NewgenONE
Data Science Studio

Release Notes

Version: 1.0

Release Date: 10th October 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
3 Supported platforms 5
4 System requirements 6
4.1 Cloud resources 6
4.2 On-premises resources 6
4.3 Hardware requirements 6
4.4 Software requirements 7
5 Feature components 8
6 Deployment steps 10
7 Rollback plan 10
8 Integration strategy and their dependencies 11
9 Software deliverables 11
10 Containerization support 11
10.1 Advantages of Containerization 11
10.2 Supported environments for Containerization 11
10.3 Container deliverables 12
11 Release artifacts 13
12 Version number 13
13 Build number 13
14 Release dates 13
15 Product development location 14

1 About this document

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

1.1 Revision history

Revision date	Description
October 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@newgen.co.in.

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 release of the NewgenONE Data Science Studio platform. It can be deployed both on-premises and in the cloud, specifically Azure Cloud. For cloud deployment, it utilizes a dedicated environment within Newgen's managed private Azure cloud, tailored to each customer. On the other hand, on-premises 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.

3 Supported platforms

NewgenONE Data Science Studio 1.0 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 platform.

4.1 Cloud resources

The table below lists the minimum cloud resource requirements:

Service Name	Quantity	Tier or SKU
Azure VM	1	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	1	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:

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 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	8.5
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

5 Feature components

The table below lists the feature components of the Data Science 1.0 release:

S.No.	Category	Sub-category	Component
1	Source	NA	wasb
			MySQL
			AutoFeature
			Snowflake
			Salesforce
			SQL Server
2	Sink	NA	wasb
			Es
			Local Sink
			Snowflake
			Salesforce
			MySQLl
			SQL Server
3	ETL	Single Data Source Operations	Cast Column
			Chunker
			Concat Columns
			Data Difference
			Delete Duplicate Rows
			Date Time Field Extract
			Dummy Column
			Delete Null Rows
			Doc Normalizer
			Drop Columns
			ETL Missing Value
			Expression
			Column Filter
			Lemmatization
			Mapping Null to Zero
			MELT
			NGram
			Column Rename
			Remove Outlier
			Random Sampling
			Sort Data Frame
			Slice String

S.No.	Category	Sub-category	Component
			String Transformation
			Stemming
			Stopword
			Timer Series Missing Value
			Value Map Operation
		Multiple Data Source Operations	Data Join
			Data Union
4	Data Transformation	NA	Data Augmentation
			Group by Function
			Jupyter Notebook
			Variable Derivation
			Rule Engine
5	Data Exploration	Data Exploration	Crosstab
			Dex (Auto and Manual)
			Manual Dex
			Dex in Action
		Data Quality	Data Validation
6	Segmentation	NA	Segment Creation
7	Data Profiling	Structured Data	Bucketizer
			Clip Outlier
			Missing Value
			Normalizer
			One Hot Encoder
			Quantile Discretizer
			String Indexer
			Scaler
		Dimensionally Reduction	PCA
			SVD
		Unstructured Data	Count Vectorizer
			Doc2Vec
			Tf-idf
		Handling Imbalance Data	Smote
8	Model Development	Deep Learning	Multiple Layer Perceptron (MLPC)
		Graph Analytics	Page Rank (PR)
			Triangle Counting (TC)
		NLP	Lda
			Word2vec
		Evaluators	Binomial Classification Evaluation (BNC-Evl)
			Clustering Evaluation (C-Evl)
			Multinomial Classification Evaluation (MNC-Evl)

S.No.	Category	Sub-category	Component
			Regression Evaluator (R-Evl)
			Split Train
	Model Development - Machiner Learning	Clustering	K
			BiK
			GMM
		Collaborative Filtering	ALS
			Affinity
		Time Series	ARIMA
			ARIMAX
			AUTOARIMA
		FP Mining	Frequent Pattern Growth (FPG)
		Survival Analysis	AFT
		Classification	RF-C
			GB-C
			LR-C
			NB-C
			SVM
			DT-C
			Ensembling-C
			Stacking-C
		Regression	LIR
			RF-R
			GB-R
			DT-R
			GLM
			Ensembling-R
			Stacking-R
9	Model Inference	NA	Model View Batch
			Open Scoring API

6 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*.

7 Rollback plan

Not applicable

8 Integration strategy and their dependencies

Not applicable

9 Software deliverables

Following are the software deliverables:

- Docker images for docker deployment and zip packages for standard deployment
- The build for the user interface (UI) in the case of standard deployment.

10 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.

10.1 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).

10.2 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:

<table>
<tr><th>Supported Environment</th><th colspan="2">On-premise</th><th colspan="2">MS Azure</th></tr>
<tr><td rowspan="4">Container Orchestration</td><td colspan="2">• Docker Compose 2.19.1 or later version
• Docker 22.0.4 or later version
Following are the supported sub-platforms:</td><td colspan="2">• Docker Compose 2.19.1 or later version
• Docker 22.0.4 or later version
Following are the supported sub-platforms:</td></tr>
<tr><td>Host operating system</td><td>Ubuntu 20.04</td><td>Kubernetes Worker Nodes Host operating system</td><td>Ubuntu 20.04</td></tr>
<tr><td>Container Base Image</td><td>Ubuntu 20.04</td><td>Container Base Image</td><td>Ubuntu 20.04</td></tr>
</table>

	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)	

10.3 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 Azure API to update Azure Blob Storage SAS	• For Mongo init • 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

11 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

12 Version number

NewgenONE Data Science Studio 1.0

13 Build number

1.0

14 Release dates

Following are the release dates:

- **Product release date** – 10th October 2023
- **Market launch date** – 1st November 2023

15 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