NewgenONE
Data Science Studio

Release Notes

Version: 1.0 Service Pack 1

Release Date: 15th December 2023

Disclaimer

This document contains information proprietary to Newgen Software Technologies Limited. The user may not disclose or use any proprietary information or use any part of this document without written permission from Newgen Software Technologies Limited.

Newgen Software Technologies Limited makes no representations or warranties regarding any software or the contents or use of this manual. It also specifically disclaims any express or implied warranties of merchantability, title, or fitness for any particular purpose. Even though Newgen Software Technologies Limited has tested the hardware and software and reviewed the documentation, it does not guarantee or imply that this document is error-free or accurate regarding any particular specification. As a result, this product is sold as it is, and the user, the purchaser, is assuming the entire risk as to its quality and performance. Further, Newgen Software Technologies Limited reserves the right to revise this publication and make changes in its content without any obligation to notify any person of such revisions or changes. Newgen Software Technologies Limited authorizes no Newgen agent, dealer, or employee to make any modification, extension, or addition to the above statements.

Newgen Software Technologies Limited has attempted to supply trademark information about company names, products, and services mentioned in this document. The trademarks indicated below were derived from various sources.

Copyright © 2023 **Newgen Software Technologies Ltd.** All Rights Reserved.
No part of this publication may be reproduced and distributed without the prior permission of Newgen Software Technologies Limited.

Newgen Software, Registered Office, New Delhi
E-44/13
Okhla Phase- II
New Delhi - 110020
INDIA
Phone: +91 1146 533 200
info@newgensoft.com

Table of contents

1 About this document 4
1.1 Revision history 4
1.2 Intended audience 4
1.3 Documentation feedback 4
2 Release summary 5
2.1 New features and enhancements 5
2.2 Resolved defects 6
3 Supported platforms 6
4 System requirements 7
4.1 Cloud resources 7
4.2 On-premises resources 7
4.3 Hardware requirements 7
4.4 Software requirements 8
5 Deployment steps 9
6 Rollback plan 9
7 Integration strategy and their dependencies 9
8 Software deliverables 9
9 Containerization support 9
9.1 Advantages of Containerization 9
9.2 Supported environments for Containerization 10
9.3 Container deliverables 10
10 Release artifacts 11
11 Version number 11
12 Build number 11
13 Release dates 12
14 Product development location 12

1 About this document

These release notes provide information on the features, supported platforms, and system requirements of NewgenONE Data Science Studio 1.0 SP1.

1.1 Revision history

Revision date	Description
December 2023	Initial publication

1.2 Intended audience

These release notes are intended for the following:

- Data engineers
- Business analysts
- Citizen and expert data scientists

1.3 Documentation feedback

To provide feedback or any improvement suggestions on technical documentation, write an email to docs.feedback@newgensoft.com.

To help capture your feedback effectively, share the following information in your email:

- Document name
- Version
- Chapter, topic, or section
- Feedback or suggestions

2 Release summary

This marks the initial service pack release of the NewgenONE Data Science Studio platform. It supports both cloud and on-premise deployment. Data Science Studio uses a dedicated environment within Azure cloud, tailored to each customer. On the other hand, on-premise deployment takes place at the customer's site. The application login is facilitated through:

- Microsoft's active directory using LDAP for on-premises deployment
- Azure Active Directory for Azure cloud deployment

User onboarding is handled in bulk, granting appropriate permissions to each user through automated processes.

NOTE:
The subsequent sections provide relevant details pertaining to both cloud and on-premises deployments of the NewgenONE Data Science Studio platform.

2.1 New features and enhancements

The following table lists the new features category and components:

S.No.	Category	Sub-category	Component
1	Model Development – Machine Learning	Classification	Automl Classifier
		Regression	Automl Regressor

The following table lists the enhanced features in NewgenONE Data Science Studio 1.0 SP1:

S.No.	Feature update
1	For dead pipelines, the logs on Job Monitoring screen displays summarized information along with description.
2	Shap is now enabled for composite join, multiple column group by function, multiple column variable derivation, and for ALS.
3	Ability to query salesforce data source.
4	Select column category option is now available in Salesforce and Snowflakes data sources.
5	Snakbar messages stay for longer time now.
6	Ability to delete a scheduled project on production without killing the jobs.
7	Hourly basis project scheduling in case of scheduling with Change Frequency being on Specific Date, is removed from model deployment.
8	Model publish: Validation is added that, on scheduling a pipeline with specific date, the end date should be at least 30 days and 7 days in case of monthly and weekly scheduling respectively.
9	Time zone localization in case of productionization. Pipeline schedule time now reflects as per the user's browser's timezone.
10	In ETL date difference the date format — yyyy/dd/mm is added.
11	Metric changes for classification and regression algorithm in case of hyperparameter tuning.

2.2 Resolved defects

The following table lists the resolved issues in NewgenONE Data Science Studio 1.0 SP1:

S.No.	Issue fix
1	Left side navibar accordion (Parent child), not working properly in Mozilla Firefox.
2	Upload JSON issue — settings are not getting reflected in the pipeline on JSON upload in pipeline options popup.
3	While editing a pipeline, on removing rule engine node from the pipeline, the pipeline doesn't get saved.
4	After updating a model pipeline with updates being — addition of any node for segmentation or data profiling, the shap and pmml options disappear from the sink node.
5	Time Series Missing Value (TSMV) — constant value input field doesn’t take the value entered in the field.
6	Production pipeline doesn't run if the production sign in is different from the model dev sign in, which has published the pipeline for productionization.
7	Group By Function (GBF) with model algorithm node(s) doesn't work, when model pipeline runs multiple times in productionization.
8	While doing model comparison on the pipeline result screen, the time stamp doesn't reflect correctly.
9	The serving pipeline doesn't execute on clicking the run option in the pipeline options popup in the workspace.
10	Go js license configuration — copyright message is appearing on the canvas screen.
11	Not able to save the pipeline with dex and data join nodes in a pipeline.
12	Pipeline count against projects in productionization workspace are not correct.
13	Not able to select the shap tree in productionization.
14	Local sink doesn't work in case of serving pipeline in productionization.
15	In multiclass evaluator, for no best model cases, model results are not available.
16	Evaluator doesn't work with single model algo in model production.

3 Supported platforms

NewgenONE Data Science Studio 1.0 SP1 supports the following environments:

Browser	• Microsoft Edge 113 and later • Chrome 113 and later • Mozilla Firefox 113 and later
Language Support	English (United States)

4 System requirements

This section describes the minimum system requirements for using the NewgenONE Data Science Studio 1.0 SP1 platform.

4.1 Cloud resources

The table below lists the minimum cloud resource requirements:

Service name	Quantity	Tier or SKU
Azure VM	2	Standard E4as v4 (8 vCPUs, 32 GB RAM, 30 GB operating system disk), 256 GB external disk Operating System: Linux (Ubuntu 20.04)
Azure Blob Storage	1	StorageV2 (general purpose V2), Globally Redundant Storage (GRS)

4.2 On-premises resources

The table below lists the minimum on-premises resource requirements:

Resource name	Quantity	Tier or SKU
Machine	2	Operating System: Linux (Ubuntu 20.04), 4 cores, 32 GB RAM, 30 GB operating system disk, 256 GB external disk

4.3 Hardware requirements

The table below lists the minimum and recommended hardware requirements for the VM/Machine hardware requirements:

Hardware	Minimum	Recommended
Processor	4 cores	4 cores
RAM	32 GB	32 GB
Hard Drive	256 GB Hard Disk Drive (HDD) for operating system	256 GB Solid State Drive (SSD) for operating system

4.4 Software requirements

This section lists the minimum software requirements for using the NewgenONE Data Science Studio 1.0 SP1 platform.

- **Operating system on virtual machine**: Ubuntu 20.04
- **Third-party components**:

Third-party component	Freeware/Chargeable/Open Source	Version
Elasticsearch	Open Source/Freeware	6.5
MongoDB	Open Source/Freeware	4.4.x
Apache Tomcat	Open Source/Freeware	10.1
kie-server	Open Source/Freeware	7.52.0
JupyterHub	Open Source/Freeware	3.1.1
Featuretools	Open Source/Freeware	0.23.3
Python	Open Source/Freeware	3.8
nltk	Open Source/Freeware	3.8.1
NumPy	Open Source/Freeware	1.23.0
pandas	Open Source/Freeware	1.5.3
scikit-learn	Open Source/Freeware	1.2.2
torch	Open Source/Freeware	1.13.1
uvicorn	Open Source/Freeware	0.20.0
fastapi	Open Source/Freeware	0.96.1
azure-storage-blob	Open Source/Freeware	12.14.1
Node.js	Open Source/Freeware	14.21.3
npm	Open Source/Freeware	9.8.0
Material UI	Open Source/Freeware	4.11.0
ReactJS	Open Source/Freeware	16.13.1
Highchart	Open Source/Freeware	8.1.1
GoJS	Chargeable	2.1.19
ag-grid	Open Source/Freeware	23.2.0
MySQL	Open Source/Freeware	8.0.33
Express	Open Source/Freeware	4.17.1
Redux	Open Source/Freeware	4.0.5
JDK	Open Source/Freeware	1.8 and 11.x
ODBC	Open Source/Freeware	17
snowflake-connector-python	Open Source/Freeware	3.1.0
simple-salesforce	Open Source/Freeware	1.12.4
axios	Open Source/Freeware	0.19.1

5 Deployment steps

The deployment process is executed through DevOps pipelines within Jenkins. For procedural details on how to perform the deployment, refer to the *NewgenONE Data Science Studio Installation and Configuration Guide*.

6 Rollback plan

Not applicable

7 Integration strategy and their dependencies

Not applicable

8 Software deliverables

Following are the software deliverables:

- Docker images for docker deployment.
- The ZIP packages are used for the application in standard deployment within a non-Docker environment.

9 Containerization support

This section describes the advantages of containerization and provides information on the supported environments and deliverables for the containerization of NewgenONE Data Science Studio 1.0 SP1.

9.1 Advantages of Containerization

The following are the key advantages of the NewgenONE Data Science Studio containerization:

- Reduce time and effort required for environment setup.
- Ensure consistency across the environments including Development, UAT, and Production.
- Auto-scaling of containers.
- Easy distribution and portability.
- Maintain the operational status of applications running within containers.
- Seamless deployment of updates (Rolling Deployment).

9.2 Supported environments for Containerization

The table below lists the combination of the supported environments for the installation of NewgenONE Data Science Studio on MS Azure:

Supported Environment	On-premise		MS Azure	
Container Orchestration	• Docker Compose 2.19.1 or later version • Docker 22.0.4 or later version Following are the supported sub-platforms:		• Docker Compose 2.19.1 or later version • Docker 22.0.4 or later version Following are the supported sub-platforms:	
	Host operating system	Ubuntu 20.04	Kubernetes Worker Nodes Host operating system	Ubuntu 20.04
	Container Base Image	Ubuntu 20.04	Container Base Image	Ubuntu 20.04
	Application Server	Node 14.21.x	Application Server	Node 14.21.x
	Web Server	Nginx 1.24	Web Server	Nginx 1.24
	JDK	OpenJDK 11x	JDK	OpenJDK 11.x
Database Server	MongoDB 4.4.x		MongoDB 4.4.x	
Language	English (United States)		English (United States)	

9.3 Container deliverables

The table below lists the deliverables for the on-prem and MS Azure container environments:

Deliverables	On-premise	MS Azure
Docker Images	• NAI User Instance Web • MongoDB • NAI backend Modelling API service • NAI backend Serving API service • NAI backend SHAP API service • Elasticsearch database • JupyterHub Service • Drools Rule Engine • Nginx	• NAI User Instance Web • MongoDB • NAI backend Modelling API service • NAI backend Serving API service • NAI backend Shap API service • Elasticsearch database • JupyterHub Service • Drools Rule Engine • Nginx • Azure API service
	NOTE: The container images are delivered through private container repositories such as Azure Container Registry and DockerHub.	
Configuration Files	• For Mongo init	• For Mongo init

Deliverables	On-premise	MS Azure
	• For Azure API to update Azure Blob Storage SAS	• For Azure API to update Azure Blob Storage SAS
YAML Files	Docker-compose.yml	Docker-compose.yml
Documents for Deployment and Configuration	NewgenONE Data Science Studio Installation and Configuration Guide for Docker	NewgenONE Data Science Studio Installation and Configuration Guide for Docker

10 Release artifacts

The release artifacts are as follows:

- User Guide
- Release Notes

S. N.	Enclosed	Yes / No / NA
1.	Installation guide (describing installation procedure for the network as well as standalone version; dependency of software on the operating system and other software; structure and paths of various *.*ini* files, minimum software, and hardware requirements)	Yes
2.	New function listing	Yes
3.	Modified functions	NA
4.	Interface modifications	NA
5.	Problem status (the bug IDs and descriptions for all the pending bugs)	NA
6.	COTS components (If any)	NA
7.	Rollback steps (in the case of separate document)	Yes
8.	Deployment steps (in the case of separate document)	Yes
9.	Others (if any, for example - freeware, open-source, and executables).	Yes

11 Version number

NewgenONE Data Science Studio 1.0 SP1

12 Build number

1.0.0 SP1

13 Release dates

Following are the release dates:

- **Product release date** – 15th December 2023
- **Market launch date** – 20th December 2023

14 Product development location

Newgen Software Technologies Limited,
9th Floor, Building No.7, Candor TechSpace,
IT/ITES SEZ of M/s Seaview Developers Pvt. Ltd.,
Plot No.20 & 21, Sector-135,
Noida, Uttar Pradesh – 201304
E-mail: helpdesk@newgensoft.com