ORCHESTRATE the Real-Time Flow of Data Across Data Pipelines

October 5, 2022

Scott Davis

Global Vice President

Ravi Murugesan

Sr. Solution Engineer

stonebranch

© Stonebranch 2022. All rights reserved.

Agenda

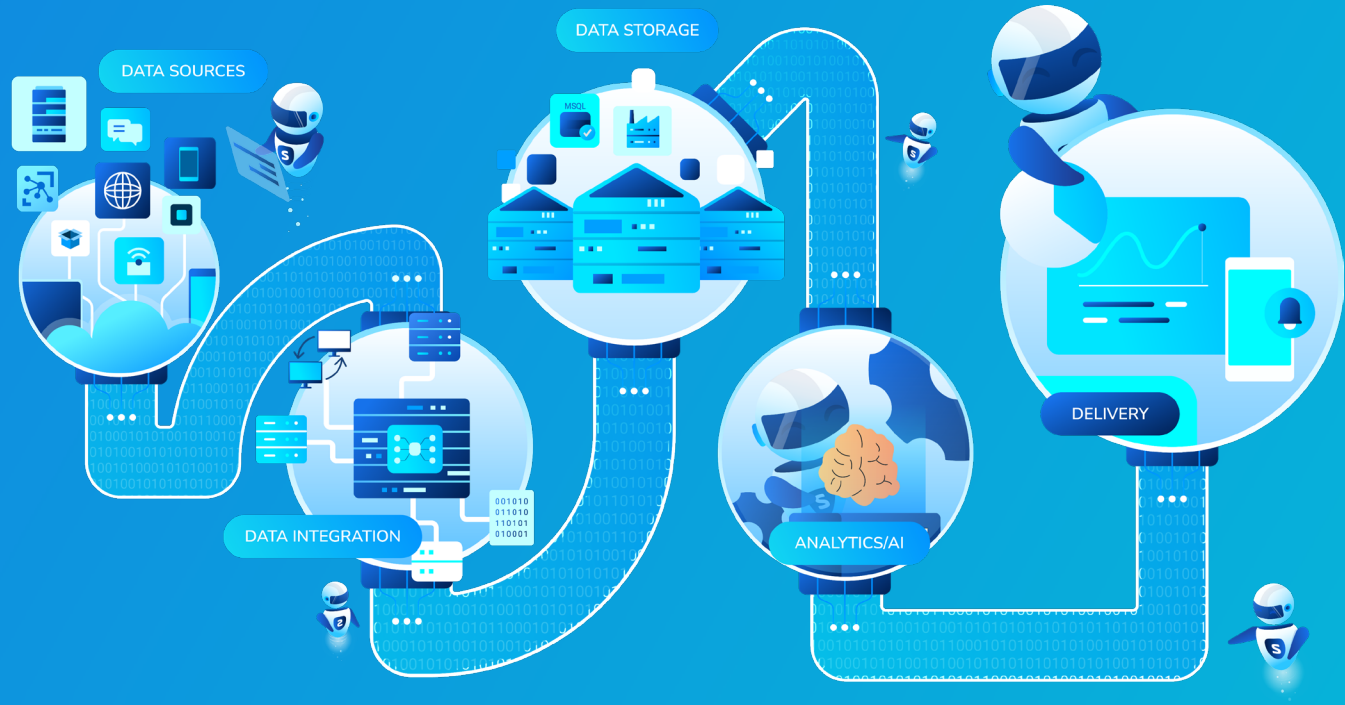
01 DevOps Orchestration Layer

02 What is a Data Pipeline

03 How to Orchestrate a Data Pipeline

04 Data Pipeline Orchestration Demo

05 Questions and Answers



stonebranch

© Stonebranch 2022. All rights reserved.

About Data Pipelines

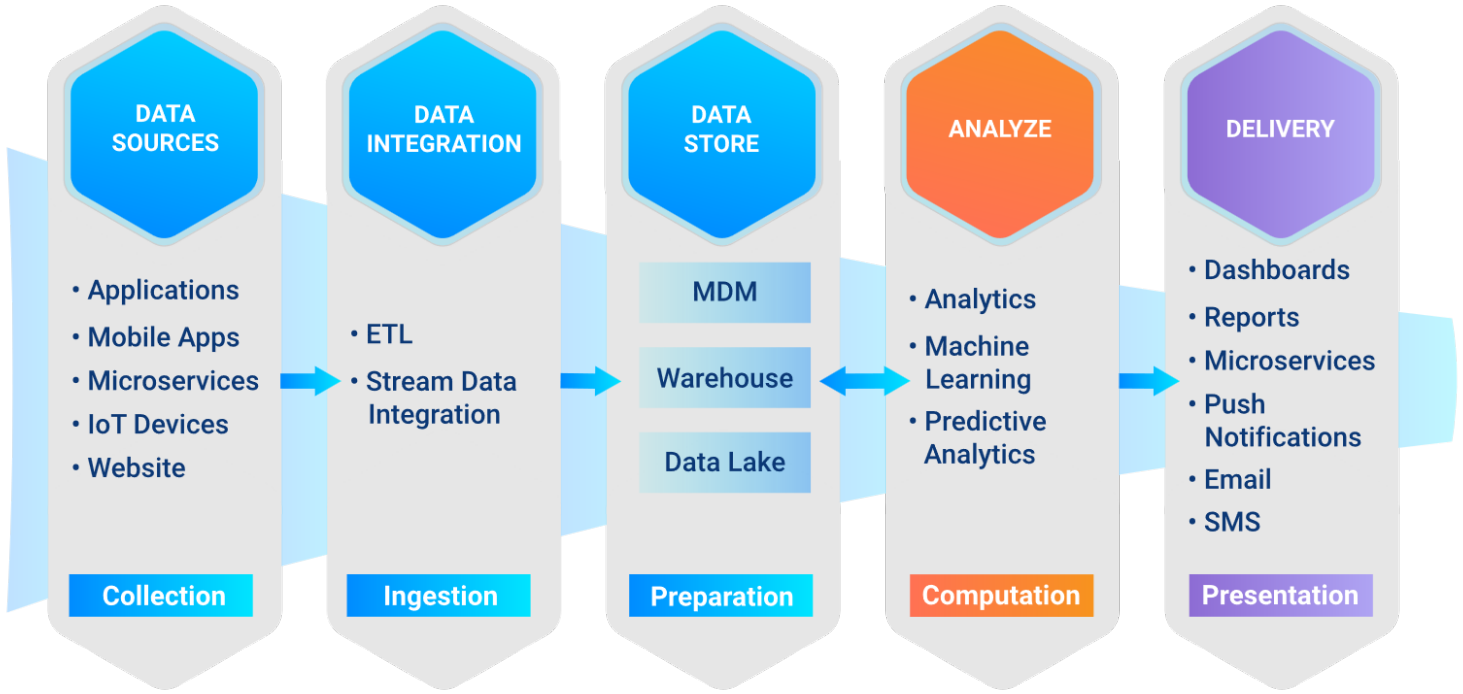
Scott Davis

Global Vice President

stonebranch

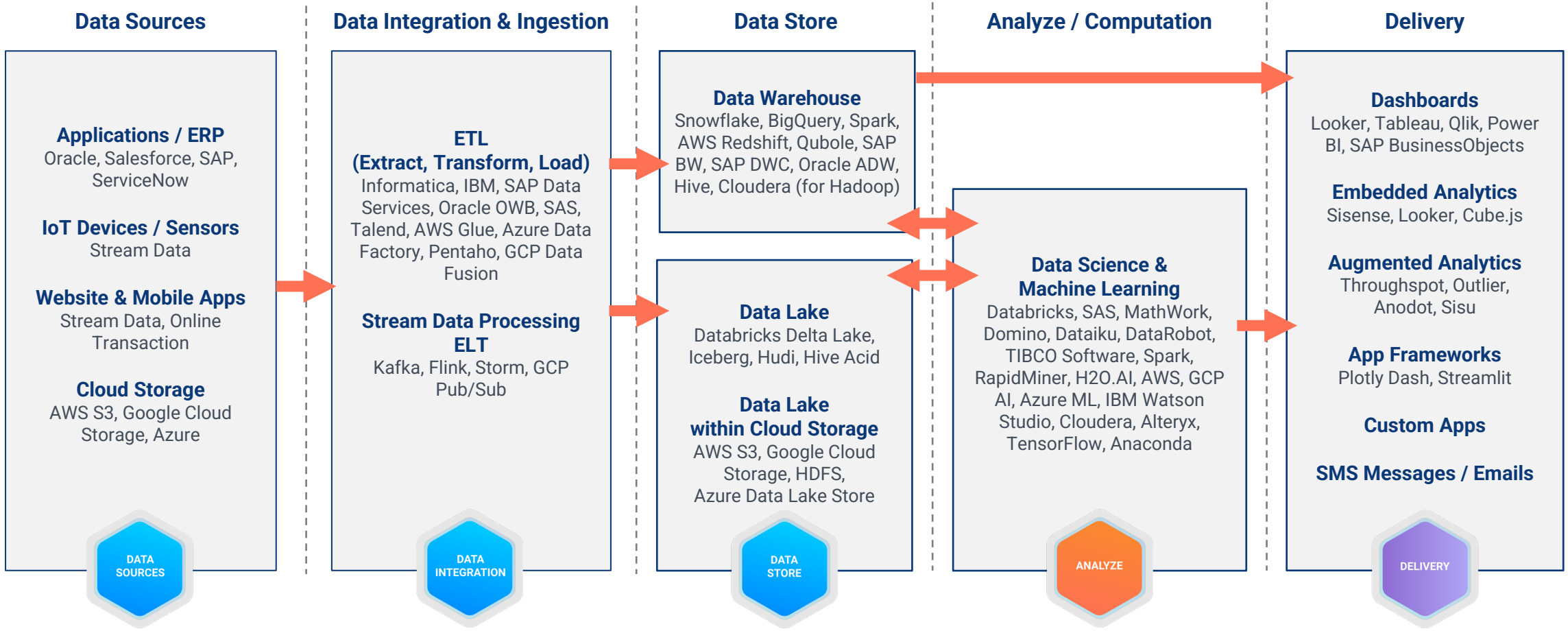
© Stonebranch 2022. All rights reserved.

Data Pipeline: Simple View



© Stonebranch 2022. All rights reserved.

Software & Tools By Stage



stonebranch

© Stonebranch 2022. All rights reserved.

The Struggle is Real

You're Not Alone

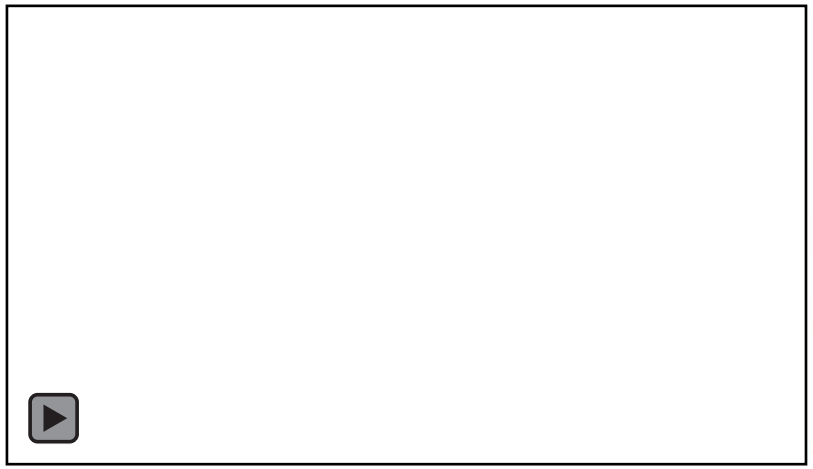
r/dataengineering

Posted by u/ohammou **Data Engineer** 12 days ago

179

Looks familiar?

Meme



10 Comments Award Share Save ...

reddit

Sort By: Best

Whack_a_mallard · 12 days ago

Too true. You gradually go from creating pipelines to just maintaining them.

14 Reply Give Award Share Report Save Follow

ThatGrayZ · 12 days ago

When I first started my job it was so exciting building everything from the ground up. I felt like I was driving innovation. Now, I'm just making sure jobs run successfully and rarely code as much as I used to in the beginning. Super depressing.

6 Reply Give Award Share Report Save Follow

Batspocky · 12 days ago

Oh my god are you future me trying to come back and give me a warning?

4 Reply Give Award Share Report Save Follow

© Stonebranch 2021. All rights reserved.

How Do Enterprises Orchestrate Today?

Common Ways to Connect Data Tools Within the Pipeline	Point-to-Point Integrations	Custom Scripts	Don't Connect (Manual Movement)

stonebranch

© Stonebranch 2022. All rights reserved.

How Do Enterprises Orchestrate Today?

Common Ways to Connect Data Tools Within the Pipeline	Point-to-Point Integrations	Custom Scripts	Don't Connect (Manual Movement)

Centralized View	Root-Cause Issues	Proactive Support	Achieve Scale	Benefits of Proper Orchestration Solutions

stonebranch

© Stonebranch 2022. All rights reserved.

How Do Enterprises Orchestrate Today?

Common Ways to Connect Data Tools Within the Pipeline	Point-to-Point Integrations	Custom Scripts	Don't Connect (Manual Movement)

Automation Pain Points

In-Built Schedulers

Can't schedule jobs in other tools

Open-Source Schedulers

Often batch- or time-based automation

Cloud Schedulers

Focus on their own ecosystems

Legacy On-Prem Focused Schedulers

Can't automate jobs in both on-prem and cloud systems, i.e., no hybrid IT automation

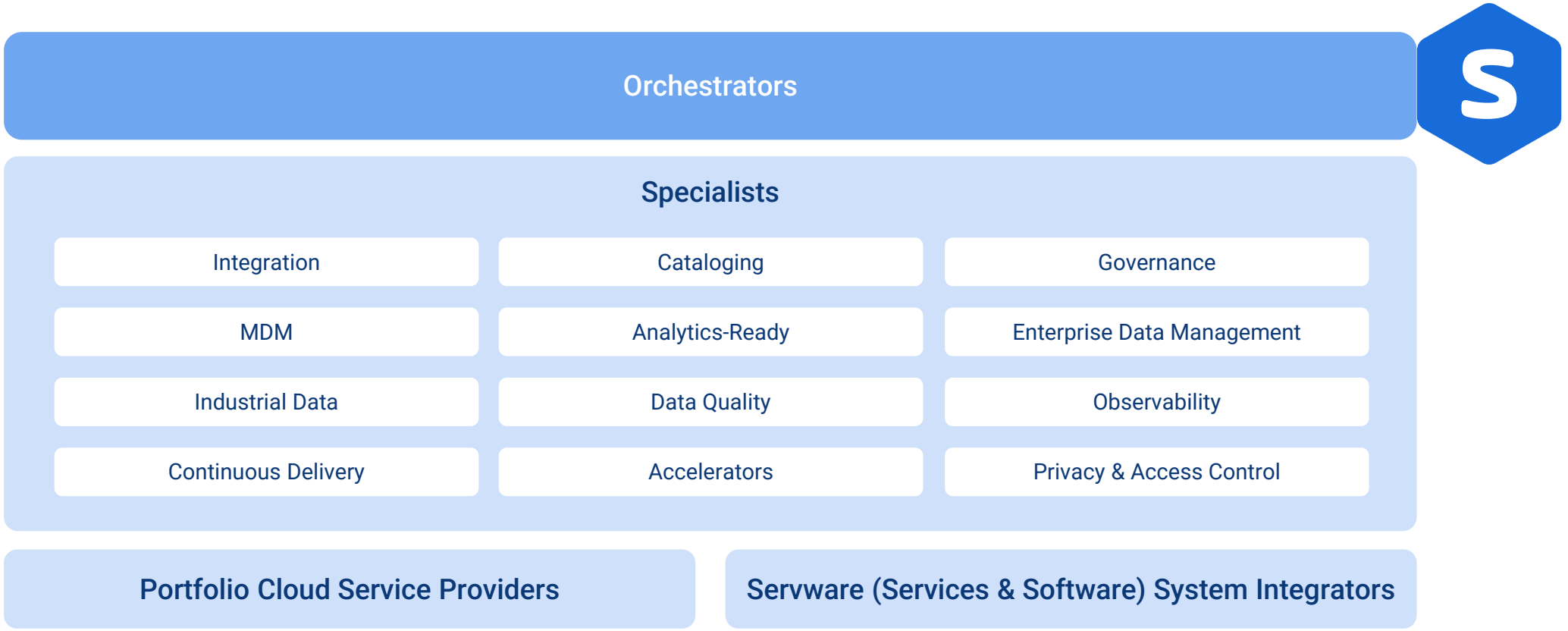
Benefits of Proper Orchestration Solutions

- Centralized View
- Root-Cause Issues
- Proactive Support
- Achieve Scale

stonebranch

© Stonebranch 2022. All rights reserved.

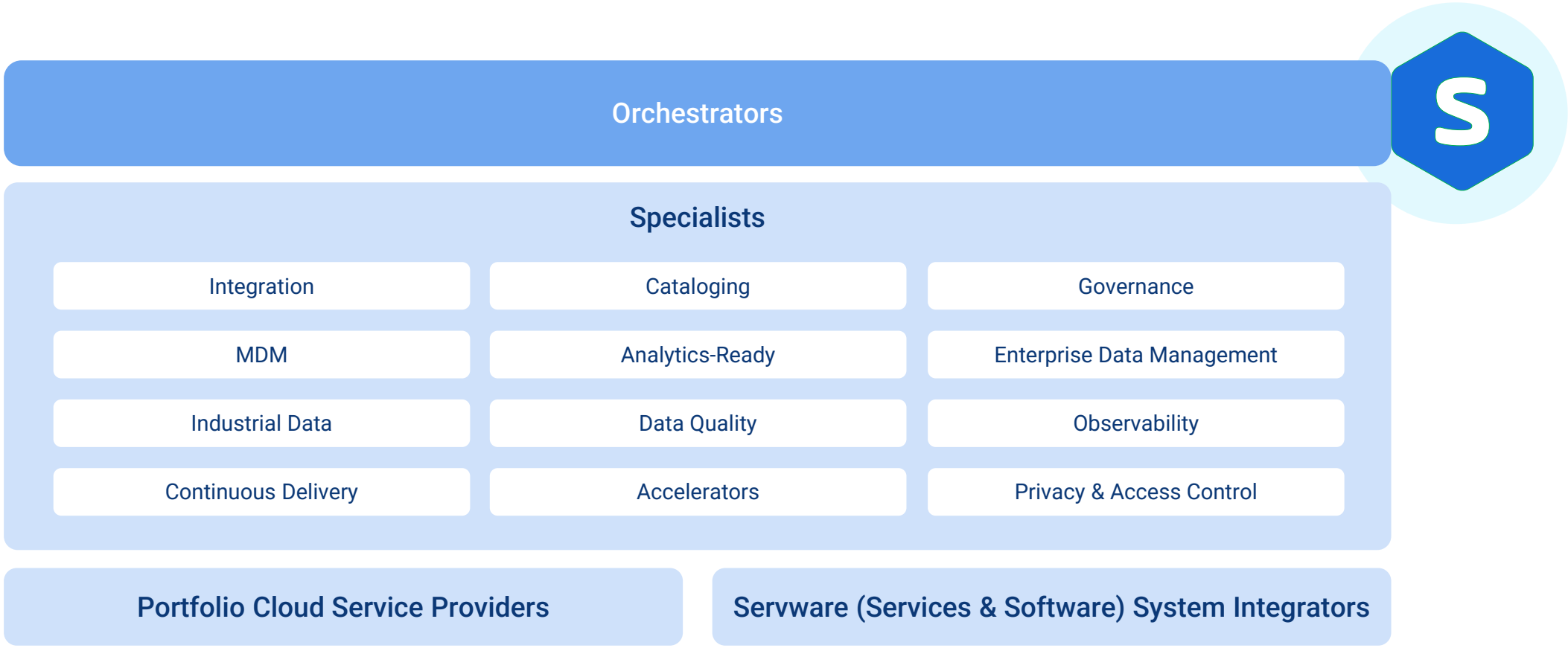
Vendor Landscape for DataOps – From Gartner



* Based on "Gartner Data and Analytics Essentials: DataOps," by Robert Thanaraj

© Stonebranch 2022. All rights reserved.

Vendor Landscape for DataOps – From Gartner



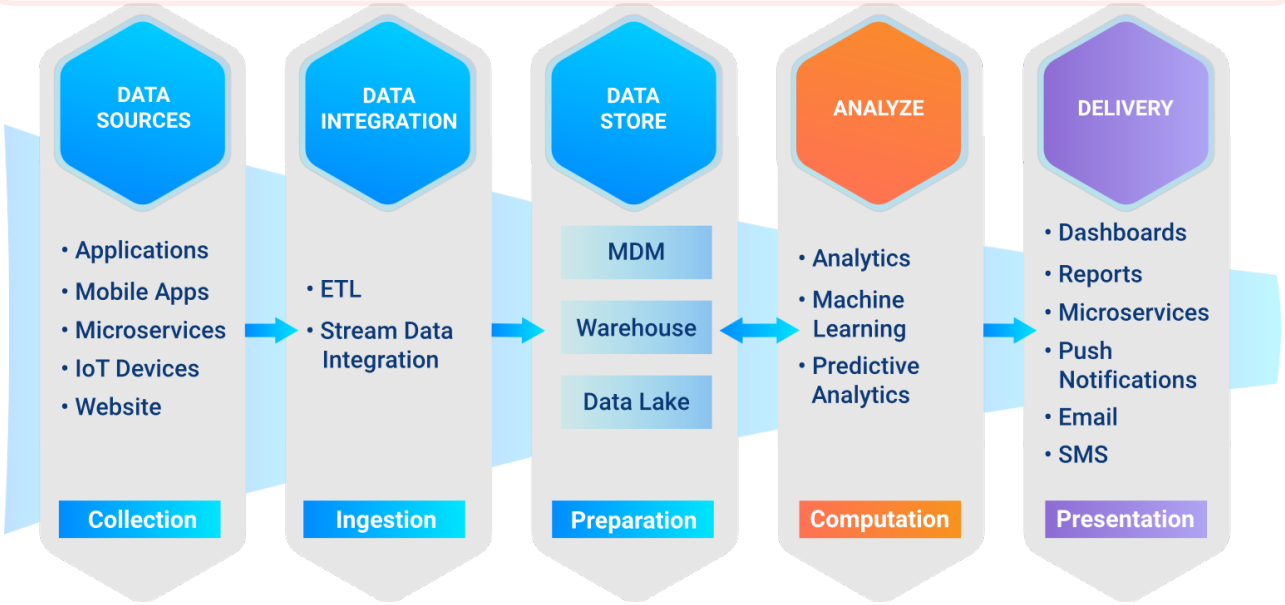
* Based on "Gartner Data and Analytics Essentials: DataOps," by Robert Thanaraj

© Stonebranch 2022. All rights reserved.

Data Pipeline Orchestration

© Stonebranch 2022. All rights reserved.

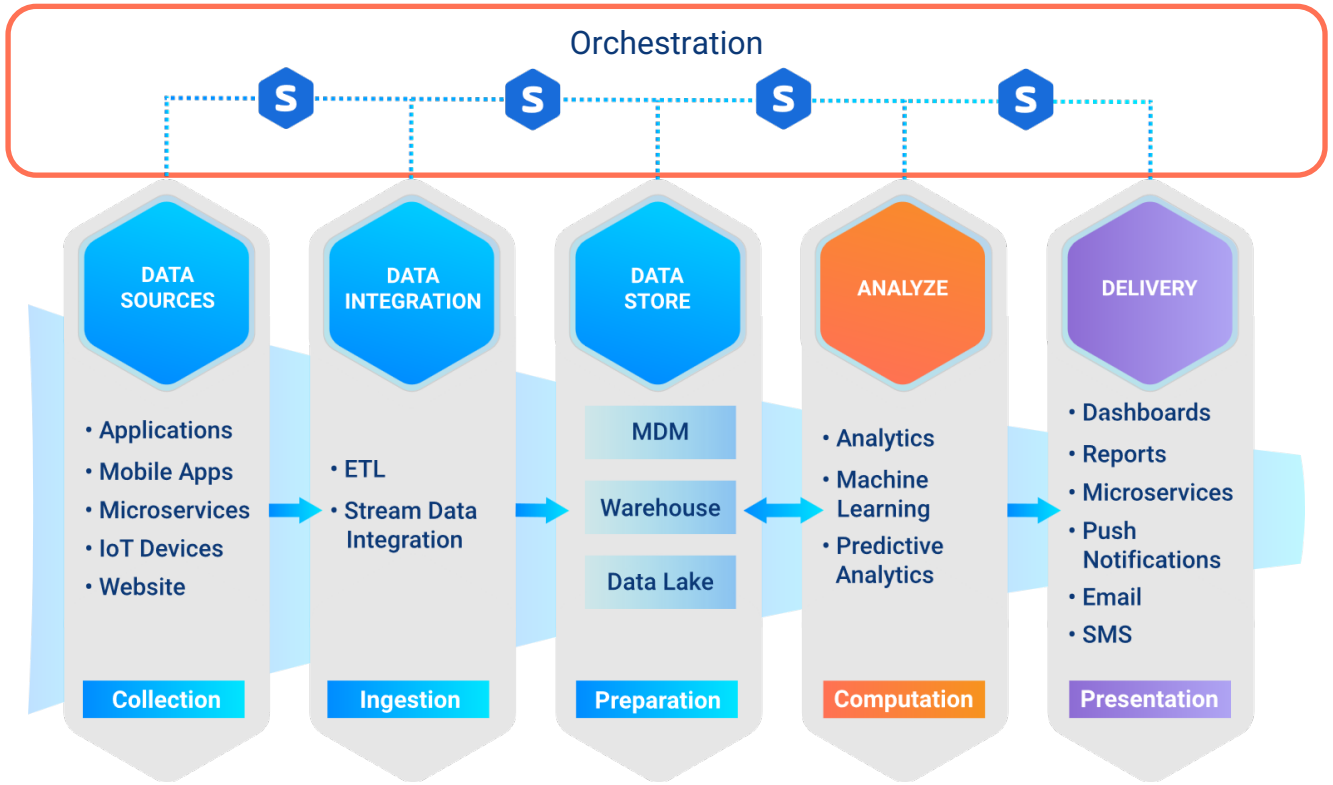
Data Pipeline Orchestration



How to accomplish the real-time automation and file transfers needed to manage the entire data pipeline.

© Stonebranch 2022. All rights reserved.

Data Pipeline Orchestration

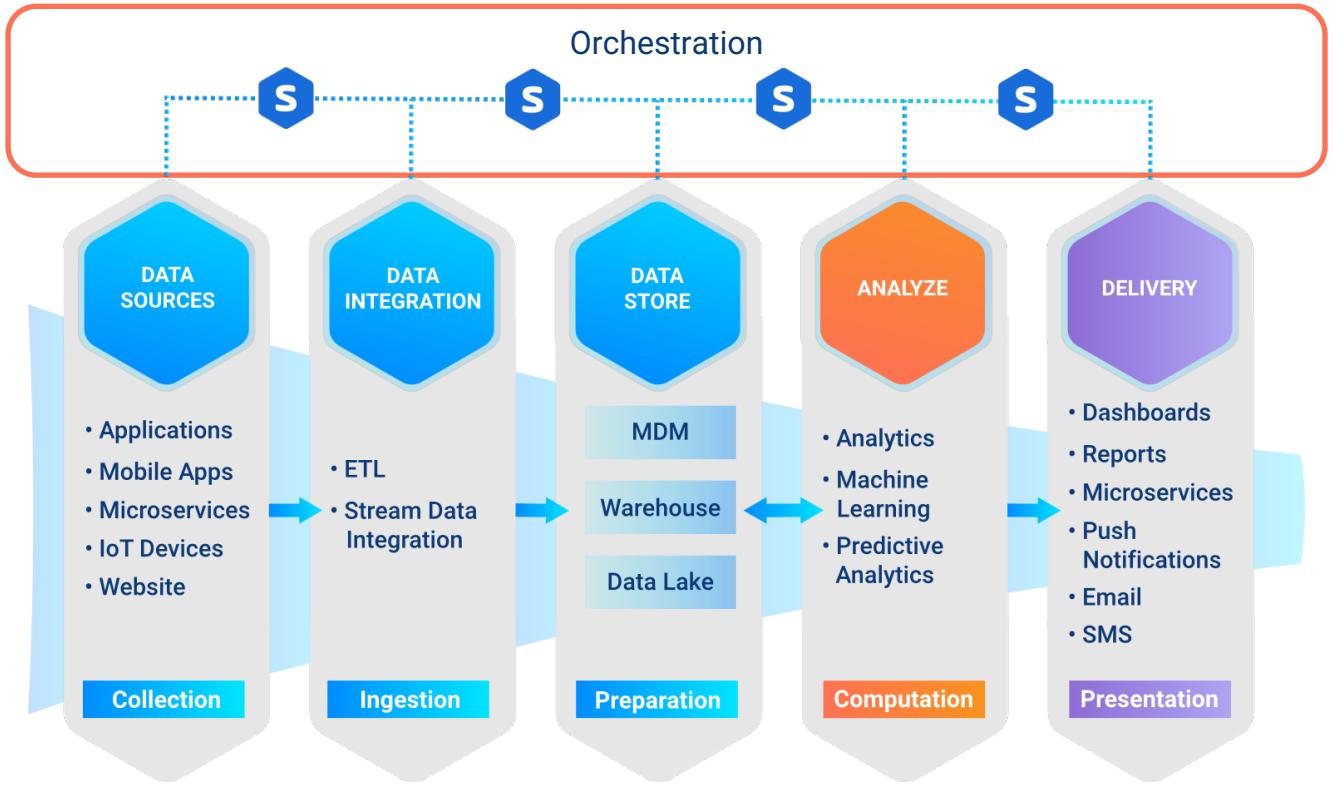


How to accomplish the real-time automation and file transfers needed to manage the entire data pipeline.

- Centrally schedule and orchestrate automated processes within each tool along the entire data pipeline
- Use APIs or Agents to control the various tools used within each stage

© Stonebranch 2022. All rights reserved.

Data Pipeline Orchestration



How to accomplish the real-time automation and file transfers needed to manage the entire data pipeline.

- Centrally schedule and orchestrate automated processes within each tool along the entire data pipeline
- Use APIs or Agents to control the various tools used within each stage

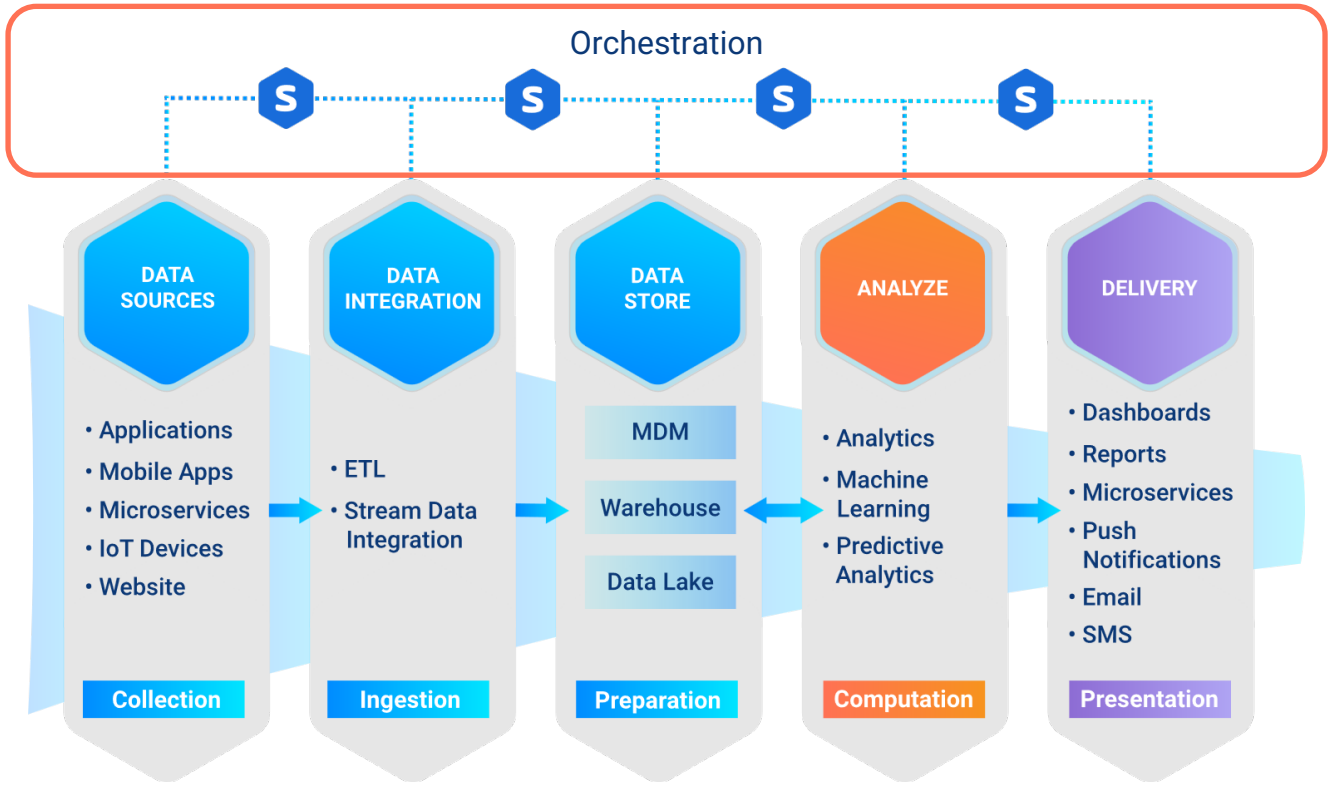
What you achieve with this approach:

- Observability of the logs and data for governance and security
- DataOps lifecycle management (Dev-Test-Prod) - including simulations
- Centralized control and visibility with visual workflows
- Quickly root-cause issues with proactive alerts when something fails

stonebranch

© Stonebranch 2022. All rights reserved.

Data Pipeline Orchestration



How to accomplish the real-time automation and file transfers needed to manage the entire data pipeline.

- Centrally schedule and orchestrate automated processes within each tool along the entire data pipeline
- Use APIs or Agents to control the various tools used within each stage

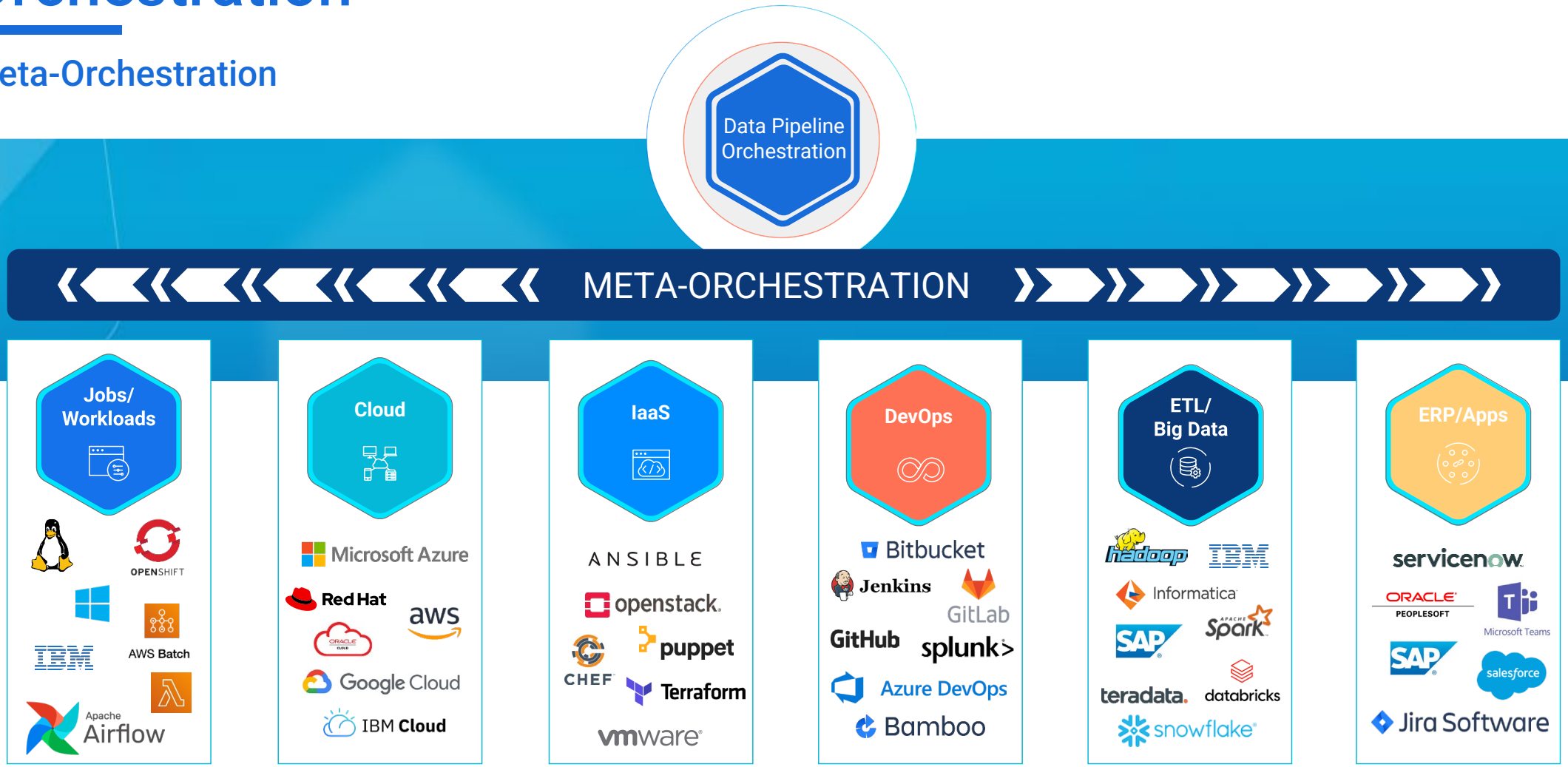
What you achieve with this approach:

- Observability of the logs and data for governance and security
- DataOps lifecycle management (Dev-Test-Prod) - including simulations
- Centralized control and visibility with visual workflows
- Quickly root-cause issues with proactive alerts when something fails

© Stonebranch 2022. All rights reserved.

Orchestration

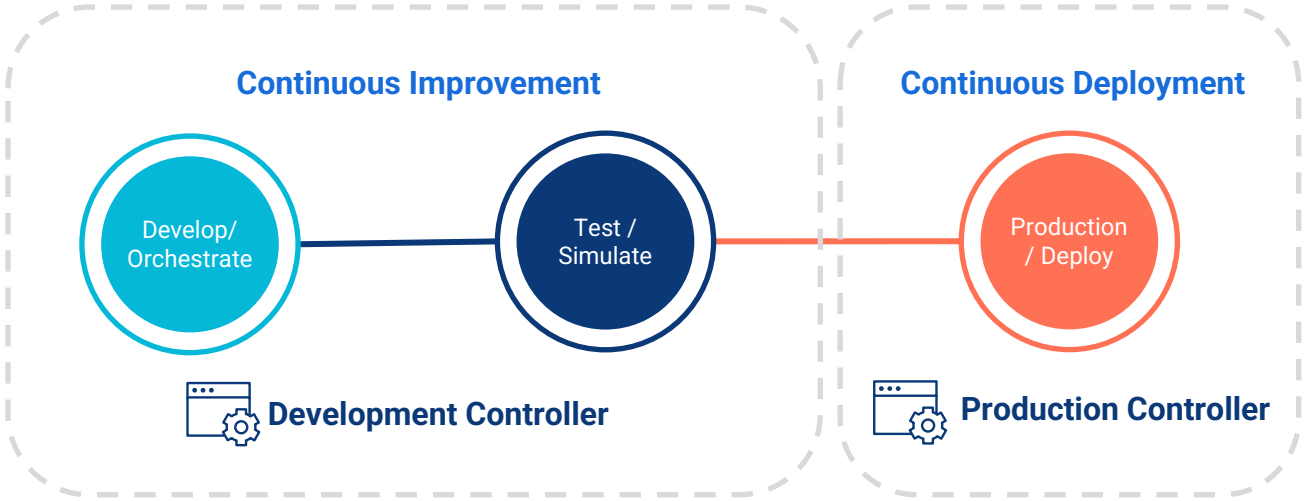
Meta-Orchestration



© Stonebranch 2022. All rights reserved.

Putting the Ops in DataOps

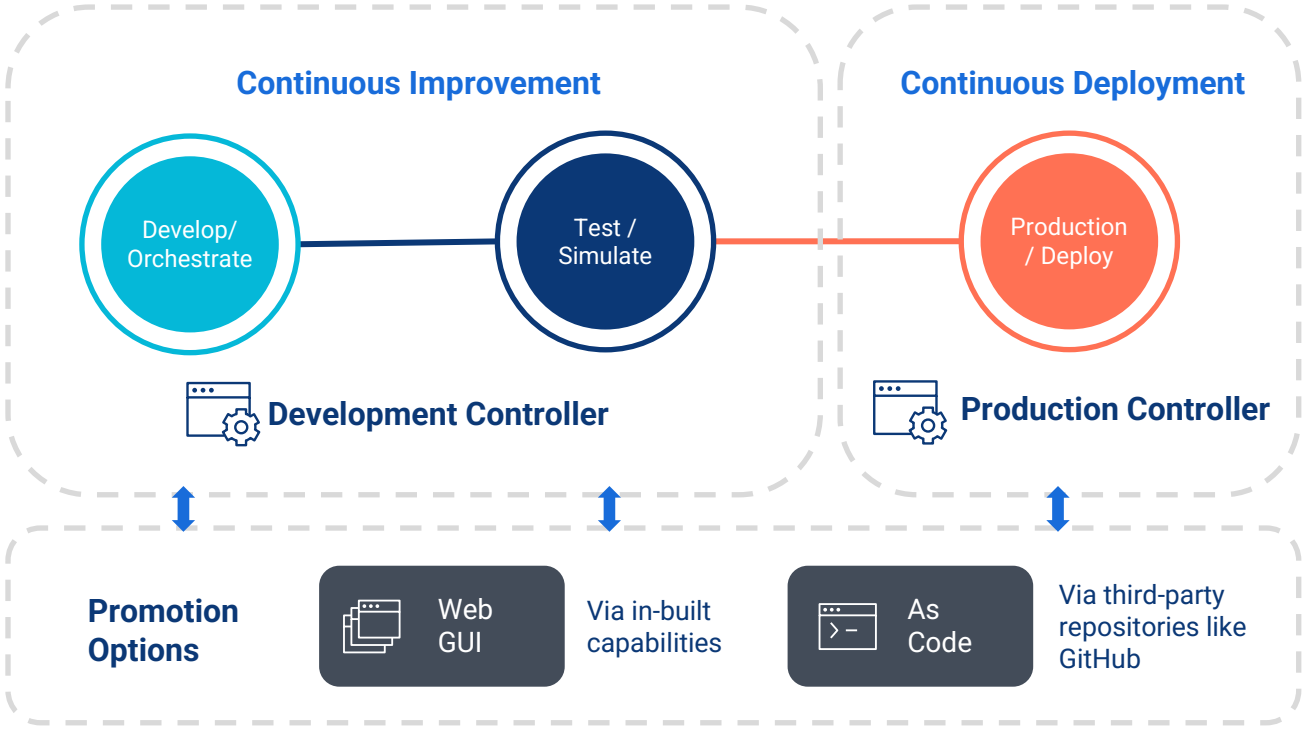
For Enterprises Ready for the Next Level of Maturity



© Stonebranch 2022. All rights reserved.

Putting the Ops in DataOps

For Enterprises Ready for the Next Level of Maturity



© Stonebranch 2022. All rights reserved.

Data Pipeline Orchestration Demo

Ravi Murugesan

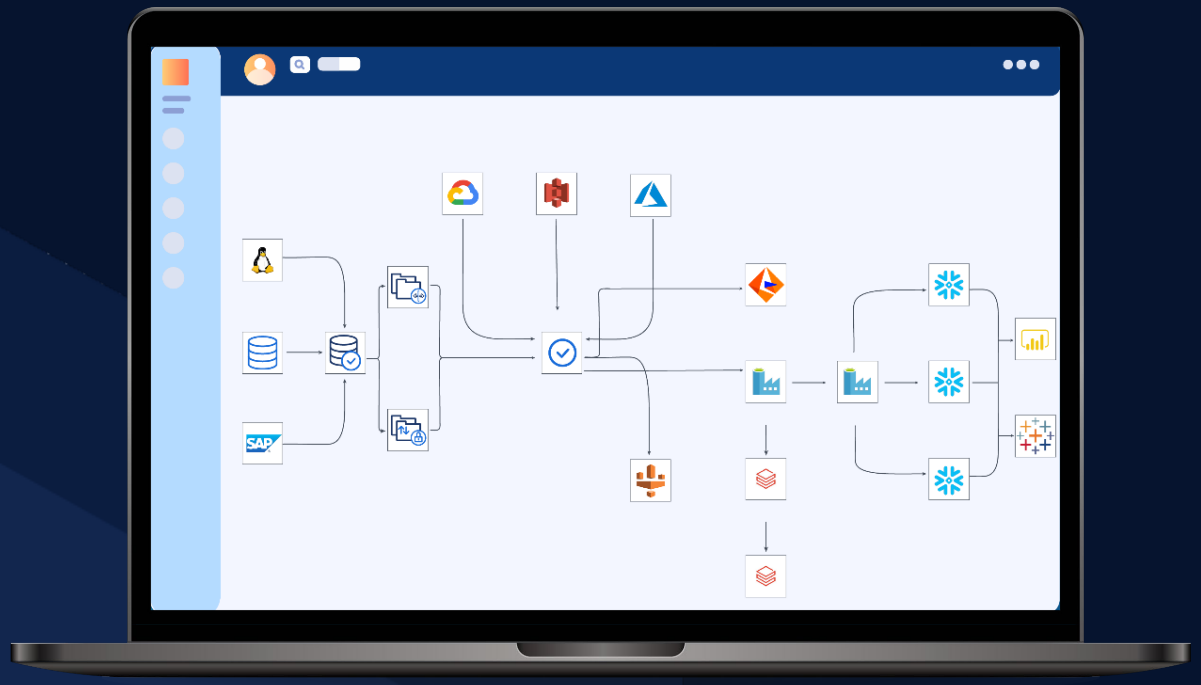
Sr. Solution Engineer

stonebranch

© Stonebranch 2022. All rights reserved.

Demonstration

Update Visual Dashboard from Multiple Data Sources (both on-prem and cloud-based)



Live orchestration of a data pipeline, including

- Sources (cloud, on-prem, apps)
- Ingestion, transformation (Informatica)
- Data Storage (Azure Blob, Snowflake, AWS S3)
- Delivery (Tableau)

© Stonebranch 2022. All rights reserved.

Q & A

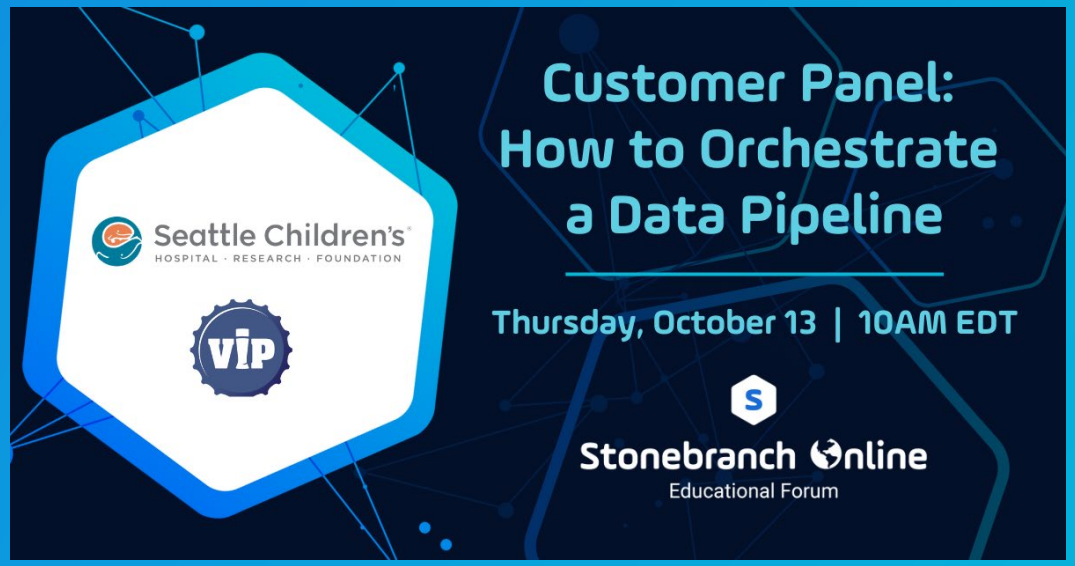
Scott Davis
Global Vice President

scott.davis@stonebranch.com
Stonebranch - Atlanta, USA

Ravi Murugesan
Sr. Solution Engineer

ravi.murugesan@stonebranch.com
Stonebranch – Frankfurt, Germany

Next Step to Learn More:



https://hubs.ly/Q01ncJj20 <---- Visit Link

© Stonebranch 2022. All rights reserved.

Thank You

stonebranch

© Stonebranch 2022. All rights reserved.

Universal Automation Center

Data Pipeline Orchestration Solution

stonebranch

© Stonebranch 2022. All rights reserved.

Real Time Hybrid IT Automation

Universal Automation Center Platform

A Platform Approach

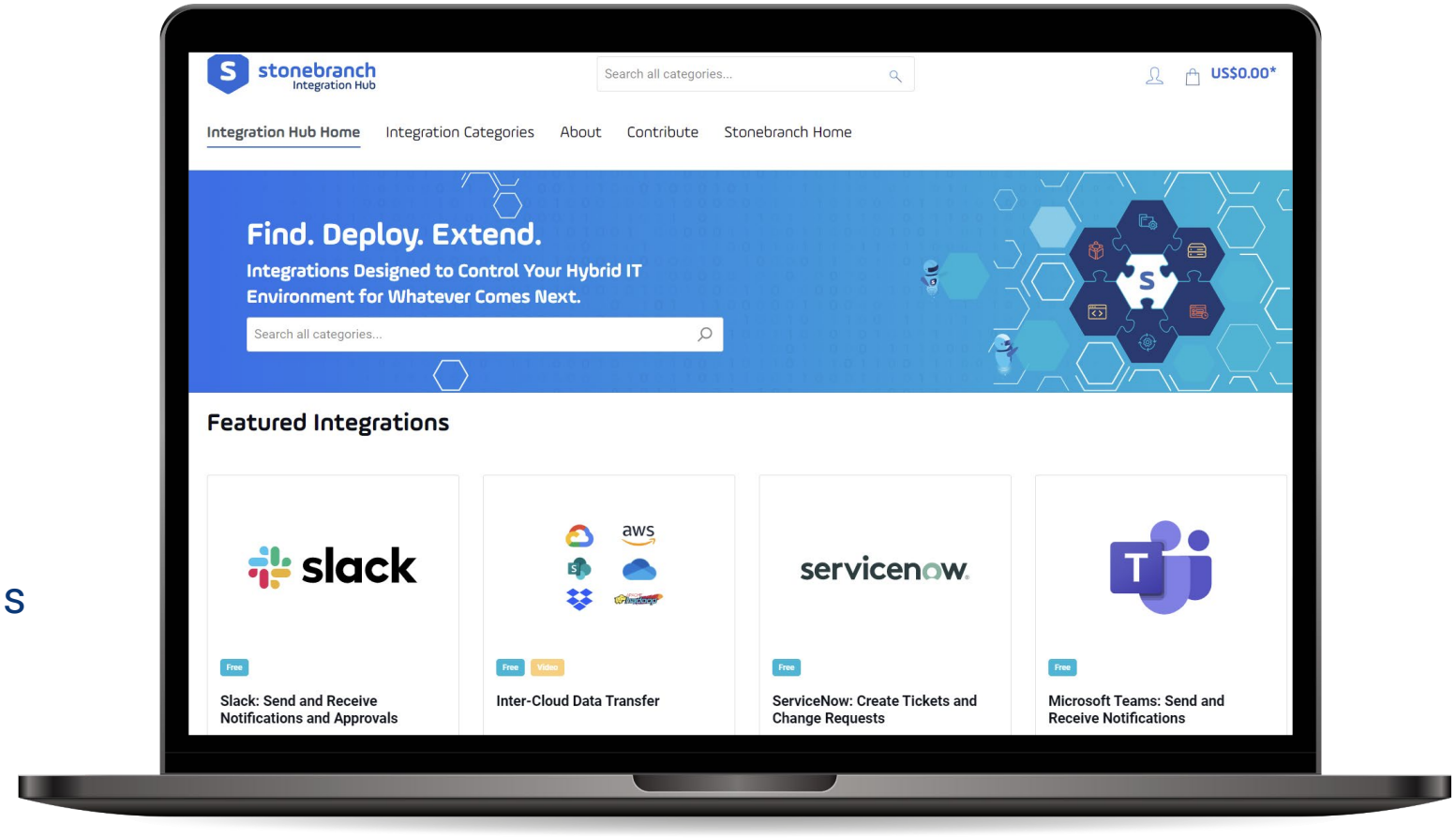
Orchestrating IT processes from on-prem,
to cloud, to containerized microservices

© Stonebranch 2022. All rights reserved.

Orchestration = Integration

Find. Deploy. Extend.

- Download extensions
- Share extensions
- Community driven
- Constant additions (monthly)
- **Large Data Pipeline Focus**
- Rapid creation of new integrations



© Stonebranch 2022. All rights reserved.

What to Look for in a Data Pipeline Orchestration Solution

Real-Time Data Flow

Use modern event-based triggers to power real-time automation of data movement across your entire hybrid IT data pipeline. Remove the need for the traditional time-based automation. Serve internal stakeholders and external customers in the moment.

DataOps Enabled—To Manage the Data Toolchain

Manage your data pipeline with a DevOps-like approach that supports pipelines-as-code. Use standard lifecycle methodologies (Dev/Test/Prod), that include versioning, for end-to-end orchestration.

Proactive Monitoring and Alerting

Visual dashboards and drill-down reports designed to enable real-time guidance, SLA monitoring and updates on your entire operation. Use proactive alerts to root-cause issues and find solutions faster than ever.

Container Technology Including Kubernetes and Docker

Move data back and forth between containers and schedule big data tools to run in containers. Create, start, stop, remove and monitor containers.

Built-in Managed File Transfer

No need to have separate tools that move source data into the data pipeline. Stonebranch's proprietary managed file transfer (MFT) is encrypted, compressed and fault tolerant, far exceeding traditional FTP.

Centralized Control

Take control of your entire IT environment from a centralized command center–style platform. Manage the data pipeline across your entire enterprise. Keep it running with visibility into every process or workflow.

© Stonebranch 2022. All rights reserved.

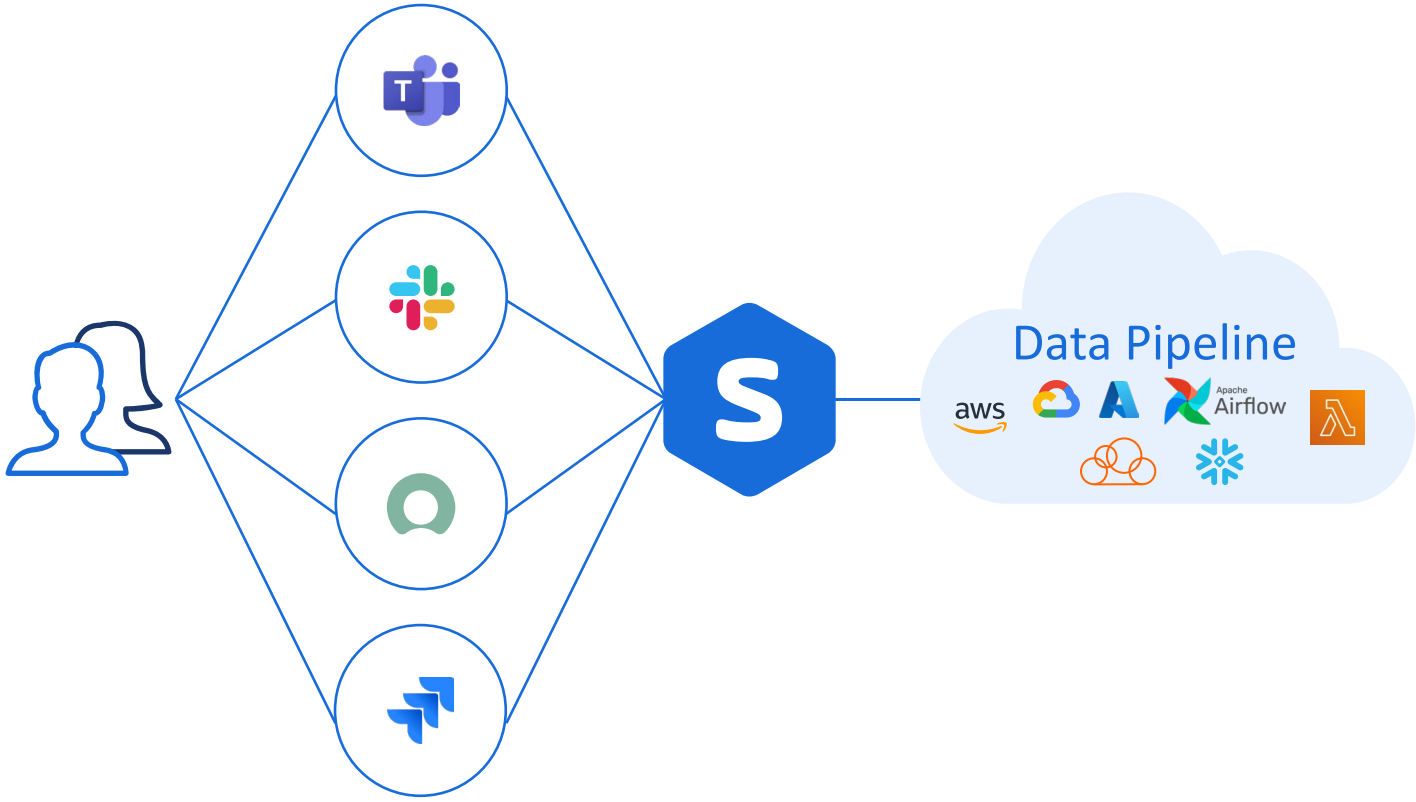
Appendix

Self-Service Automation

Data teams approve and trigger **automated workflows & pipelines** from common business applications

Centralized collaboration platform for data, developers, and operations

IT ops teams gain operational visibility



© Stonebranch 2022. All rights reserved.

Summary

Who is this for?

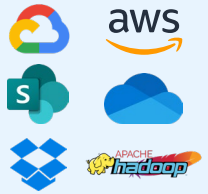
- Want to keep using **existing data tools,** but are ready to graduate from opensource schedulers to enterprise grade platforms
- Would like a single platform to connect Data Teams, Developers, IT Ops, and Cloud Ops teams – to help **scale their data program**
- Need to operationalize **DataOps methodologies** to gain speed and improve data quality
- Want to gain **full visibility** across the entire pipeline – to move quickly when issue arise
- Have a growing or changing data tool landscape, and need the ability to **rapidly build new integrations** (or download pre-existing integrations)
- Need to enable data scientists or business users with simple **self-service capabilities** via the platform or third-party tools like ServiceNow, Microsoft Teams, or Slack
- **Bonus:** Want a central IT automation and orchestration platform (beyond data pipeline orchestration) to support cloud automation, on-prem automation, traditional job scheduling, and DevOps orchestration

stonebranch

© Stonebranch 2022. All rights reserved.

Integration Highlights

Examples



Multi-Cloud Data Transfer Between Any Cloud Storage Provider

Removes the need for intermediate storage systems that are typically leveraged to flow data between cloud service providers.

- No longer need to copy data in two places
- Transfers data between any cloud storage provider
- In addition to traditional cloud service providers, transfer data to Hadoop File Systems, SharePoint, and Drop Box

kafka

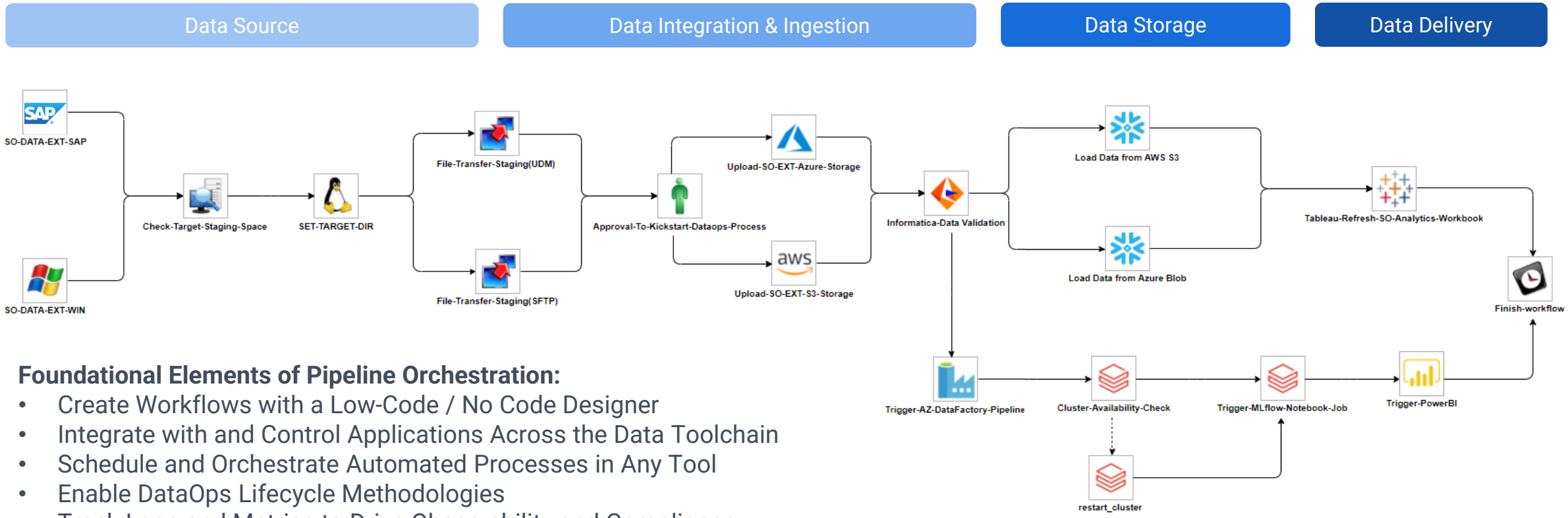
End-users may add Kafka to UAC workflows. At its core, UAC integrates with Kafka to broadcast events, or “wait” for events, in support a variety of end-user needs. Additionally, users gain visibility into what is happening (successes or failures).

Use cases include:

1. Publish event to Kafka
2. Kafka event monitor
3. Trigger and wait

© Stonebranch 2022. All rights reserved.

Recap: Demo Alignment to Stages



Foundational Elements of Pipeline Orchestration:

- Create Workflows with a Low-Code / No Code Designer
- Integrate with and Control Applications Across the Data Toolchain
- Schedule and Orchestrate Automated Processes in Any Tool
- Enable DataOps Lifecycle Methodologies
- Track Logs and Metrics to Drive Observability and Compliance

© Stonebranch 2022. All rights reserved.

Customer Use Case

One of the Largest Global Food & Beverage Manufacturers in the World

Customer Use Case: Overview

One of the Largest Global Food & Beverage Manufacturers in the World

Original Approach

- Their data pipeline for the enterprise data management environment with **Azure Data Factory**
- Azure Data Factory worked well in an Azure environment
- It served as an entry point for the project
- **The Challenge:** Data Factory did not integrate with their full stack of solutions used along the data pipeline

Evolution & Goal

- **Goal:** Orchestrate the full pipeline end-to-end
- **Objective:** Identify a platform that could connect all their critical data tools

Overall Strategy

- On-prem to cloud digital transformation
- Implemented an enterprise analytics data management environment
- Hub-and-spoke model to help keep regional resource groups and services segregated
- Approved services are first developed and deployed at the hub level, with further spoke deployment via containers

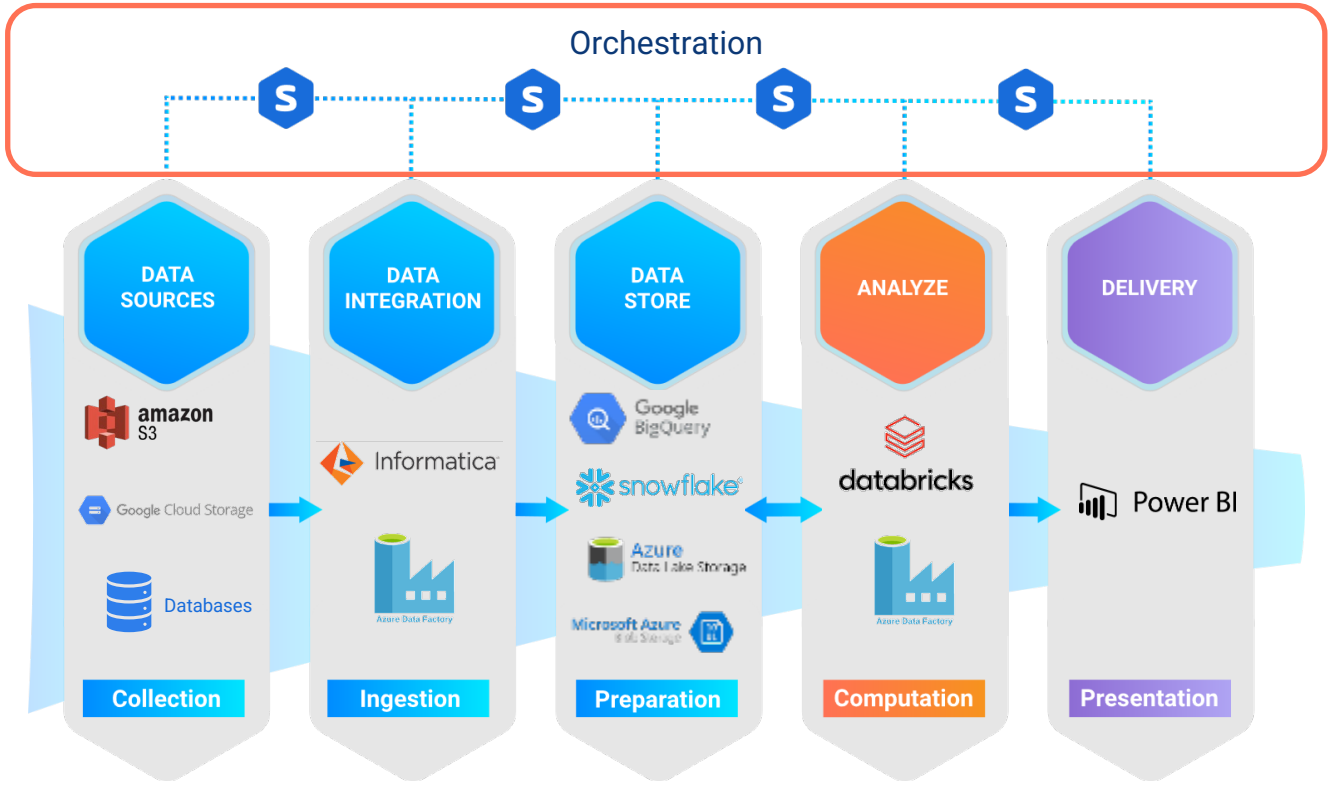
© Stonebranch 2022. All rights reserved.

Data Pipeline Orchestration

One of the Largest Global Food & Beverage Manufacturers in the World

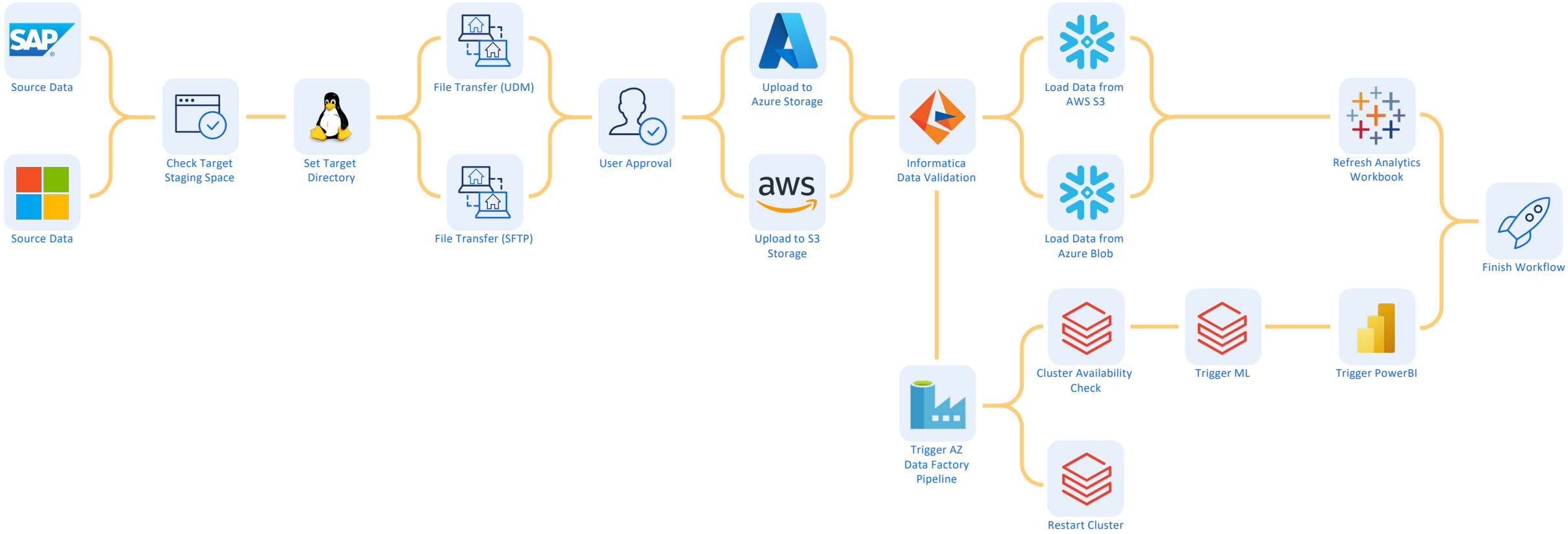
Achieving Their Goal

- Secure and robust file transfer
- DataOps: define pipelines as code and gain lifecycle management (test/dev/prod) capabilities
- Integrate diverse data pipelines that are built using various cloud-based and on-prem services and tools
- For operations: visibility into the process, improve SLAs, real-time monitoring, alerting
- Unified view to design and orchestrate workflows across multiple cloud and on-prem applications



© Stonebranch 2022. All rights reserved.

Big Data Pipeline Workflow Example



© Stonebranch 2022. All rights reserved.