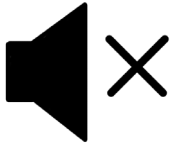


Sep 26, 2023

CDI-Elastic and Advanced Serverless

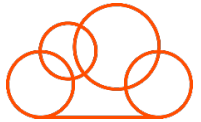
- Sreehari Desai, Senior Solutions Architect, CSA

Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our [Success Portal](#) - where you can also download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

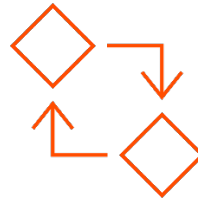
Feature Rich Success Portal



Bootstrap trial and
POC Customers



Enriched Customer
Onboarding
experience



Product Learning
Paths and Weekly
Expert Sessions

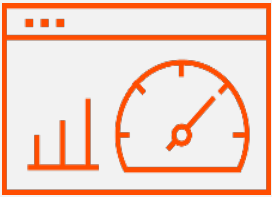


Informatica
Concierge



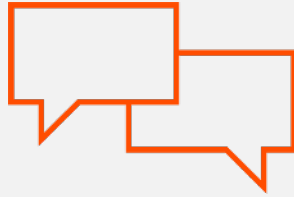
Tailored training and
content
recommendations

More Information



Success Portal

<https://success.informatica.com>



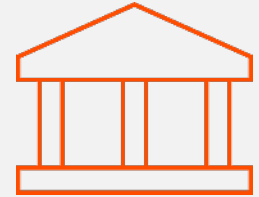
Communities & Support

<https://network.informatica.com>



Documentation

<https://docs.informatica.com>



University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

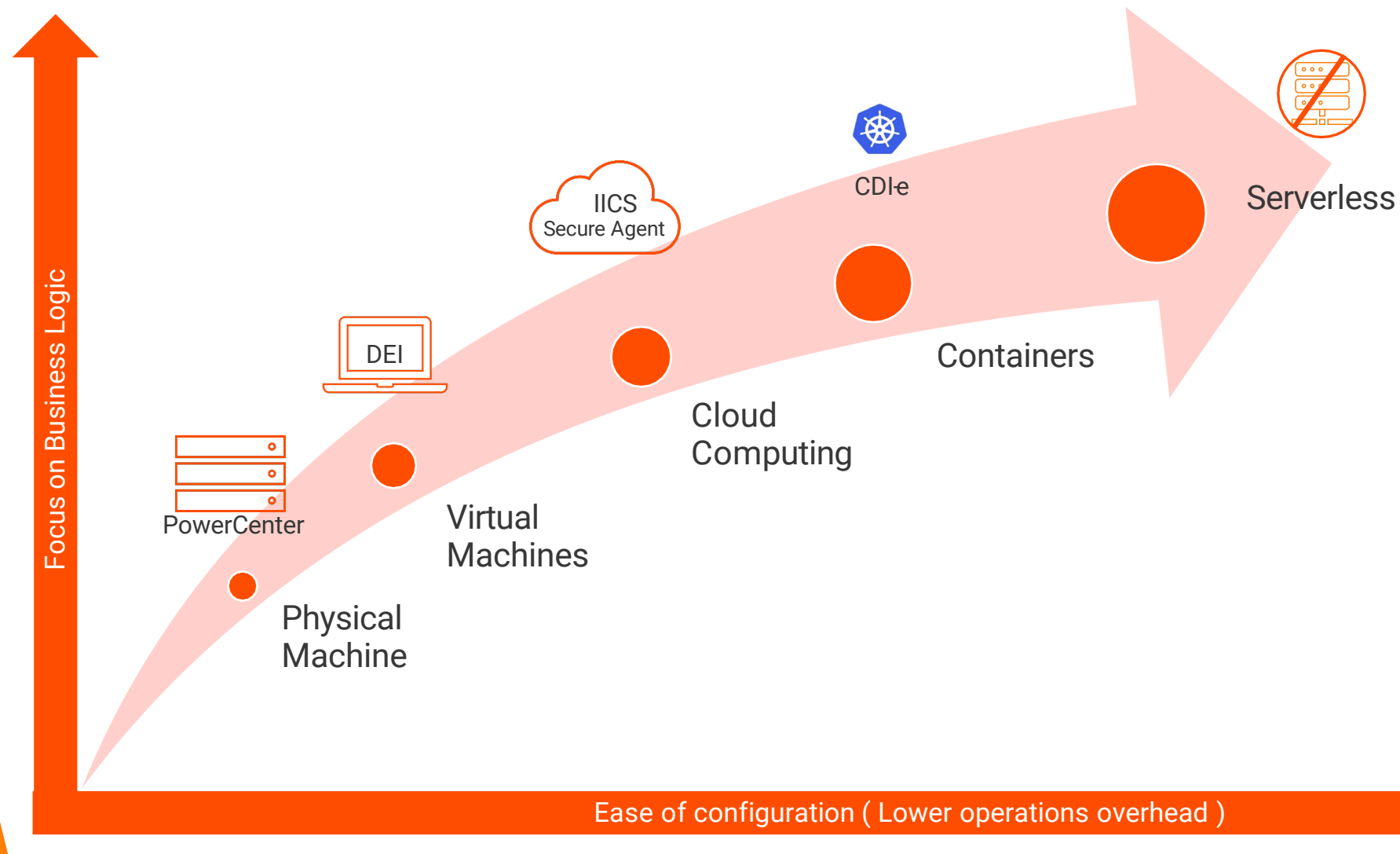
Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Agenda

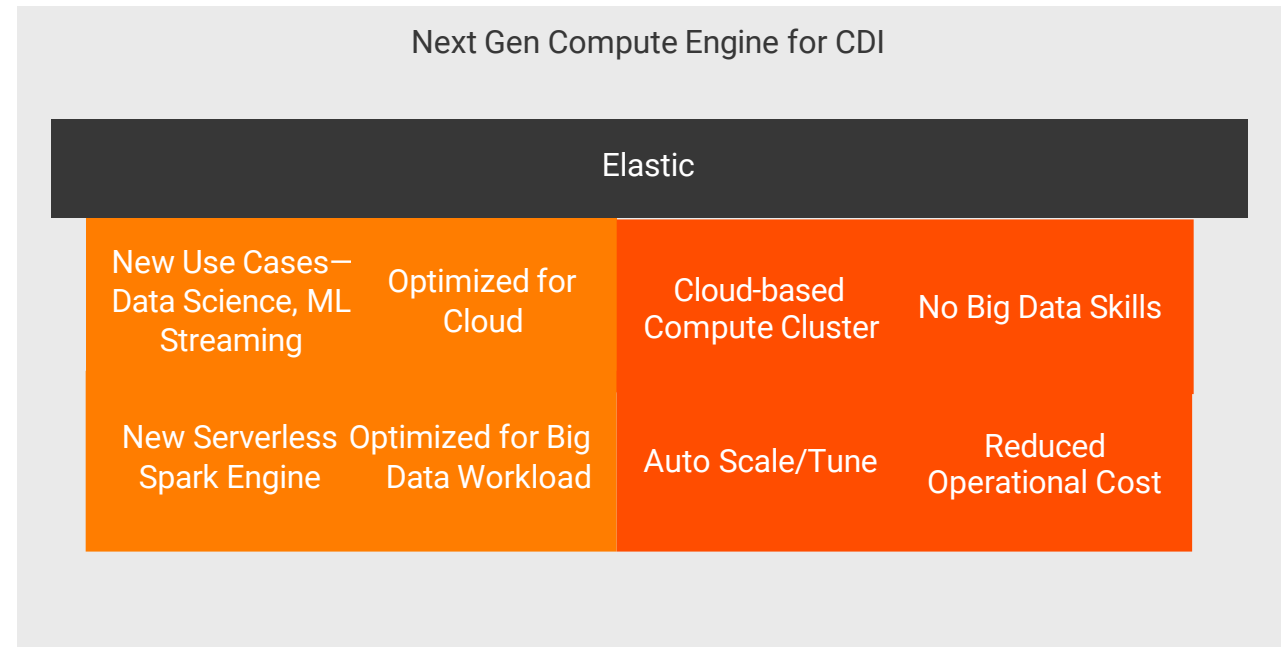
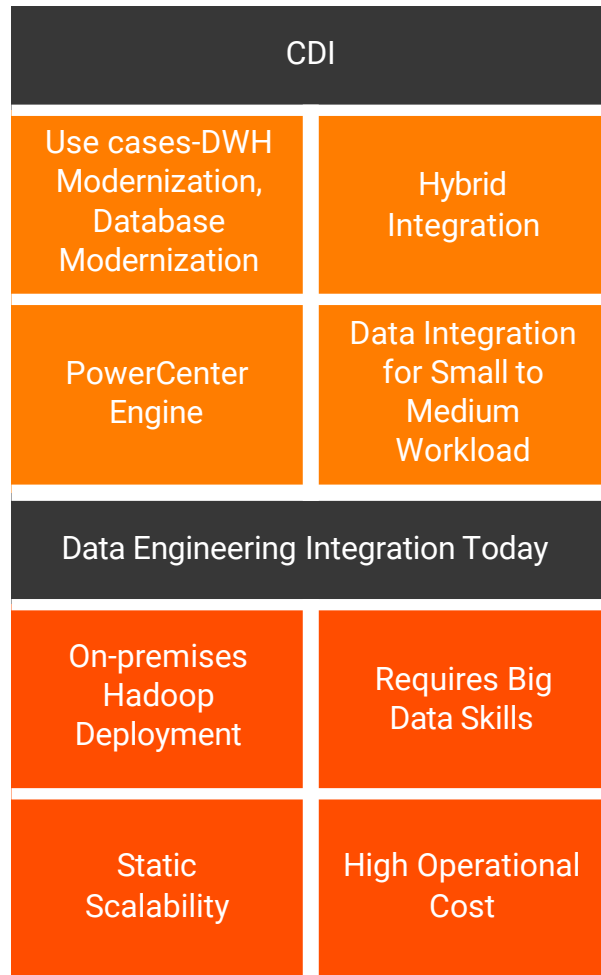
- Introduction
- Advanced Clusters (CDI-E) & Advance Serverless
- Advanced Cluster architecture & Execution Flow
- Lifecycle of Advanced cluster
- Advance Serverless
- Auto Scaling
- Demo
- Q&A

Introduction

Paradigm Shift in Computing Evolution

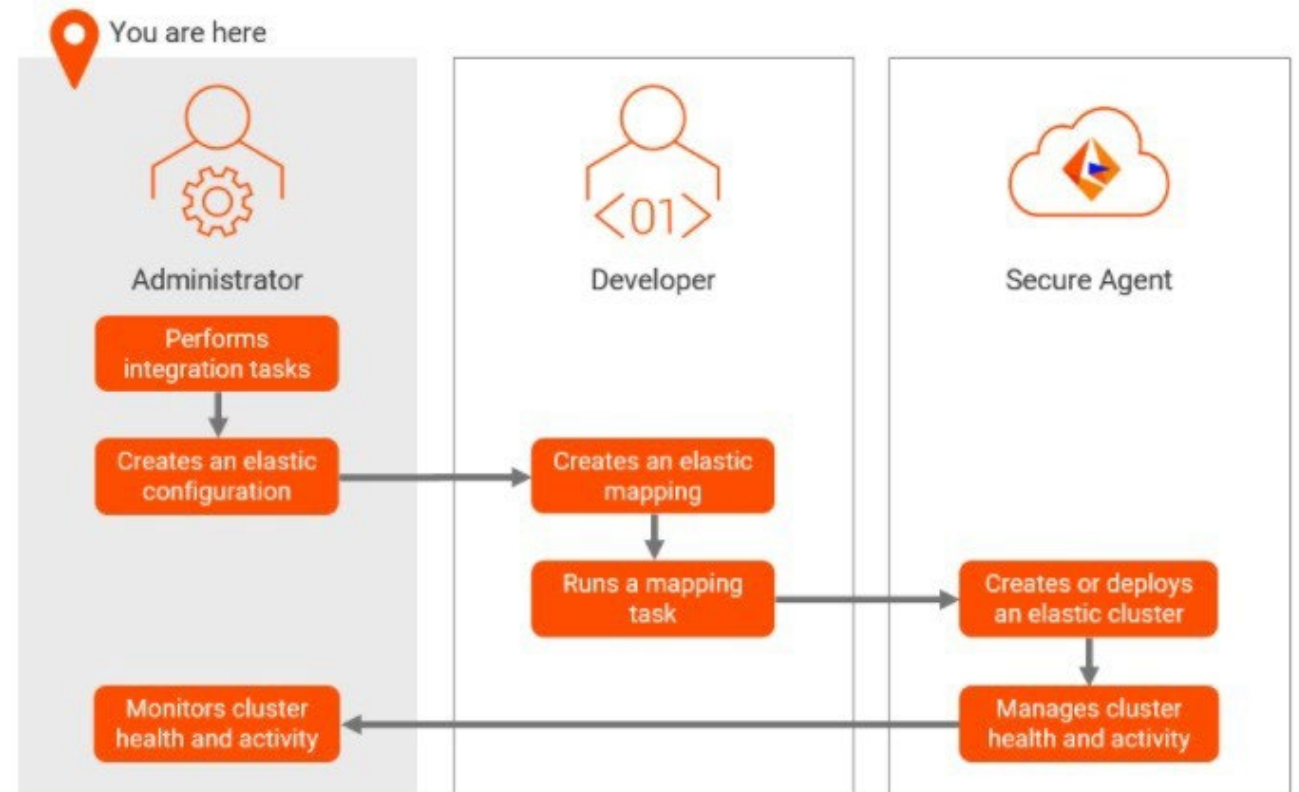


Advanced Clusters (CDI-Elastic) evolution



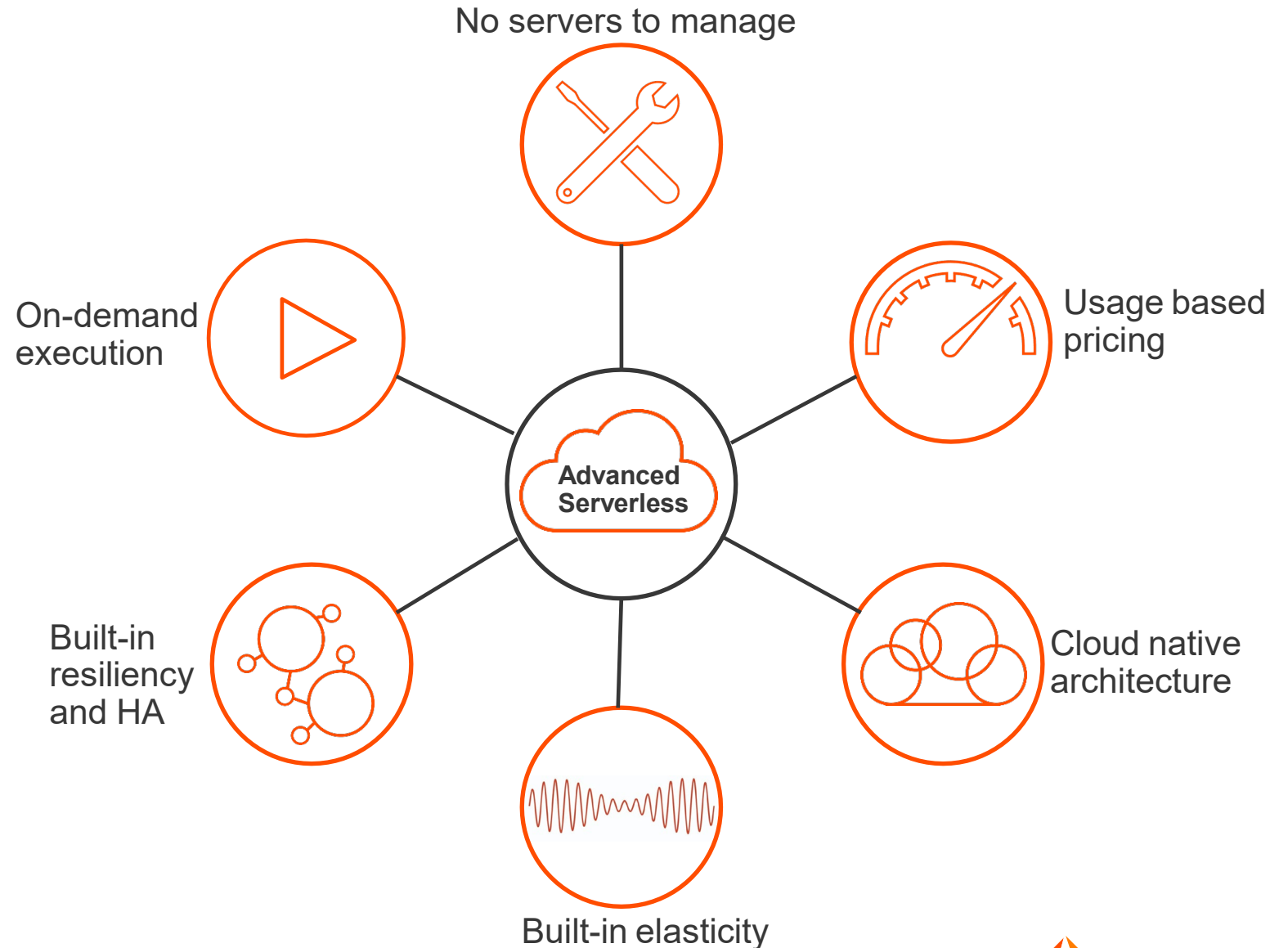
Advanced Clusters (CDI-Elastic)

- Data Integration Elastic allows an user to run elastic jobs in a fully managed environment.
- The Elastic clusters allow Auto Scaling thereby reducing server maintenance activities and operational costs.
- Developers need to only design & run the flow, the management would be done by an auto scale cluster
- Preferred for Med-High workload data volume
- Available on AWS, Azure & GCP
- Supports CDI-E, CDQ-E

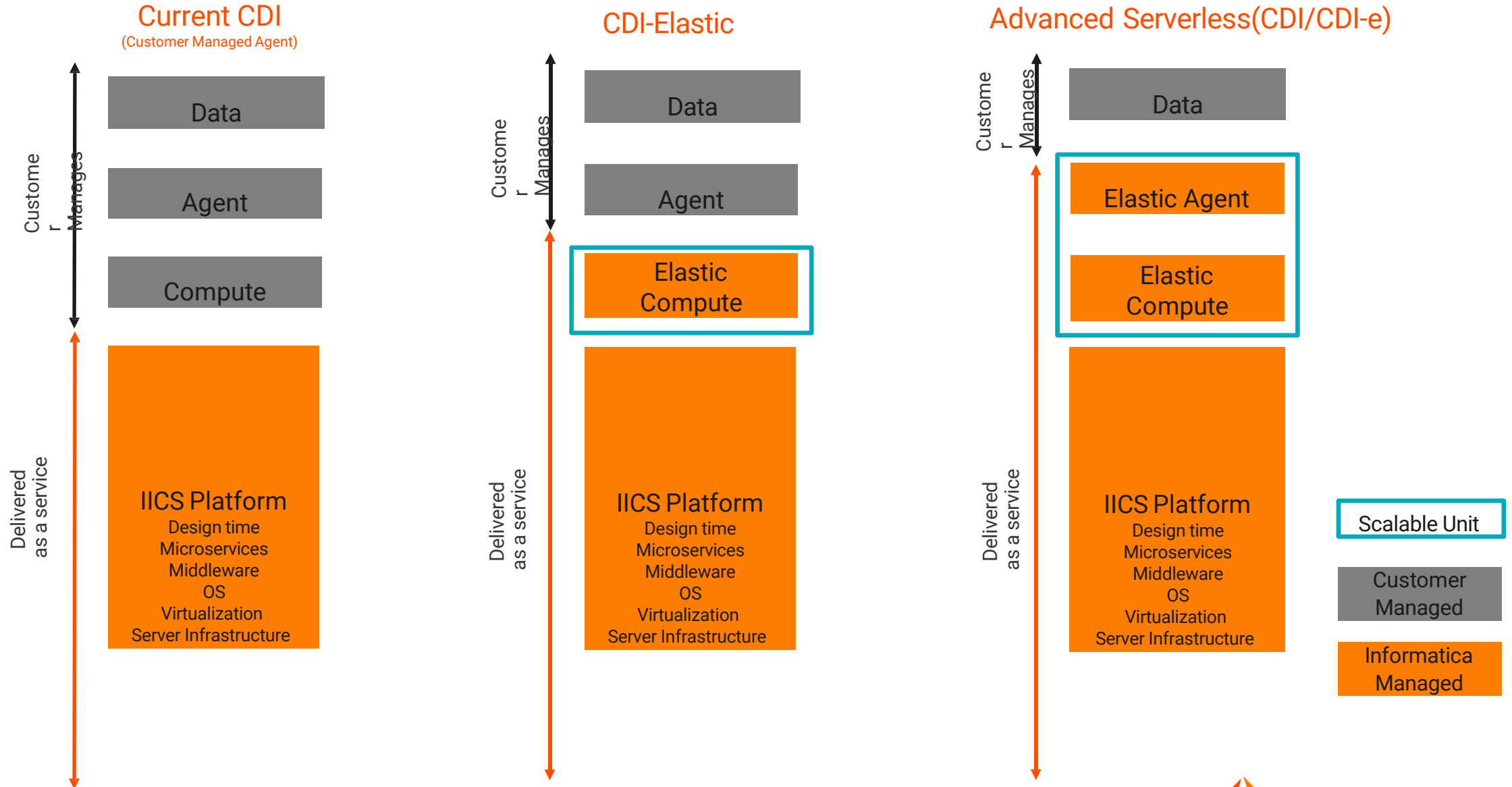


Advanced Serverless

- No servers to manage. Informatica manages both agent & server
- Demand/Event driven Continuous Scaling
- Built-in resiliency and high availability
- Consumption based pricing
- Built in elasticity
- Cloud native architecture
- Preferred for both Low-Med & Med-High workload volume
- Supports CDI, CDI-E, CDQ



CDI-Elastic & Serverless



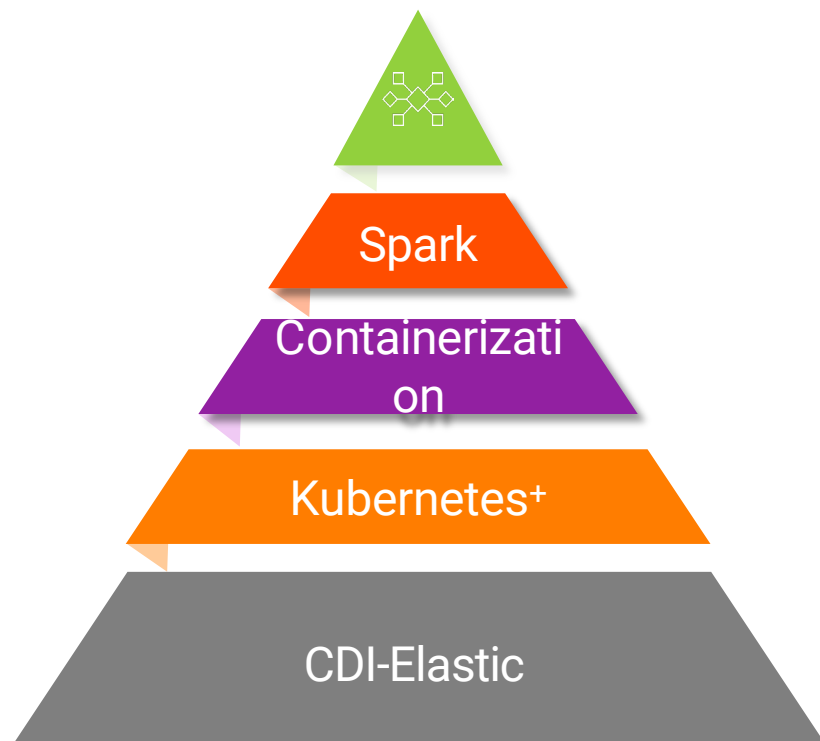
Advanced Cluster

Architecture

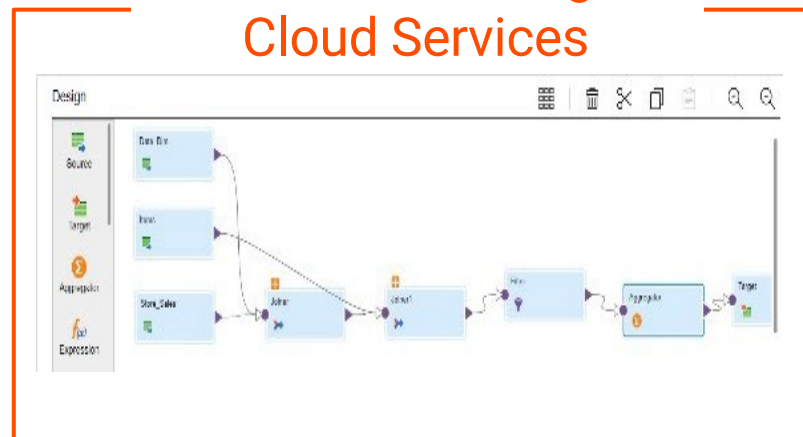
Lifecycle of Cluster

Prerequisites

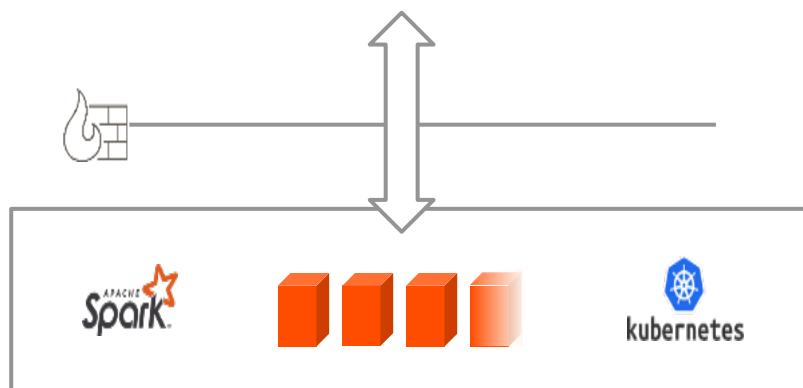
Advanced Cluster Elastic Architecture



Informatica Intelligent Cloud Services



Same, familiar Informatica design-time



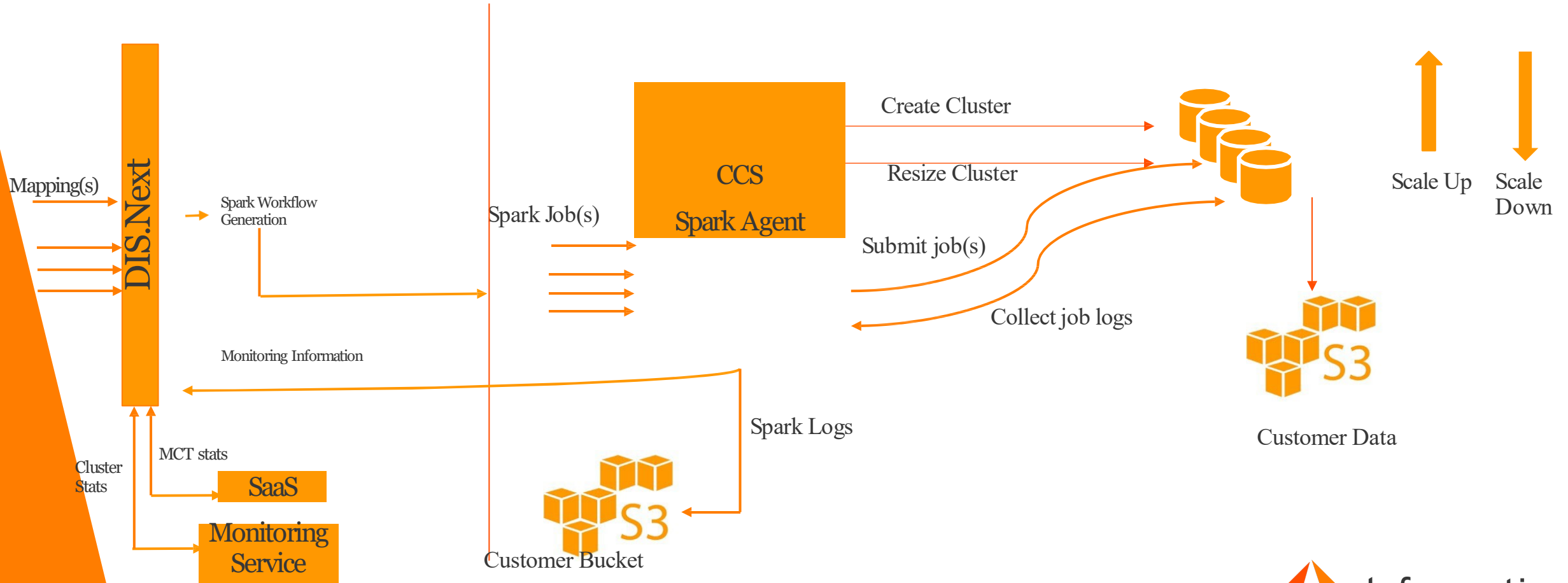
Auto-scaling Spark cluster

Deployed on Cloud network

Advanced Cluster Execution Lifecycle

INFORMATICA VPC

Compute Cluster



© Informatica. Proprietary and Confidential.

Elastic Cluster

- Auto-scaling technology scales the cluster up or down based on the size of the workload.
- The elastic cluster consumes resources only while running jobs.
- CLAIRE®, Informatica's AI engine, uses machine learning to auto-tune the jobs that run on the cluster to achieve optimal job performance.
- High availability, recovery, and resilience ensure that jobs can continue running smoothly during interruptions.
- The data remains in the customer's cloud environment.

Elastic Cluster prerequisites

- Verify that you have the correct privileges in your IDMC organization.
- Verify that you have the necessary AWS subscriptions
- Create storage locations for cluster files
- Create the VPC and subnets (optional)
- Create user-defined security groups for Amazon EC2 (optional)
- Download and install a Secure Agent on an Amazon EC2 instance.
- Create the cluster operator role and policies
- Create or reuse the Secure Agent role and add necessary policies
- Configure the trust relationship for the cluster operator role to include the Secure Agent role
- Assign the Secure Agent role to EC2 machine.

Advanced Cluster Configuration

How to configure the Advanced cluster

Advanced Cluster Configuration

- The Advanced cluster configuration is done by the Administrator of the IICS cloud org
- Based on the cloud formation setup, the configuration is done
- Along with these basic config details, there are advance properties also available.

Basic Configuration

Name: *	<input type="text" value="Puneeth_aws_cdie"/>
Description:	<input type="text"/>
Runtime Environment:	<input type="text" value="CDI-E-AWS"/>
Cloud Platform:	<input type="text" value="Amazon Web Services (AWS)"/>
Region: *	<input type="text" value="US West (N. California)"/>
Master Instance Type: *	<input type="text" value="t3.xlarge"/>
Master Instance Profile: ?	<input type="text"/>
Worker Instance Type: *	<input type="text" value="t3.2xlarge"/>
Number of Worker Nodes: *	Min: <input type="text" value="1"/> Max: <input type="text" value="3"/>
Worker Instance Profile: ?	<input type="text"/>
Enable High Availability: ?	<input type="checkbox"/>
Availability Zones:	<input type="text" value="Select Availability Zones"/>
EBS Volume Type:	<input type="text" value="SSD (gp2)"/>
EBS Volume Size: * ?	Min GB: <input type="text" value="100"/> Max GB: <input type="text" value="100"/>
Cluster Shutdown: * ?	<input type="radio"/> Smart Shutdown <input checked="" type="radio"/> Idle Timeout <input type="text" value="30"/> <input type="text" value="minutes"/>
Mapping Task Timeout: ?	<input type="text"/> <input type="text" value="minutes"/>
Staging Location: * ?	<input type="text" value="infatest/puneeth/cdie/staging"/>
Log Location: * ?	<input type="text" value="infatest/puneeth/cdie/logs"/>
Encrypt Data: ?	<input type="checkbox"/>

Advanced Configuration

VPC: ?	<input type="text" value="v"/>
Subnets: ?	<input type="text" value="s"/>
AWS Tags: ?	<input type="text"/>
Custom Properties: ?	<input type="text"/>
Initialization Script Path: ?	<input type="text"/>
ELB Security Group: ?	<input type="text"/>
Master Security Group: ?	<input type="text"/>
Worker Security Group: ?	<input type="text"/>



Lifecycle of Advanced Cluster

- A cluster lifecycle is the sequence of events that occurs on the elastic cluster.



1. The agent creates an elastic cluster – Based on the elastic configuration.
2. Jobs run on the cluster -The agent pushes the elastic job to the cluster and leverages the Serverless Spark engine to process the data logic in the job
3. The agent stops the cluster -After the Secure Agent stops the cluster, the agent verifies that all cluster resources are deleted.
 - The agent restarts the cluster when another elastic job is submitted

Major Advantages of Advanced Cluster

Simplicity

- Easy drag-and-drop design
- Hierarchy Processing Transformation
- Dynamic Data Partition
- High-performance, metadata-aware connectors
- Intelligent Structure Discover
- Integrated data quality and data profiling

Productivity

- Next best recommendation
- Dynamic mapping
- Auto tuning
- Automatic data quality rules
- 70% time savings
- AI-powered, CLAIRE-based

Scalability

- Elastic Spark on Kubernetes-based processing engine
- Advanced Push Down Optimization
- Move workloads freely between clouds
- Purpose-built tools and experiences for all user personas
- Run elastic mapping on GPU

Advanced Serverless

Cloud Data Integration Processing Anywhere

Cloud Data Integration

Cloud-native and codeless data integration with intelligence and automation

Cloud Data Integration Elastic

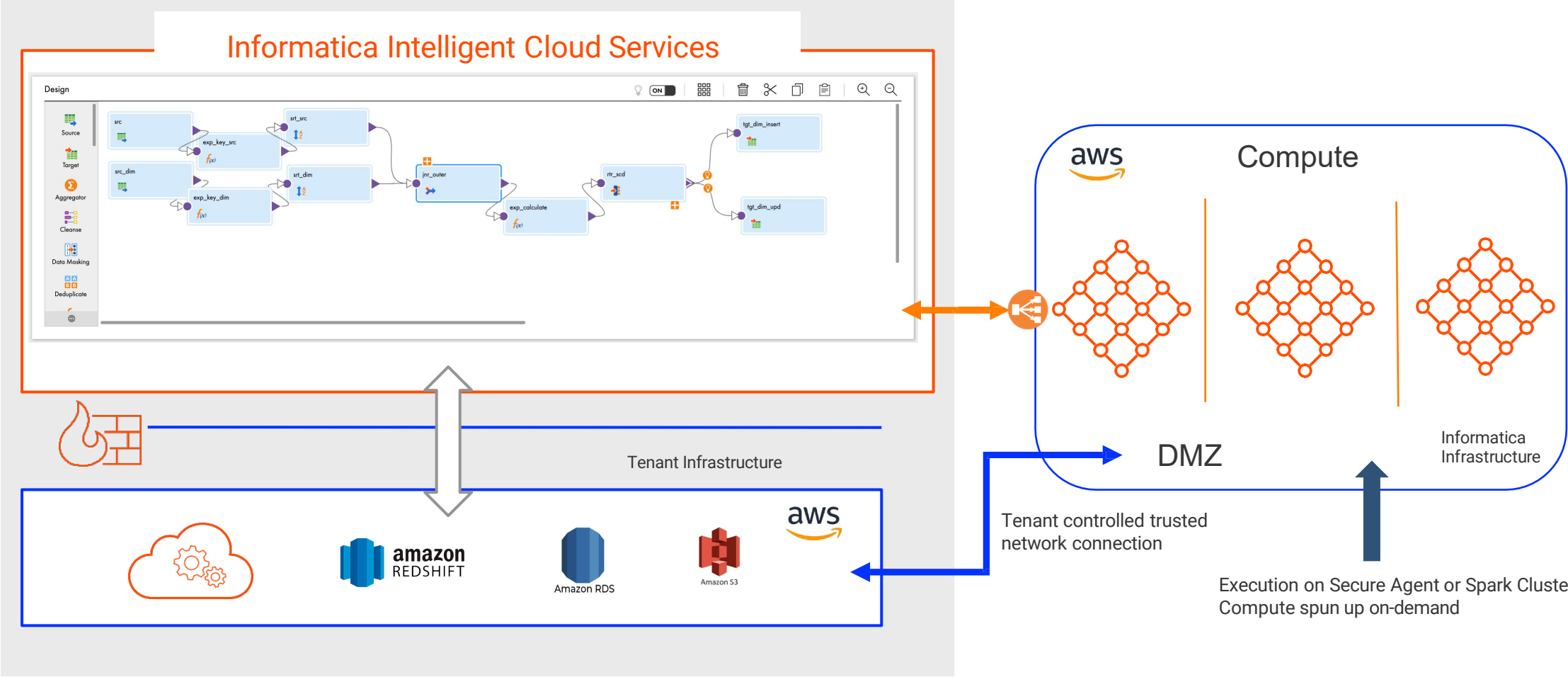
Adds Spark processing with auto tuning and auto scaling capabilities

Advanced Serverless Deployment

Fully managed scale-out environment with no clusters or software to manage and high-performance data integration with built-in elasticity

Advanced Serverless Execution Model

CDI / CDI-Elastic



When to use Advanced Serverless?

When Users have a need to ...

1. Reduce infrastructure maintenance and secure agent administration overhead
 - No more procuring, provisioning, patching
2. Dynamically or seasonally changing infrastructure needs
 - Flexibility to meet changing workloads
3. Jump start projects or do quick POCs
 - Start building data pipelines on Day 1 instead of waiting to secure, provision infrastructure
4. Process lots of concurrent workloads
 - Serverless scales to process infinite concurrent workloads whereas with SA you are limited to processing capacity of the host

Auto Scaling

Cluster Auto Scaling

- CDI-E utilizes auto scaling to address the challenges like over or under utilization of cluster, thus saving data analysts valuable time and operational costs without sacrificing performance
- The minimum & maximum number of nodes are responsible for maintaining the cluster scaling
- Auto scaling adds or removes worker nodes to the cluster based on the demand of the workload
- CDI-Elastic provides idle time-based or smart termination of clusters (cluster lifecycle management) as a part of cluster management and scale-down for auto scaling clusters

Auto Scaling in Elastic Mapping

How the cluster scales up when the load is high

- The Elastic mapping is used to convert a csv file with 10M rows to a json file

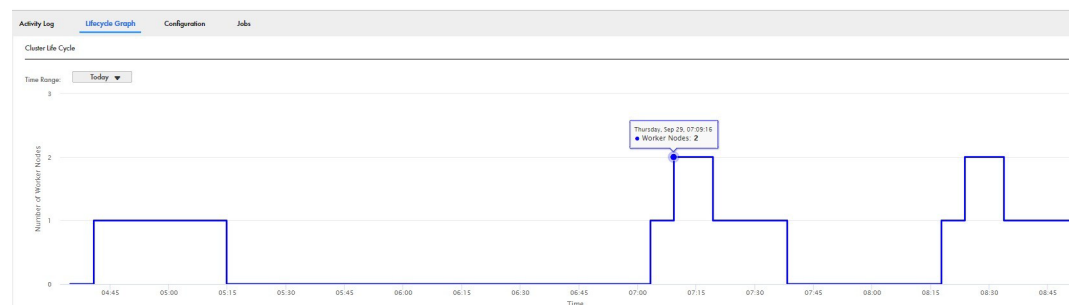
Spark Job Name	Duration (HH:MM:SS)	Total Tasks	Failed Tasks	Input Size / Records	Output Size / Records	Status
▼ Job-0	00:00:53	7	0	776.75 MB / 10426213	1.51 GB / 10426213	✓ Success
Stage-0	00:00:53	7	0	776.75 MB / 10426213	1.51 GB / 10426213	✓ Success

Cluster Activity Log (15) ✓ Up to date

Updated 8:38:54 AM PDT

Time ▼	Activity	Master Instance Type	Worker Instance Type	Master Nodes	Worker Nodes	Total Nodes
Sep 29, 2021, 8:33:45 AM	Scale Down	t3.xlarge	t3.2xlarge	1	1	2
Sep 29, 2021, 8:23:45 AM	Scale Up	t3.xlarge	t3.2xlarge	1	2	3
Sep 29, 2021, 8:17:45 AM	Start	t3.xlarge	t3.2xlarge	1	1	2
Sep 29, 2021, 8:12:45 AM	Starting			0	0	0

- Lifecycle of the cluster while the task is in progress



Demo

References

- CDI-E Elastic

https://network.informatica.com/onlinehelp/IICS/prod/admin/en/index.htm#page/oo-iics-administrator_cdie/Introduction_to_Data_Integration_Elastic_administration.html

- Serverless

https://network.informatica.com/onlinehelp/IICS/prod/admin/en/index.htm#page/gg-iics-rte/Serverless_runtime_environments.html

- Pre requisites

AWS - https://network.informatica.com/onlinehelp/IICS/prod/admin/en/index.htm#page/oo-iics-administrator_cdie/AWS_integration_tasks.html

Azure - https://network.informatica.com/onlinehelp/IICS/prod/admin/en/index.htm#page/oo-iics-administrator_cdie/Microsoft_Azure_integration_tasks.html

GCP - https://network.informatica.com/onlinehelp/IICS/prod/admin/en/index.htm#page/oo-iics-administrator_cdie/Google_Cloud_integration_tasks.html



Thank You