

# GCS Tech Tuesdays: Enterprise Data Catalog 10.5

Stijn Carion  
Informatica Global Customer Support



# Agenda

1 10.2 / 10.4 EOL

2 High Level  
Architecture

3 Informatica  
Cluster Service

4 Upgrade

5 10.5 Version  
Numbers

6 Useful resources

7 Q & A

# 10.2 /10.4 End Of Life

# 10.2 /10.4 End Of Life

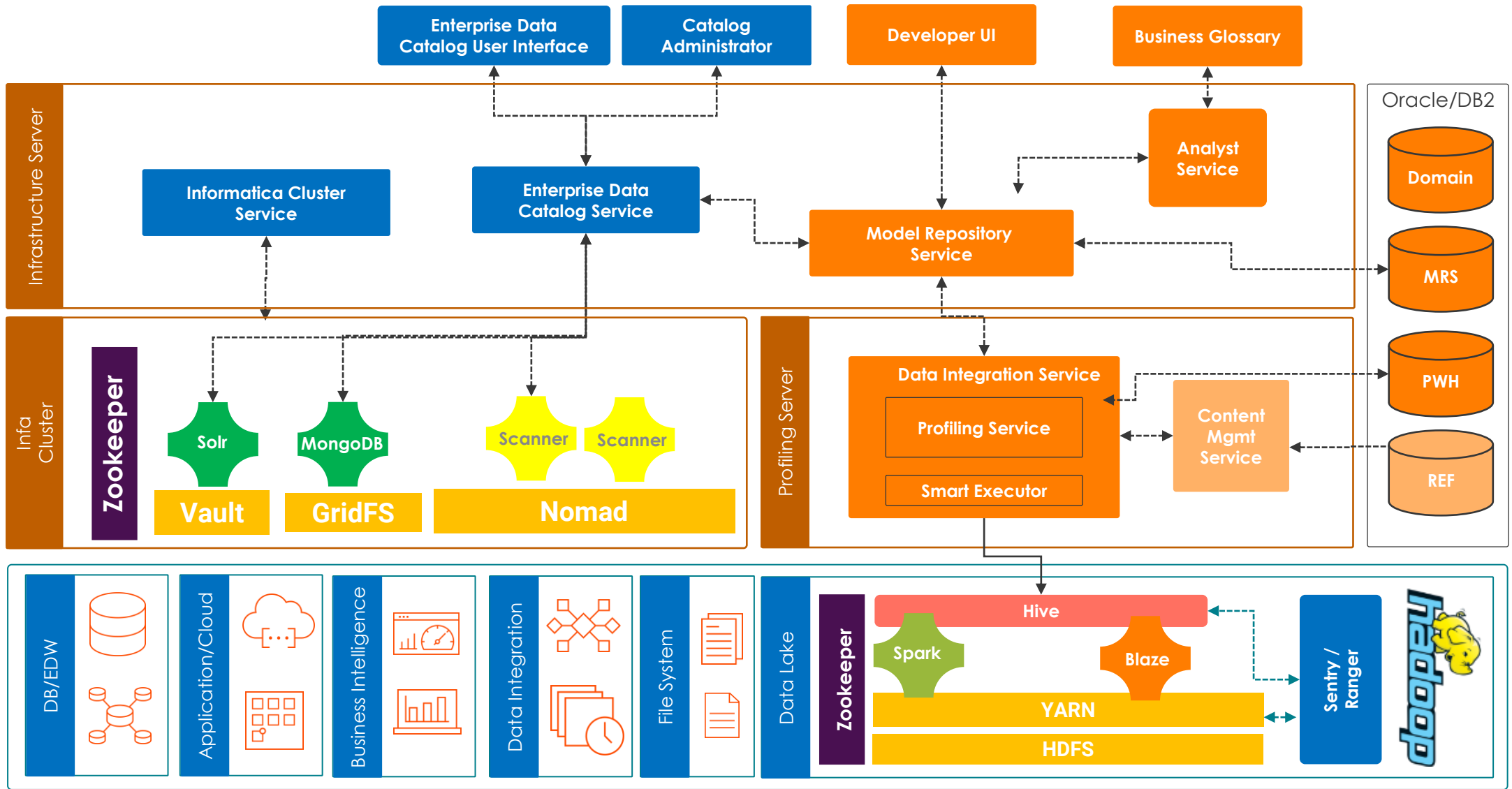
- 10.2 / 10.4 will no longer be supported as of 31 March 2022
- These versions are dependent on Cloudera HDP which will be EOL December 2021
- Informatica has chosen MongoDB and HashiCorp Nomad as the main 3rd party components for replacement of HDP.
- Informatica recommends all EDC Customers to upgrade to 10.5 well in advance of March 2022

# High Level Architecture

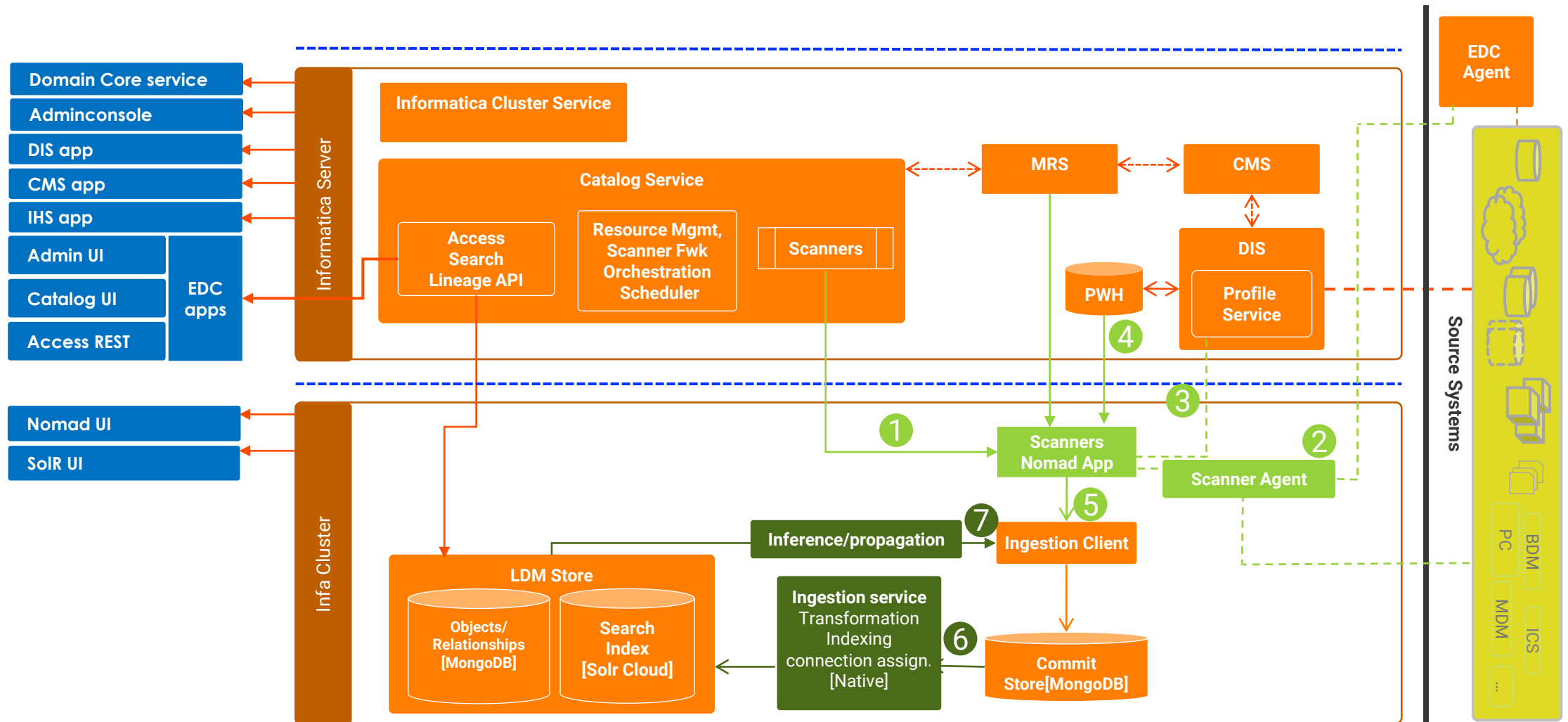
## EDC Architecture 10.5 – What’s Changed?

Store/Engine	Before (10.4.1)	After (10.5.0)
Asset Store	HBase	MongoDB
Graph Store	<u>JanusGraph</u>	MongoDB
Index Store	SOLR, Elastic (DPM)	SOLR, Elastic (DPM)
Event Store (DAA)	Relational (MRS)	Relational (MRS)
Stage Store	HBase	MongoDB
Scan Content Store	HDFS	MongoDB <u>GridFS</u>
Similarity Store	HBase	Postgres
Config Store	Relational (MRS)	Relational (MRS)
Monitoring Store	Relational (MRS)	MongoDB
Compute	Native, Spark (YARN)	Native on Nomad, Spark Standalone (DPM)

# EDC Architecture 10.5 – Services



# EDC Internals 10.5 – Scanner process





# Informatica Cluster Service

# 10.5 Informatica Cluster Service - architecture

## Deployment Nodes

- Nomad (**Multiple Nodes**) – Scanner Execution / Processing
  - MongoDB / MongoDB GridFS (**Multiple Nodes**) – Storage
  - Solr (**ALL Nodes**) – Index Storage
  - Zookeeper (**ALL Nodes**) – Service Coordination
  - PostgreSQL DB (**Single Node**) – Similarity Storage
- 
- New License Key is required for 10.5 (Otherwise Cluster Service Startup fails)
  - \*\* All components are packaged with Informatica Binary – No need to install any component manually

# Informatica Cluster Service - architecture

## Deployment Nodes

**New Informatica Cluster Service - Step 3 of 4** ✕

Fields marked with an asterisk (\*) are required.

**Specify the properties for this new Informatica Cluster Service**

**Informatica Cluster Options**

Gateway Host *	<input type="text" value="infa-demo-edccl1.project.com"/>
Data Nodes *	<input type="text" value="infa-demo-edccl1.project.com,infa-demo-edccl2.project.com,infa-demo-edccl3.project.com"/>
Processing Nodes *	<input type="text" value="infa-demo-edccl1.project.com,infa-demo-edccl2.project.com,infa-demo-edccl3.project.com"/>
Gateway User *	<input type="text" value="infa"/>
Cluster Custom Directory	<input type="text" value="/opt/informatica/ics"/>
Cluster Shared File System Path	<input type="text" value="/infa/ics"/>

Enable Advanced Configuration

Enable Service

**Nomad Service Options**

**Zookeeper Service Options**

**Solr Service Options**

**Mongo DB Service Options**

**Postgres Service Options**

? < Back Next > Finish Cancel

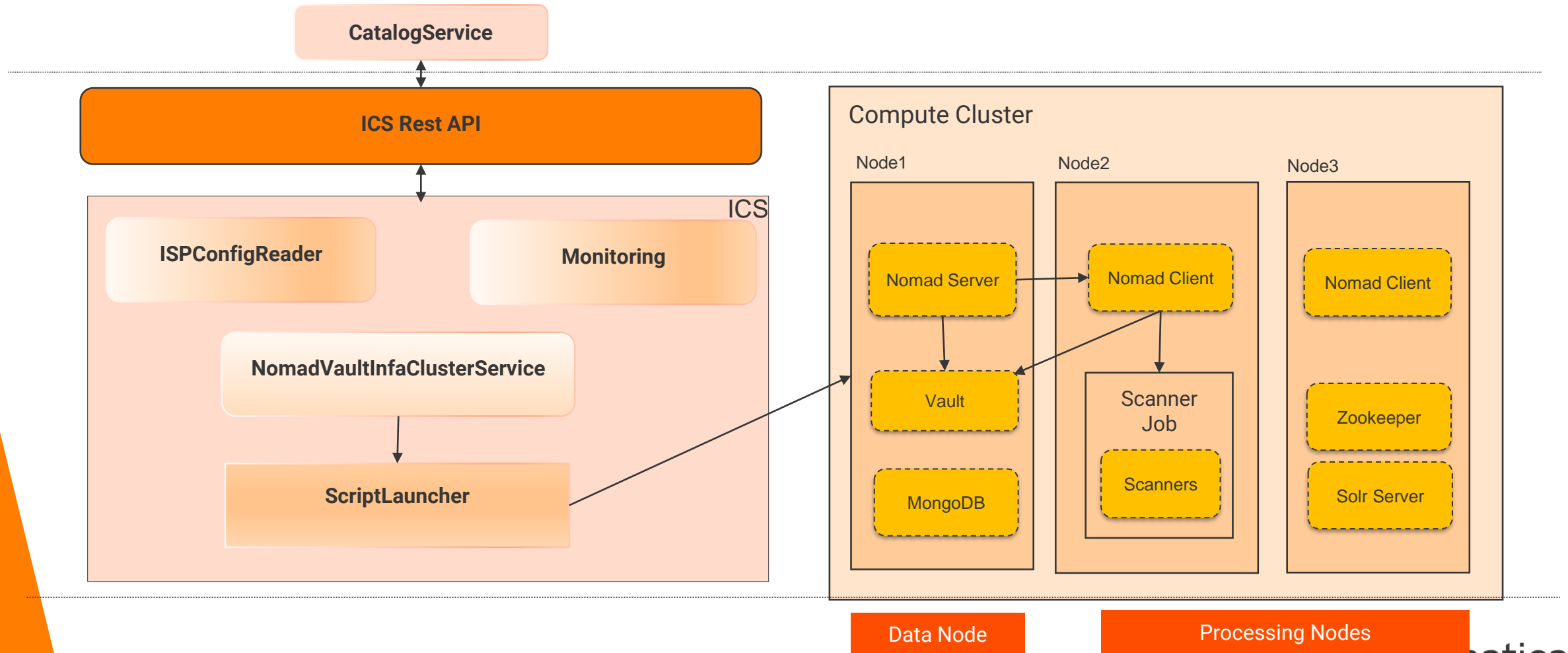
MongoDB

Nomad

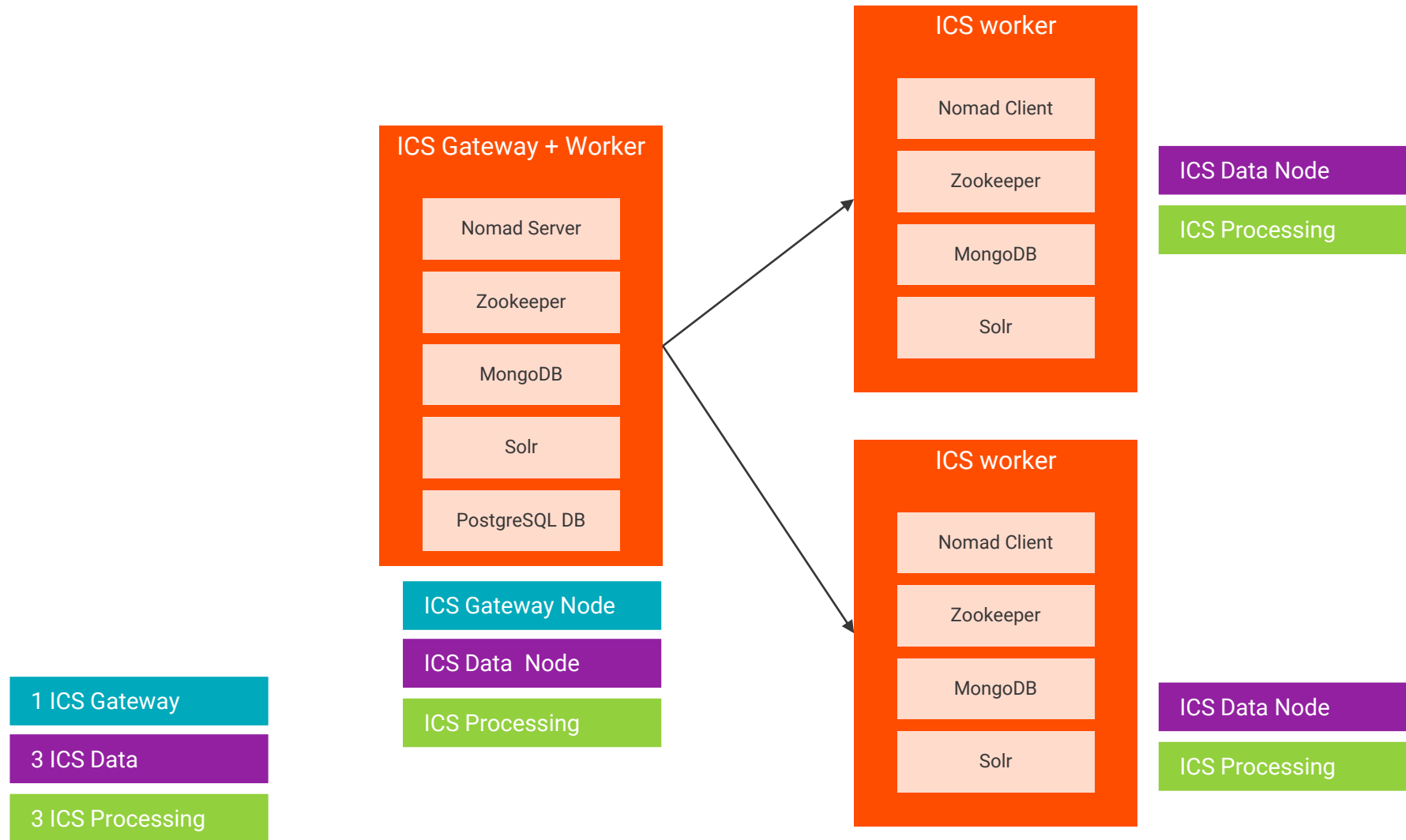
Solr and Zookeeper  
is deployed on all  
nodes



# Informatica Cluster Service (ICS)



# ICS – 3 Nodes (Reference Recommendation)



# ICS creation

## EDC Informatica Cluster Service

- If domain is SSL enabled – you cannot create ICS without SSL (like in previous versions)
  - sudo access is required during installation or whenever you run clean cluster (to redeploy)
  - ICS contains : Gateway Node , Data Node and Processing Node
    - Single Node
      - Provide Same Node as Gateway Node , Data Node and Processing Node
    - 3 Nodes
      - Provide One Node as Gateway Node
      - All 3 Nodes as Data Nodes and Processing Nodes
    - 6 Nodes
      - Provide One Node as Gateway Node
      - 3 Nodes as Data Nodes
      - Remaining 3 as Processing Nodes
- \*\* Separate node is recommended for Similarity Discovery (PostgreSQL DB)

# ICS creation

### New Informatica Cluster Service - Step 3 of 4

Fields marked with an asterisk (\*) are required.

**Specify the properties for this new Informatica Cluster Service**

**Informatica Cluster Options**

Gateway Host \*

Data Nodes \*

Processing Nodes \*

Gateway User \*

Cluster Custom Directory

Cluster Shared File System Path

Enable Advanced Configuration

Enable Service

**Nomad Service Options**

**Zookeeper Service Options**

**Solr Service Options**

**Mongo DB Service Options**

**Postgres Service Options**

Processing Nodes

Zookeeper and Solr should be on All Nodes

Data Nodes

## Advance Configuration

### Nomad Service Options

Nomad Server Hosts *	<input type="text"/>
Nomad HTTP Port *	<input type="text" value="4646"/>
Nomad Serf Port *	<input type="text" value="4648"/>
Nomad RPC Port *	<input type="text" value="4647"/>
Nomad Server Working Directory *	<input type="text" value="/opt/informatica/ics/nomad/nomadserver"/>
Nomad Client Working Directory *	<input type="text" value="/opt/informatica/ics/nomad/nomadclient"/>
Nomad Custom Options	<input type="text"/>

### Zookeeper Service Options

Zookeeper Hosts *	<input type="text"/>
Zookeeper Port *	<input type="text" value="2181"/>
Zookeeper Peer Port *	<input type="text" value="2888"/>
Zookeeper Leader Port *	<input type="text" value="3888"/>
Zookeeper Installation Directory *	<input type="text" value="/opt/informatica/ics/zk/install"/>
Zookeeper Data Directory *	<input type="text" value="/opt/informatica/ics/zk/data"/>
Zookeeper Custom Options	<input type="text"/>

### Solr Service Options

Solr Hosts *	<input type="text"/>
Solr Port *	<input type="text" value="8983"/>
Solr Installation Directory *	<input type="text" value="/opt/informatica/ics/solr/install"/>
Solr Data Directory *	<input type="text" value="/opt/informatica/ics/solr/data"/>
Solr Custom Options	<input type="text"/>

### Mongo DB Service Options

Mongo DB Hosts *	<input type="text"/>
Mongo DB Port *	<input type="text" value="27017"/>
Mongo DB Log Directory *	<input type="text" value="/opt/informatica/ics/mongo/log"/>
Mongo DB Data Directory *	<input type="text" value="/opt/informatica/ics/mongo/data"/>
Mongo Custom Options	<input type="text"/>

### Postgres Service Options

Postgres DB Host *	<input type="text"/>
Postgres DB Port *	<input type="text" value="5432"/>
Postgres Installation Directory *	<input type="text" value="/opt/informatica/ics/postgres/install"/>
Postgres DB Log Directory *	<input type="text" value="/opt/informatica/ics/postgres/log"/>
Postgres DB Data Directory *	<input type="text" value="/opt/informatica/ics/postgres/data"/>
Postgres Custom Options	<input type="text"/>

# Informatica Cluster Service - Performance

- Overall Improvement in scanner execution time
- Better and smooth catalog UI performance
- Start and Stop of Cluster Service takes less than ~5 minutes.
- Cluster Services are more stable than in HDP cluster.
- \*\* Cancelling scan is almost immediate.
- \*\* Do not have to manually go to Nomad UI to stop running jobs (like in HDP cluster yarn application)



# EDC Sizing guideline (summary)

## Updated requirements for 10.5.0

- Refer to Sizing and Performance Tuning Guide for sizing recommendations, parameter tuning and more.

Env. Size	# of conc. (total) users	# of objects	Infrastructure			Metadata processing						Infa Cluster Services						
			Catalog Service			Discovery			Adv. Scanners*			Data/Compute (MongoDB, Nomad)				Similarity Store (PostgreSQL)		
			CPU	RAM (GB)	Disk (GB)	CPU	RAM (GB)	Disk (GB)	CPU	RAM (GB)	Disk (GB)	# of nodes	CPU	RAM (GB)	Disk (GB)	CPU	RAM (GB)	Disk (GB)
Small	20 (200)	2 Million	16	32	200	8	16	20	4	12	50	1	8	24	120	8	16	200
Medium	50 (500)	20 Million	24	32	200	32	64	100	4	16	100	3	24	72	2 TB	8	16	500
Large	75 (750)	50 Million	48	64	300	32	64	500	4	32	200	3	48	144	6 TB	8	16	1 TB
X-Large	100 (1000)	100 Million	48	64	300	64	128	500	4	32	200	6	96	288	12 TB	8	16	2 TB



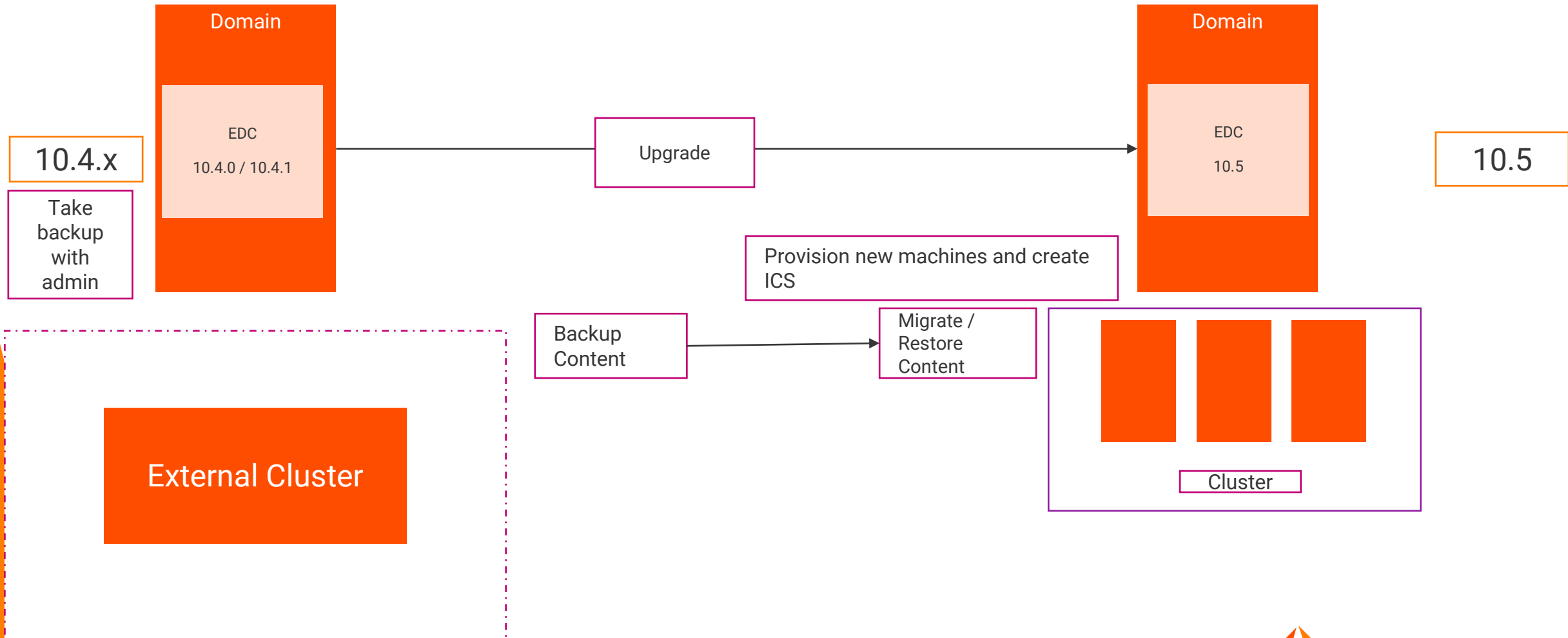
Upgrade

# 10.5 Upgrade Path and License Key

- Direct upgrade path from
  - 10.4.0
  - 10.4.1
- Other versions first upgrade to 10.4.0 or 10.4.1
- New License Key is required for 10.5 (Otherwise Cluster Service Startup fails)

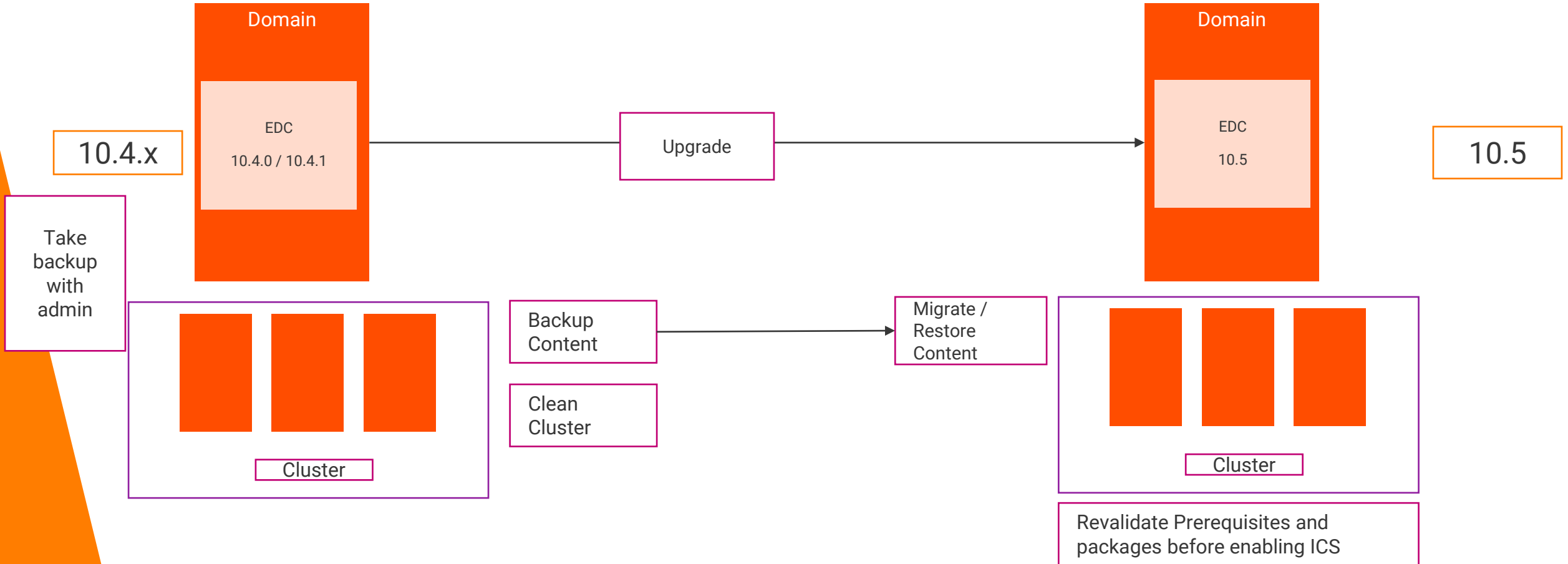
# Upgrade from 10.4.0 or 10.4.1

## EDC – External Cluster (High Level)



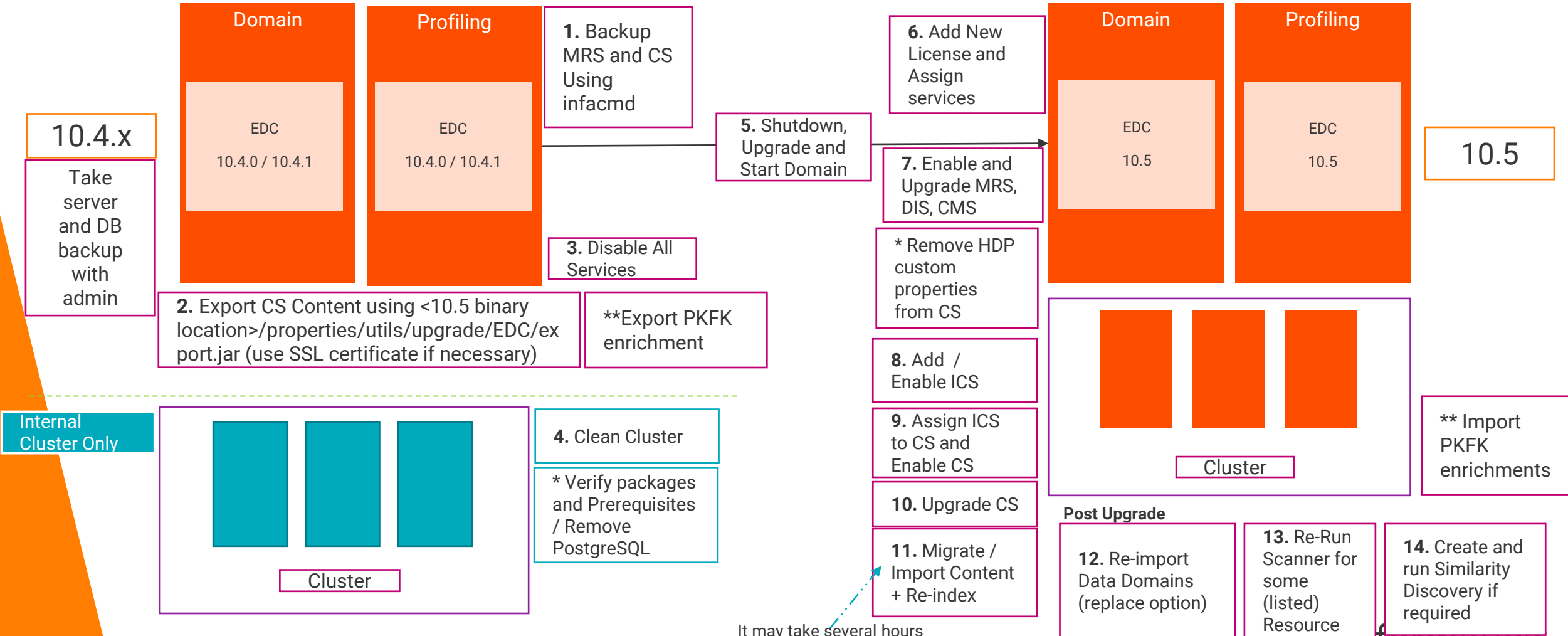
# Upgrade from 10.4.0 or 10.4.1

## EDC – Internal Cluster (High Level)



# Upgrade from 10.4.0 or 10.4.1

## EDC – Internal or External Cluster



It may take several hours

\* Every ~5million asset – 1.5hrs

# Upgrade from EDC 10.4.0 or 10.4.1

Select one of the following options to install or upgrade:

## 1. Install and configure Informatica version 10.5 products.

Choose this option if one or more of the following conditions is true:

\* You want to install and configure Informatica domain services. You can install:

- Only traditional products such as PowerCenter and Informatica Data Quality
- Only the Data Engineering products for Integration, Quality, and Streaming
- Traditional products and the aforementioned Data Engineering products

\* You want to install Enterprise Data Catalog. You can install:

- Enterprise Data Catalog binaries in an existing domain
- Data Engineering products and Enterprise Data Catalog

\* You want to install Enterprise Data Preparation. You can install:

- Enterprise Data Preparation binaries in an existing domain with Enterprise Data Catalog
- Enterprise Data Catalog and Enterprise Data Preparation binaries in an existing domain
- Data Engineering products, Enterprise Data Catalog, and Enterprise Data Preparation

\* You want to install Data Privacy Management. You can install:

- Data Privacy Management in an existing domain with Enterprise Data Catalog
- Data Engineering products, Enterprise Data Catalog, and Data Privacy Management

## 2. Upgrade Informatica products to version 10.5.

Choose this option if one of the following conditions is true:

\* You want to upgrade a previous version of Informatica domain and application services. You can upgrade:

- Only traditional products such as PowerCenter and Informatica Data Quality
- Only the Data Engineering products for Integration, Quality, and Streaming
- Traditional products and the aforementioned Data Engineering products

\* You want to upgrade to Enterprise Data Catalog version 10.5. When you choose this option, the installer upgrades the domain and the Enterprise Data Catalog binaries.

\* You want to upgrade Enterprise Data Preparation and Enterprise Data Catalog. When you choose this option, the installer upgrades the domain, the Enterprise Data Catalog binaries, and the Enterprise Data Preparation binaries. The binaries are installed if Enterprise Data Preparation is not installed.

\* You want to upgrade to Data Privacy Management version 10.5. When you choose this option, the installer upgrades the domain, Enterprise Data Catalog, and Data Privacy Management binaries.

## 3. Configure the Enterprise Data Catalog or Enterprise Data Preparation version 10.5 services on a node in the domain.

Choose this option if you installed the binaries and one of the following conditions is true:

\* You want to configure Enterprise Data Catalog services on a node in the domain.

\* You want to configure Enterprise Data Preparation services on a node in the domain with Enterprise Data Catalog.

## 4. Install or upgrade Data Transformation Engine.

Option to Upgrade EDC to 10.5



# Upgrade process

## High Level Steps

- Upgrades supported from EDC 10.4 and 10.4.1 (latest CPs)
  - Backup catalog content (external / internal cluster) : *infacmd.sh Idm backupContents + using upgrade utility <binary location>/properties/utils/upgrade/EDC/export.jar*
  - Clean embedded cluster hosts : *infacmd.sh ics cleanCluster*
  - Prevalidation of Cluster Machines *java -jar InformaticaClusterValidationUtility.jar -in input.properties*
  - Upgrade the platform : *install.sh*
  - Increase Heap Size (MRS, CS before upgrade)
  - Upgrade MRS and DIS (*GUI or command infacmd.sh mrs upgradeContents*)
  - Deploy new ICS service
  - Enable and Upgrade Catalog service
  - Restore / migrate catalog content : *infacmd.sh Idm migrateContents*
  - Re-index the content



# Upgrade Process

## Heap Size (export and migrate content)

- It is recommended to Increase Heap Size in content export and content migrate commands.
- It is to avoid out of memory issues
- In export.jar / infacmd.sh ldm migratecontent command pass JVM parameters

Assets	Heap Size Recommendation	Time Estimation (hrs)
< 5 MILLION	6 GB	1.5
10 MILLION	10 GB	3
20 MILLION	10 GB	6
50 MILLION	15 GB	15
100 MILLION	20 GB	30

# Rerun required for Resources (Upgrade)

## \*\*Correction

Since publication of this Webinar the Product Team has corrected the official documentation, confirming that a rerun for resources is not required when upgrading from 10.4.X to 10.5.X

Please note the following:

- Similarity Discovery will have to be rerun
- When upgrading from 10.2.X to 10.5.X, via 10.4: Please refer to the 10.4. Upgrade Guide for any Scanners that have to be re-run. If these Scanners were not re-run in 10.4.X, they will have to be re-run in 10.5.X

# Similarity Results (Upgrade)

- Similarity results will not be migrated
- Create New Similarity Discovery Resources with Resource Group.
- Rerun all resources where similarity discovery is required.
- Run Similarity Discovery Resource after upgrade.

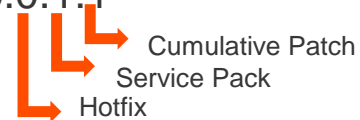
# 10.5 Version Numbers

# 10.5 Version Numbers / Patching

- Monthly patches through Hotfix / Service Packs / Cumulative Patches
  - Hotfix: Informatica platform (EDC/DQ/DEI-Q + Client)
  - Service pack: EDC + DEI-Q + Client
  - Cumulative Patch: EDC Only

- Tracked in Version Number:

- Latest Patch level: 10.5.0.1.1



- Next Cumulative Patch(10.5.0.1.2) release this week
- Next Hotfix(10.5.1) release end of September

Useful resources

# Useful resources

- Enroll on the free OnDemand eLearning to see what's new in EDC 10.5 : <https://now.informatica.com/Whats-New-in-Enterprise-Data-Catalog-105-onDemand-eLearning.html>
- EDC 10.5 General Overview Presentation  
<https://www.youtube.com/watch?v=PbbSrYpbUOk>  
<https://www.informatica.com/content/dam/informatica-cxp/techtuesdays-slides-pdf/Enterprise%20Data%20Catalog%2010.5%20Overview.pdf>
- EDC 10.5 Upgrade Experience Presentation  
<https://www.youtube.com/watch?v=VYCeGIEBcSI>
- Customer Success EDC 10.5 Upgrade Planner:  
<https://success.informatica.com/upgrade-kit-10-5/Enterprise-Data-Catalog.html>
- EDC 10.5 Documentation Set  
<https://docs.informatica.com/data-catalog/enterprise-data-catalog/10-5.html>
- Installation Guide  
<https://docs.informatica.com/data-catalog/enterprise-data-catalog/10-5/installation-for-enterprise-data-catalog.html>
- Upgrading from Versions 10.4.0 and Later  
<https://docs.informatica.com/data-catalog/enterprise-data-catalog/10-5/upgrading-from-versions-10-4-0-and-later.html>
- Additional information regarding External CA Certificates configuration  
[https://knowledge.informatica.com/s/article/HOW-TO-Configure-custom-SSL-for-Enterprise-Data-Catalog-with-external-CA-signed-certificates-using-custom-SSL-scripts?language=en\\_US](https://knowledge.informatica.com/s/article/HOW-TO-Configure-custom-SSL-for-Enterprise-Data-Catalog-with-external-CA-signed-certificates-using-custom-SSL-scripts?language=en_US)

Q&A