StreamSets

BENEFITS

Automated pipeline deployment.

Shared pipeline repository for team collaboration.

Deployment via Kubernetes.

Integration with leading governance tools.

StreamSets Control Hub™

Collaborative development, automated deployment and integrated governance of dataflow pipelines.

When it comes to data movement, size matters. Building pipelines to simply forward data to a limited number of systems is straightforward. However, as pipelines grow in number, complexity and importance, and then become deployed across a wide sweeping number of systems or endpoints, keeping track of which pipelines are deployed where, when they need to be upgraded, or diagnosing pipeline failures can require a lot of time and effort.

Also, as groups of developers build individual pipelines, the risk of duplication becomes an issue. Often times teams are building similar dataflow logic, resulting in redundant development cycles, slowing project delivery and costing you time and money. If these teams were able to share and collaborate on pipeline development, then productivity would rise, and efficiencies and use of best practices would be the standard.

Furthermore, as dataflow becomes a critical part of your architecture, so do does keeping tabs on data movement, including who is accessing it, where data originated and where it's going. Governance teams often have to rely on "data at rest" systems to infer where data has been, but they don't always have a direct view into this lineage information.
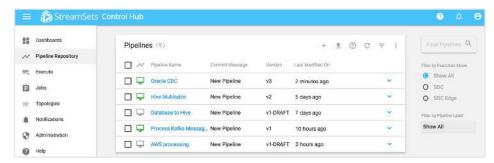
StreamSets Control Hub (SCH) gives enterprises control over the dataflow pipeline lifecycle. It allows teams to work collaboratively, deploy and scale automatically and integrate data movement into broader governance initiatives.

As part of the StreamSets Data Operations Platform, SCH is a central point of control for all your dataflow pipelines. A shared repository allows groups of teams to publish, subscribe and collaborate on pipeline development. Full lifecycle management allows you to track the version history of pipelines, giving you full control over how the development process evolves.

A cloud-based dataflow designer simplifies building and deploying pipelines to diverse systems and end points in the cloud, on premises or at the edge (using StreamSets Data Collector Edge). SCH offers full automation and provisioning capabilities regardless of system location, so it's easy to deploy and maintain complex dataflow architectures at scale.

For automated deployment, SCH provides support for Kubernetes across private and public clouds. For teams looking to leverage containers as part of their deployment, dataflow pipelines built in StreamSets can be deployed and elastically scaled from a central location.

Finally, SCH integrates directly with leading governance tools, namely Cloudera Navigator™ and Apache Atlas™ so teams can provide data movement metadata in support of data governance initiatives.

StreamSets Control Hub pipeline repository helps development teams collaborate on ongoing development

StreamSets

FEATURES

Cloud based design tool & shared pipeline repository

- Build robust dataflow pipelines via an easy-to-use cloud service.
- Test pipelines wherever execution happens — on cloud, on edge or on premises.
- Controlled publish and version management including rollback.
- Tag pipelines to dataflow topologies for end-to-end live visibility.

Architecture wide visibility and control

- End-to-end topology view.
- Visualize multiple pipelines in an overarching architecture.
- Integrates with StreamSets Dataflow Performance Manager™ for live dataflow metrics and SLA enforcement.

Automated deployment and provisioning

- Automatically deploy pipelines created in StreamSets Data Collector (SDC) or SDC Edge.
- Leverage Kubernetes for elastic scaling across private and popular public clouds.

Data governance support

- Flow metadata throughout dataflow pipelines.
- Built-in processors to store metadata at any point in a dataflow.
- Metadata pushdown integration with Cloudera Navigator™ and Apache Atlas™.
- Metadata exposed for use in other 3rd-party systems.

The StreamSets Data Operations Platform is designed to simplify the entire dataflow lifecycle, including how to build, execute and operate enterprise dataflows at scale. Developers can design batch and streaming pipelines with a minimum of code, while operators can aggregate dataflows into topologies for centralized provisioning and performance management.

ABOUT

StreamSets is headquartered in San Francisco. Our mission is to help enterprises harness their data in motion. StreamSets software is in use at hundreds of organizations and we're backed by top-tier Silicon Valley venture capital firms, including Accel Partners, Battery Ventures, Ignition Partners, and New Enterprise Associates.

LEARN MORE

Get up and running with StreamSets in minutes. Visit us at:

www.streamsets.com