/Sub

## Supported Destinations*:

Amazon Kinesis, Splunk Enterprise/Cloud, SignalFx (Metrics/Traces), Azure Event-Hubs, S3

## SVC Based Pricing

## Requirements:

Region

- US East
- Others coming soon

New and Existing Customers

## Sign-up:

http://splunk.com/DSPSaaS

* *More sources and destinations to come in future releases*

# Get Started Today! (On-Prem)

## Supported Data Sources*:

Apache Kafka®, Amazon Kinesis, Ingest REST APIs, Splunk (Universal Forwarder, Heavy Weight Forwarder), Azure Event Hubs, Azure Monitor Metrics, Syslog (Splunk connect for Syslog), HEC, Amazon S3, Amazon Metadata, Amazon CloudWatch Metrics, Microsoft 365, Google Cloud Monitoring Metrics, Pub/Sub

## Supported Destinations*:

Apache Kafka®, Amazon Kinesis, Splunk, SignalFx (Metrics/Traces), Amazon S3, Azure Event-Hubs

## Infrastructure Based Pricing (vCPUs)

* *More sources and destinations to come in future releases*

## Hardware Requirements:

Minimum Node Requirement

- CPU: 8 core (16 recommended)
- Memory: 64GB (128GB recommended)
- Network: 10Gbps
- Storage: 1TB

Minimum 3 Node Cluster

## Sign-up:

https://www.splunk.com/en_us/software/stream-processing.html


# Thank You

Please provide feedback via the

**SESSION SURVEY**