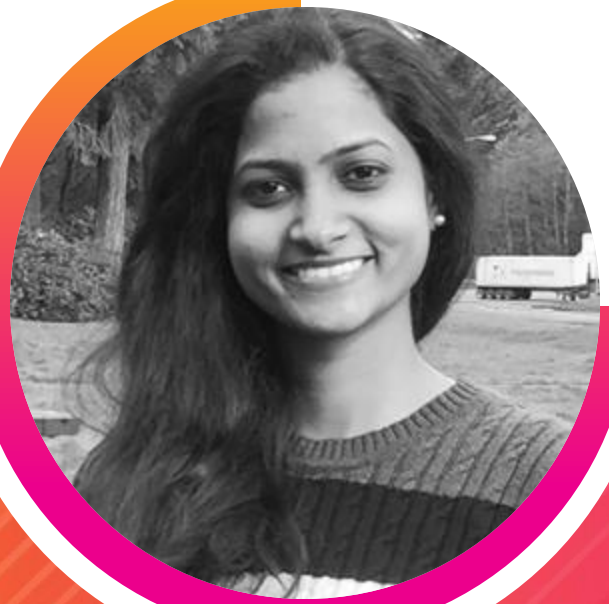
Collect Service: Introducing a New Ingest Model

splunk> .conf19



Poornima Devaraj
Technical Product Manager | Splunk

Jove Zhong
Engineering Director | Splunk

splunk> .conf19

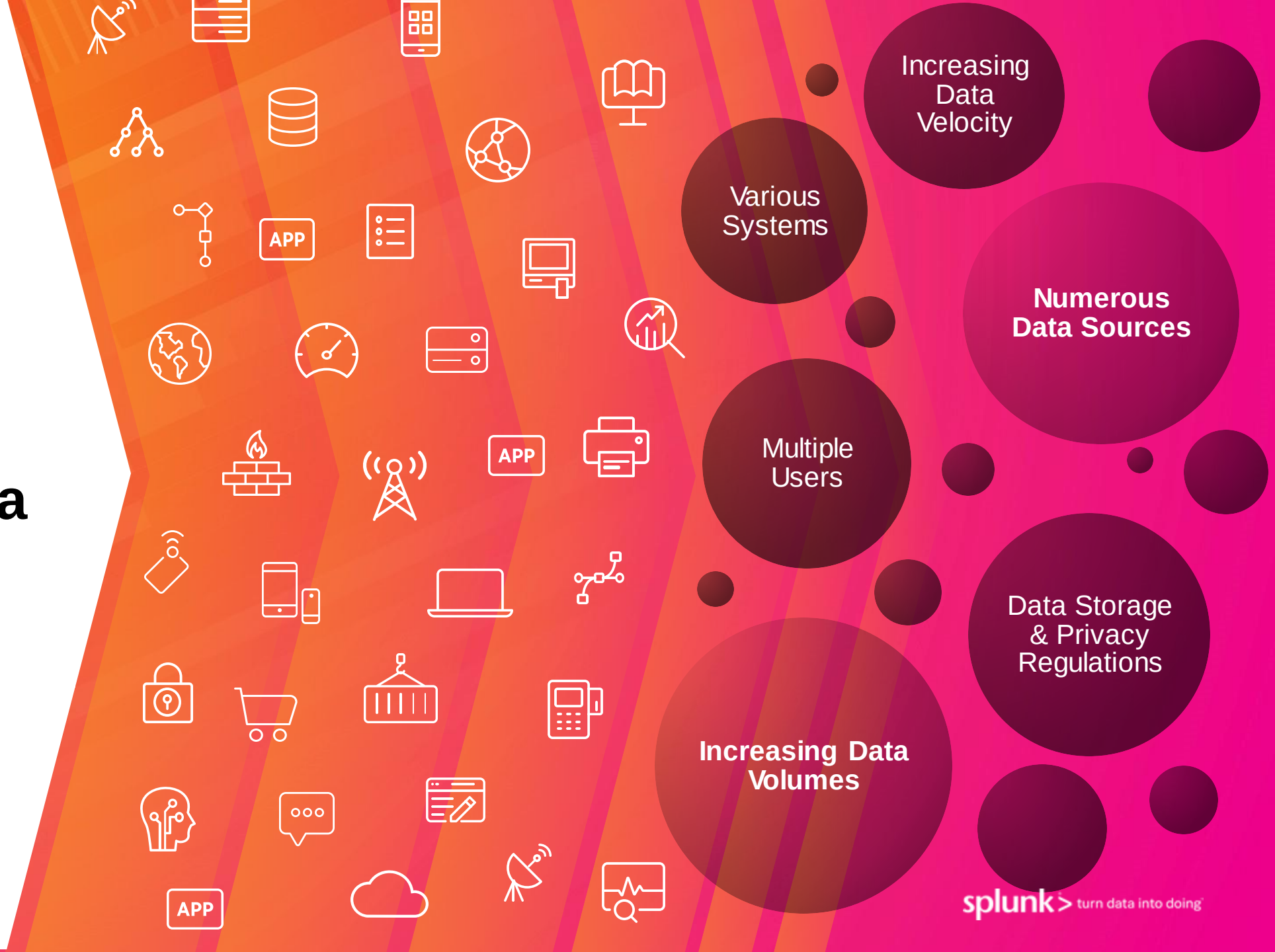
Forward-Looking Statements

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.

splunk> turn data into doing

Turning Real-time Data Into Action is Hard



splunk> turn data into doing

Splunk Data Stream Processor

Splunk Data Stream Processor

Filter
Enrich
Normalize
Transform
Aggregate
Format
Mask Sensitive Data
Detect data patterns or conditions
Track and monitor pipeline health

Turn raw data into high-value information

Take action on data in motion

Protect sensitive data

Distribute data to Splunk or other destinations

splunk>

Data Warehouse

Public Cloud

Message Bus

splunk> .conf19

Get Data In Splunk

splunk> .conf19

Get Data in Splunk

Get started

Files and Directories

Upload a file or index a local file or monitor an entire directory

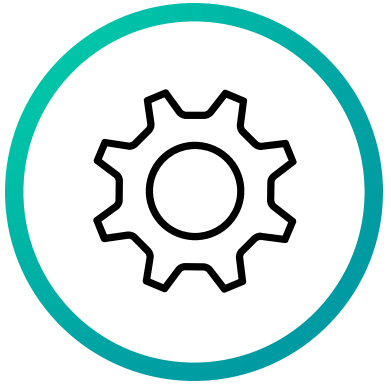
HTTP Event Collector

Configure tokens that clients can use to send data over HTTP or HTTPs

TCP/UDP

Configure the Splunk platform to listen on a network port

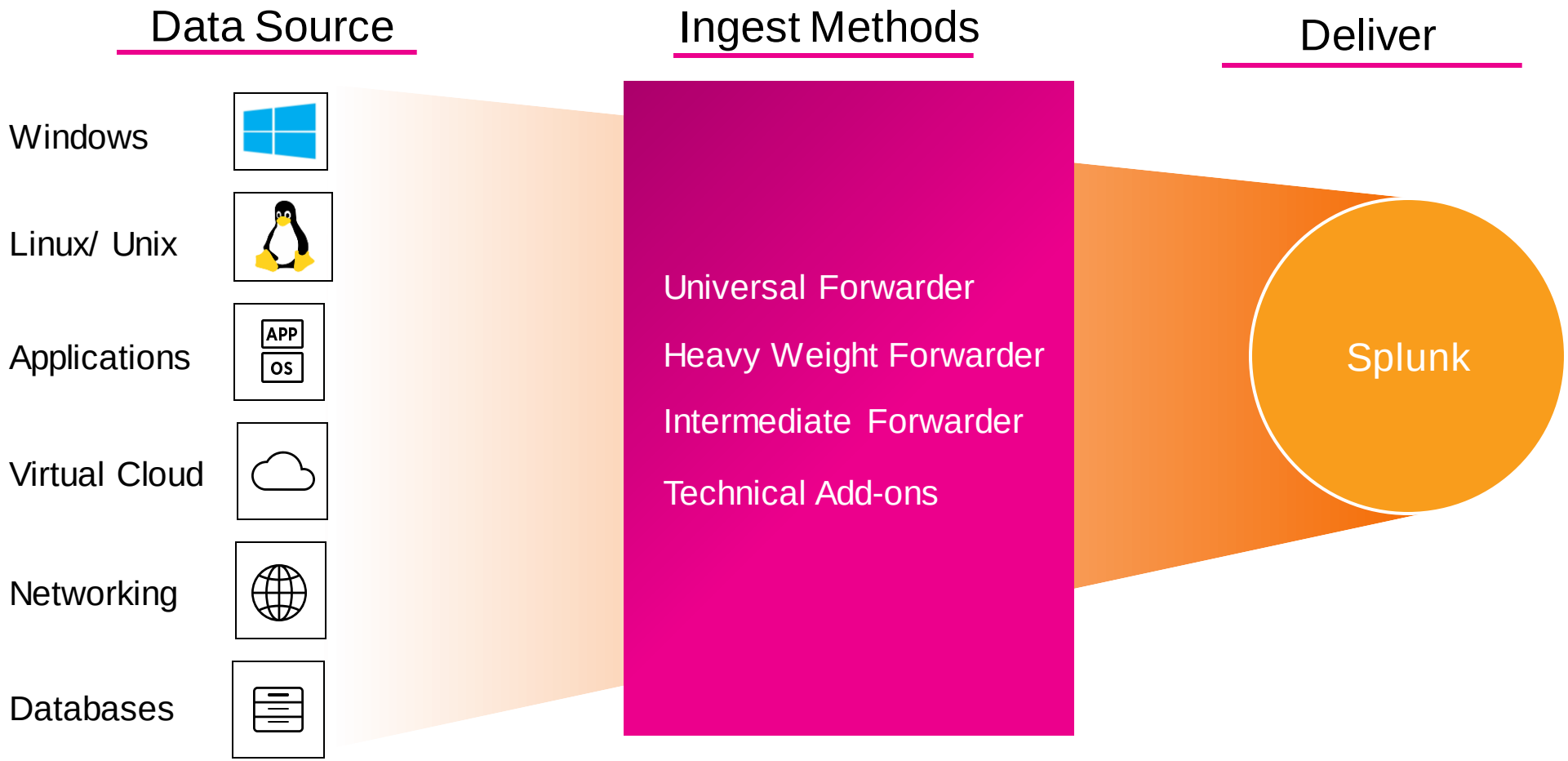
Scripts or modular inputs



Get data from any API, or database with a script

Get Data into Splunk

Current Tool Kit



Get data into Splunk

DSP Ingest Methods

splunk> .conf19

Get Data into Splunk

Data Stream Processor – Ingest Methods

- Upload a file
- Ingest API
- Forwarder Service
- Message bus

Use-case

Persona : IT admin

splunk> .conf19

Jason

Splunk IT admin

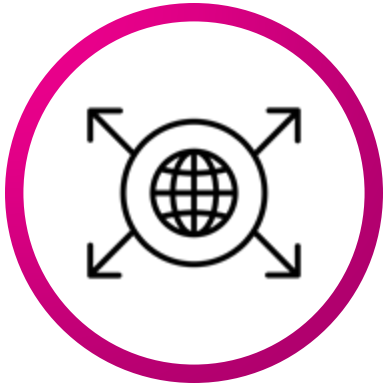
Az company

I want to Monitor Infrastructure Performance

- Deployed across multiple regions
- Approx. ~50 different services
- Collect metrics for individual services
- Amazon CloudWatch Metrics
- Metrics pulled over very recent time windows, say the last 30 or 60 min

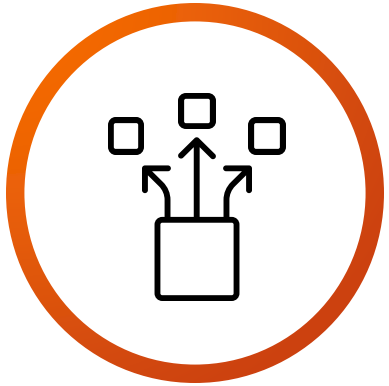
3 Key Infrastructure Adaptability Requirements

Scalability



- Can the system scale automatically?
- Can it scale to collect data from multiple regions?
- Can I scale up & down?
- Can I manage multiple jobs within a specific time interval?

High Availability



- Can I have maximum resiliency to meet my business continuity needs?
- Can it recover automatically without manual intervention?

Performance

- How fast can I collect the metrics?
- What throughput will I get?
- Can I collect metrics for recent time window, say the last 30 or 60 min?

Get Data into Splunk

Data Stream Processor – Ingest Methods

- Upload a file
- Ingest API
- Forwarder Service
- Message bus

+

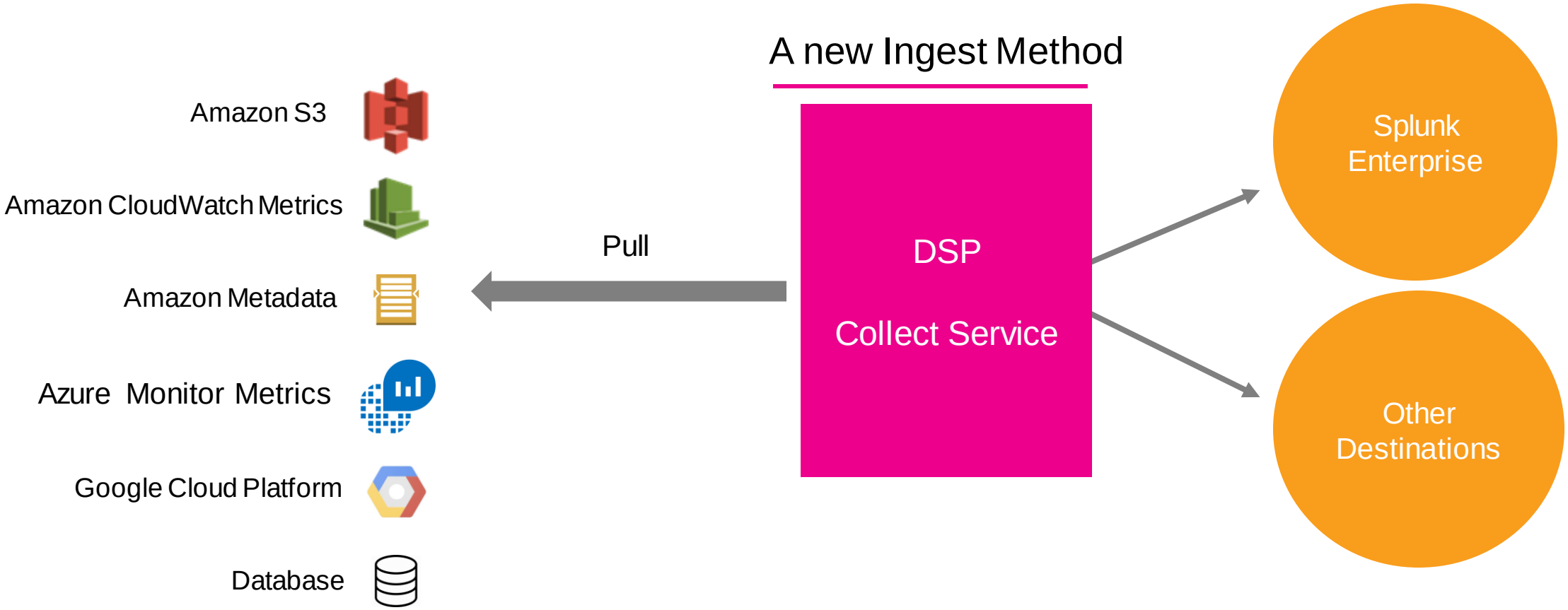
Introducing Collect Service

Collect Service

In a nutshell

- Scalable data collection service to pull large volumes of data from external data sources
- Focus on scheduled data collection at Scale and leverage DSP for data transformation
- Available in Splunk Cloud Services and Splunk Data Stream Processor (DSP) on-prem

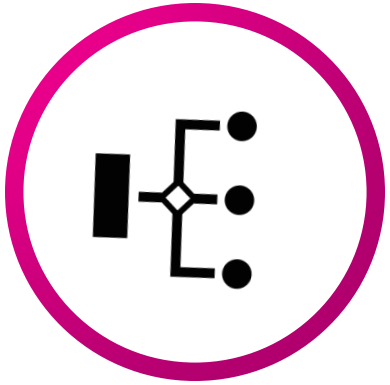
Collect Data: A New Scalable Ingest Mechanism



Key Concepts

In Collect service

Connector



Extension of the Collect Service that can collect events or metrics from a data source

Schedule a job

Defines when, what, and how to collect data

Execution

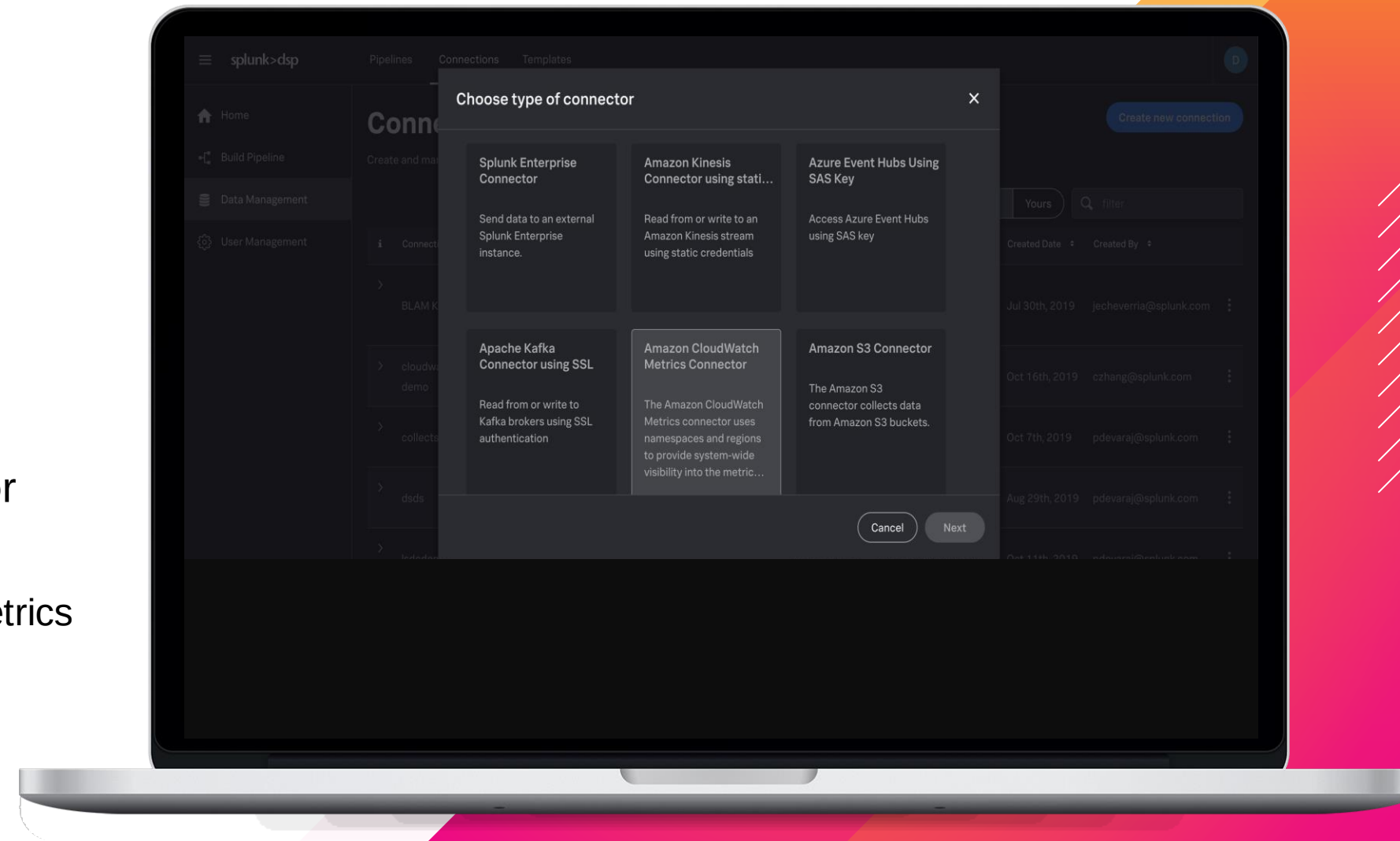
One iteration of a scheduled job

Worker

Worker provides the abstraction for underling computing resources

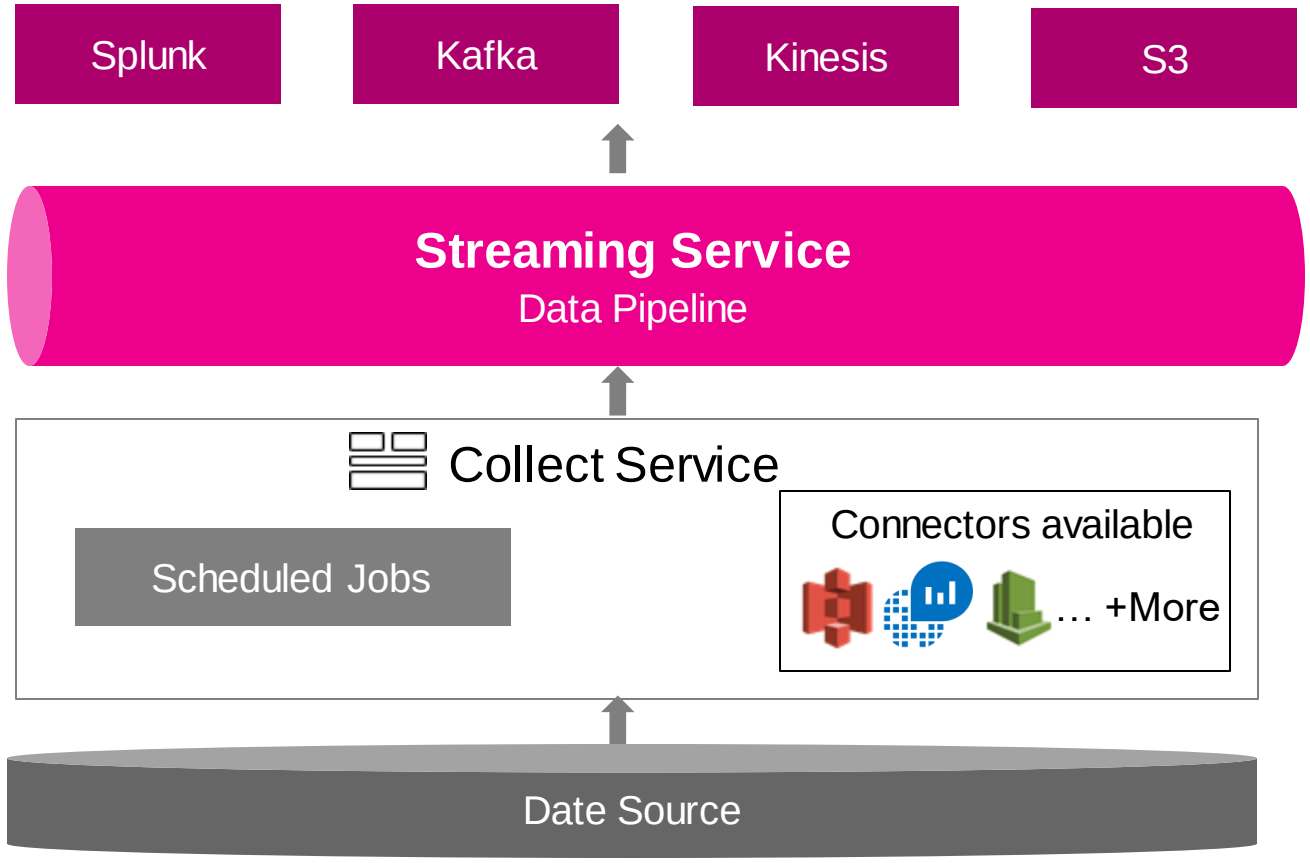
Available Collect Service Connectors today!

- Amazon S3 Connector
- Amazon CloudWatch Metrics
- Amazon Metadata
- Azure Monitor Metrics



Architecture Overview

Data collection service designed for the “pull” scenario



How it works?

How it works?

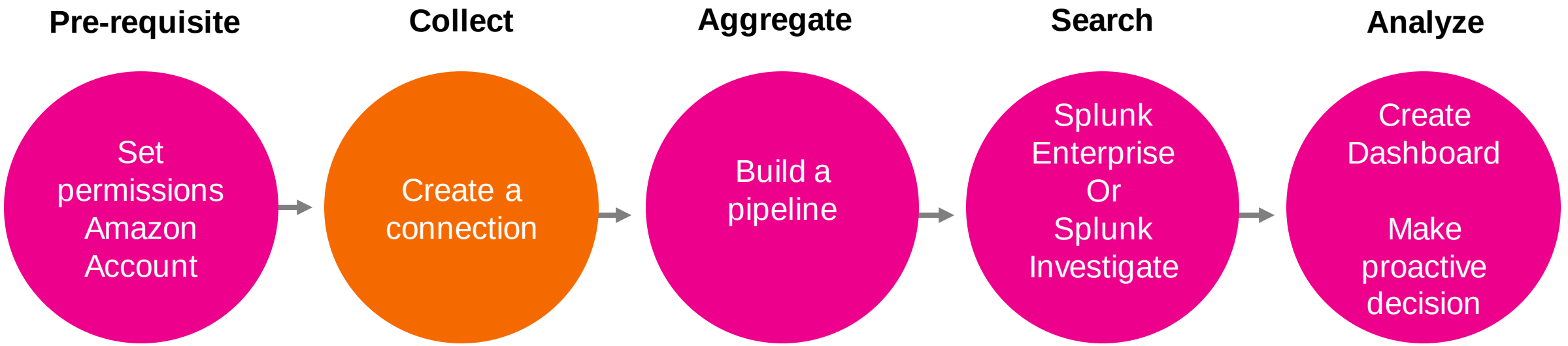
Collect Service in action

1. Connectors (Pull-Based)
   - Know how to connect to external systems
   - Scheduled to pull new data from an external source
   - Each time they run, they connect and fetch data
   - Connectors then write data to the DSP Pipeline

2. Connection
   - A configured instance of a connector
   - What kind of connector?
   - Where to connect? (i.e. which AWS account, which region ..)
   - How to authenticate? (i.e. AWS access keys)

Amazon CloudWatch Metrics Connector in DSP

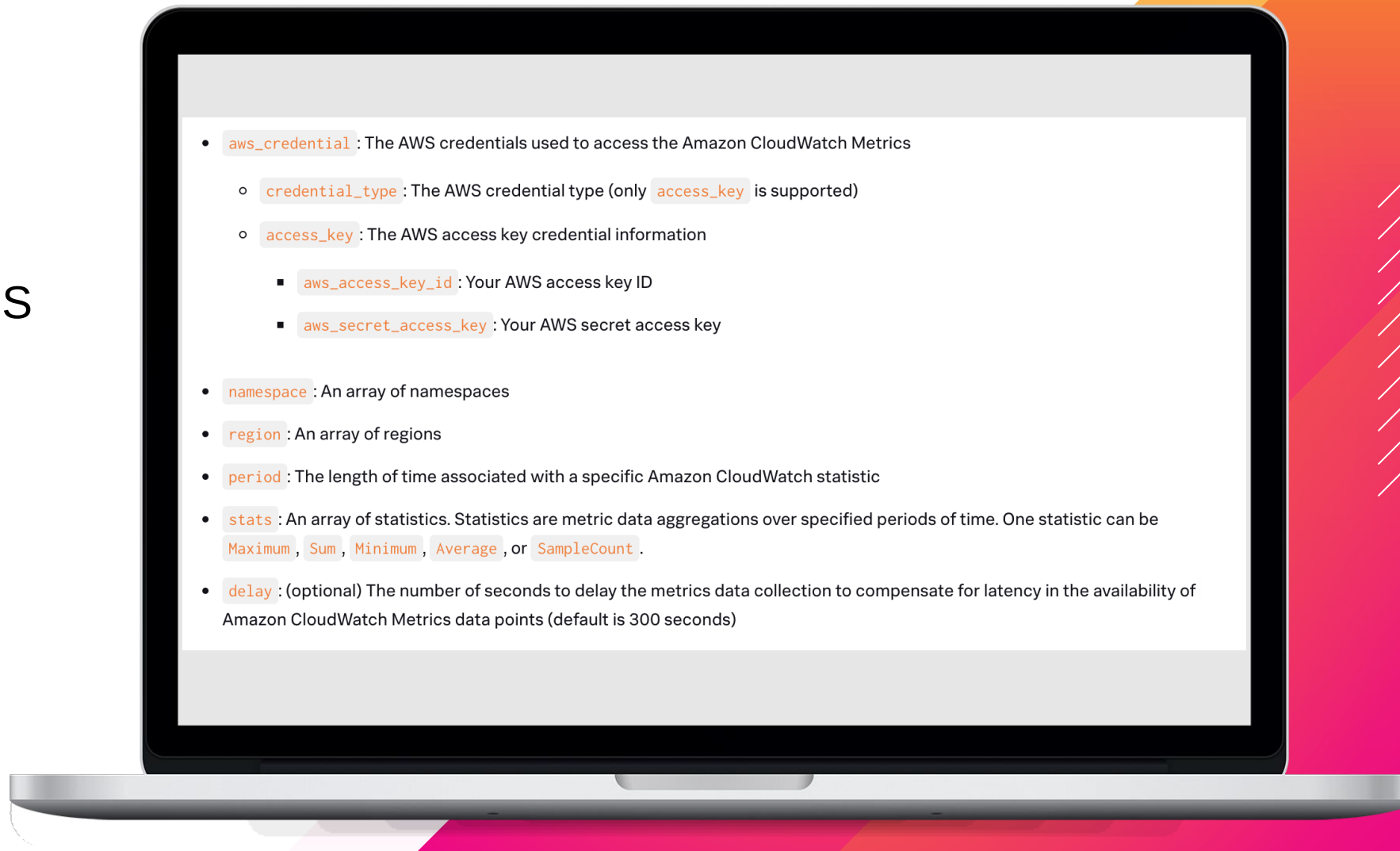
Monitor Infrastructure Performance



Pre-requisite

Pre-requisite

- Set Permissions for AWS
- Set Permissions for the AWS/EC2 namespace
- Set Parameters used in the Splunk AWS CloudWatch Metrics Connector



Create a connection

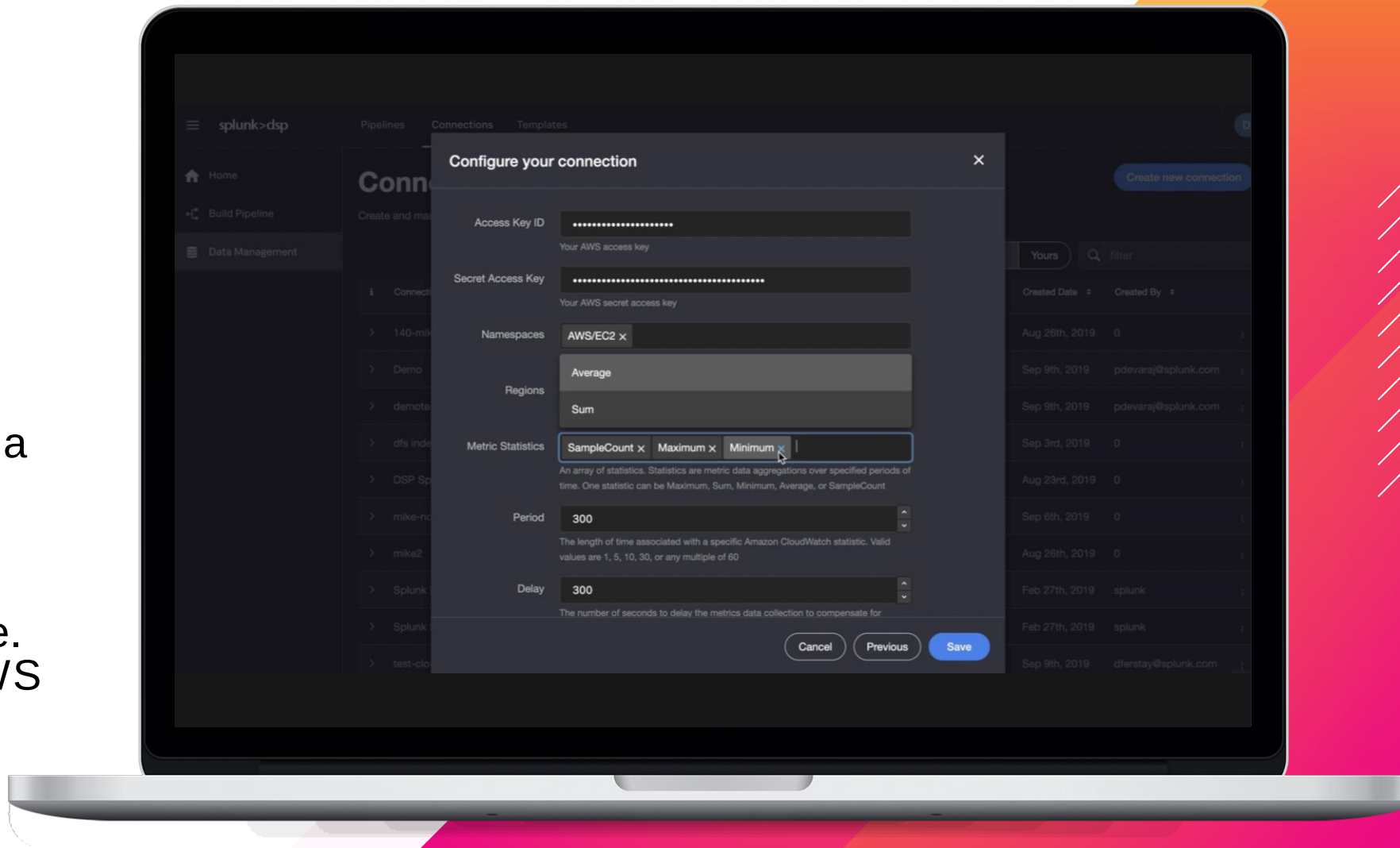
Collect Service - Connectors

splunk> .conf19

Create a connection using UI

Connection

- A configured instance of a connector
- Where to connect?
- How to authenticate? (i.e. username/password, AWS access keys, client certificates)



Create a Connection Using API

```
curl -X POST "https://api.scp.splunk.com/<tenant>/collect/v1beta1/jobs/" \
  -d '{
        "name": "your connection name",
        "connectorID": "aws-s3",
        "schedule": "45 * * * *",
        "sqs_queue_url": "http(s)://sqs(-fips).<REGION>.amazonaws.com/<YOUR_ACCOUNT_NUMBER>/<YOUR_QUEUE_NAME>",
        "aws_credential": {
            "access_key": {
                "aws_access_key_id": "your AWS access key",
                "aws_secret_access_key": "your AWS secret access key"
            }
        },
        "filetype": "json",
        "json": {
            "field": "Records"
        },
        "scalePolicy": {
            "static": {
                "workers": 2
            }
        }
    }'
```

https://sdc.splunkbeta.com/reference/api/collect/v1beta1#endpoint-createJob

Connections Management

Data Stream Processor

splunk> .conf19

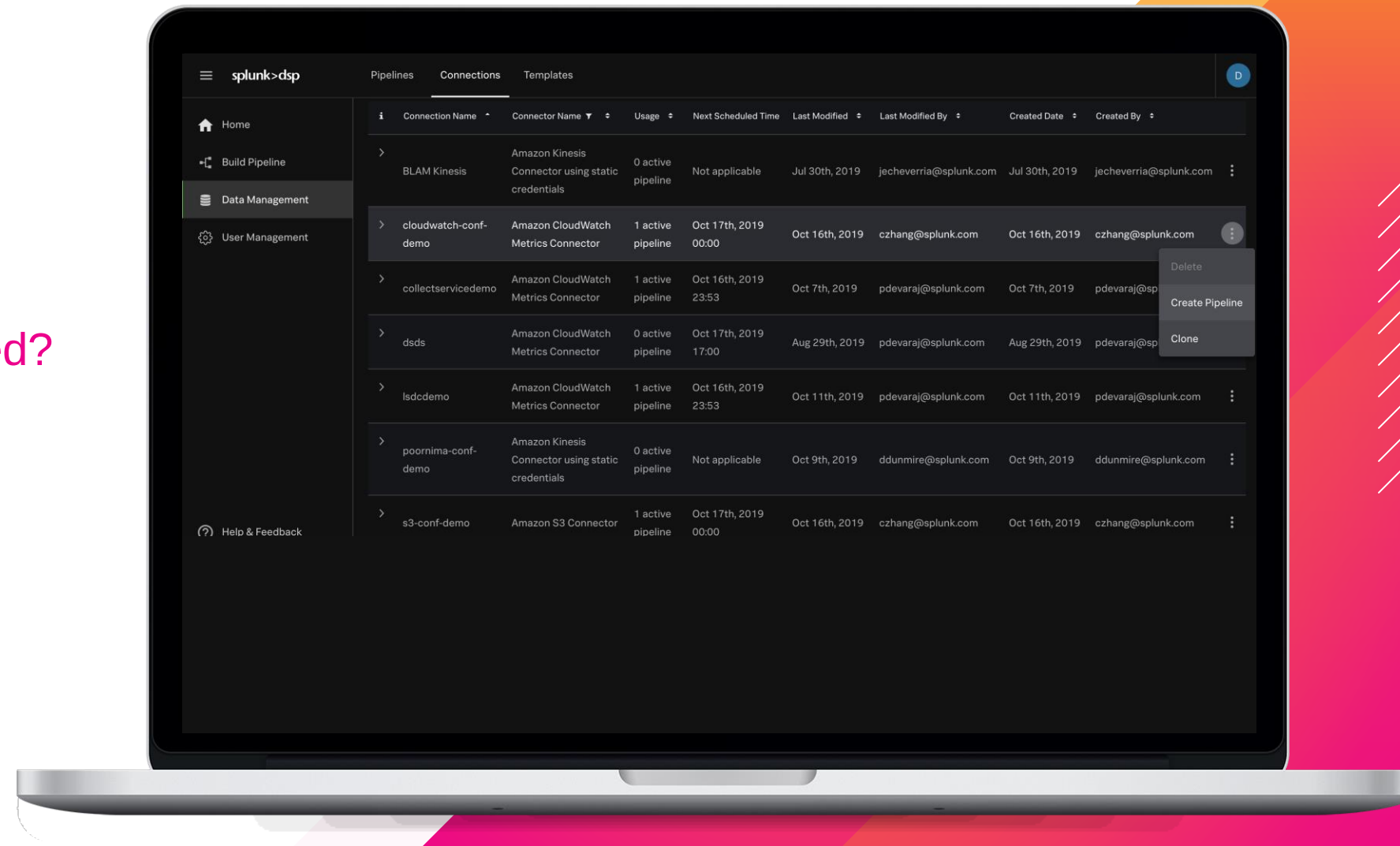
Connections Management

When is next job scheduled?

Create a pipeline

Actions

- Delete
- Edit
- Clone



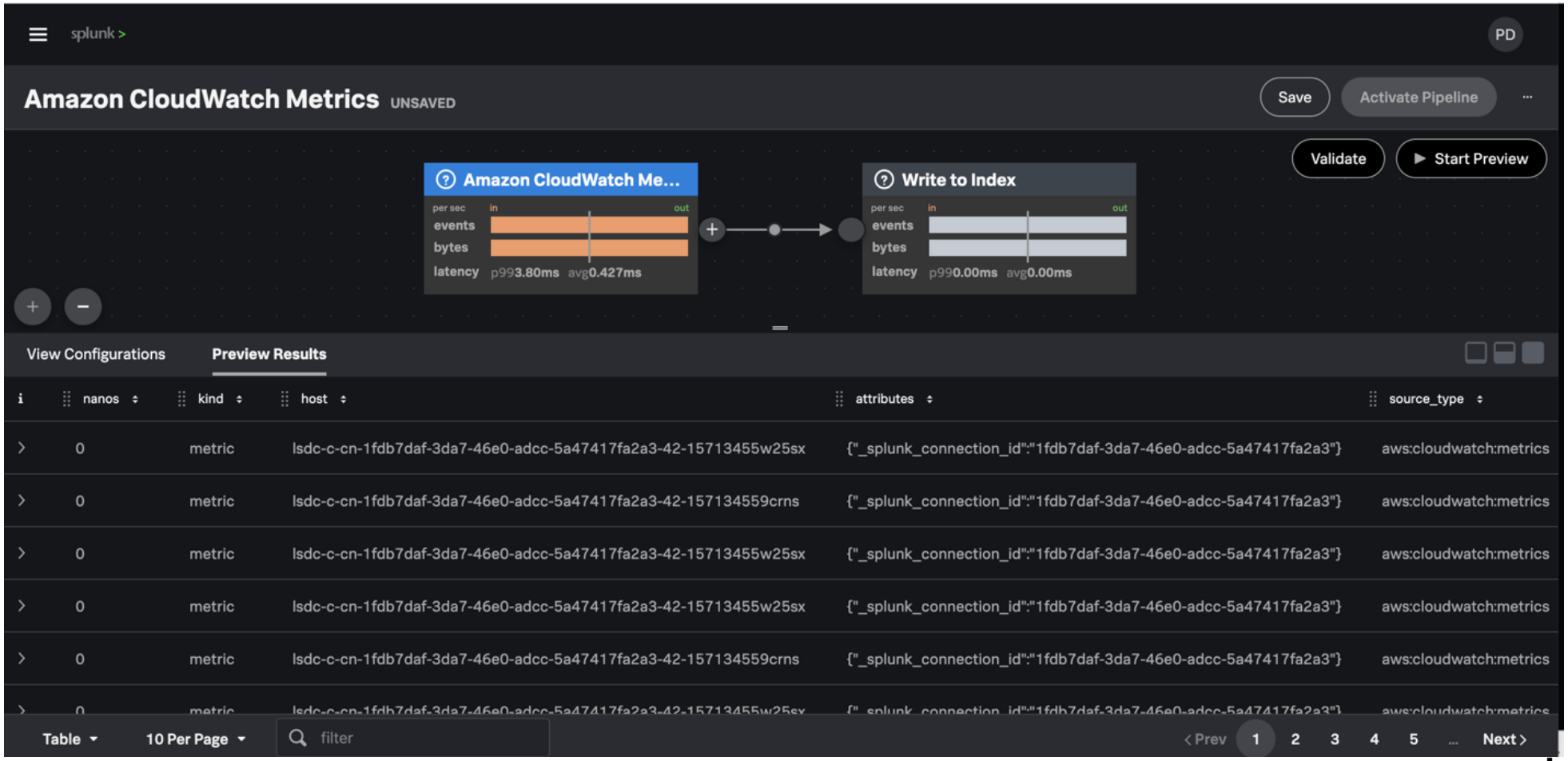
Build a pipeline

DSP

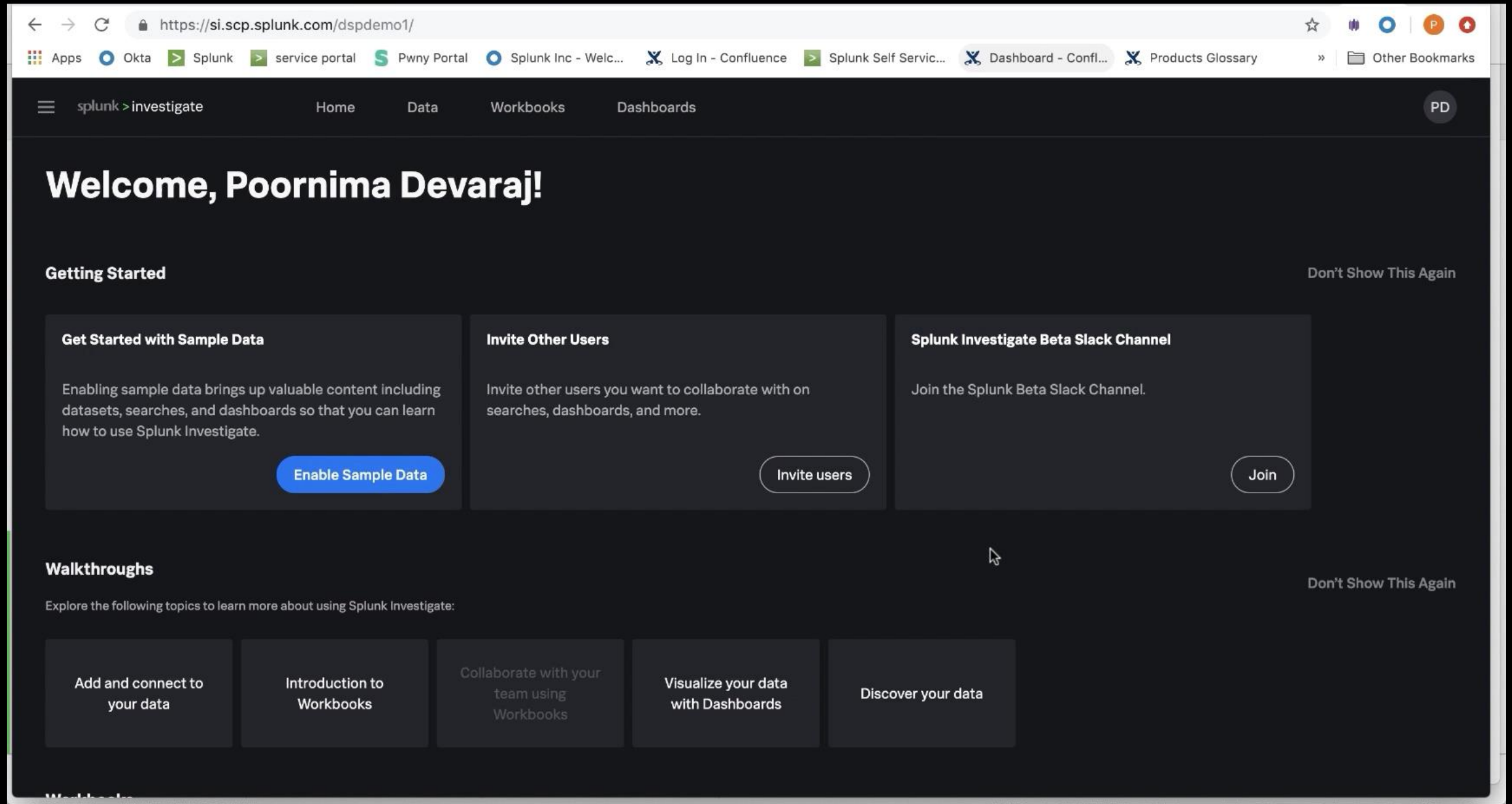
splunk> .conf19

Create a Pipeline > Preview Results

DSP > Build Pipeline > Preview results

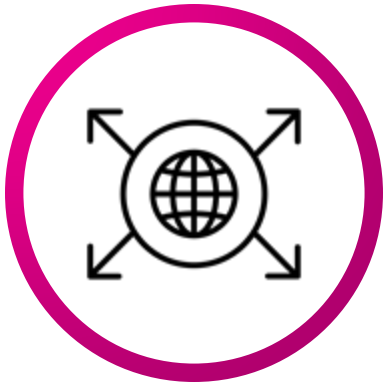


Demo


https://si.scp.splunk.com/dspdemo1/
Apps
Okta
Splunk
service portal
Pwny Portal
Splunk Inc - Welc...
Log In - Confluence
Splunk Self Servic...
Dashboard - Confl...
Products Glossary
Other Bookmarks
splunk > investigate
Home
Data
Workbooks
Dashboards
PD
Welcome, Poornima Devaraj!
Getting Started
Don't Show This Again
Get Started with Sample Data
Enabling sample data brings up valuable content including datasets, searches, and dashboards so that you can learn how to use Splunk Investigate.
Enable Sample Data
Invite Other Users
Invite other users you want to collaborate with on searches, dashboards, and more.
Invite users
Splunk Investigate Beta Slack Channel
Join the Splunk Beta Slack Channel.
Join
Walkthroughs
Don't Show This Again
Explore the following topics to learn more about using Splunk Investigate:
Add and connect to your data
Introduction to Workbooks
Collaborate with your team using Workbooks
Visualize your data with Dashboards
Discover your data

How Collect Service Meets Adaptability Requirements

Scalability

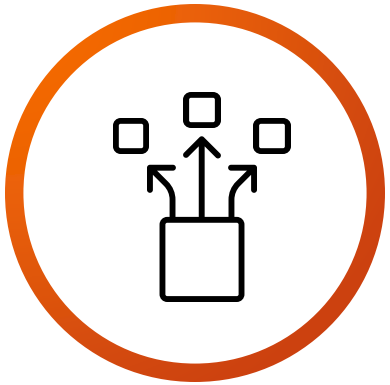


Kubernetes

Scale-out Architecture

- Workers to scale
- Load Balance
- Resource pool for multiple jobs

High Availability



Multiple replicas

Kubernetes

- Recovery from failure
- Isolation of failures/ errors

Avoid single point of failure

Centrally manage the checkpointing for data discovery and collection

Performance

Map Reduce Model

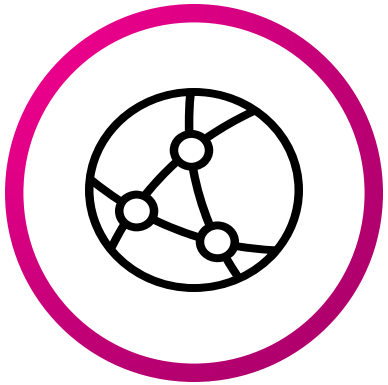
- Collect data in-parallel

Load Balance

Golang-based data collection engine and connectors

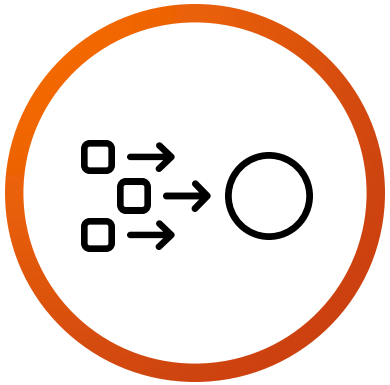
More Value from Collect Service

Resource Optimization



- Pull data on-demand
- Shared computing resources across jobs/executions

Extensibility

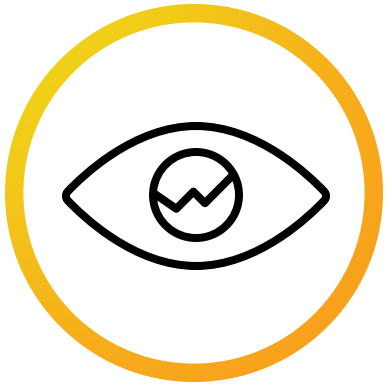


- Extend by implementing connectors in Go/Java/Python or any other preferred languages/framework
- Same connector works on cloud and on pre
- Improved development experience to test/debug connectors without DSP

Registry service not available

- Considering options
- Let's talk if you want to build a new connector

Visibility and Management



- Collect service centrally manages all inputs
- Admin know where the data is coming from and the throughput/error/warning
- Connections Management

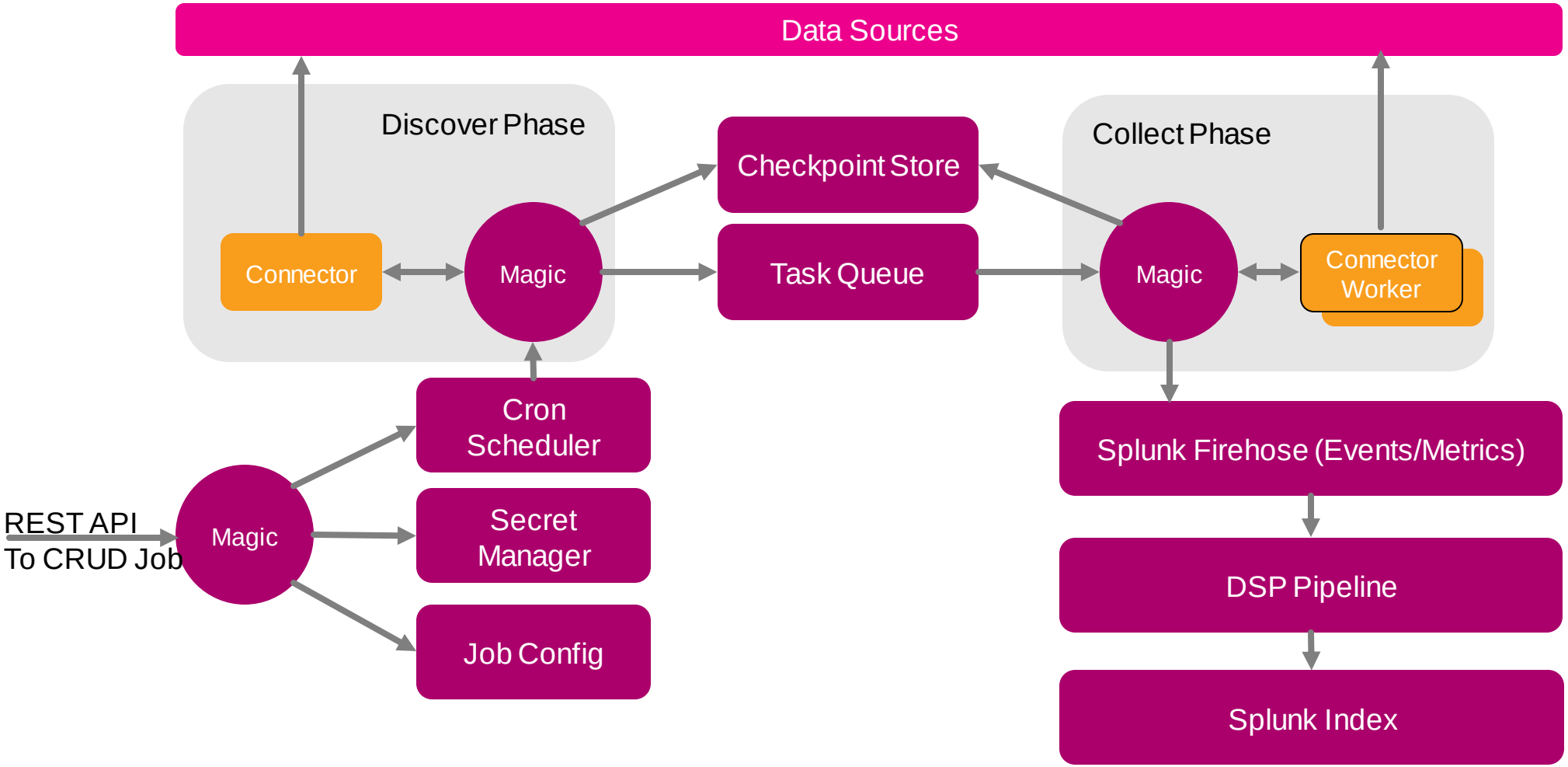
Sneak Preview: How to Build New Connector

- Any chance to reuse current pull-based or streaming connectors?
- Does your use case fit the Collect Service?

- Understand the connector workflow
- Build it with Go/Python/..
- Test/debug locally

- Register the new connector in DSP
- Create data collection job via REST or web UI

Connector Workflow



Next Steps

Sign up today! for Splunk Investigate Free Trial!

Pull data into DSP using connectors available

DSP is also available On-prem

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

Let's get connected.

- Need more connectors
- Build your own connector on collect service

Resources & Contacts

Where can I find more information?

Developer Guide:
https://dev.splunk.com/scs/docs/add/collect
https://dev.splunk.com/scs/docs/add/collect/cloudwatchmetrics
https://dev.splunk.com/scs/docs/add/collect/s3
https://dev.splunk.com/scs/docs/add/collect/awsmetadata
https://dev.splunk.com/scs/docs/add/collect/azuremonitormetrics

Information on DSP: https://docs.splunk.com/Documentation/DSP

Poornima Devaraj, Technical Product Manager (pdevaraj@splunk.com)
Thor Taylor, Director Product Management (ttaylor@splunk.com)
Dan McBreen, AVP New Product Sales (dmcbreen@splunk.com)

Q&A

Poornima Devaraj | Product Manager
pdevaraj@splunk.com

Jove Zhong | Engineering Director
jzhong@splunk.com

splunk> .conf19

.conf19

splunk>

Thank You!

Go to the .conf19 mobile app to

RATE THIS SESSION