Congratulations! Your project has been successfully on boarded on Enterprise Data Lake (EDL) platform. The application has been setup on EDL SBOX Environment.

**Application Name: SBOX Testing**

**Application Code:** TEST

1. **Access**

## Steps to gain access to the EDL platform and your project workspace:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Steps | Access Area | |  | Detail | How To |
|  | |  |  | |  |
| Step-1 | | Apply access to Sbox Zones | Pre-requisite to gain access to EDL Sbox Linux cluster. This process registers your NT ID with the EDL SBOX cluster. This doesn’t provide access to any business data. | | [Request to get added to EDL SBOX Linux system](https://mykai.jnj.com/index.php/Enterprise_Data_Lake/How_To_Documents/All_Cloudera_EDL_How_To_Documents/How_to_Get_Added_to_a_Zone_(SANDBOX)) |
| Step-2 | | Apply access to the project specific AD groups | This process provides access to the project specific work area on the cluster. Please apply to only the applicable AD Group – Do not apply to all of them. | | [Request to get membership of project Active Directory (AD) group](https://mykai.jnj.com/index.php/Enterprise_Data_Lake/Access_and_Security/Steps_for_Access_to_EDL) |
| **ITS-APP-EDL-DEV-TEST1-DVLP-USR**  (This AD group will be present only the SBOX environment)  **\*81** refers to the allocated project number. | All the Sboxelopers (only individual users; no service accounts) |
| **ITS-APP-EDL-DEV-TEST1-SUPP-USR** | All the support users (only individual users; no service accounts) |
| **ITS-APP-EDL-DEV-TEST1-ETL-USR** | All the ETL (Extract Transform Load) related Service Accounts (only service accounts; no individual users) |
| **ITS-APP-EDL-DEV-TEST1-RPT-USR** | All the Reporting related  related Service Accounts (only service accounts; no individual users) |

1. **Project Work Space**

## Below are the **Folders**/**Databases** that shall be provisioned.

* 1. **Databases**

|  |
| --- |
| **STRUCTURED DATA** |

|  |  |
| --- | --- |
| Databases | Purpose |
| TEST\_STG | The staging database stores raw data extracted from the disparate source data systems. This data needs be cleansed, processed and transformed before presenting to business. This database is not usually exposed to business.  Path : /sbx/edl/test/str/test\_stg |
| TEST\_WRK | The work database is used to store the temporary tables, log tables and error tables that shall be created/used during the data transformation process from STG database to CORE database. This database is not usually exposed to business.  Path : /sbx/edl/test/str/test\_wrk |
| TEST\_CORE | CORE database is where the business ready data, which is processed, enriched and is available for business to perform Analytics and Reporting.  Path: /sbx/edl/test/str/test\_core |

**Note**: Based on the additional business needs from the Application team, additional Database capabilities can be made available for provisioning. Please contact EDL Admin Team for more information.

* 1. **HDFS Folders**

|  |
| --- |
| **HDFS Folders** |

|  |  |
| --- | --- |
| HDFS Folder Location | Purpose |
| /sbx/edl/test/appcode/scripts | **“Scripts”** folder shall hold all the Application related programs/scripts. Application team can chose to create sub folders under this for further consolidation.  **Ex**: Under **“Scripts”** folder one can choose to create, **“Hive”, “Pig”, “Impala, “Java”, “Python”** to consolidate and organize respective scripts. |
| /sbx/edl/test/appcode/configs | **“Configs”** folder shall hold all the configuration related files that are need to run the programs and scripts. Files that consist of the parameter values that can be passed to the programs/scripts present in the **“Scripts”** folder so that they can be more dynamic in nature. |
| /sbx/edl/test/appcode/logs | **“Logs”** folder shall hold all the logs generated by the programs/scripts which are useful to understand/debug the status. Application team needs to implement cleanup process on this folder in order to keep it sane. |

* 1. **Linux Folders**

|  |
| --- |
| **Linux Folders on Edge Node** |

|  |  |
| --- | --- |
| Linux Folder Location (Edge Node) | Purpose |
| /data/data01/sbx/edl/test/appcode/scripts | **“scripts”** folder shall hold all the Application related programs/scripts. Application team can chose to create sub folders under this for further consolidation.  **Ex**: Under **“Scripts”** folder one can choose to create, **“Hive”, “Pig”, “Impala, “Java”, “Python”** to consolidate and organize respective scripts. |
| /data/data01/sbx/edl/test/appcode/configs | **“configs”** folder shall hold all the configuration related files that are need to run the programs and scripts. Files that consist of the parameter values that can be passed to the programs/scripts present in the **“Scripts”** folder so that they can be more dynamic in nature. |
| /data/data01/sbx/edl/test/appcode/logs | **“logs”** folder shall hold all the logs generated by the programs/scripts which are useful to understand/debug the status. Application team needs to implement cleanup process on this folder in order to keep it sane. |
| /data/data01/sbx/edl/test/landing | **“landing”** should be leveraged for storing the Application data temporarily incase App Team cannot directly land on to HDFS. Data in this folder is a temporary stop gap and should be cleared once moved into respective HDFS Folder/Database.  **Ex:** Data from **“MBOX”** or other storage mechanisms can be pulled using **“SFTP”** and stored on this folder temporarily and then moved to HDFS. |

1. **EDL Environment Details**

|  |
| --- |
| Sandbox |
| HUE web UI:  [*https://itsusralsp07931.jnj.com:8888/hue/accounts/login/?next=/hue/*](https://itsusralsp07931.jnj.com:8888/hue/accounts/login/?next=/hue/)  [*https://itsusralsp07930.jnj.com:8888/hue/accounts/login/?next=/hue/*](https://itsusralsp07930.jnj.com:8888/hue/accounts/login/?next=/hue/) |
| Edge Node:  *ITSUSRALSP07948.jnj.com* |
| Oozie Server:  *ITSUSRALSP07928.jnj.com* |
| Oozie Server:  *beeline -u 'jdbc:hive2://itsusralsp07930.jnj.com:10000/default;principal=hive/itsusralsp07930.jnj.com@NA.JNJ.COM'* |
| Impala Daemon’s:   |  |  |  | | --- | --- | --- | | *ITSUSRALSP07930.jnj.com* | *ITSUSRALSP07931.jnj.com* | *ITSUSRALSP07932.jnj.com* | |
| Hue Web UI is a set of web applications that enable you to interact with a CDH cluster. Hue applications let you browse HDFS and work with Hive and Cloudera Impala queries, MapReduce jobs, and Oozie workflows. |
| Edge Node is Linux host with all the gateways installed on it for the Cloudera Hadoop components |