

FN20602

Data Stream Processor: How to get the most out of your data!





Blaine Wastell

Product Management Director, Splunk



Thor Taylor

Product Management Director, Splunk

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.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

Our World Never Stops Evolving.

////////////////////////////////////
New Ideas. New Devices. New Processes.

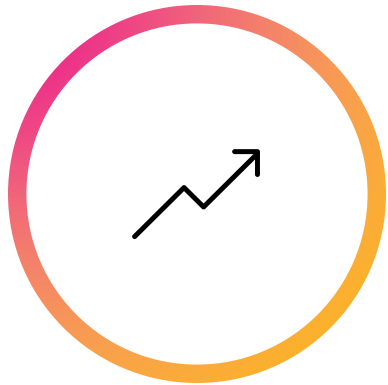


Turning Real-time Data Into Action is Hard



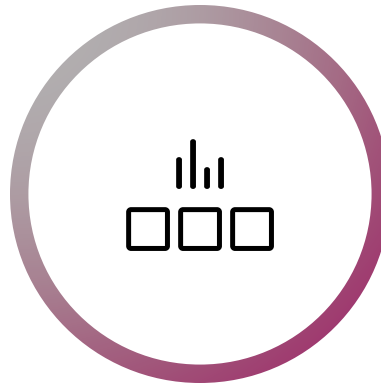
Organizations need a solution to guarantee efficient data delivery across the enterprise

Real-time Data Transformation



Format and remove unnecessary or noisy data in the stream

Continuous, Real-time Insights



Detect specific conditions or data patterns before data reaches its destination

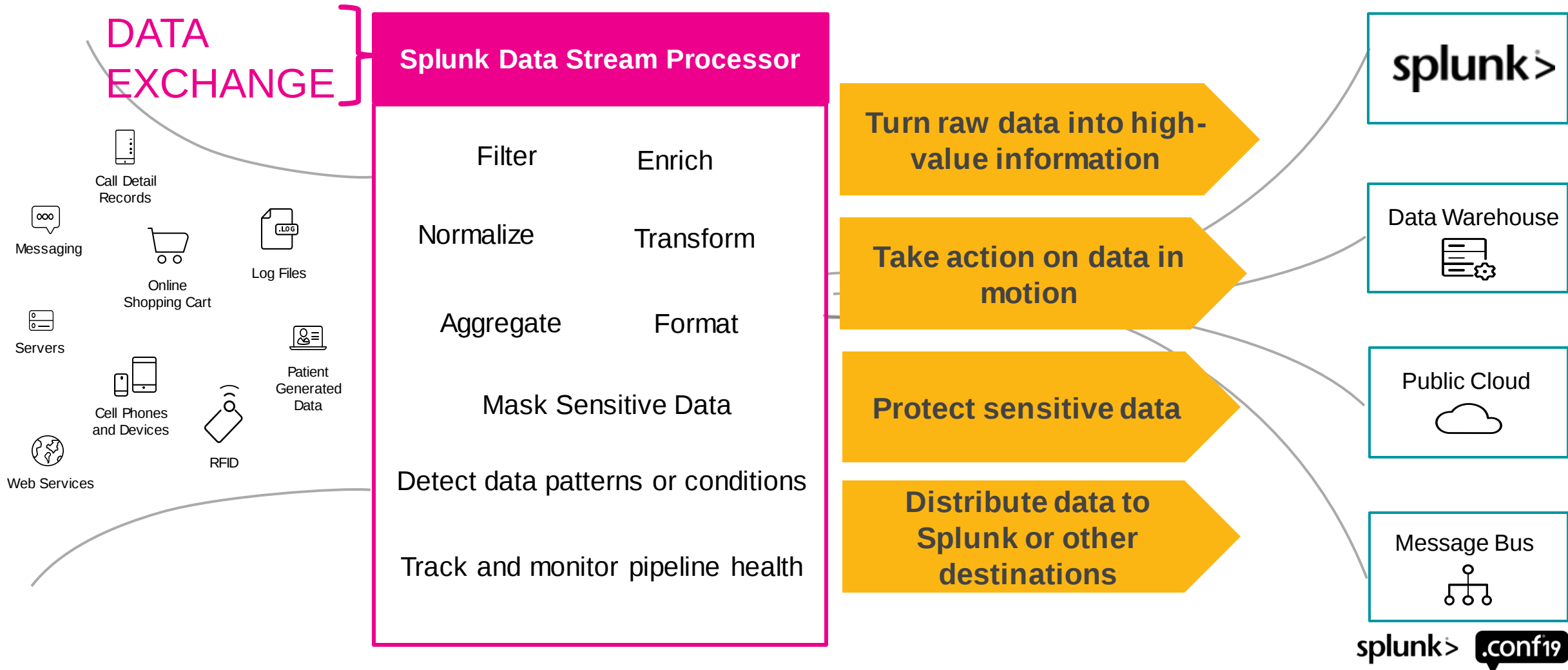
Enterprise-Wide Data Delivery



Guarantee delivery of high-volume, high-velocity data to multiple systems

Splunk Data Stream Processor

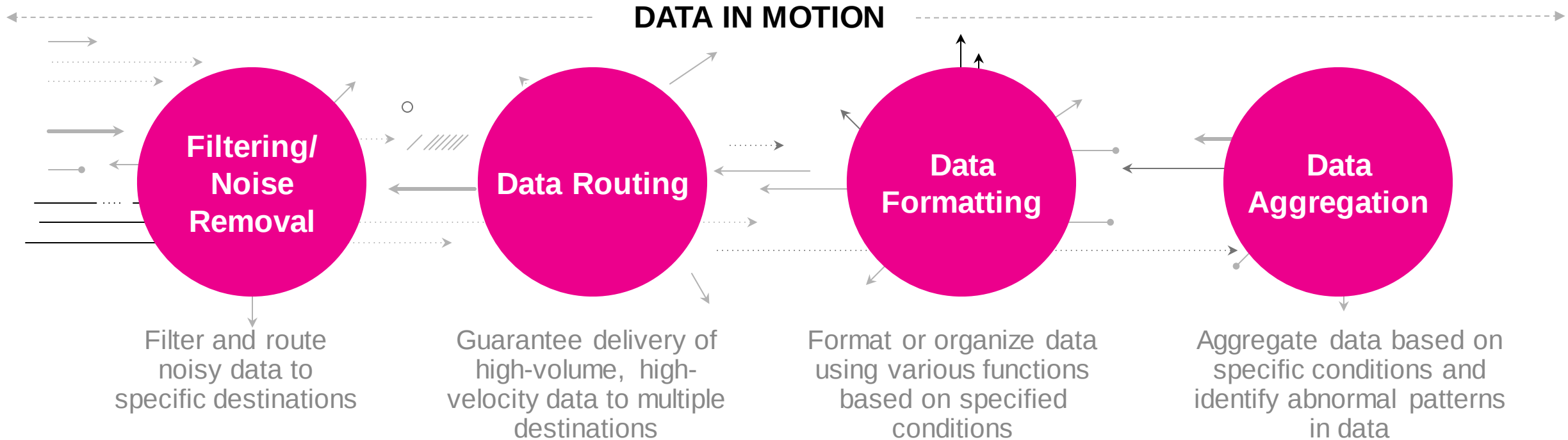
A near real-time stream processing solution that collects, processes and delivers data to Splunk and other destinations in milliseconds



The Business Impact

Compliance & Data Privacy	Improve compliance with data related regulations such as GDPR	Mask sensitive data such as user identifying information or credit card information
Immediate Business Insights	Detect abnormal activity or behavior as data is in flight, before data is sent downstream	Summarize individual events that don't meet specific thresholds or conditions and immediately send to analysts for further investigation
Operational Efficiency + Worker Productivity	A single data exchange for the entire enterprise that allows users to spend more time on the data that matters	One place to collect data from diverse sources, turn data into valuable information or insights, then distribute results to multiple systems for various users to consume

Use Cases



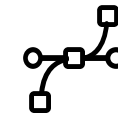
Filtering + Data Routing

SCENARIO: BILLIONS OF CALL DETAIL RECORDS NEED TO BE DELIVERED TO VARIOUS TEAMS FOR DEEP ANALYSIS

DSP BENEFIT: ONE PLACE TO GATHER ALL DATA, FILTER OUT LOW VALUE RECORDS AND DISTRIBUTE HIGH-VALUE RECORDS TO RELEVANT TEAMS FOR QUICK ANALYSIS

How do I quickly send relevant data to various users while also storing all call detail records for future use?

50TB
EVENTS
7
DESTINATIONS



Filter & Route to various destinations with DSP

Filter low value records and route high value records to relevant destinations, including a long-term storage destination for the low value records.

Data Formatting

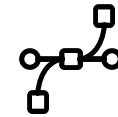
SCENARIO: 30TB OF DATA WITH PRIVATE INFO ON BILLIONS OF CUSTOMERS NEED TO MASKED BEFORE SENDING TO VARIOUS DOWNSTREAM SYSTEMS

DSP BENEFIT: CONTINUOUSLY REDACT SENSITIVE USER INFO BEFORE SENDING TO DOWNSTREAM SYSTEMS

How can we quickly deliver large datasets to multiple systems while ensuring sensitive user info isn't sent to the wrong systems

30 TB
Events

18
Destinations



**Mask Private Data
with DSP**

DSP **masks** or hides **sensitive information** on the fly, **before sending** data to other systems, ensuring efficient data delivery while **maintaining compliance** standards.

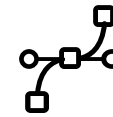
Data Aggregation

SCENARIO: ENERGY COMPANY NEEDS TO MONITOR HEALTH OF 8M ENERGY METERS ACROSS VARIOUS REGIONS

DSP BENEFIT: SUMMARIZE METERS THAT AREN'T PERFORMING, SEND INSIGHTS DOWNSTREAM AND PREVENT
DOWNSTREAM SYSTEM OVERLOAD

Which meters are healthy vs. which ones need to be investigated for abnormal behavior?

120TB
EVENTS/METRICS
8M
SMART METERS/ DATA
SOURCES



Summarize
billions of events
with DSP

Aggregate meter health stats, compare against predefined conditions, then send metrics downstream for analysts to quickly take action.

Why Splunk for Stream Processing?

Out-of-the-Box Stream Processing Technology

Complete stream processing solution built with advanced functions such as data aggregation

Complements the Splunk Platform

Built to complement Splunk's portfolio of products that solve IT, Security, IOT, and business analytics use cases

Custom Functions

Build your own processing functions to meet your specific business needs

Integration with Other Systems

Easily integrate with other systems such as Kafka, S3, CloudTrail, Kinesis, Event Hubs, Universal Forwarders, Heavy Weight Forwarders, HEC plus more



University of Illinois

Efficiently delivers data campus-wide while protecting sensitive student information

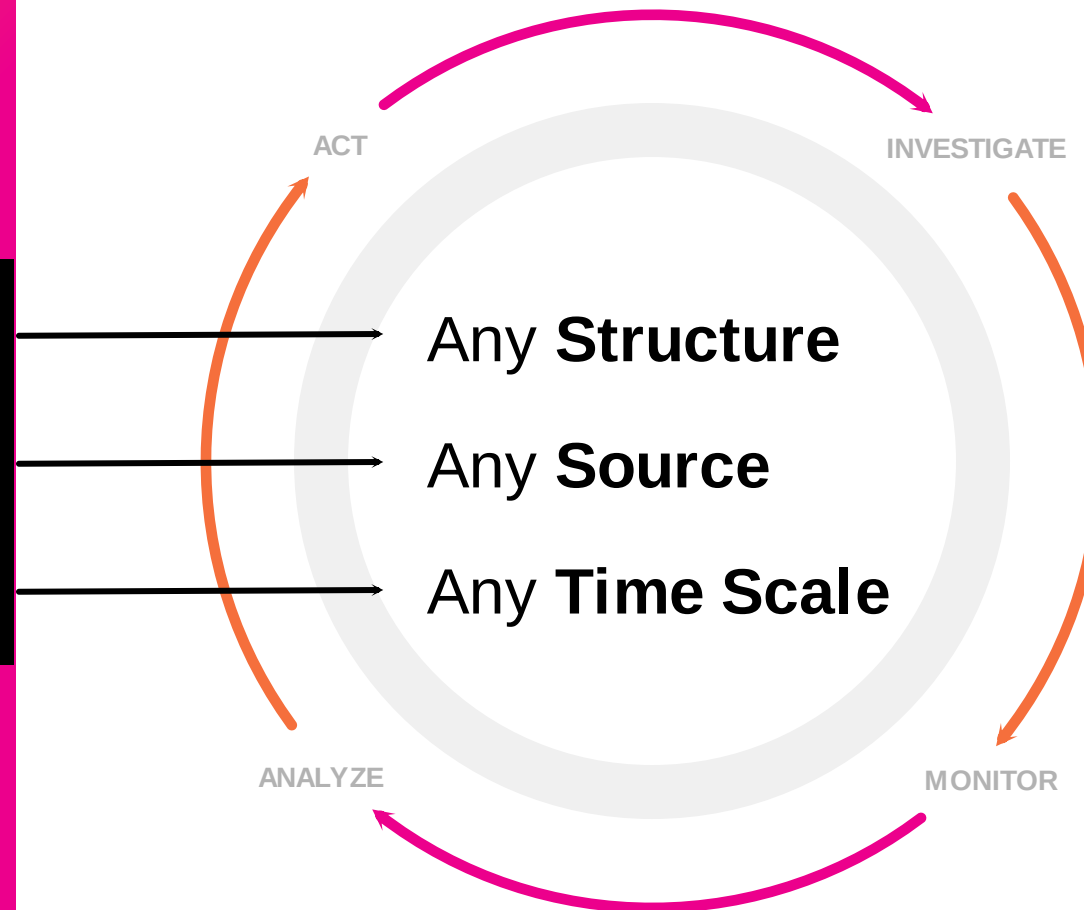
“We are committed to being a leader in teaching, research and public engagement and these were important considerations that went into our selection of Splunk Cloud as our campus-wide data platform. Splunk Cloud helps us manage everything from IT and Security to providing personalized student experiences both in and out of the classroom. **Splunk Data Stream Processor’s ability to mask private student information in real-time gives us a very effective tool to create a more safe and connected campus.**”

—Joe Barnes, Chief Privacy and Security Officer

Splunk Data Stream Processor

The Data-to-Everything Platform

- ✓ High-velocity, high-volume **unstructured** or **structured** data
- ✓ Deliver data to Splunk or **any destination**
- ✓ **Millisecond** data processing and insights



IT

Security

IoT

Biz
Analytics

Data Stream Processor Explained

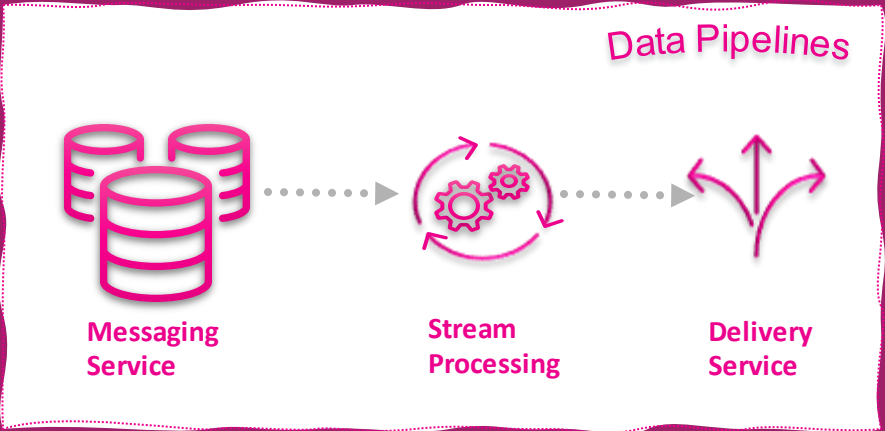


DSP Architecture

Data Pipelines



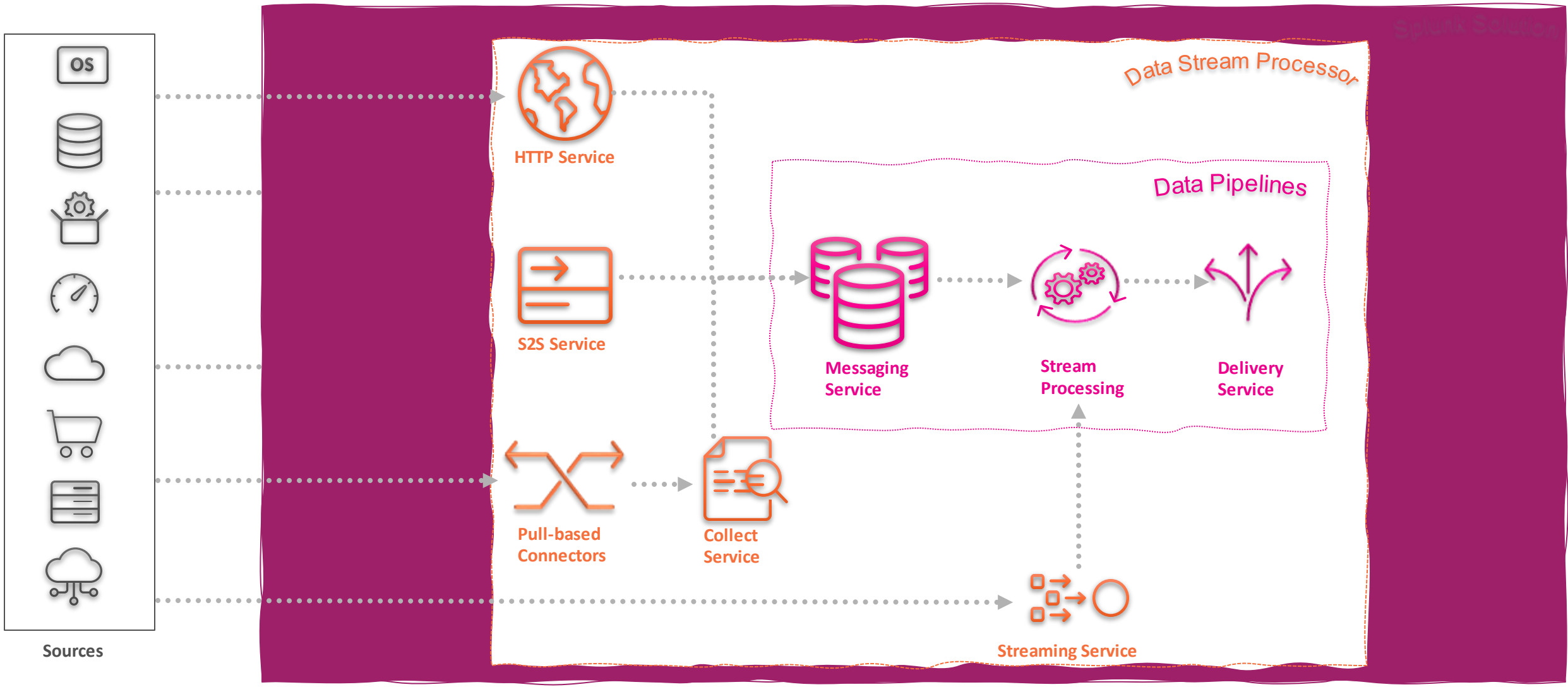
Sources



Splunk Solution

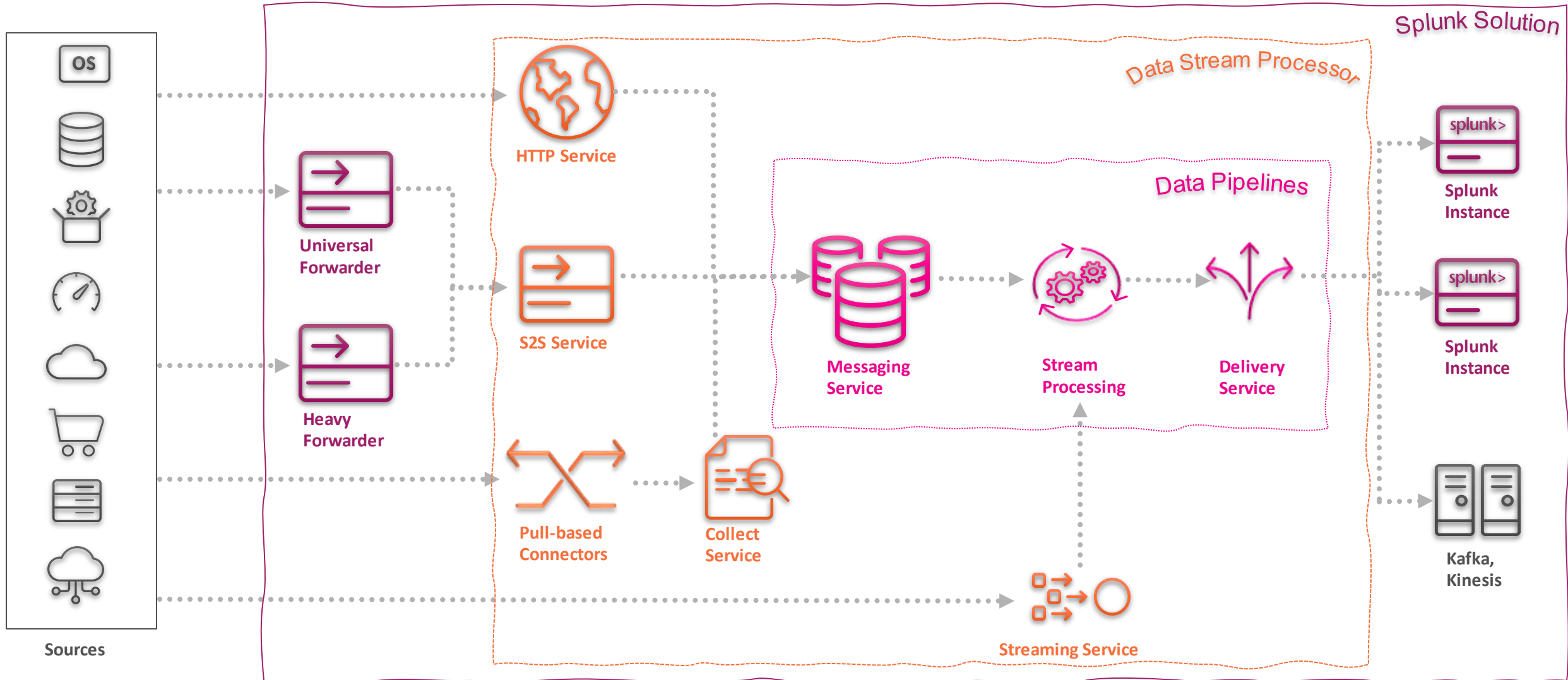
DSP Architecture

DSP



DSP Architecture

Splunk Solution



Use Case Grouping



Filter and Forward - Basic

Filter and forward to both Splunk destinations as well as 3rd party destinations is by far the most common use case - *Data Delivery*

- Improve data resiliency
- End to end visibility
- Future proof data ingest
- Deliver data to any destination



Filter, Format and Forward - Intermediate

Adding in formatting to enrich, normalize and parse events as they move downstream improving data quality for the business - *Data to Information*

- Detect and act on sensitive data
- Normalize data across the enterprise
- Capture perishable insights



Filter, Format, Aggregate and Forward - Advanced

Providing state-based operations to garner better insights and providing greater control for high throughput, low value data sets - *Data Insights/Reduction*

- Detect insights across all data
- Extract high value information from noisy, low value data
- Summarize billions of data points

Packaging

Features

Use Case
Distribution

Support

Pricing



Splunk Edition

Send to Splunk only
destinations

Basic – 20%
Intermediate – 60%
Advanced – 20%

Included with DSP
License

Discuss with Account
Team



Universal Edition

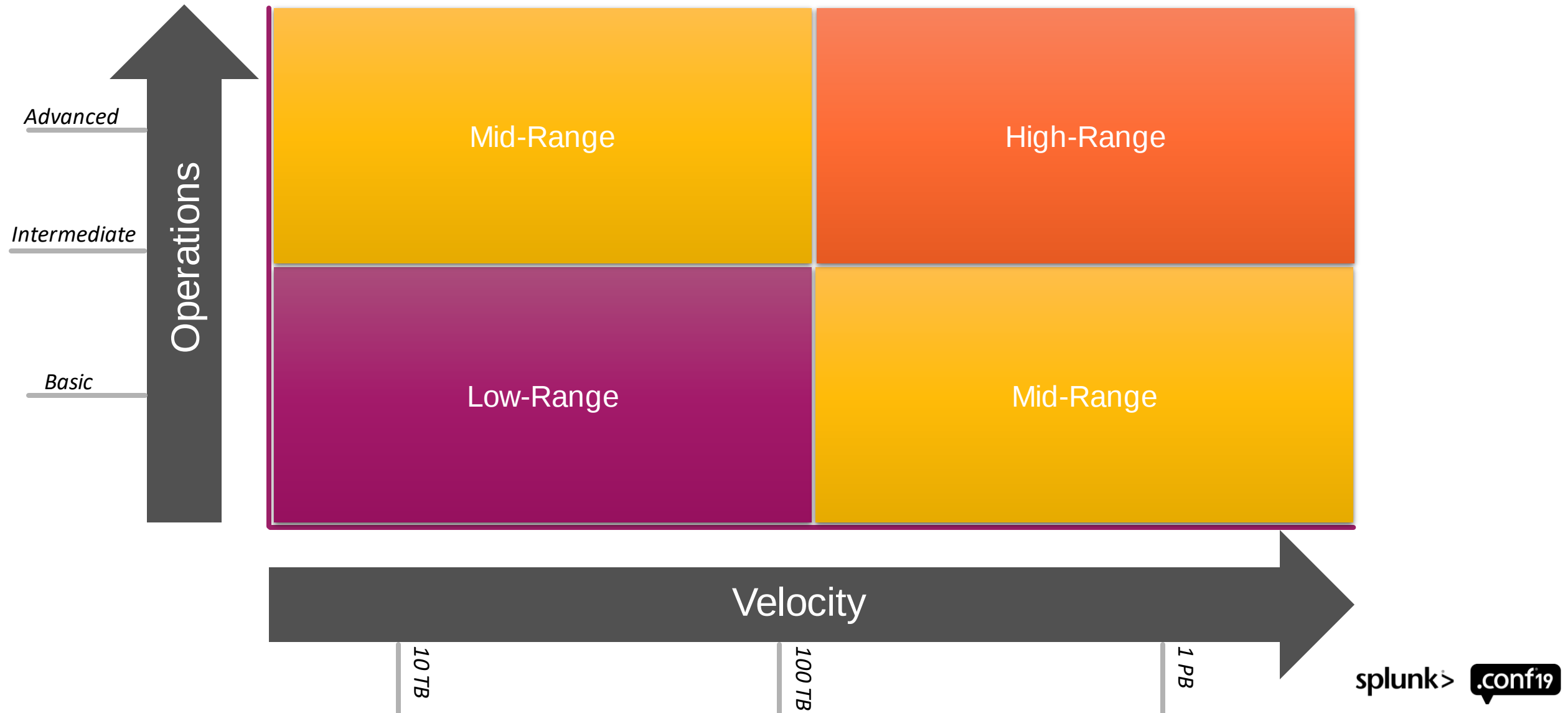
Send to any
destination

Basic – 10%
Intermediate – 70%
Advanced – 20%

Included with DSP
License

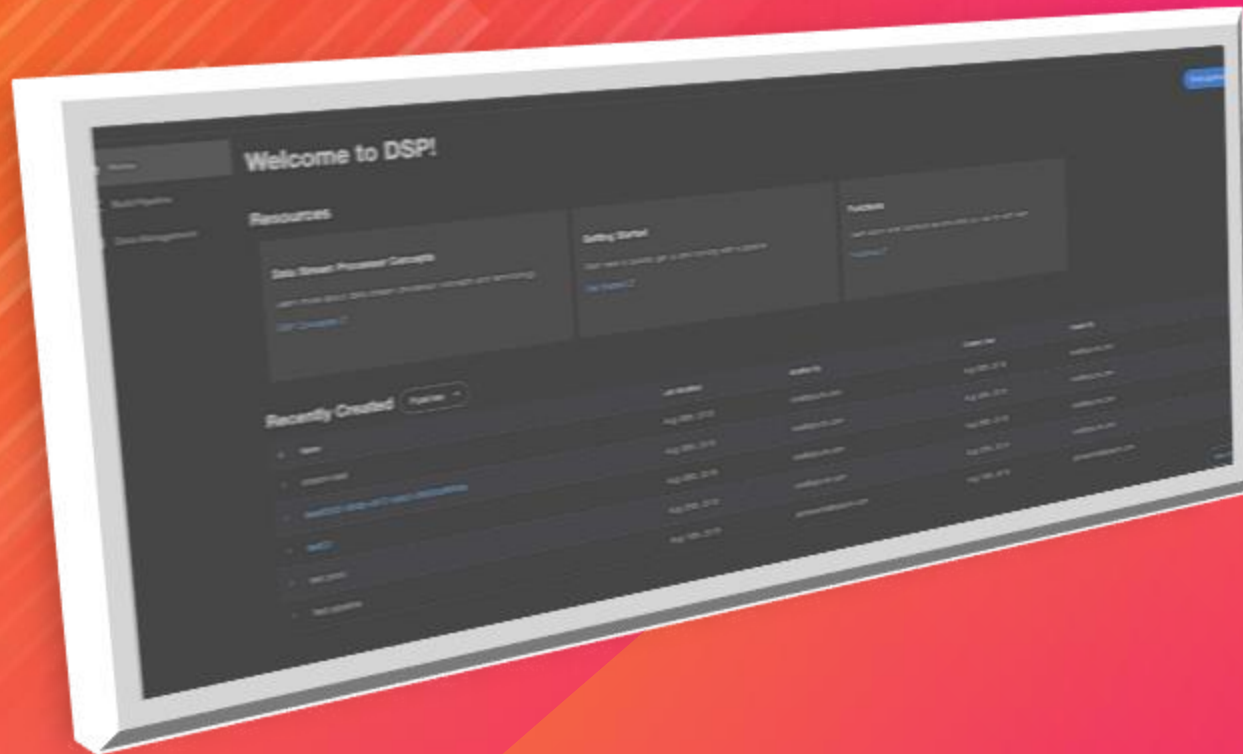
Discuss with Account
Team

vCPU Expectations



.conf19

splunk>



Demo

splunk> turn data into doing™

Key Takeaways

Use Data Stream
Processor to help make
data driven decisions

1. Ingest data with a **scalable** and resilient architecture
2. Route data to **Splunk** indexes or **external** destinations
3. Transform data **before writing to indexes**: remove noise, shape data, enrich data, redact personal identifiable information
4. Send data to either **DSP** or **Splunk indexes** using Universal or Heavy Weight Forwarders

Data Stream Processor Sessions

1. DEV1317 - Data Stream Processor: Architecture and SDKs
2. FN1786 - Using Splunk Data Stream Processor for advance stream management
3. FN1987 - Using Splunk Data Stream Processor as a streaming engine for Apache Kafka
4. FN2033 - Using Splunk Data Stream Processor as a data transformation, altering, and action engine
5. FN2062 - Data Stream Process: How to get the most out of your data!

Get Started Today!

Q&A

- ✓ **Supported Data Sources*:** Apache Kafka®, Amazon Kinesis, Amazon S3, AWS CloudTrail, Azure Event Hubs, REST APIs, Splunk (Universal Forwarder, Heavy Weight Forwarder)
- ✓ **Supported Destinations:** Apache Kafka®, Amazon Kinesis, Splunk
- ✓ **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 5 Node Cluster**
- ▶ **Daily Ingestion:**
- ▶ **7 – 25 TB**





**Thank
You!**

Go to the .conf19 mobile app to

RATE THIS SESSION

Appendix

Ingestion Methods

Method	Splunk Enterprise	DSP
Monitor file or path Through Universal Forwarder or Heavy Weight Forwarder	✓	✓
Upload a file	✓	
TCP & UDP Not recommended; better to use Syslog-ng	✓	
HTTP Event Collector Send to HTTP for App logs or Splunk Connect for Kubernetes, Docker, Kafka	✓	Ingest API (similar to HEC)
Scripts & mod inputs Collects data from API or generate data from system or service	✓	✓ Use Universal Forwarder or Heavy Weight Forwarder
Add-ons	✓	Connector + Templates + Groups
Splunk Connect Kafka, Kubernetes, Docker, PCF	✓	
LSDC connectors (for DSP) Pull based connector		✓
Push based native DSP connectors Kinesis, Azure Event Hubs, Apache Kafka		✓

data into doing™