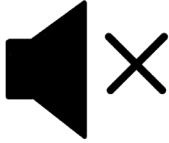


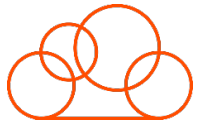
Enterprise Data Catalog 10.5 : Upgrade Experience

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 **INFASupport YouTube channel** and **Success Portal** - where you can 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



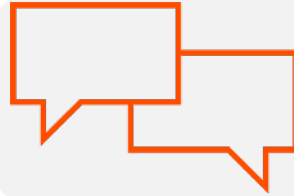
Tailored training and
content
recommendations

More Information



Success Portal

<https://success.informatica.com>



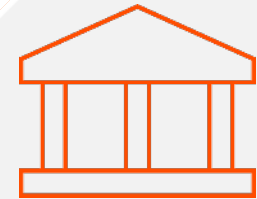
Communities & Support

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

Statements made herein are based on information currently available, 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.

Speaker



Devashish Sharma

Sr. Manager- Products,
Enterprise Data Catalog



Informatica®

EDC Architecture Change in 10.5

Deprecated support for internal and external Hadoop clusters with 10.5 onwards

Replacing



With



Motivations

- HDP going EOL end of 2021
- Customers can continue to use old stack on 10.4.1 till March'22
- New architecture to keep-up with market trend

What to expect?

- No additional hardware or software prerequisites
- No functional loss, similar or improved performance and scale
- Seamless upgrade and content migration
- Continued and improved full support on EDC and all deployed services
- Better support EDC customers on the longer term
- Faster turn around for OS support, security patches

NOTE: EDC 10.4.x and earlier versions are end-of-support by 31 March 2022. Must upgrade to 10.5 prior to 31 March 2022.

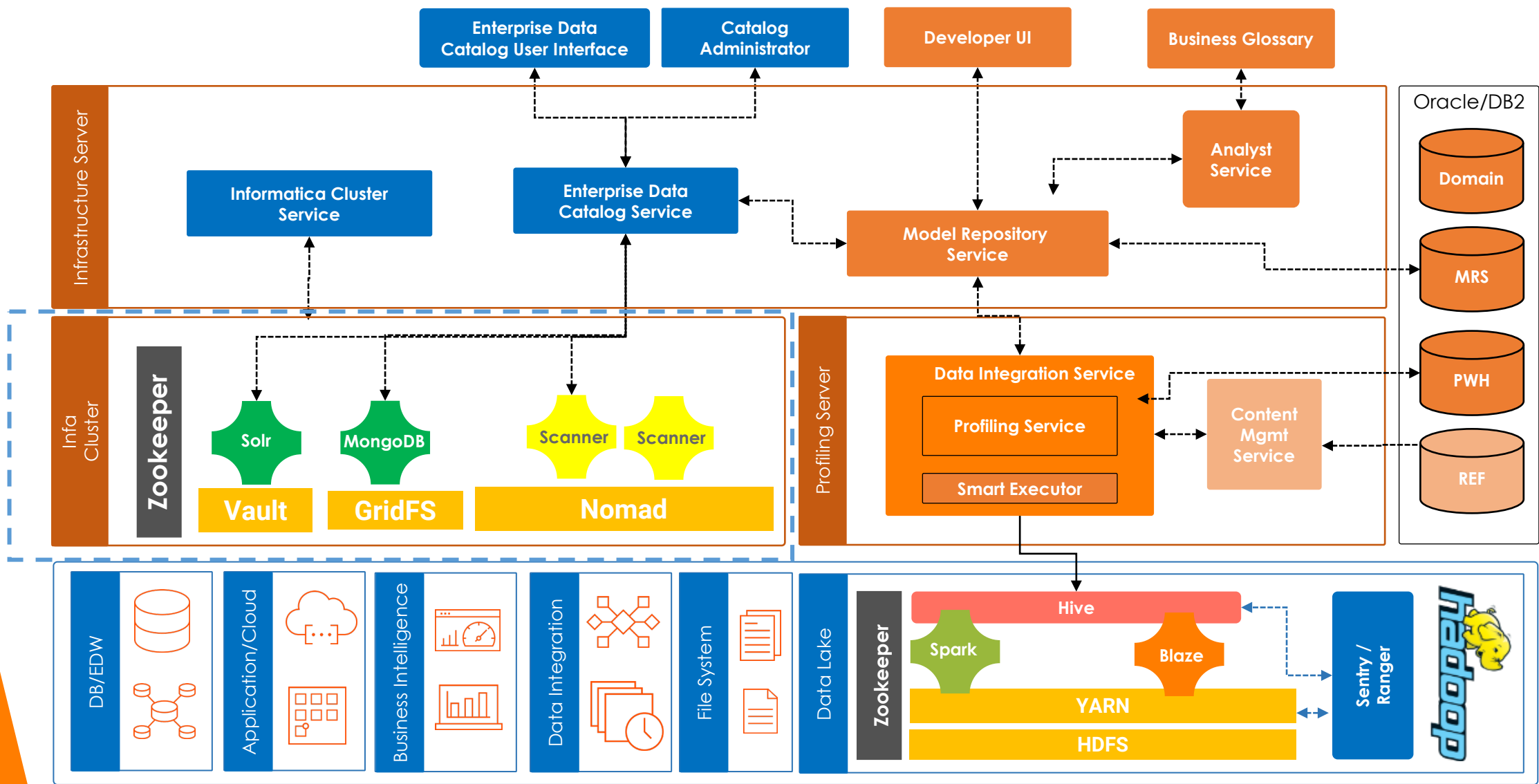
EDC 10.5 New Tech Stack

- For Storage:
 - MongoDB, MongoDB GridFS
 - PostgreSQL
 - SOLR
- For orchestration and security:
 - Nomad
- For compute:
 - Native Java (Scanners, ingestion service)

Note: Software license for the EDC platform components (MongoDB, PostgreSQL, SOLR, Nomad) are included with EDC

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

EDC Architecture 10.5 – Services



Upgrade Process

Like previous version upgrade process

Upgrade 10.x to up to 10.4.1

- Backup catalog content
- Clean embedded cluster hosts – (run `infacmd.sh ihs cleanCluster`)
- Upgrade the platform
- Auto Deploy new version of embedded cluster service
- Restore catalog content
- Upgrade the content
- Re-index the content

Upgrade to 10.5

- Backup catalog content
- Clean embedded cluster hosts – (run `infacmd.sh ihs cleanCluster`)
- Upgrade the platform
- Auto Deploy new ICS service
- Restore/migrate catalog content
- Upgrade the content
- Re-index the content

Advanced Scanners Integration in EDC10.5

Installer integration

- Binaries is part of the Informatica platform
- Deployment of advanced scanner services is part of the platform installer wizard

Admin UI integration and configuration simplification

- Automation of creation, execution control of EDC resources from Advanced Scanner UI
- Default configuration inline with standard scanners (transformation logic, etc.)

Native models

- Advanced scanners leverage full EDC features. Re-using and extending existing EDC models

Support for EDC features

- Less administration overhead (Connection less metadata extraction)
- Connection assignment with other resources
- Reference objects

Shared license

- Advanced scanners service uses platform license mechanism (license option to be enabled as part of the platform license file)

Enterprise Readiness

Security

- Authentication enable for all services (mTLS), Vault to manage internal certificates
- Encryption of data in transit with TLS
- Encryption of data at rest with platform encryption mechanism (AES-256 in 10.5)

Highly available (HA)

- No more Single Point Of Failure (SPOF)
- Each component is deployed in HA mode across the node in the cluster
- Possibility to provide a PostgreSQL HA custom setup

Disaster Recovery (DR) support

- Hot backup support for regular replication to the DR site

**Refer to Implement Disaster Recovery and High Availability Guide for EDC 10.5*

© Informatica. Proprietary and Confidential.

EDC SIZING GUIDELINE SUMMARY

UNCHANGED REQUIREMENTS FOR 10.5.0

Refer to [Sizing and Performance Tuning Guide](#) for sizing recommendations, parameter tuning and more

EDC 10.5 Sizing				Infa Cluster 10.5				Infa Hadoop 10.4.X Embedded Cluster			
Env. size	# of conc. (total)	Metadata resources	# of objects	# of nodes	CPU	RAM	Disk	# of nodes	CPU	RAM	Disk
Low	20	30-40	1 million	1	8	24 GB	120 GB	1	8	24 GB	120 GB
Medium	50	200-400	20 million	3	24	72 GB	2 TB	3	24	72 GB	2TB
High	75	500-1000	50 million	6	48	144 GB	6 TB	6	48	144 GB	12 TB
Very High**	100	1100-1500	100 million	6	96	288 GB	12 TB	NA	NA	NA	NA

Unchanged Requirements for ICS

** New Env Sizing in 10.5

EDC Sizing Guideline Advanced Scanners and Similarity Profiling

EDC 10.5 Sizing	Advanced Scanner Server + Per Advanced Scanner Job *				Similarity Profiling #			
Env. size	# of nodes	CPU	RAM	Disk	# of nodes	CPU	RAM	Disk
Low	-	4	12 GB	50 GB	1	8	16GB	200GB
Medium	-	4	16-24 GB	100 GB	1	8	16GB	500GB-1TB
High	-	4	32 GB	200 GB	1	8	16GB	1TB
Very High#	-	4	32 GB	200 GB	1	8	16GB	2TB

* - Advanced Scanner Server & Service run on domain with additional HW (RAM, CPU & Disk) and not as separate node

- Similarity Profiling to be deployed on additional node

EDC10.5 Performance Highlights

- Faster upgrade to 10.5
 - 2-4M catalog assets in 1 hour
 - 20M catalog assets in 6 hours (10 hours previously)
 - 100M catalog assets in 20 hours (a few days previously)
- Growth up-to 500M assets as compared to 100M on previous version
- Similarity performance is 1.5x to 63x faster with new the similarity configuration support
- Performance coverage verified for 12 Advanced scanners
- New architecture scales better even with higher concurrency
- Improved application startup by ~20% with 5x reduction in war file size

FAQs

Q: Is there an automatic procedure to migrate the current internal cluster data to new mongoDB?

A: We will provide automated migration script. However, we also would have step-by-step guide for the deployment.

Q: Will nomad+mongodb will be shipped with EDC install and do like HIS for internal cluster today?

A: Yes, it will be one installation for overall new architecture deployment.

Q: Will it be bundled with embedded Mongo? Will it still support Hadoop pushdown when scanning customers Hadoop env?

A: we will bundle MongoDB and continue to pushdown to source cluster

Q: How will this impact our AWS, Azure, & GCP Marketplace offerings?

A: New 10.5 marketplace listing will be published after 10.5 release

Q: Will there be Kerberos in Nomad cluster?

A: No. Kerberos security will be handled through other mechanism; mTLS auth and encryption implemented.

Q: What happens to EDP for 10.5? Is it also going to use MongoDB?

A: EDP customers would just have to upgrade the EDC deployment. The EDP deployment will work as is.

More FAQs: https://knowledge.informatica.com/s/article/Enterprise-Data-Catalog?language=en_US

Detailed Upgrade Process

Upgrade Prerequisite

- 100 GB of free disk space on the machine where Informatica domain runs.
- Java Development Kit (JDK) 1.8 is installed on all cluster nodes.
- If you are upgrading from a cluster enabled for Kerberos and SSL, take a backup the Informatica domain truststore and default keystore files.
- Verify that ntpd is synchronized between the Informatica domain node and the cluster nodes.
- The cluster is up and running
- No jobs are running in Enterprise Data Catalog.
- The Catalog Service is up and running
- Enable the read-only mode for the Catalog Service.
- Install the following applications and packages on all nodes before you upgrade Enterprise Data Catalog:
 - Bash shell
 - libtirpc-devel
 - rsync
 - libcurl
 - xz-libs
- The default ports for the Informatica Cluster Service must be available.
- Clear the configured values for the INFA_TRUSTSTORE and INFA_TRUSTSTORE_PASSWORD environment variables if the domain is enabled for Secure Sockets Layer (SSL)

The default ports for the ICS

	Default port
HTTP/HTTPS	9075
Nomad Serf	4648
Nomad HTTP	4646
Nomad RPC	4647
ZooKeeper	2181
ZooKeeper peer	2888
ZooKeeper leader	3888
SOLR	8983
MongoDB	27017
PostgreSQL	5432

Compute Cluster Pre-Requisite

Sudo Privileges

- Commands require sudo: mkdir, chown, chmod, tee, echo, systemctl, cp, mv, tar, find, unzip, sysctl, rm, umount

*Sudo privilege is only required at the first time when we enable ICS. After all services are installed, no sudo privilege is required anymore unless any of service is required re-install/re-start

TLS requirements

- Java JDK needs to be installed. (JRE) version 1.8
- Openssl needs to be installed
- JDK keytool utility is required on PATH
- Openssl is required on PATH

What is all sudo commands used for:

- sudo mkdir, chmod, chown: 1. create all required folders if they doesn't exists, including customDir, allServiceLogDir, allServiceDataDir, allServiceOtherDir, certsDir, tmpDir. 2. set proper permission for all service configuration file, status file, all certs file, env tmp file.
- sudo tee: create status file of ICS
- sudo echo: check sudo privilege in pre-validation part.
- sudo systemctl: dameon-reload, start/stop/enable/restart all service, including nomad, solr, zk, mongod, discoverypostgres, elasticsearch, spark
- sudo tar/unzip: to untar or unzip all service binary file.
- sudo cp/mv: to copy service configuration file to system configuration folder.
- sudo sysctl: only used in ES, sudo sysctl -w vm.max_map_count="262144"
- sudo rm : clean customDir, allServiceLogDir, allServiceDataDir, allServiceOtherDir, /etc/infastatus.conf, tmpWorkingDir, cliInUsrBin, all service configuration file
- sudo umount: umount nomad server job working dir when gc is not executed completely. Only used in clean cluster.
- sudo find: find service related file to clear. Only used in clean cluster.

During Upgrade – Things to Remember

- If Domain is custom SSL then certificate for ICS and CS has to be generated before hand
- If Hadoop gateway user is root then non root user with sudo permission has to be assigned to Infa Cluster Service as gateway user
- Make sure non root user must have read and write permission to Informatica Cluster Service custom directory
- If Advance scanner feature need to be enabled then select advance scanner option during upgrade
- Point Informatica Cluster Service and Catalog Service to new 10.5 Default.keystore (Optional)

Overall Upgrade process

- Get the new license keys for Informatica Cluster Service and Informatica domain. **Due to the migration from IHS to ICS, lic. Key before 10.5 will not work.**
- Back up the Model repository.
- Back up the catalog using infacmd to be used in cases of roll back.
- Back up the catalog using the export utility to restore data post upgrade.
- Disable all application services
- Clean the cluster using the infacmd ihs cleanCluster command.
- Upgrade the Informatica domain.
- Enable and upgrade the Informatica application services, except the Informatica Cluster Service and the Catalog Service.
- Enable the Informatica Cluster Service. (Create Informatica Cluster Service incase of pre upgrade catalog service pointing to External Cluster)
- Enable and Upgrade the Catalog Service.
- Restore the backed-up catalog data using the infacmd.sh LDM migrateContents command.
- Verify the restored content using the infacmd migrateContents -verify command.

Support via Knowledge Base

Use the link to look for community support for an error:

[https://search.informatica.com/KBSearch#t=All&sort=relevancy&f:@athenaproduct=\[Enterprise%20Data%20Catalog\]](https://search.informatica.com/KBSearch#t=All&sort=relevancy&f:@athenaproduct=[Enterprise%20Data%20Catalog])

"IngestCmd - Terminal Error Encountered, Exiting" while using the migrateContents command during data migration in EDC

Description

In Enterprise Data Catalog (EDC), the migrateContents command fails with the following error messages:

```
IngestCmd - Terminal Error Encountered, Exiting
java.lang.RuntimeException: java.lang.RuntimeException: Error resolving
resource: inf_inl ERROR: "Res Delphi_ACM APPRV_LATEST_ROUND 15774
Killed" while performing catalog backup using the export
```

This issue occurs due to a duplicate

Solution

To resolve this issue, use the -re:

Description

In Enterprise Data Catalog (EDC), when the user attempts to perform a catalog backup using the export command, it fails with an error message:

```
Res Delphi_ACM APPRV_LATEST_ROUND 1577425911330000001
```

Solution

To resolve this issue, increase the memory value for the backup.jar utility to the recommended value using the following command:

```
java -Xmx2096m -jar export.jar export
```

Primary Product
Enterprise Data Catalog

Product Version
10.5

Product Component
Install & Upgrade

The CreateContents command fails during Postgres content creation

Description

The CreateContents command fails with the following error message when Enterprise Data Catalog (EDC) is installed on a SUSE Linux Enterprise Server 15.

```
/data/customdir/postgres/install/bin/psql: error while loading shared
libraries: libtinfo.so.5: cannot open shared object file: No such file
or directory
```

Solution

To resolve this issue, install the libncurses5 package using the following command:

```
zypper install libncurses5
```

Primary Product
Enterprise Data Catalog

Product Version
10.5

Product Component
Install & Upgrade

Post Upgrade Task

- Configure the Advanced scanner and set global variables before using it.
- Download the compatible versions of the SAP transports for SAP BW, SAP BW/4HANA, and SAP S/4HANA resources.
- In Catalog Administrator, rerun all the existing resources that contains similarity results to view column similarity results of the previous Informatica version in the current version
- In the resources to be grouped field, select the resources from which you want to view column similarity results

Questions?



Thank You