



cQube – Operational Details Document

July 2021

Current Version 3.0

Document released by:
Sreenivas Nimmagadda

Document reviewed by:
Radhika Prabhu

Document acceptance by:
Arvind Gopalakrishnan

Version History

Release Version	Release Date	Covered Topics	Affected Sections
V 1.0	04.06.2020	<ul style="list-style-type: none"> One step Installation Data emission API Role based login for Admin, Report viewer and Emission users Change Password Create New User Student Attendance Report CRC App data Report Semester Assessment Report 	<p>Document was created with the below sections.</p> <ul style="list-style-type: none"> 1. Purpose of this document 2. Background 4. cQube Requirements 6. cQube - Product Installation 7. Data emission process 8. Database Structure 9. cQube NIFI data file processing and data validations 12. Reports
V 1.1	16.07.2020	<ul style="list-style-type: none"> Infrastructure reports with configuration stage implementation API development for downloading log files and files, metrics from S3 buckets Data validations Admin screens Access management using Keycloak 	<ul style="list-style-type: none"> 5. cQube network setup 9.1.4 Data Validation after Ingestion 10. Admin login process & Features 11. cQube release upgradation process 12.2 Scatter plot visualization
V1.2	20.08.2020	<ul style="list-style-type: none"> Diksha data content play visualization Telemetry implementation and telemetry visualization Enhancement of the process scheduler cQube Configurations Keycloak integration automation User management through admin screens 	<ul style="list-style-type: none"> 3. Prerequisites for cQube installation 6.2 cQube - Configurations 10.1.1 Admin - Features - Point 3 10.1.1 Admin - Features - Point 1 12.3 Stacked bar report
V1.2.1	26.08.2020	<ul style="list-style-type: none"> Overall usage count for textbooks, courses in Diksha charts, Metrics for diksha usage count, Diksha report in summary statistics crc report metric issue 	<ul style="list-style-type: none"> 12.4 Table format visualization 12.5 Bar Chart visualization

V1.3	11.09.2020	<ul style="list-style-type: none"> UDISE data map based visualization 	4.1 cQube product prerequisites Updated 6.2 cQube - Configurations Updated 6.3 Workflow Updated 11.3 Configurations at cQube upgradation process 12. Reports Updated
V1.4	30.09.2020	<ul style="list-style-type: none"> Composite report configuration Scatter plot visualization using the composite data 	6.2.1 Enable / Disable the process groups 6.2.5 U-DISE configuration 6.2.6 Composite report configuration 7.1.2 Emission order 7.2 Emission file naming conventions & structure 7.2.1 static File Name Structure 12. Reports 12.5 Bar Chart visualization
V1.4.1	09.10.2020	<ul style="list-style-type: none"> Login page with new design Landing page with new design Percentage representation at Diksha column chart 	
V1.5	16.10.2020	<ul style="list-style-type: none"> UDISE Report - Configuration stage implementation DIKSHA api call Telemetry dashboard implementation PAT report enhancement - addition of time range 	6.2.5 U-DISE configuration
V1.6	06.11.2020	<ul style="list-style-type: none"> PAT Heat Chart Report Separation of Diksha "Usage by Location & Usage by Content Name" Report Teacher course enrolment (Diksha TPD) Responsive code development for Landing page Keycloak code changes when the Postgres restarted Nifi Process Optimization for memory issue Admin - UI Screen updates for the Nifi processor status 	4.2 cQube Limitations 10.1.1 Admin - Features - Nifi scheduler 12.5 Heat Map Report visualization

V1.6.1	13.11.2020	<ul style="list-style-type: none"> Heat map changes for PAT and TPD Reports Cropping of India map to State level map for all Map reports Negative values issue in Nifi processing Addition of PAT and TPD data sources in the admin statistics screen 5,6,7 and 8 grades not displaying defect fix at the download file of PAT map school level Adding the i-icon for all landing page dashboard icons Other UI text alignments for all reports Digital certificate add to the Emission and S3 output file download APIs 	3.1 Subdomain for API certificate based authentication 12.1 Landing page dashboard 12.2 Map based visualization
V1.7	08.12.2020	<ul style="list-style-type: none"> Load balancer changes from ELB to ALB Nifi Scheduler for Daily, Weekly, Monthly and Yearly State specific configurations State specific Static table configurations from UDISE data Diksha TPD Enrolment/Completion reports API Token based authentication Configuration of session out time Configuration of Diksha Columns 	3. Prerequisites for cQube installation - Application Load balancer 3.1 API Token based authentication 6.2.1 State specific configurations 6.2.4 Session time-out 6.2.9 Diksha API Configuration 6.2.10 Diksha Columns Configurations 10.1.1 Admin - Features - Nifi schedulers
V1.8	24.12.2020	<ul style="list-style-type: none"> Progress card Index Enhancements to existing reports <ul style="list-style-type: none"> a. Pagination in heatmap 	6.2.6 Infrastructure configuration 12.8 Progress card Report 12.7 Heat Map Report visualization

		<ul style="list-style-type: none"> b. Display indicator wise metrics in PAT heatmap c. Configuration of infra weights API d. Multiple select for list of courses and districts in TPD course progress report 	
V1.8.1	06.01.2021	<ul style="list-style-type: none"> • Details of School level Information was cleaned • Removed the info card color and added the color coding as a circular marker at the card heading • Tooltip Information was sorted, details were cleaned and categorised. • Static tables were configured as UDISE / State specific • Diksha TPD UUID school_id mapping. 	6.2.11 State/Udise table configuration 11.2 Steps for cQube release Upgradation
V1.9	13.01.2021	<ul style="list-style-type: none"> • Homepage for Progress card. • Teacher attendance map report • Nifi default schedules with different timings • Hyperlink for course level metrics download at Diksha reports 	10.1.1 Admin - Features - Nifi schedulers 12. Reports Updated with Teacher attendance report
V1.10	08.02.2021	<ul style="list-style-type: none"> • New UI design Progress card report • New UI design Composite report • Student attendance exception report • Teacher attendance exception report 	11.2 Steps for cQube release Upgradation 12. Reports Updated with Teacher attendance report

		<ul style="list-style-type: none"> • Periodic Assessment Test exception report • Student attendance time range selection • Diksha Performance optimisation 	
V1.10.1	12.02.2021	<ul style="list-style-type: none"> • Time Selection (Last day, 7 days, 30 days & Overall) implementation for CRC & Teacher Attendance • The file names for downloaded files have been standardized and the names of the headers have been changed • Recommendations on infrastructures in districts/ clusters/ blocks/ schools when the values calculated are lower when compared with the state mean values for the respective infra details. 	11. cQube release upgradation process 11.1 Prerequisites 12. Reports Updated with Teacher attendance, CRC time series. Download naming conventions.
V1.11	03/03/2021	<ul style="list-style-type: none"> • Time selection implementation <ul style="list-style-type: none"> ◦ Teacher exception report ◦ Progress card • Configuration of semester assessment • Data replay for a selected period of time • Download functionality for all months data <ul style="list-style-type: none"> ◦ All the months data in one Excel sheet for Student attendance report ◦ All the months data in one Excel sheet for Teacher attendance report 	6.2.12 Semester Assessment Test (SAT) Configurations 7.2 Emission file naming conventions & structure 7.2.1 static File Name Structure 7.2.2 Transactional File Name Structure 10.1.1 Admin - Features - Added the 7th feature for data replay.

V1.12	31/03/2021	<ul style="list-style-type: none"> • Time selection implementation <ul style="list-style-type: none"> ◦ CRC Report ◦ PAT map based report • School management selection • Trend Line charts for student attendance report • Progress card update - CRC visits in the progress card is displayed based on the 'number of visits' instead of 'no visits' • Nifi upgradation to the latest stable version 	4.1.2 Software Upgradations 6.3 Workflow 12.9 Trends Chart
V1.12.1	22/04/2021	<ul style="list-style-type: none"> • Adding of Total Students, Total schools in PAT Map report & Heatmap tooltip • Adding of the Total Students, Total schools in PAT Map report footer • Adding of the Total Students, Total schools in SAT Map report & Heatmap tooltip • Adding of the Total Students, Total schools in SAT Map report footer • Addition of the Pagination & Search functionality to the table reports • Enabled the school level metrics in CRC Report 	
V1.13	12/05/2021	<ul style="list-style-type: none"> • Periodic assessment query performance optimization • Semester assessment query performance optimization 	7.2.1 static File Name Structure 12. Reports

		<ul style="list-style-type: none"> ● Student attendance query performance optimization ● UI Changes <ul style="list-style-type: none"> ○ Removal of the Diksha TPD teachers percentage report ○ Addition of enrolment and completion number in the tooltip for TPD Total Enrolments /Completion ○ Move the Exception Segment at the very end in the cQube home page dashboards ○ Addition of Color indicators to all the heatmap reports ○ Addition of cell color and tooltip in PAT LO table 	
V1.13.1	20/05/2021	<ul style="list-style-type: none"> ● PAT performance improvement ● Update to PAT LO table with percentage instead of marks ● Defect fixes - composite report and diksha progress report 	5.2 User Actions in creation of Nginx 12. cQube Use-Case Configuration 13. cQube Private Data Center Installation 14. Reports

Release Version	Release Date	Changed Topics	Affected Sections
V 2.0	11.06.2021	<ul style="list-style-type: none"> ● Installation / Upgradation was split into below 2 part as indicated below <ul style="list-style-type: none"> ○ Base Installation/Upgradation ○ Workflow Installation/Upgradation ● Data retention policy ● Data filtration configuration 	<p>Document was created with the below sections.</p> <p>6.1.2 cQube 2.x Installation 10.1.1 Admin - Features - 8. Data retention</p>
V 3.0	21.07.2021	<ul style="list-style-type: none"> ● cQube product divided into two use cases <ul style="list-style-type: none"> ○ Educational Use-case ○ Test Use-case ● Single code base for S3 and in house implementation of cQube ● Configurable NFS/ SFTP for cQube data emission 	

Table of Contents

1. Purpose of this document	14
2. Background	14
2.1 Use of this document	14
2.2 State of this document	14
2.3 Acronyms	14
3. Prerequisites for cQube installation	15
3.1 API Token based authentication	16
4. cQube Requirements	17
4.1 cQube product prerequisites	17
4.1.1 Software Requirements	18
4.1.2 Software Upgradations	19
4.1.3 Data Storage Locations	19
4.1.4 Software security requirements	19
4.1.5 Hardware requirements	20
4.2 cQube Limitations	21
4.2.1 Nifi process scheduler	21
4.2.1 Emission file size and record count	21
4.2.2 Backlog file processing	21
4.2.3 Key cloak Restart when the Postgre restarts	21
5. cQube network setup	22
5.1 Private Subnet	23
5.2 User Actions in creation of Nginx	24
5.3 Public Subnet	25
5.4 AWS Load Balancer	25
5.5 IAM user and Role creation for S3 connectivity	26
6. cQube - Product Installation	27
6.1.1 cQube 1.x Installation	27
6.1.2 cQube 2.x Installation	28
6.2 cQube - Configurations	29
6.2.1 State specific configurations	29
6.2.2 Enable / Disable the process groups	30
6.2.3 Keycloak Two-Factor Authentication	30
6.2.4 Session time-out	31
6.2.5 Nifi query parameter configurations	31
6.2.6 Infrastructure configuration	33
6.2.7 U-DISE configuration	35

6.2.8 Composite report configuration	37
6.2.9 Diksha API Configuration	38
6.2.10 Diksha Columns Configurations	38
6.2.11 State/Udise table configuration	38
6.2.12 Semester Assessment Test (SAT) Configurations	39
6.3 Workflow	40
Data reprocessing instructions:	40
7. Data emission process	41
7.1 The steps involved in the data Emission Process	42
7.1.1 Emission APIs	43
7.1.2 Emission order	43
7.2 Emission file naming conventions & structure:	43
7.2.1 static File Name Structure:	45
7.2.2 Transactional File Name Structure:	47
8. Database Structure	49
8.1 Types of database tables	49
9. cQube NIFI data file processing and data validations	50
9.1 S3 bucket partitioning	51
9.1.1 S3 emission bucket partitions	51
9.1.2 S3 input bucket partitions	51
9.1.3 S3 output bucket partitions	51
9.1.4 Data Validation after Ingestion	51
10. Admin login process & Features	53
10.1 Admin login process:	53
10.1.1 Admin - Features	54
10.2 Report Viewer	59
11. cQube release upgradation process	60
cQube upgradation is the updating of the version of the cQube. As of now cQube allows users to upgrade to the latest version from the previous version only.	60
At the time of a new release users might choose either the installation of the latest version or the upgrade to the latest version.	60
11.1 Prerequisites	61
11.2 Steps for cQube release Upgradation	61
11.3 Configurations at cQube upgradation process	62
11.4 Limitations	63
12. cQube Use-Case Configuration	63
12.1 New Use-Case creation	63
12.2 New use-case integration with cQube	64

13. cQube Private Data Center Installation	65
13.1 Data Storage Locations	65
13.2 Software security requirements	65
13.3 cQube network setup	65
13.4 Private Subnet	66
13.5 Public Subnet	67
13.6 Data emission process	69
14. Reports	70
14.1 Landing page dashboard	70
14.2 Map based visualization	71
14.2.1 Map exception reports	72
14.3 Scatter plot visualization	73
14.4 Stacked bar report	73
14.5 Table format visualization	74
14.6 Bar Chart visualization	74
14.7 Heat Map Report visualization	75
14.8 Progress card Report	75
14.9 Trends Chart	76
(i) cQube common Issues and Resolutions	78
(ii) List of NIFI processor groups & UI Components	78

1. Purpose of this document

The purpose of this document is to describe the operational details of the cQube product. The operational details include description of functionalities like the cQube product installation, admin functionalities and upgradation of the cQube product. Some of the technical details like the cQube network setup and the data emission process have been added to this document to understand the operations clearly.

2. Background

cQube is an analytics and decision making tool for the education system, this product can be used for monitoring the education system in a state and on a broader scale for monitoring the schools across the state/ districts/ clusters/ blocks/ villages and schools.

2.1 Use of this document

This document can be used by anyone who works with the cQube product and whoever has the permission to view the documents of the completed project. This document helps the user understand the design and architecture, as well as the operational and technical details of the cQube.

2.2 State of this document

This document is in a final draft version and is under change control. To request a change to the technical document please contact the system architect or the project manager.

2.3 Acronyms

The following is a list of acronyms which will be used throughout this document:

Table – 1: Acronyms

Acronym	Description
S3	Simple Storage Service
NIFI	Apache NIFI
PK	Primary Key
FK	Foreign key
RDAC	Restricted Database Access Control

NSQF	National Skills Qualification Framework
PGI	Postgraduate Institute
CWSN	Children with special needs

3. Prerequisites for cQube installation

Mentioned below are the prerequisites for the installation of the cQube product:

- The cQube product has to be installed in the AWS environment
- Firstly an instance with Ubuntu 18.04 OS has to be created.
- All the hardware requirements mentioned in section 4.1.4 have to be adhered to
- Before starting installation, the network set-up has to be completed as mentioned in section 5 of this document.

The below table describes the Infrastructure details to install the cQube

Activity	Infrastructure used for cQube installation	Required Skills
1-2 Weeks	2 Days	
Filing of the requirements in the required format and obtaining relevant approvals for procurement of the Infrastructure and securing funding for monthly/yearly spend.	AWS account 1. cQube Server (CPU - 8 Core RAM - 32 GB Storage - 500 GB) 2. S3 Buckets (S3 emission 100GB S3 input - 750 GB S3 Output - 250 GB)	AWS Basic Operation Skills - EC2 - S3 -Application Load Balancer - IAM - VPC - Route53 - Certificate Manager (SSL)
	OpenVPN Access Server (EC2 instance CPU - 1 Core RAM - 2 GB Storage - 10 GB)	VPN admin Operations
	NGINX Reverse proxy(EC2 instance CPU - 2 Core RAM - 4 GB Storage - 30 GB)	NGINX configuration skills
	NAT gateway	AWS NAT gateway
	Ubuntu 18.04	Will be taken care while

		creating EC2 instance
	Java jdk 1.8	Will be installed through One-Step cQube Installation
	Python 3	Will be installed through One-Step cQube Installation
	NIFI 1.11.4	Will be installed through One-Step cQube Installation
	Angular 9.1.6, Chart.js 2.9.3, Leaflet 1.6.0	Will be installed through One-Step cQube Installation
	PostgreSQL 10.12	Will be installed through One-Step cQube Installation

The table below gives the estimated time for the cQube network setup and product installation

S No	Task Details	Estimated Time(Approx.)
1	To create the EC2 Instance with Ubuntu 18.04	5 mins
2	For overall network setup	30 mins - 40 mins
3	cQube Installation	30 mins
4	Data emission for Student attendance, CRC and Semester assessment data	1 min with AWS otherwise it depends on the internet speed
5	Data processing time for Student attendance, CRC and Semester assessment data, Infrastructure, Diksha, UDISE	150 -180 mins

For more details of the data processing time please refer section 6

3.1 API Token based authentication

The cQube emission users and Ad-hoc users should connect with cQube by API token based authentication. The below steps were involved in the token based authentication.

- Admin should create the token from the “create user” page of the admin console while the user creates.
- Admin shares the token to the emission users.
- Users get the API token and configure it in the emission scripts.
- Users emit the files by sending the token as a parameter to each request.
- The same token will be valid forever for the lifetime of cQube usage.

Below is the example with the API token.

```
curl --location --request POST 'https://cqube.tibilprojects.com/data/upload-url' \
--header 'Authorization: Bearer
eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJpc3MiOiJtTFYzbk9KbnhybzFZcDhSY2t4a3BHY1N6RXMy
VVZHMStJ9.mtYc683qSxp2o5nR8t9on3ZV6v6pXqmtvoQKkN3c' \
--header 'Content-Type: application/json' \
--data-raw '{"filename": "upload_test.zip"}'
```

4. cQube Requirements

This section describes the prerequisites to install and configure the cQube product setup and the cQube product setup process. This section also describes the cQube network setup and the setup process

4.1 cQube product prerequisites

The cQube product installation process has a few system prerequisites that have to be followed. The system software, hardware, and security requirements have been derived and have to be adhered to before installation. The figure below gives an overview of the software requirements, security requirements, hardware requirements and the data storage locations:

[Figure – 1: System hardware, software requirements, security requirements and data storage](#)

Software Requirements

Ubuntu 18.04: Supported OperatingSystem for cQube

Java jdk 1.8: Provides environment to install NIFI

Python3: Execution of Ansible scripts and emission API

NIFI 1.11.4 : A central orchestration tool connected to all units from S3 emission data to data visualization.

Angular 9.1.6, chart.js 2.9.3 & Leaflet 1.6.0:To create user reports

Keycloak 10.0.2 : Keycloak will allow single sign-on with Identity and Access Manaqement services to cQube

Security requirements

NIFI 1.11.4: NIFI will be secured by restricting port access

PostgreSQL 10.12 : RDAC, user/role management, row level security implementation

S3 : Access keys and secret keys

Data storage locations

S3 emission Data: The location to store the data files emitted daily.

S3 Input Data: Raw data storage.

S3 Output Data: A source where the processed data resides.

PostgreSQL 10.12 : RDBMS database to store processed data in objects format used by visualization tools

Hardware Requirements

CPU: 8 core

RAM: 32 GB

STORAGE: 1 TB

S3 Input: 500GB (extended based on input data)

S3 Output: 250GB (extended based on metrics file)

4.1.1 Software Requirements

- Ubuntu 18.04: This is the operating system that supports the cQube product
- Java JDK1.8: This provides an environment for NIFI installation
- Python3: Python plays a role in the execution of Ansible scripts and data emission API using a virtual environment.
- NIFI 1.11.4: A central orchestration tool which is connected to all the units and all the different sections where the data flows, starting from the S3 emission data location to the data visualization stage
- PostgreSQL 10.12: An RDBMS database to keep all the processed data in relational data format. The data stored here is used by NIFI to prepare the JSON format files which can be used in the visualization charts.
- Angular 9.1.6, ChartJS 2.9.3, Leaflet 1.6.0, Highcharts 8.2.2: Angular, ChartJS and Leaflet are used to create dashboards/ user reports. The data stored in the S3 output bucket can directly be used to create the reports.

4.1.2 Software Upgradations

- NIFI upgradation: NIFI was upgraded with the latest stable version 1.12.1 which is released on 28th September 2020 by Apache Software Foundation. NIFI 1.12.0 was handled the memory issues

Highlights of the 1.12.1 release include:

- Restored support for keystores containing multiple certs.
- Toolkit generates SAN entries in certs again.
- UnpackContent if unable to determine file permissions just behaves as it used to
- Registry URL saves properly whether a port is present or not.

Below are the reference of the NIFI 1.12.0 & NIFI 1.12.1 releases. A full list of issues that were resolved can be found in the below documents.

For NIFI 1.12.0:

<https://issues.apache.org/jira/secure/ReleaseNote.jspa?projectId=12316020&version=12346778>

For NIFI 1.12.1:

<https://issues.apache.org/jira/secure/ReleaseNote.jspa?projectId=12316020&version=12348757>

4.1.3 Data Storage Locations

- **S3 emission data bucket:** The data emitted from the state education system has to be stored until the NIFI reads the data. S3 emission input storage buckets are the storage locations where the emitted data files are stored.
- **S3 Input data bucket:** This is a location where the raw data resides for all future references.
- PostgreSQL: All the transformed and aggregated data will be stored in PostgreSQL tables.
- **S3 Output data bucket:** A location where the processed data resides in JSON format.

4.1.4 Software security requirements

- cQube product security will be provided by implementing the private subnet with AWS load balancer as described in the section 2 of this document.
- All the ports will be accessed by Nginx server only, So those ports will not be accessed directly from the internet.

- S3 buckets will be secured by the AWS default security.
- PostgreSQL will be secured by the following ways

(The database security will be implemented in the future versions of cQube)

- **RDAC:** Postgres provides mechanisms to allow users to limit the access to their data that is provided to other users. Database super-users (i.e., users who have `pg_user.usesuper` set) silently bypass all of the access controls described below with two exceptions: manual system catalog updates are not permitted if the user does not have `pg_user.usecatupd` set, and destruction of system catalogs (or modification of their schemas) is never allowed.
- **User and Role Management:** the user roles will be granted with one or more cQube database will have three different types or roles:
 1. role role (identified by prefix `r_`)
 2. group role (identified by prefix `g_`)
 3. user role (generally personal or application names)
- **Row Level Security:** In addition to the SQL-standard privilege system available through GRANT, tables can have row security policies that restrict, on a per-user basis, which rows can be returned by normal queries or inserted, updated, or deleted by data modification commands. This feature is also known as Row-Level Security. When row security is enabled on a table all normal access to the table for selecting rows or modifying rows must be allowed by a row security policy.

4.1.5 Hardware requirements

Listed below are the minimum hardware specifications/ requirements in order to install the cQube product

- 8 core CPU
- 32 GB RAM
- 500 GB - 1 TB harddisk
- Three S3 Buckets will be used - for cQube data emission, cQube data input and cQube data output

(Currently, the S3 buckets will be created automatically during installation. This will be revised in future releases and the AWS Admin will have to create the S3 buckets before starting the installation)

4.2 cQube Limitations

4.2.1 Nifi process scheduler

- The nifi jobs are designed for sequential processing, the jobs need to run one by one sequentially.
- Jobs need to be scheduled based on the cpu and memory.
- As of release-1.5 the nifi schedule used to be lost whenever an admin application restart occurred. From release-1.6 this issue has been rectified by maintaining the schedule information in the base directory specified during the installation.

4.2.1 Emission file size and record count

- The emission file size for the PAT report supports a maximum of 20 million records in the dataset and the csv file size was 1.9GB.
- The file emission for student attendance file supports a maximum of 12.5 million dataset and the csv file size was 1.5GB.
- The file emission for semester assessment supports a maximum of 3.5 million dataset and the csv file size was 0.6GB.

4.2.2 Backlog file processing

- There will be a few chances of the backlog getting accumulated in the S3 emission bucket due to huge load or due to failures. The below points to be taken care of in this situation.
- If there is a huge backlog i.e., several files accumulated in S3 emission bucket then it needs to be processed manually one after the other by passing one file at a time.
- The summary statistics screen in the admin console shows the processed file information and the S3 files screen in the admin console shows the unprocessed files in cQube.

4.2.3 Key cloak Restart when the Postgre restarts

- Keycloak needs to be restarted if there is a restart of postgres or in case of any connection drop. This has been handled from cQube side if there is a restart of postgres, keycloak will automatically get restarted.

- Users will not be able to login to the cQube dashboard if there is a connection drop from keycloak to postgres database.
- In case of connection drop the keycloak needs to be restarted by issuing “**sudo systemctl restart keycloak.service**”.
- After restart the users will be able to login back to the cQube dashboard.

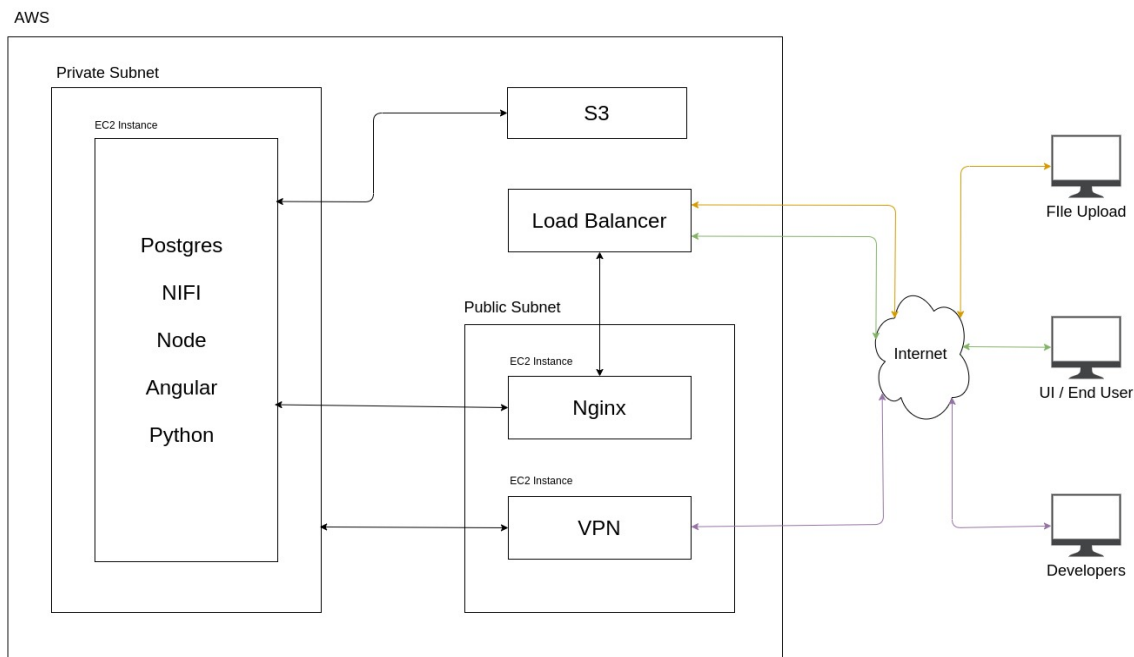
5. cQube network setup

The following steps define how the cQube setup and workflow is completed in AWS. cQube mainly consists of the areas mentioned below:

1. Private Subnet
2. Public Subnet
3. AWS Load Balancer
4. IAM user and Role creation for S3 connectivity.

The cQube network setup process is described in the block diagram below:

Figure – 2: cQube network setup diagram



- The Yellow arrows in the network diagram indicate the file upload users' connectivity through the load balancer.

- The Green arrow in the network diagram indicates the end user's connectivity through the load balancer.
- The purple colour arrow in the network diagram indicates the developers connectivity through the VPN.

5.1 Private Subnet

Private subnet is used to secure the cQube server from public access. The instance will not be having the public ip. An EC2 instance will be created in a private subnet and all cQube components will be installed in this.

Note: For the concurrent users between 100 to 1000, The recommended Nginx machine type would be 'm' or 'c' series 2 core machine to use. And need to increase the machine size according to the concurrent users traffic.

The steps involved to create EC2 instance in private subnet

- Create a virtual private cloud(VPC) in AWS
- Create a subnet in the created VPC with no Routing Table attached to Internet gateway
- Create an EC2 instance to install all the cQube software components.

EC2: cQube server

- create an ec2 instance with below mentioned configuration
- 8 core cpu, 32 gb ram and 1 tb storage
- security_group:
 - port 4200, 3000, 8000 inbound from nginx
 - port 4201, 3001, 9000, 8080, 8096, 22 inbound from openvpn server

Load Balancer:

- create a load balancer
- configure it to connect to nginx on port 80, 443

Domain Name:

- create a domain name
- configure CNAME of load balancer to domain name

SSL:

- create a certificate from AWS Certificate Manager
- attach it to load balancer

Security Group:

- port 80, 443 inbound from 0.0.0.0/0

5.2 User Actions in creation of Nginx

Addition of Gzipping to the UI -

Have to enable the compression in the proxy server for the content type 'application/javascript' and 'text/css'.

Below is the sample configuration [for reference] to add in nginx conf.

```
server {  
    gzip on;  
    gzip_types    application/javascript text/css;  
    gzip_min_length 1000;  
    gzip_static on;
```

Cache implementation -

Have to enable caching in the proxy server for /api location directive.

Below is the sample configuration [for reference] to add in nginx conf.

```
location /api {  
    proxy_cache my_cache;  
    proxy_buffering on;  
    proxy_cache_methods POST;  
    proxy_cache_key "$request_uri$request_body";  
    proxy_ignore_headers Expires Cache-Control X-Accel-Expires;  
    proxy_cache_valid any 30m;  
    proxy_ignore_headers "Set-Cookie";  
    proxy_cache_use_stale error timeout updating http_500 http_502 http_503  
    http_504;  
    add_header X-Proxy-Cache $upstream_cache_status;
```

5.3 Public Subnet

Public subnet will contain two EC2 instances, one is for OpenVPN and another is for Nginx, which will act as a reverse proxy. It is used to provide connectivity with the private subnet.

The steps involved to create the public subnet

- Create a subnet in that same VPC where the private subnet was created with a Routing Table attached to the Internet Gateway.
- Create the first EC2 instance with OpenVPN AWS AMI and configure it to connect with the private subnet.
- Create the second EC2 instance with Ubuntu 18.04 AWS AMI and install Nginx to connect with the cQube server which is in the private subnet.

EC2: Openvpn server

- create an ec2 instance with openvpn ami
 - create users to access the instances and for cQube admin pages
- security_group:
- port 22 inbound to 1 public ip (admin's)

EC2: Nginx server

- create an ec2 instance with ubuntu 18.04 ami
 - configure the nginx to connect Load balancer and cQube server
 - configuration file will be provided
- security_group:
- port 80, 443 inbound to Load balancer
 - port 22 inbound to openvpn server

NAT Gateway:

- create a nat gateway / nat instance
 - configure the connection to cQube server
- security_group:
- port 80, 443 inbound from private_subnet

5.4 AWS Load Balancer

AWS load balancer is used in cQube to avoid the security risks and also to control the traffic among the servers to improve uptime and easily scale the cQube by adding or removing

servers, with minimal disruption to cQube traffic flows. cQube is using the Classic Load Balancer

The steps involved to create the load balancer

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/v2/home>
2. On the navigation bar, choose a Region for your load balancer. Be sure to select the same Region that you selected for your EC2 instances.
3. On the navigation pane, under LOAD BALANCING, choose Load Balancers.
4. Choose Create Load Balancer.
5. For Classic Load Balancer, choose Create.

The steps involved to configure the load balancer with EC2 instance

- On the Configure Health Check page, leave Ping Protocol set to HTTP and Ping Port set to 80.
- For Ping Path, replace the default value with a single forward slash ("/"). This tells Elastic Load Balancing to send health check queries to the default home page for your web server, such as index.html.
- On the Add EC2 Instances page, select the Nginx instance to register with the load balancer.

5.5 IAM user and Role creation for S3 connectivity

An AWS Identity and Access Management (IAM) user is an entity that you create in AWS to represent the person or application that uses it to interact with AWS. A user in AWS consists of a name and credentials. An IAM user with **administrator permissions** is not the same thing as the AWS account root user.

To provide the connectivity between EC2 and S3 we need to create an IAM user with a supported role. The role should have the List, Read and Write permissions

We have multiple ways to create the IAM user account, But in cQube we created the IAM user from the AWS GUI by following the below points.

AWS S3:

- create 3 buckets for input, output and emission

IAM User:

- create an IAM user
- assign iam_policy to user
- download the aws access_key and secret_key

IAM Policy:

- create a policy from AWS IAM
- provide access to list, read and write the object to s3 buckets

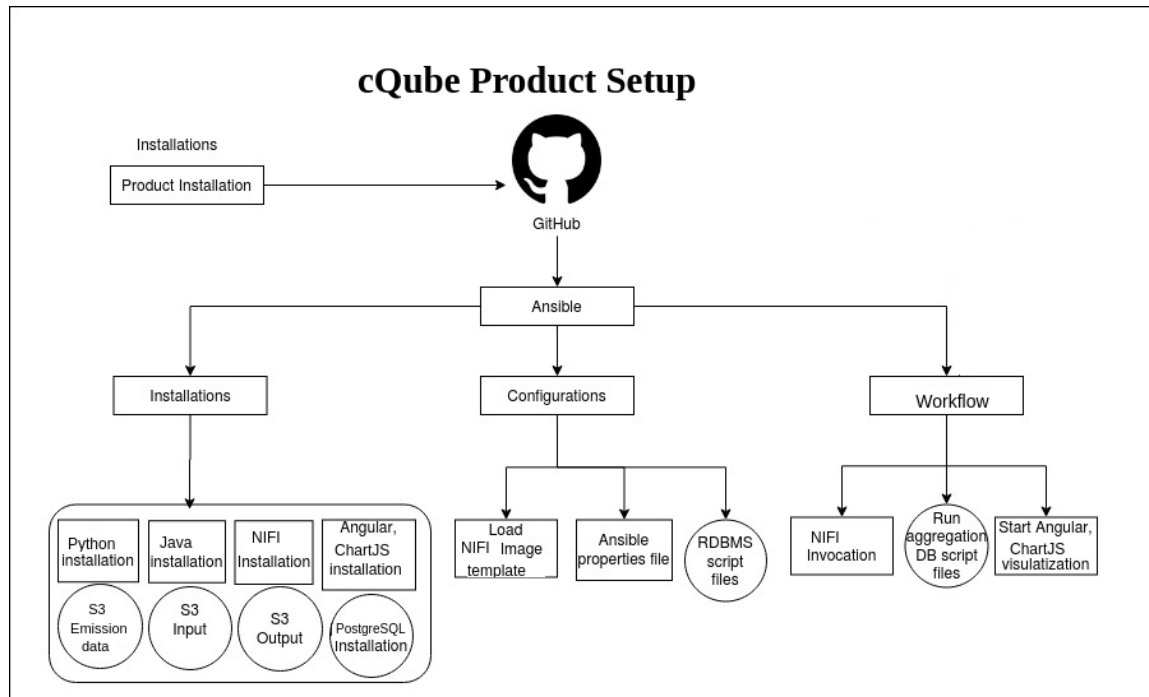
Note: By adding vpc endpoint to connect to s3 buckets, the data transmission between ec2 and s3 happens within aws network. This helps to increase the network speed by 15%.

6. cQube - Product Installation

The diagram below describes the cQube product installation process, the product configuration process and an overview of all the different software that have been used in the cQube tech stack.

6.1.1 cQube 1.x Installation

Figure – 3: Flow diagram for the cQube product setup



The cQube product can be installed with a one-step installation process. The one-step process installs the complete cQube stack, which includes Ansible automation scripts to install Java,

Python, NIFI, Angular, ChartJS, Leaflet, S3 emission data, S3 input bucket, S3 output bucket and PostgreSQL installations.

Steps for Installation:

- As a first step, clone the files from GitHub using the following command:
`$ git clone https://github.com/project-sunbird/cQube.git.`
- This downloads the cQube installation files. The cQube GitHub has all the installation files required for installing cQube.
- The README.md file has all the instructions that have to be followed for the cQube product installation.
- The install.sh script installs the complete cQube stack.
- The Install.sh file calls the Ansible playbooks in the background which will complete the cQube installation setup.
- The complete installation process takes approx 30 minutes.

6.1.2 cQube 2.x Installation

In the cQube 2.x the cQube product can be installed in a two-step installation process. cQube installation package was divided into below 2 components

- cQube base installation
- cQube workflow Installation

cQube base installation

The cQube base installation installs the complete cQube stack, which includes Ansible automation scripts to install Java, Python, NIFI, Angular, ChartJS, Leaflet and PostgreSQL installations. All the softwares will be installed without any data or data processing units.

S3 emission data, S3 input bucket, S3 output bucket and the remaining S/W configurations will be taken care by cQube base installation.

cQube Workflow installation

The cQube Workflow installation is the top-up layer of the cQube base installation. Once the cQube base installation is completed, Admin has to start the cQube workflow installation.

From cQube workflow installation, All the database tables will be created at PostgreSQL using config.sql file and NIFI data processing units will be created by using the NIFI image file which is taken from the github repository.

Below are the github locations for the config.sql & NIFI image

https://github.com/project-sunbird/cQube_Workflow/blob/main/development/postgres/cQube.sql

https://github.com/project-sunbird/cQube_Workflow/tree/main/development/nifi

All the step by step installation is clearly explained in the [README.md](#) file at the below location

https://github.com/project-sunbird/cQube_Base/blob/release-2.0/README.md

6.2 cQube - Configurations

Configurations are done as a part of the installation process by the ansible scripts. All the properties of the sensible information like URLs and passwords are saved in the ansible properties file which is encrypted by default. Configuration is an automated process, no manual interactions are required in this stage.

cQube configurations through Ansible code:

- Nifi template would be uploaded & instantiated with the variables & parameters.
- All the Nifi controllers will be enabled and all the Nifi processors will be started.
- Postgres static, transaction, aggregation tables will be created in the newly created database.

The below configurations should be done in cQube while the installation and the data process.

- Enable / Disable the process groups - At installation time
- Keycloak two factor authentication - Need to verify for the roles Admin and Report viewer to go with 2 factor authentications at the time of login.
- Nifi process configurations - Change the query parameters for Nifi process and verify whether those are affecting the process or not
- Infrastructure configuration - Need to check if the selected infra values are affecting or not.

The above four configurations in explanation,

6.2.1 State specific configurations

This configuration should be performed at the installation stage of cQube. The state will be specified in the config.yml file. The state code will be specified with 2 letters of uppercase characters which are referred from the below file.

https://github.com/project-sunbird/cQube/blob/release-1.7/ansible/installation_scripts/state_list

Only upgrade of 1.7 from 1.6.1 will have the state specific entry where other releases will not avail this feature. Configuration at upgrade of 1.7 will be entered at `upgradation_config.yml`

6.2.2 Enable / Disable the process groups

This configuration should be performed at the installation / Up-gradation stage of cQube. The cQube user may select the process groups which they wanted to include into the cQube. While the installation / Up-gradation of the cQube user will be available to select the existing process group by giving the true / false in the `datasource_config.yml` document which is available at the link below.

https://github.com/project-sunbird/cQube/blob/release-1.2.1/ansible/installation_scripts/datasource_config.yml

All the configuration should be given like below,

```
nifi_telemetry: true
nifi_crc: true
nifi_attendance: true
nifi_semester: true
nifi_infra: true
nifi_diksha: true
nifi_udise: true
nifi_pat: true
nifi_comp: true (If composite report is required)
```

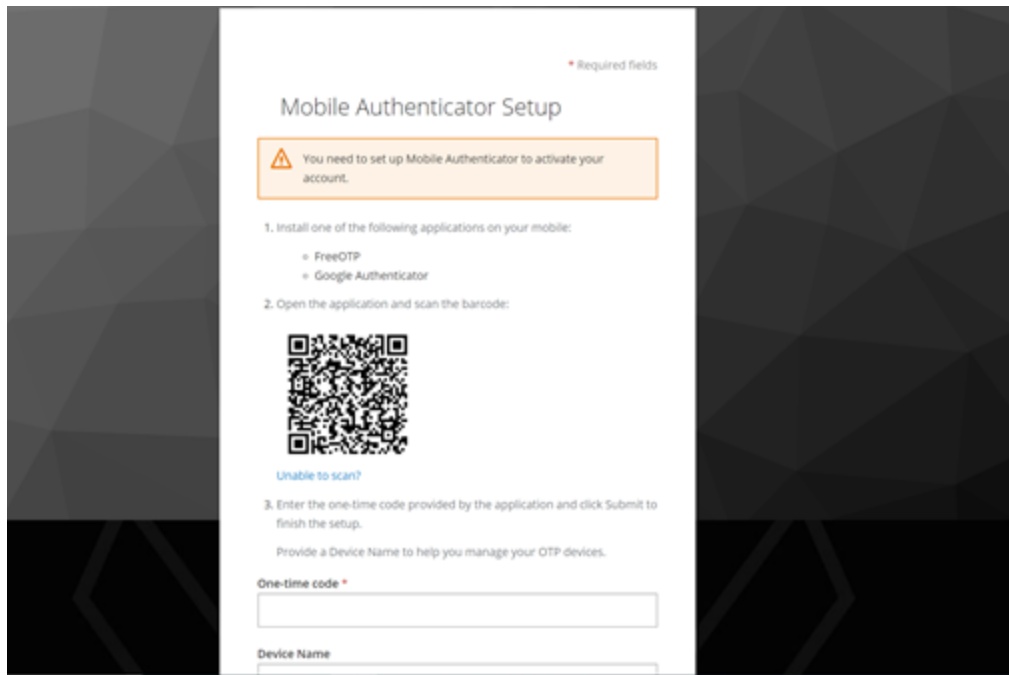
As of now there are a total of 7 processor groups available in cQube.

6.2.3 Keycloak Two-Factor Authentication

Keycloak will allow single sign-on with Identity and Access Management services to cQube.

Keycloak will be configured to cQube at the time of installation in the execution process of `install.sh` command. If two factor authentication is set up the below screen would appear after first time login for all the users. Users need to set up two factor authentication for the first time from his/her mobile using the Google authenticator.

Figure – 4: QR code registration for 2 factor authentication



Below are steps:

- Download Google Authenticator app into your mobile.
- Scan the QR code to integrate cQube dashboard with Google authenticator app.
- Provide the OTP to login.

6.2.4 Session time-out

Session time should be mentioned in the `config.yml` File during the installation / Upgradation. The default time for the session expiry would be 7 days. Users have the facility to decrease the session out time to a minimum of 30 minutes and maximum of 3650 days.

Minutes should be mentioned as 'M' and Days mention 'D'.

Example: 7D for 7 days

30M for 30 minutes

6.2.5 Nifi query parameter configurations

The Nifi configuration process allows us to change the query parameters before the installation process starts. To change the query of a nifi processor, we need to update the query parameter of that processor.

- Query should be re constructed outside the cQube.
- Query parameters should be defined in the configuration files of the nifi processor group.
- Additional filters can be added in the query parameter.
- Cast, round other functions can be used to update the query in the parameter configuration.

- The query in the NIFI process should be replaced with the new Query.
- The Query results should be affected on UI Reports.
- Below are a few examples to add or remove filters & apply functions to the query.

Example:

1. Adding a filter to the existing query for the static processor

By default the static_get_invalid_names parameter in static_data_parameters.txt would be

The Original Query:

```
"static_get_invalid_names": "select school_id,school_name,block_id,district_id,cluster_id from
school_hierarchy_details where cluster_name is null or block_name is null or district_name is null
```

The Updated Query:

```
"static_get_invalid_names": "select school_id,school_name,block_id,district_id,cluster_id from
school_hierarchy_details where cluster_name is null or block_name is null or district_name is null
or school_name is null",
```

Output: By adding the additional filter to not allow the school_name with null records into cQube reports.

2. Updating infra parameter based on active infrastructure attributes

By default the infra_normalize parameter in infra_parameters.txt would be

The Original Query parameter:

```
"infra_normalize": "select school_id ,
case when HaveDrinkingWater <>1 then 0 else 1 end as drinking_water,
case when NoOfToilet=0 or NoOfToilet is null then 0 else 1 end as toilet,
case when HaveCWSNToilet <>1 then 0 else 1 end as cwsn_toilet,
case when HaveElectricity <>1 then 0 else 1 end as electricity,
case when HaveCCTV <>1 then 0 else 1 end as cctv,
case when HaveLibrary <>1 then 0 else 1 end as library from flowfile"
```

The above query parameter can be changed according to the data fields activated by the state.

If only three of the fields choose to be activated by the state the query needs to be updated as below.

The Updated Query parameter:

```
"infra_normalize": "select school_id ,
case when NoOfToilet =0 then 0 else 1 end as toilet,
case when HaveElectricity <>1 then 0 else 1 end as electricity,
case when solarpanel_yn=TRUE then 1 else 0 end as solar_panel from flowfile"
```

We need to update the query in the infrastructure configuration file i.e., `infra_parameters.txt` file after that the installation process can be started. This will map with the infrastructure input data with the cQube database tables.

6.2.6 Infrastructure configuration

cQube is having the flexibility to support multiple state data with minimal changes in UI code for the fluctuated data sources like Infrastructure, Infrastructure score weights, Semester assessment subjects.

Below Steps will be performed to implement the database configuration stage:

- Before cQube is installed, the infrastructure data source needs to be configured depending on the emission data fields required by the state in the infrastructure configuration file.
- Depending on the infrastructure attributes and its datatype the infrastructure query parameter, a case statement is written to map the input data field and the cQube infrastructure table.
- The case statement needs to be updated in the infrastructure configuration file [infra_parameters.txt](#) file by updating the parameter: `infra_normalize`.
- From the below table the `infra_normalize` parameter will look like :

```
"infra_normalize":"select school_id ,
case when NoOfToilet =0 then 0 else 1 end as toilet,
case when HaveElectricity<>1 then 0 else 1 end as electricity,
case when solarpanel_yn=TRUE then 1 else 0 end as solar_panel from flowfile"
```

- Example table :

datatype	input data	infrastructure_master.csv	case statement
integer	NoOfToilet	Toilet	case when NoOfToilet =0 then 0 else 1 end as toilet,
bit	HaveElectricity	Electricity	case when HaveElectricity<>1 then 0 else 1 end as electricity,
boolean	Solar Panel	Solar Panel	case when solarpanel_yn=TRUE then 1 else 0 end as solar_panel,

Note :- For infrastructure_master column the values should be converted to lowercase and spaces are converted to underscore(_) example : Solar Panel converted to solar_panel

- Based on the configuration, the changes are made to handle state specific data fields
- After successful validations, data tables are created with only active and required data fields for the active data fields are processed and metrics are generated
- The metrics generated will be stored in JSON files, which is used for visualization

- Change only the score & status of the infrastructure in [infrastructure_master.csv](https://github.com/project-sunbird/cQube/blob/release-1.8/development/postgres/infrastructure_master.csv), don't update any other column values.
- Make sure the sum of score of all infrastructure is 100 (For status=1(active))

To select the infrastructure fields, please fill the details in infrastructure_master.csv file which is available at the link below.

https://github.com/project-sunbird/cQube/blob/release-1.8/development/postgres/infrastructure_master.csv

The Infrastructure calculation will happen as in the below example.

If a school has drinking water, handwash, electricity, toilet, playground, hand pumps, library then the score would be calculated as below

$$20*1 + 10*1 + 10*1 + 20*1 + 20*1 + 10*1 + 10*1 = 100$$

The total infrastructure score would be 100, if the school does not have any infrastructure available, then for that infrastructure it would be awarded with 0.

If the school does not have playground, handpump, library, the score would be calculated as

$$20*1 + 10*1 + 10*1 + 20*0 + 20*1 + 10*0 + 10*0 = 60$$

The total infrastructure score would be calculated to 60.

If we need to calculate the score at different levels such as district, block, cluster, the school infrastructure count would be added for all the blocks, clusters, districts.

To get the infrastructure score at district, block, cluster add all the infrastructure of all the schools available.

Weight * (infrastructure of all the school/ Total number of schools)

For example, if there are two schools in cluster the infrastructure score would be calculated as

$$20*(1+1)/2 + 10*(1+0)/2 + 10*(1+1)/2 + 20*(0+0)/2 + 20*(1+1)/2 + 10*(0+0)/2 + 10*(0+0)/2 = 55$$

Infrastructure score for the cluster would be 55.

The Metrics will be stored in the JSON files for infrastructure visualization. An example JSON document for the district infrastructure is given below:

```
{
  "district": {
    "id": 2401,
    "value": "Kachchh",
    "block": {
      "value": "Abdasa",
      "id": 240107,
      "infra_score": {
        "value": "79.00",
        "average": {
          "value": "170",
          "percent": "79.81",
          "total_schools": {
            "value": 213,
            "total_schools_data_received": {
              "value": "213",
              "hand wash": {
                "value": "213",
                "percent":
```

```
"100.00"},"solar_panel": {"value": "203","percent": "95.31"},"library": {"value": "210","percent": "98.59"},"drinking_water": {"value": "13","percent": "6.10"},"tap_water": {"value": "13","percent": "6.10"},"hand_pumps": {"value": "158","percent": "74.18"},"playground": {"value": "159","percent": "74.65"},"news_paper": {"value": "213","percent": "100.00"},"digital_board": {"value": "3","percent": "1.41"},"electricity": {"value": "207","percent": "97.18"},"toilet": {"value": "213","percent": "100.00"},"boys_toilet": {"value": "213","percent": "100.00"},"girls_toilet": {"value": "213","percent": "100.00"}}
```

Steps to be taken in case of upgradation:

1. In case of upgradation the infrastructure score and status needs to be updated in the cQube/development/postgres/infrastructure_master.csv as per the existing active fields and weights configured in the previous release.

6.2.7 U-DISE configuration

cQube is having the flexibility to support configuring multiple indices and its metrics with minimal changes in UI code for udise indices and metrics, udise score weights and there are 32 input data files updated in 7.2 section of this document that need to be emitted to visualize the UDISE report.

Below Steps needs to be performed to implement the UDISE configuration stage:

- Before cQube is installed, UDISE data source needs to be configured depending on the emission data fields required by the state in the udise configuration file.
- During configuration users can activate/deactivate the indices and metrics status and their corresponding weights, but not the key columns (ex : id,description,column,type,indice_id should not be modified/edited). Only fields like status and score are updated based on the active indices,metrics and their weights.
- In configuration stage fields will be activated/deactivated for that state, based on requirement
- We can also create our own indices and metrics during the udise configuration stage.
- There is an [exhaustive list](#) of calculated metrics available by default in cQube which is calculated from the UDISE raw input tables.
- The [exhaustive](#) calculated metrics can be used for creation of a normalised metric.
- We can create any number of normalised metrics from the exhaustive list by choosing to add any of the calculated metrics.
- The normalised metric can be defined under

1. Newly created index with new normalized metrics
 2. Newly created index with few new normalized metrics and few existing normalized metrics with different name and metric_id.
 3. Existing index with new normalized metrics
- After successful validations the active metrics and indices are generated.
 - The metrics/indices generated will be stored in JSON files, which is used for visualization
 - There are 3 types of directions and they are No, Forward, Backward, based on the metric the direction can be configured. Based on this normalization takes place between 1 to 0 for backward metrics and between 0 to 1 for forward metrics.
 - If the schools are not having any particular metrics, the total weights(denominator) are considered for the available schools. Here is an [example](#) in the Metric level configuration sheet.

Check [udise_config_example](#) sheet for examples of above three possible configurations.

To select the Udise indices,metrics please fill the details in the udise_config.csv file which is available at the link below.

https://github.com/project-sunbird/cQube/blob/release-1.5/development/postgres/udise_config.csv

Note :-

1. For providing weights the sum of score of all active indices should be 100
2. And also the score of all active metrics of each active indices should be 100
3. Change only **status** and **score** column from the [udise_config.csv](#) file, depending on the use case.
4. While opening [udise_config.csv](#) file in excel , please use '|' as a delimiter.
5. For newly created metrics/indices the field metric_config is updated as value 'created' in udise_config.csv file to differentiate the existing static metrics/ indices with newly created metrics/indices

The indices,metrics normalization and calculations will happen as in the below example.

If a state selected community participation and medical index indices and their metrics as below

id	description	column	type	indice_id	status	score
3000	community participation	Community_Participation	indice		1	60
3001	% SMC members provided training	cp_smc_members_training_provided	metric	3000	1	40

300	2	Total meetings held by SMC	cp_total_meetings_held_smc	metric	3000	1	40
300	3	SMDC in school	cp_smdc_school	metric	3000	1	20
700	0	medical index	Medical_Index	indice		1	40
700	1	Medical Check-up Conducted	med_checkup_conducted	metric	7000	1	60
700	2	De-worming Tablets	med_deworming_tablets	metric	7000	1	20
700	3	Iron Tablets	med_iron_tablets	metric	7000	1	20

The total infrastructure score will be calculated using below formulas :

1. Before start doing calculation, we will normalize the data by grouping them in quartiles as in the following document ,[udise metric calculation logic](#)
2. Community participation (cp)= $(cpm1*40)/100+(cpm2*40)/100+(cpm3*20)/100$
3. Medical index (med)= $(medm1*60)/100+(medm2*20)/100+(medm3*20)/100$
4. infra_score= $(cp*60)/100+(med*40)/100$

6.2.8 Composite report configuration

cQube has the facility to combine the information of all reports and showing it in a single scatter plot and such report is called Composite report, Composite report is the combination of multiple reports which is used to co-relate the information between data sources in school, cluster, block, district level. Example : We can compare the attendance and semester performance of the school in a single scatter report.

1. This configuration should be performed at the installation / Up-gradation stage of cQube.
2. The user may select the process groups which they wanted to include into the cQube (As in [6.2.1](#)) . The metrics available in the selected processor groups will be used in the composite report.
3. If composite report is required to be enabled, make sure the [nifi_comp: true] processor group is enabled in [composite_enable](#) similar to as in [6.2.1](#)

6.2.9 Diksha API Configuration

Diksha configuration for progress-exhaust & summary-rollup dataset:

We have integrated Diksha progress-exhaust & summary-rollup dataset through the API with cQube. During the installation/upgradation user needs to configure the Diksha production base_url,token, encryption key to progress-exhaust & summary-rollup datasets. We need to configure the cQube/development/python/cQube-raw-data-fetch-parameters.txt file before upgradation or installation process with the production parameters.

Diksha Summary-rollup dataset:

The diksha summary-rollup dataset contains the course & textbook usage data and daily files are downloaded from the Diksha API configured in the above step directly to cQube s3 emission bucket. Once the summary-rollup CSV file is downloaded to the emission bucket the data processing takes place and the visualization reports will populate.

Diksha progress-exhaust dataset:

The diksha progress-exhaust dataset contains the course enrollment & completion data. This dataset is downloaded based on the batch_id passed to the diksha API. The list of batch_id's needs to be emitted in a CSV file through the emission API. All the data related to batch_id's are downloaded to the emission bucket. Once the progress-exhaust zip file is downloaded it is unencrypted using the configured encryption password and stored into an emission bucket. The data processing takes place and the visualization reports will populate.

6.2.10 Diksha Columns Configurations

Diksha columns will be configured at the Installation / Upgradation stage. The summary-rollup dataset has two columns 1. content_gradelevel 2. collection_gradelevel. If the columns are available in the summary-rollup dataset exhaust, the "diksha_columns" parameter needs to be true in the config.yml & upgradation_config.yml else if the columns are not available the diksha_columns parameter needs to be false.

If the columns are true,Diksha "Usage by user profile" report will be enabled from the dashboard else the report will be disabled if the configuration value is false at config.yml & upgradation_config.yml.

6.2.11 State/Udise table configuration

Static table configuration (Udise/State tables): During the installation Udise/State specific tables can be configured.

The static table configuration is enabled for below tables

1. district_master
2. block_master
3. cluster_master
4. school_master

Based on the availability of state/udise tables, accordingly it needs to be configured in config.yml before installation.

Once after installation, if the configuration needs to be changed, the user can reinstall cQube.

All the data will be lost if the user reinstall's cQube. the aggregated table data & s3 output files will be wiped off.

Users need to remit the data files to generate the aggregate tables & s3 output files.

During the upgrade the configuration cannot be performed. Only for release-1.8.1 the configuration has been enabled to change from state to udise and vice versa. For subsequent release this option would be disabled.

6.2.12 Semester Assessment Test (SAT) Configurations

SAT configuration will happen from the emission file.

There are 2 files which need to be emitted to configure the Semester Assessment Test, they are semester_exam_mst, semester_exam_qst_mst.

The semester_exam_mst file needs to be updated with the following details: exam_id, assessment_year, standard, subject, exam_code, total_questions, total_marks & semester information.

The other file is semester_exam_qst_mst and it needs to be updated with question_id, exam_id, question_title, question_marks, indicator_id, indicator_title, indicator.

Please refer to the [emission fields document](#) for the fields details semester_exam_mst, semester_exam_qst_mst.

Users can emit these 2 files with incremental load whenever there is a new exam conducted. Once a user emitted these files, the user can emit the marks scored by each student for all the schools for newly conducted exam's.

Nifi will process the data based on the semester_exam_mst & semester_exam_qst_mst and the data file to aggregation tables and S3 output files.

6.3 Workflow

cQube workflow will be done in two parts

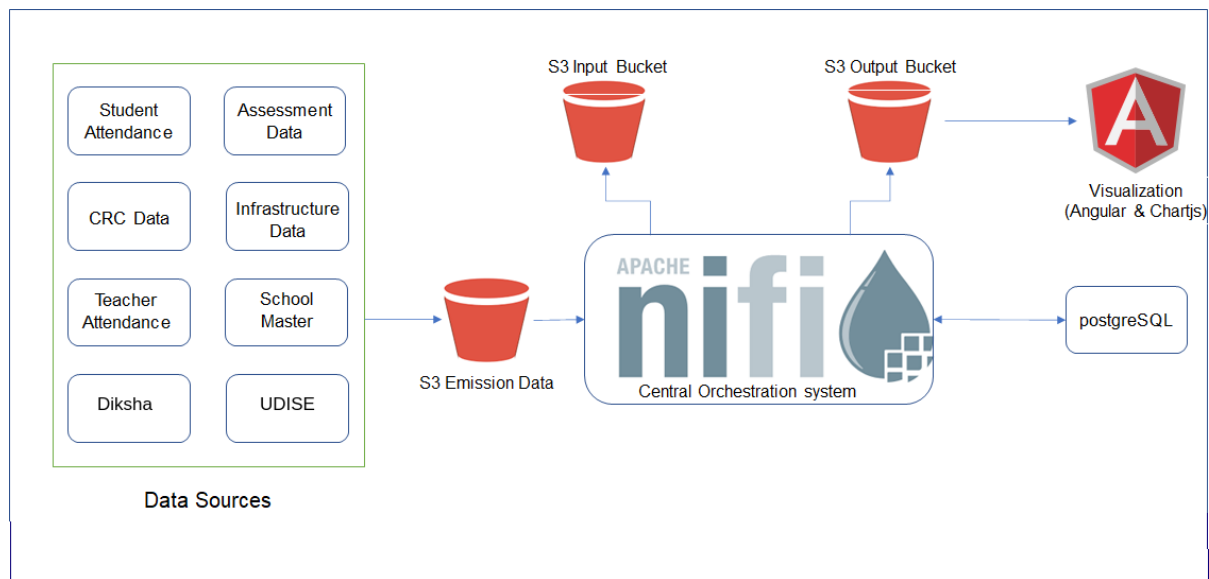
1. Data emission
2. Data processing

Data emission: Data from the state education system will be emitted into the S3 emission data folder using the cQube data emission API
emission API is described in the section 7

Data processing:

NIFI is the central orchestration system which handles the complete workflow process and generates JSON files. These JSON files are used for creation/ generation of the visualization reports. Data from the state education system will be emitted into the S3 emission data folder using the cQube data emission API. The below image will give more clarity on the data processing.

Figure – 5: cQube Architecture diagram



Data reprocessing instructions:

- Order of emission is school_master, school_management_master and management names are updated in the school_hierarchy_details.
- The data from the school_hierarchy_details is used as reference for updating the aggregation tables of all the reports. Whenever the particular data source is processed the

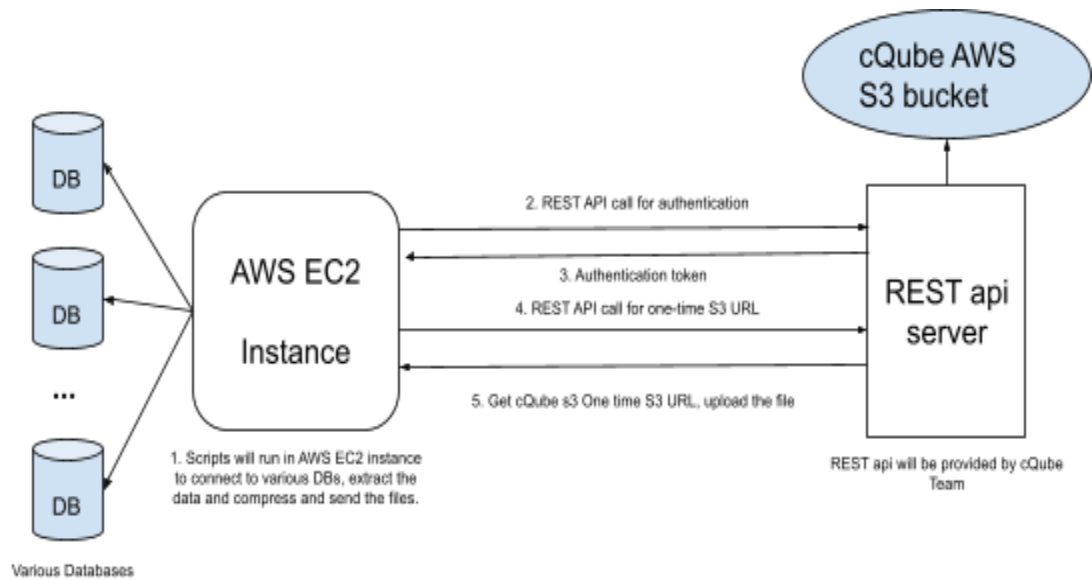
metrics would be generated based on the school management at school, block, cluster, district level.

- If any of the below data issues exists, they need to be rectified before data processing. If there is special character ']' in school name column in following tables school hierarchy table, periodic_exam_school_result, Periodic_exam_school_qst_result, semester_exam_school_result, semester_exam_school_qst_result. They need to be removed before the data processing for the particular data source.
- PAT & SAT aggregation tables shouldn't have the year(2020 or 2021) in the academic_year column. They should have (2020-21,2019-20).
- District names in diksha_total_content, diksha_content_trans(derived_loc_district) should be aligned with school_hierarchy_details. The explicit name handling for the district_name has been removed and districts which don't match will be in 'others'.
- The Infrastructure fields needs to be correctly updated in the infra_normalize parameter in infrastructure_parameters.txt & diksha parameters in cQube-raw-fetch-parameters.txt
- Diksha tpd workflow has been optimized there by reducing the processing time.
- Emission users have to emit the school_grade_enrolment and subject_details for PAT & SAT exception reports.

7. Data emission process

cQube provides a Rest API for the data ingestion. Authenticated API call provides onetime S3 URL and emits the data into the S3 emission bucket.

Figure – 6: Data emission process diagram



7.1 The steps involved in the data Emission Process

- The data emission users will work from the data source location.
- Data emission users develop the automation extraction process which is outside cQube to extract the data from the sources.
- Data emission will be performed periodically as per the specified time interval from different data sources with the help of the above automated extraction process.
- Once the automation extraction process extracts the data fields which are required by cQube(Mentioned in section 4.2), it is converted into csv pipe delimited formatted files.
- The CSV data files will then be placed into the state data center by the automation extraction process.
- Emission process will be invoked by the automation extraction process once the extraction is completed.
- The emission process will be authenticated, An emission API will be invoked by authenticating the API token.
- The emission process makes an API call to generate AWS S3, one time presigned URL.
- API calls with the data file as a parameter to emit the data files using the https protocol into cQube.
- The emission details will be captured by the log files and will be available to the admin for review.

7.1.1 Emission APIs

- Data into cQube can be ingested through emission API.
- After extracting the data from the source, save the file with the data source name. Below are the naming conventions.
- For configuring the emission job, fill the details in config.py like emission URL, AWS token, file path which is available in the configuration file at below location [cQube/development/python/config.py](#)
- After configuration to emit the files into the cQube system, invoke the [cQube/development/python/client/client.py](#) program.

7.1.2 Emission order

- Static files which are defined in section 7.2.1 in this document need to be emitted first.
- For the CRC report, the inspection_master needs to be emitted prior to user_location_master to visualize the CRC reports.
- Similarly for PAT (Periodic assessment report), periodic_exam_mst and periodic_exam_qst_mst needs to be emitted prior to periodic_exam_result_trans

7.2 Emission file naming conventions & structure:

- All the files should be in a CSV format with PIPE("|") separated.
- The data files should be emitted individually in zip format.
- The list of columns which have to be emitted for a data source are provided in the [file](#).
- The header should be in the same order as described in the [file](#), for that respective data set.
- Data emission overview of data flow can be viewed [here](#)
- Below is the list of data sources used in various reports.

Source Name	Used in report	cQube Data Source Name
District Master	All reports	district_master
Block Master	All reports	block_master
Cluster Master	All reports	cluster_master
School Master	All reports	school_master
CRC Inspection Master	CRC Report	inspection_master
CRC User Location Master	CRC Report	user_location_master

Student Attendance	SAR Report	student_attendance
Infrastructure Master	Infrastructure Report	infra_trans
Semester	SEMESTER Report	semester
Diksha	DIKSHA	diksha
School Facility	UDISE	sch_facility
NSQF Class 12 Placement	UDISE	nsqf_plcmnt_c12
NSQF Class 10 Placement	UDISE	nsqf_plcmnt_c10
NSQF Class Condition	UDISE	nsqf_class_cond
School Exam Result Class 12	UDISE	sch_exmres_c12
School Exam Result Class 10	UDISE	sch_exmres_c10
School Exam Result Class 5	UDISE	sch_exmres_c5
School Incentives CWSN	UDISE	sch_incen_cwsn
School Enrollment by stream	UDISE	sch_enr_by_stream
School Enrollment by CWSN	UDISE	sch_enr_cwsn
School Enrollment by Medium of Instruction	UDISE	sch_enr_medinstr
School Enrollment by Age	UDISE	sch_enr_age
School Enrollment by Report	UDISE	sch_enr_reptr
School Enrollment by Fresh	UDISE	sch_enr_fresh
School Enrollment by New Admission	UDISE	sch_enr_newadm
School staff position	UDISE	sch_staff_posn
School Exam Class 10 marks	UDISE	sch_exmmarks_c10
School Exam Class 12 marks	UDISE	sch_exmmarks_c12
NSQF Exam Class 10 Results	UDISE	nsqf_exmres_c10
NSQF Exam Class 10 Results	UDISE	nsqf_exmres_c12

NSQF Training Provision	UDISE	nsqf_trng_prov
NSQF Faculty	UDISE	nsqf_faculty
School Exam class 8 results	UDISE	sch_exmres_c8
School Profile	UDISE	sch_profile
Teacher Profile	UDISE	tch_profile
School Receipt Expenditure	UDISE	sch_recp_exp
NSQF Basic information	UDISE	nsqf_basic_info
NSQF Enrollment by Caste	UDISE	nsqf_enr_caste
NSQF Enrollment by Sub section	UDISE	nsqf_enr_sub_sec
School PGI Indicators	UDISE	sch_pgi_details
School Safety	UDISE	sch_safety
School Incentives	UDISE	sch_incentives
Periodic Exam Master	Periodic Exams	periodic_exam_mst
Periodic Exam Question Master	Periodic Exams	periodic_exam_qst_mst
Periodic Exam Results	Periodic Exams	periodic_exam_result_trans
Semester Exam Master	Semester Exams	semester_exam_mst
Semester Exam Question Master	Semester Exams	semester_exam_qst_mst
Semester Exam Results	Semester Exams	semester_exam_result_trans

Naming conventions of the files and the file structure should be followed as mentioned below:

7.2.1 static File Name Structure:

cqube_emission

```

|
|— block_master
|   └─ block_mst.zip
|       └─ block_mst.csv
|— cluster_master
|   └─ cluster_mst.zip

```

```

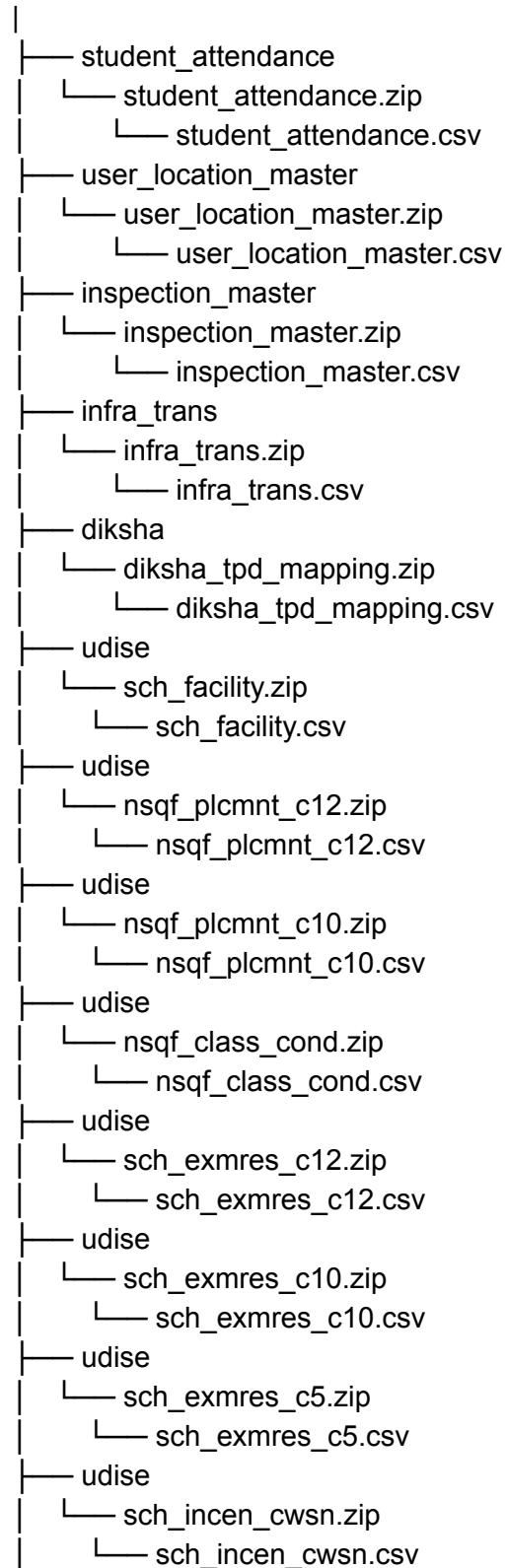
|      └─ cluster_mst.csv
|─ district_master
|   └─ district_mst.zip
|      └─ district_mst.csv
|─ school_master
|   └─ school_mst.zip
|      └─ school_mst.csv
|─ infrastructure_score
|   └─ infrastructure_score.csv
|─ pat
|   └─ periodic_exam_mst.zip
|      └─ periodic_exam_mst.csv
|─ pat
|   └─ periodic_exam_qst_mst.zip
|      └─ periodic_exam_qst_mst.csv
|─ sat
|   └─ semester_exam_mst.zip
|      └─ semester_exam_mst.csv
|─ sat
|   └─ semester_exam_qst_mst.zip
|      └─ semester_exam_qst_mst.csv
|─ sat
|   └─ semester_exam_subject_details.zip
|      └─ semester_exam_subject_details.csv
|─ sat
|   └─ semester_exam_grade_details.zip
|      └─ semester_exam_grade_details.csv
|─ pat
|   └─ periodic_exam_subject_details.zip
|      └─ periodic_exam_subject_details.csv
|─ pat
|   └─ periodic_exam_grade_details.zip
|      └─ periodic_exam_grade_details.csv
|─ school_category
|   └─ school_category_master.zip
|      └─ school_category_master.csv
|─ school_management
|   └─ school_management_master.zip
|      └─ school_management_master.csv

```

Note: school_grade_enrolment file can be emitted either as periodic_exam_grade_details or semester_exam_grade_details. The data will be updated to school_grade_enrolment and will be used in school_hierarchy_details. The count from school_hierarchy_details will be used in pat & sat reports. Before the sat & pat workflow the school_grade_enrolment needs to be emitted.

7.2.2 Transactional File Name Structure:

cqube_emission



- └─ udise
 - └─ sch_enr_by_stream.zip
 - └─ sch_enr_by_stream.csv
- └─ udise
 - └─ sch_enr_cwsn.zip
 - └─ sch_enr_cwsn.csv
- └─ udise
 - └─ sch_enr_medinstr.zip
 - └─ sch_enr_medinstr.csv
- └─ udise
 - └─ sch_enr_age.zip
 - └─ sch_enr_age.csv
- └─ udise
 - └─ sch_enr_newadm.zip
 - └─ sch_enr_newadm.csv
- └─ udise
 - └─ sch_enr_reptr.zip
 - └─ sch_enr_reptr.csv
- └─ udise
 - └─ sch_enr_fresh.zip
 - └─ sch_enr_fresh.csv
- └─ udise
 - └─ sch_staff_posn.zip
 - └─ sch_staff_posn.csv
- └─ udise
 - └─ sch_exmmarks_c10.zip
 - └─ sch_exmmarks_c10.csv
- └─ udise
 - └─ sch_exmmarks_c12.zip
 - └─ sch_exmmarks_c12.csv
- └─ udise
 - └─ nsqf_exmres_c10.zip
 - └─ nsqf_exmres_c10.csv
- └─ udise
 - └─ nsqf_exmres_c12.zip
 - └─ nsqf_exmres_c12.csv
- └─ udise
 - └─ nsqf_trng_prov.zip
 - └─ nsqf_trng_prov.csv
- └─ udise
 - └─ nsqf_faculty.zip
 - └─ nsqf_faculty.csv
- └─ udise
 - └─ sch_exmres_c8.zip

```

|   └─ sch_exmres_c8.csv
├─ udise
|   └─ sch_profile.zip
|       └─ sch_profile.csv
├─ udise
|   └─ tch_profile.zip
|       └─ tch_profile.csv
├─ udise
|   └─ sch_recp_exp.zip
|       └─ sch_recp_exp.csv
├─ udise
|   └─ nsqf_basic_info.zip
|       └─ nsqf_basic_info.csv
├─ udise
|   └─ nsqf_enr_caste.zip
|       └─ nsqf_enr_caste.csv
├─ udise
|   └─ nsqf_enr_sub_sec.zip
|       └─ nsqf_enr_sub_sec.csv
├─ udise
|   └─ sch_pgi_details.zip
|       └─ sch_pgi_details.csv
├─ udise
|   └─ sch_safety.zip
|       └─ sch_safety.csv
├─ udise
|   └─ sch_incentives.zip
|       └─ sch_incentives.csv
├─ pat
|   └─ periodic_exam_result_trans.zip
|       └─ periodic_exam_result_trans.csv
├─ sat
|   └─ semester_exam_result_trans.zip
|       └─ semester_exam_result_trans.csv

```

8. Database Structure

This section describes the database structure, the types of database tables and the naming conventions

8.1 Types of database tables

The data tables are divided into static tables and transactional tables. All the data tables can be divided under these four table types:

1. Static Tables
2. Hierarchical Tables
3. Dynamic Tables
4. Aggregated Tables

All the four types of tables are updated in the data dictionary below.

Click on the link below for the data dictionary:

https://github.com/project-sunbird/cQube/blob/release-1.11/documents/cQube_data_dict.pdf

Click on the link below for the latest ERD:

<https://github.com/project-sunbird/cQube/tree/release-1.11/documents/ERD>

9. cQube NIFI data file processing and data validations

This section describes the cQube data file processing calculations and data validations. The files should be loaded into S3 emission bucket as per the emission process schedule timings.

The file processing will be controlled by NIFI and the process will be run as batch mode within the scheduled timing. By default each data source is having different schedule timings which is referred to in [10.1.1 Admin - Features](#) - Nifi schedulers. The administrator can schedule for the data sources.

All the NIFI process details and the processing time will be captured into the NIFI log files which will be available for the Admin to review.

cQube performance summary					
Stage	Type	Zipped File name	Zip file Size	Records	Time taken
Files processed and time taken from API to S3 Output bucket	Static/Master files	school_master	3.2 kb	54728	
		cluster_master	32 kb	3247	
		block_master	1.3 kb	254	
		district_master	1.3 kb	38	
	CRC files	inspection_master	22.4 mb	497075	5 minutes
		user_location_master	57.8 mb	3054698	
	Semester files	semester	128.8 mb	3746754	10 minutes

	Student attendance files(one month)	student_attendance	250.4 mb	10473331	20 minutes
Overall time taken					~35 minutes

9.1 S3 bucket partitioning

S3 buckets will contain partitions for the data files to store. The partitions are created at the S3 input bucket & the S3 Output bucket.

9.1.1 S3 emission bucket partitions

- All the files in the S3 emission bucket will be in the CSV format.
- S3 emission bucket follows the folder hierarchy based on the data sources.
S3 -> Bucket name -> Data source -> emitted zip files with timestamp
- The folders and the files will be removed from the S3 emission bucket once NIFI copies the data.
- Unprocessed data files will remain in the S3 emission bucket for one week and then they will be deleted automatically at the end of the week.

9.1.2 S3 input bucket partitions

- All the files in the S3 input bucket will be in the CSV format.
- S3 input bucket follows a hierarchical partitioning based on the Data source, Year, Month, date and timestamp
S3 -> Bucket name -> Data source -> Year -> Year - Month -> date_Source name

Example for the S3 input bucket:

`S3/cqube-gj-input/student_attendance/2020/2020-05/2020-05-29_student_attendance`

9.1.3 S3 output bucket partitions

- All the files in the S3 Output bucket will be in the JSON format.
- S3 output bucket follows the hierarchical partitioning based on the data source, Year, Month, date and timestamp, similar to the partitioning that the S3 input bucket follows.
- Metadata files will have information of the latest updated output files which helps cQube to consider the latest output file during the visualization stage.

9.1.4 Data Validation after Ingestion

File size mismatch: NIFI does not allow the file into the S3 Input bucket if the file size does not match. A notification is sent through an email to the data emitter to check the file size and re-emit the data file.

Record count check for the emitted file: Nifi checks if the emitted file records count matches with the original file records count. If the file records do not match, NIFI will not process the file into the S3 Input bucket. A notification email is sent to the data emitter to re-emit the data file.

NIFI fetches the data from the S3 emission data bucket and sends it to the S3 input bucket, after performing the following validations:

- Column level validations: Column datatype mismatch, Number of columns, Data exceeding the column size.
- Improper Data handling: Missing/ null data values for mandatory fields, Empty data files, Special characters, Blank lines in data files.
- Duplicate records validation: NIFI validates the duplicate records by grouping the same kind of records together.

Validation for duplicate records: Records which have duplicate values for all fields (mirror image record) then NIFI will consider the first record and the rest of the records will be eliminated.

Records with duplicate ID: For the rest of the duplicate records where the records are having the same ID (student ID/ assessment ID/ infra ID/ CRC visit ID) and different values will not be inserted into the database tables as ID is the primary key.

Example: For the Duplicate records with different lat long details, NIFI eliminates the records which have the same id and different lat long details or different names or different values.

Records with same values: For semester report, The records which are having the same values for fields Student ID, School ID, semester, studying class and different values for the subjects then NIFI will eliminate those records.

Overlapping data validation: Overlapping data validation takes place based on the data source.

- The NIFI process for student attendance reports will check the last updated day's record from the transactional table and will process the records from the day after the last updated date. The records from all of the previous days will not be considered for NIFI processing
- For the other data sources, duplicate records where the records are having the same ID (student ID/ assessment ID/ infra ID/ CRC visit ID) and different values will not be inserted into the database tables as ID is the primary key.

Other data issues: Data handling in cases like job failures, missing data for certain days and late receipt of the data (receiving data after a few days), updating the wrong data, upon request (when issue identified at the report)

All the validation results will be captured into the NIFI log files as per the table below,

Sl No	Validation type	Error message
1	File size mismatch	"ERROR: file size mismatch for <filename>"
2	Record count check for the emitted file	"ERROR: record count mismatch for <filename>"
3	Validation for duplicate records	"ERROR: duplicate key value violates the data_source" along with record
4	Records with duplicate ID	"ERROR: Detail: Key (school_id)=(<id>) already exists"
5	Records with same values	"ERROR: duplicate key value violates data_source" along with record
6	Overlapping data validation	"ERROR: duplicate key value violates data_source" along with record

Other errors will be displayed with appropriate "ERROR: <error message>".

Admin dashboard will contain the details of the results of the validations as

1. The last files uploaded with the file name and timestamp
2. Total number of records in the file
3. Number of successfully uploaded records
4. Number of duplicate records
5. Number of invalid records.

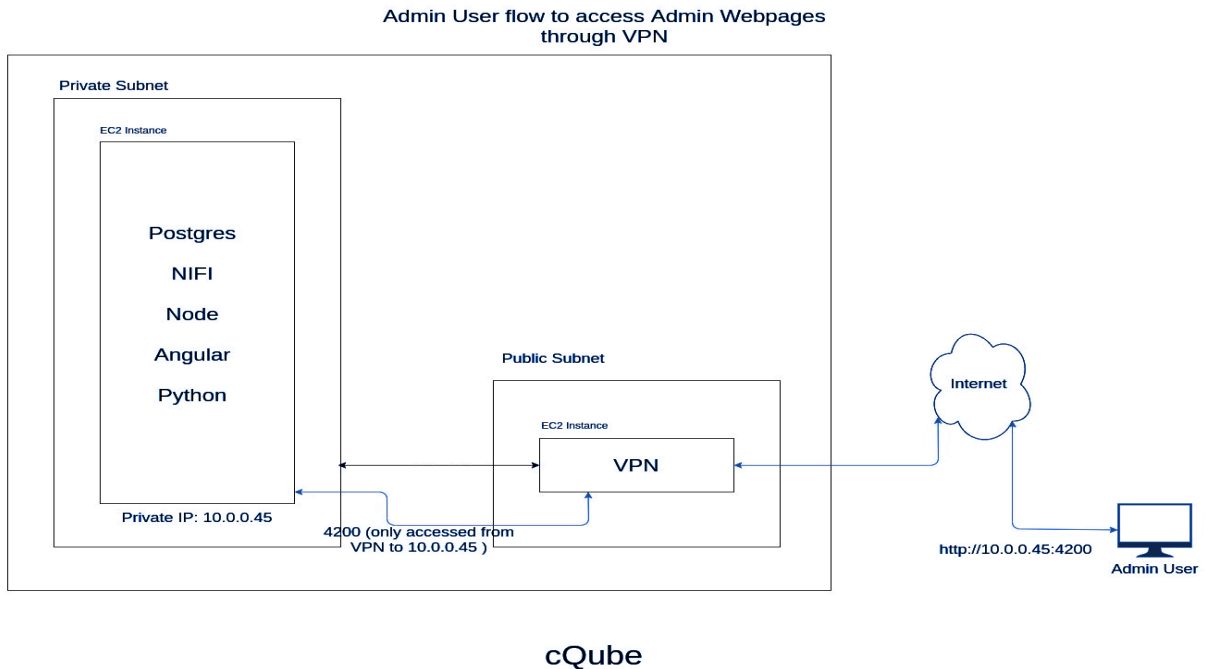
10. Admin login process & Features

cQube has two different interfaces, an interface for the cQube administrator and another interface/dashboard for the cQube reports:

- Administrator pages: Administrator activities will run separately in the VPN and admin must login through the 2 factor authentication process to perform the admin tasks. Change password functionality will be included in this login.
- cQube dashboard for report: This is a normal login which will have the cQube insights of the metrics.

10.1 Admin login process:

Figure - 7: Admin user flow through VPN



Admin must follow the 2 factor authentication process to perform the admin tasks. The Admin user should connect to the VPN to avail the 2 factor authentication by following the steps mentioned below:

- Admin has to download the google authenticator app to his phone and register the app with the QR code showing in the OPENVPN Access server page.
- By providing the credentials & Google authenticator code to download the user-locked profile(client.ovpn) file.
- Admin has to install the client openvpn-connect application
- Admin has to create the OpenVPN profile from the client.ovpn file
- Admin has to validate the user authentication and Google authenticator code to login to the cQube VPN.
- With the successful authentication admin can access the cQube admin features by opening the local ip in the browser.

10.1.1 Admin - Features

1. Create users:

- Admin can create new cQube users from the user interface page.
- Admin will assign the role to the user while creation by choosing the role drop down value.

2. **S3 buckets**

Admin can perform view files and download actions in the S3 buckets by the following steps:

- By using the user interface admin is able to view the raw data files from the S3 input bucket, Metrics files from the S3 output buckets, Transformed and Aggregated data files from the S3 output bucket files.
- Admin can able to download the metrics files in the JSON format, transformed and Aggregated data files in CSV format from the S3 output buckets
- Apart from the user interface Admin users can call the cQube API to download the metrics files in JSON format.
- They can use the same download API to download the transactional data files and aggregated data files in CSV format from the S3 output bucket.
- The admin users use the API to download the raw files from the S3 input bucket in the zip format also.

Admin can enter the AWS console and perform the files cleanup, files Download activities

- Raw data will be downloaded in CSV format and the metrics data will be downloaded in JSON format.

3. **Nifi Scheduler:**

Admin can control the Nifi processes by choosing the start and stop times.

- All the Nifi processes will be listed.
- The Schedule Time column has a drop down box with the hours. By selecting the hour the process will be automatically started at the scheduled time every day.
- The Stopping Hours column has a drop down box with the numbers. By selecting the number the scheduled running Nifi process will be stopped after the selected number of hours.
- Nifi process information was added to the schedule table to understand the status of Nifi processes whether they are in the running state or stopped state.
- Nifi schedules added another feature of scheduling for Daily, Weekly, Monthly and Yearly.
- If the user wanted to select the Daily schedule - The user has to mention the time of the schedule

If the user wanted to select the Weekly schedule - The user has to mention the Day of the Week

If the user wants to select the Monthly schedule - The user has to mention the Date of the Month. If the month is not having the specified date then the schedule will be run on the previous day of selected day

If the user wants to select the Yearly schedule - The user has to mention the Date & Month of the Year.

If the month is not having the specified date then the schedule will be run on the previous day of selected day

Below are the default schedules for each data source.

Data Source	Start Time	Active Run Time(In Hours)
Static data	10AM	1 hour
PAT Data	11AM	2 hours
Infrastructure Data	1:20PM	1 hour
CRC Data	2:30PM	2 hours
Teacher Attendance Data	4:40PM	1 hour
UDISE Data	5:50PM	2 hours
Student Attendance Data	8PM	2 hours
Semester Data	10:10PM	2 hours
Diksha Data	12:20AM	4 hours
Composite Report	5:40AM	1 hour
Progress card	6:50AM	1 hour

4. **View Logs:**

- All the log files will be consolidated into a common folder using soft links.
- New log files will be created every day and the old logs are retained for 7 days. At any point of time, there are 6 log files in the folder, old files are deleted when new files are created.
- Admin will have an option to view the last 200 lines on the browser with the help of the user interface provided. There will be a download option provided to download the entire log file.
- Log files will be downloaded in .log format.

- The list of log files provided to the admin are as follows:
 - Application Logs (Error and Info logs)
 - Admin Logs (Angular and Node logs)
 - System logs
 - NIFI logs
 - Data emission process logs
 - Database logs

5. **Grafana Dashboard Connectivity**

Admin has an user interface screen "Monitoring Details" screen in the admin dashboard, which connects to Grafana. And also Admin can use the Prometheus APIs to integrate with Grafana for viewing the cQube application health. The following details are displayed in the grafana dashboard:

- RAM usage
- CPU Usage
- Data storage disk usage
- Nifi heap memory usage
- Data source file metrics
- JVM usage

6. **Summary Statistics:**

- Summary statistics screen is having the processed data information.
- The records count will be displayed in the Total Records column.
- The total Number of not processed records are divided into the different validation types such as Blank lines, Duplicate Records, Null records.
- Total process records count will be displayed in the Processed Records column.
- Process Start time and Process End time is represented in the remaining columns.

7. **Data Replay:**

Admin can be able to delete some portion of the records from the database tables and can perform the re-computation by emitting the new data for the particular records. The data replay process will be performed as explained below

- The admin screen should be prepared with a select option having all the data sources
- All the enabled data sources should be picked up dynamically from the `datasource_config.yml` file from the installation directory
- Admin will select the data source and clear the data as like below

1. CRC: Year and Month
 2. Diksha ETB: From date - to date
 3. Diksha TPD: Batch Ids
 4. Infrastructure: Delete all data
 5. PAT: Exam code
 6. Semester: Academic year and the semester
 7. Static: Delete all data
 8. Student Attendance: Year and Month
 9. Teacher Attendance: Year and Month
 10. UDISE: Delete all data
- Admin will select any one option from the above list and click on the submit button
 - Reset button will be placed to clear the selected values

Data deletion operations from backend

- When admin submit the selected data source, A JSON file will be created with the selected values
- NIFI reads the file and clear the data from the transaction table and aggregation table for the selected period
- The new file should be emitted for the particular period again
- The new computation should happen and replace the values into the transactional table and aggregation table.
- The S3 out file should be affected with the new values.
- UI should show the new metrics on the visualization.

8. Data Retention:

Data retention has been implemented for the transaction tables of the below data sources.

- 1.Semester Assessment Test - (semester_exam_result_trans)
- 2.Periodic Assessment Test - (periodic_exam_result_trans)
- 3.Diksha Summary rollup - (diksha_content_trans)
- 4.CRC - (crc_inspection_trans,crc_location_trans)

By default the data retention period is set to 90 days. The retention period can be set at a 30 day interval, 30 days, 60 days, 90 days. or Unser specified days until 360 days in the data retention/replay screen.

Based on the retention period set, the data will be cleared from the transaction tables for the above sources. The data older than the retention period cannot be emitted for the above data sources, such data will be cleared and the information is logged in nifi. If the data replay is initiated for the data older than the retention period, within 5 days data re-emission is allowed.

10.2 Report Viewer

- They are defined in a hierarchical manner,

For example: If the users belong to the State administrative office, they can view the insight statistics of the districts that belong only to that state.

If the users belong to the district administrative office, they can view the statistics of the blocks which belong to the district.

The hierarchy is as shown below:

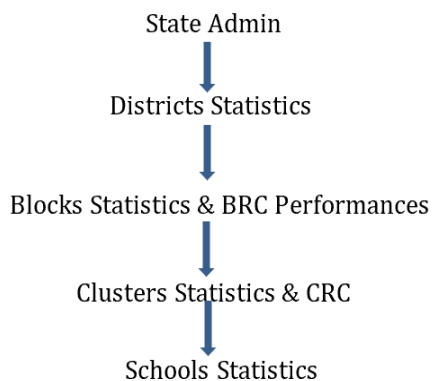


Figure – 6: Report viewer hierarchy

- They can view static dashboards according to their access scope. The selection option dropdown or drilldown options will be provided according to the access scope.
- They will not be able to access any other dashboards which are outside their scope.
- They can have insights into the performance and identify the low and high performers based on the benchmark in the dashboard.

- They can communicate with the users at lower hierarchical levels and provide them with the actionable items or exceptional performance via email (functionality provided in the dashboard)
 - They can alert the low performing district/ block/ cluster/ school based on their performance
 - They can notify the average performing district/ block/ cluster/ school based on their designation, to improve the performance
 - They can recognize the high performing district/ block/ cluster/ school based on their designation, to maintain the performance

11. cQube release upgradation process

cQube upgradation is the updating of the version of the cQube. As of now cQube allows users to upgrade to the latest version from the previous version only.

At the time of a new release users might choose either the installation of the latest version or the upgrade to the latest version.

Steps for upgradation:

- Clone the files from GitHub using the following command:
For example if the cQube upgradation is happening for version 1.3 then the command would be as
\$ git clone <https://github.com/project-sunbird/cQube/tree/release-1.3>
- This downloads the cQube files. The cQube GitHub has all the upgradation files required for upgrading cQube.
- The README.md file has all the instructions that have to be followed for the cQube product upgradation at “Upgradation:” section
- The update.sh script updates the complete cQube stack.
- The update.sh file calls the Ansible playbooks in the background which will complete the cQube upgrade setup.
- The complete upgradation process takes approx 30 minutes.

If the existing users choose the installation of the latest version there is the chance of losing the previous data. To overcome the data lost issue users has to follow the below instructions,

1. Users has to take the previous data backup and keep at some location (either at local machine or store at some S3 location)
2. Data has to be cleaned according to the table changes

For example:

- If any column is made as the primary key, the data has to be cleaned of the null values and duplicates.
 - If any columns were removed from the previous tables those column data would be removed.
 - If the total structure of any table was changed then there will not be any provision to restore the data
3. Data should be restored manually once the installation process has been completed.
 4. As part of the release-1.10.1 the composite key was altered to collection_id,uuid,batch_id, so before upgrading the duplicate records from the table should be deleted keeping the latest record of the user for a course.

11.1 Prerequisites

- The latest release branch will be created in Github.
- The latest code of cQube should be committed to the new branch of Github repository
- The existing data structure should be placed as SQLscripts in the new branch of Github repository.
- The latest data structure for cQube should be placed as SQLscripts in the new branch of Github repository.
- All the emission process should be in a stop state, there shouldn't be any data processing in nifi before the upgradation starts.
- The parameters for diksha need to be updated and the batch_id should be emitted for Diksha before running the upgrade.

11.2 Steps for cQube release Upgradation

- Each release will have to execute an upgrade.sh file.
- upgrade.sh file will call the scripts of ansible.
- Ansible scripts will perform the updates as mentioned below:
- An ansible script will run to take data backup from existing data.
- An ansible script will run the NIFI toolkit. By executing the NIFI toolkit, the NIFI will update the latest flow version to the release environment.
- An ansible script will be run to update the AngularJS code by getting the latest code from Github
- An ansible scripts will be execute the latest SQL scripts by comparing the below database structural changes

- SQL scripts will check each table from the SQL backup file with the Database schema
- If the table is already available, All the table columns will be compared.
- If there are no changes in the existing table & table fields, then the table will be skipped and no changes will be made on the existing table.
- If there are any changes in the existing table fields, the Ansible scripts run the alter query to perform the changes on the table
 - If the table does not exist, Ansible scripts perform the create the new table using create table query.
- S3 buckets data will not be touched if there are no changes in the existing tables. All the metrics and the calculated data will remain the same.
- Users can re-upgrade to the existing version of cQube. We need to re-execute the upgrade.sh
- Users can re-install the cQube by changing the configuration, but all the existing data will be erased. User can re-execute the install.sh

11.3 Configurations at cQube upgradation process

Configurations are done as a part of the installation process by the ansible scripts. All the properties of the sensible information like URLs and passwords are saved in the ansible properties file which is encrypted by default. Configuration is an automated process, no manual interactions are required in this stage.

Enable / Disable the process groups:

This configuration should be performed at the installation / Up-gradation stage of cQube. The cQube user may select the process groups which they wanted to include into the cQube. While the installation / Up-gradation of the cQube user will be available to select the existing process group by giving the true / false in the datasource_config.yml document which is available at the link below.

https://github.com/project-sunbird/cQube/blob/release-1.3/ansible/installation_scripts/datasource_config.yml

All the configuration should be given like below,

```
nifi_telemetry: false
nifi_crc: true
```

```
nifi_attendance: true
nifi_semester: true
nifi_infra: true
nifi_diksha: true
nifi_udise: true
```

As of now there are a total of 7 processor groups available in cQube.

11.4 Limitations

- Metrics will be regenerated if there are any changes in the existing table fields such as adding/deleting the new columns to the existing tables.
- Metrics will be regenerated if there are any structural changes at the data source.
- The output files should be regenerated with the recalculated metrics by having the modified fields information.

12. cQube Use-Case Configuration

From the cQube Release 3.0, cQube supports multiple use cases which will provide the data solutions for Sam/different sectors with various types of different metrics.

12.1 New Use-Case creation

Ansible Files Creation:

Create a new folder for the new use-case. Recommended name for the folder would be UC<Running Number>_Platform type.

For example, UC1_Education, UC2_Test, UC3_Healthcare

The below new files have been created for the new use-case. Keep all the files in the newly created folder

1. **config.yml**
2. **validate.sh**
3. **<new usecase_name>_datasource_config.yml**
4. **<new usecase_name>_datasource_validate.sh**

In explanation of the above files are like,

Config.yml: This file will contain the Base directory, State code, Diksha columns, Static data source, management specific details and Session timeout configuration details will be placed in config.yml file

Below is the link for reference to create the config.yml file

https://github.com/project-sunbird/cQube_Workflow/blob/cQube-release-qa/workflow_deploy/test_usecase_config.yml.template

Validate.sh: The validation which is mentioned in the config.yml will be placed in Validate.sh file.

For example code please refer the below link

https://github.com/project-sunbird/cQube_Workflow/blob/cQube-release-qa/workflow_deploy/test_usecase_validate.sh

<new usecase_name>_datasource_config.yml: The data sources which are related to the use-case will be maintained in this file. The data sources will be created with the boolean value True/False.

NIFI processor groups and the SQL queries will be invoked according to the selected Data sources which are mentioned as True.

For example code please refer the below link

https://github.com/project-sunbird/cQube_Workflow/blob/cQube-release-qa/workflow_deploy/education_usecase_datasource_config.yml

<new usecase_name>_datasource_validate.sh: All the validations related to the data sources which is mentioned in the datasource_config.yml will be placed in this file

For example code please refer the below link

https://github.com/project-sunbird/cQube_Workflow/blob/cQube-release-qa/workflow_deploy/validation_scripts/education_usecase_datasource_config_validation.sh

For the new use-case the Dataset, NIFI processors and the Visualization creations are mentioned in the technical document, Please refer to section 10 in the technical document.

12.2 New use-case integration with cQube

Ansible code changes:

- Modify the Ansible Roles in the install.yml file according to the given data sources for the Postgres, NIFI & UI to select the respective data source related stuff.
- Run the install.sh first to run all the modified components.
- For the state code to mention in the config.yml file please refer the below link

https://github.com/project-sunbird/cQube_Workflow/blob/cQube-release-qa/workflow_deploy/state_list

For the new use-case the Dataset, NIFI processors and the Visualization Configurations are mentioned in the technical document, Please refer to section 10 in the technical document.

13. cQube Private Data Center Installation

Mentioned below are the prerequisites for the installation of the cQube product:

- This document describes the requirements hardware & software details to install the cQube in the Data center.
- 3 VMs will be needed to install the cQube product
- Before starting installation, the network set-up has to be completed as mentioned in the section 3 of this document.

13.1 Data Storage Locations

- **SFTP Location(Emission directory):** The data emitted from the state education system has to be stored until the NIFI reads the data. S3 emission input storage buckets are the storage locations where the emitted data files are stored.
- **Data input files directory:** This is a location where the raw data resides for all future references.
- **PostgreSQL:** All the transformed and aggregated data will be stored in PostgreSQL tables.
- **Data output files directory:** A location where the processed data resides in JSON format.

13.2 Software security requirements

- cQube product security will be provided by installing the server at a private subnet with Nginx-Proxy server and load balancer as described in the section 2 of this document.
- All the ports will be accessed by Nginx server only, So those ports will not be accessed directly from the internet.
- Storage locations will be secured and are recommended with SFTP security.
- PostgreSQL will be secured by the following ways

13.3 cQube network setup

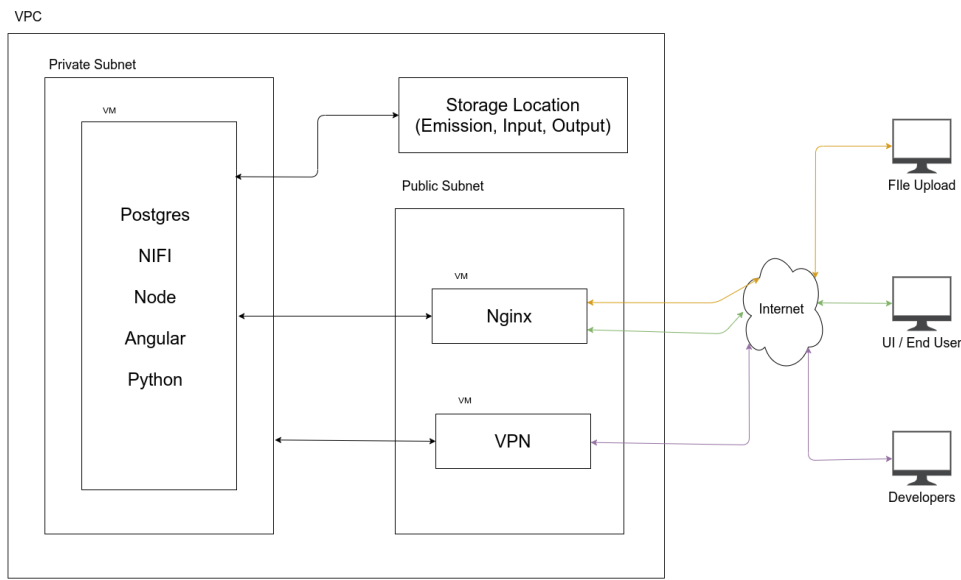
The following steps define how the cQube setup and workflow is in the data center. cQube mainly consists of the areas mentioned below:

1. Private Subnet

2. Public Subnet
3. Nginx Load Balancer

The cQube network setup process is described in the block diagram below:

Figure – 2: cQube network setup diagram



- The Yellow arrows in the network diagram indicate the file upload users' connectivity through the Nginx proxy and load balancer.
- The Green arrow in the network diagram indicates the end user's connectivity through the Nginx proxy and load balancer.
- The purple colour arrow in the network diagram indicates the developers connectivity through the VPN.

Before starting the setup, All the servers are required to have internet access and later can move the cQube server to the private subnet.

13.4 Private Subnet

Private subnet is used to secure the cQube server from public access. The instance will be created in a private subnet (will not be having the public ip) and all cQube components will be installed in this.

The steps involved to create instance in private subnet

- Create a virtual private cloud(VPC)
- Create a subnet in the created VPC with no Routing Table attached to Internet gateway
- Create an instance to install all the cQube software components.

VM1: cQube server

- create an instance with below mentioned configuration
- 8 core cpu, 32 gb ram and 1 tb storage

Firewall ports to be opened:

- port 4200, 3000, 8000 inbound from nginx
- port 4201, 3001, 9000, 8080, 8096, 22 inbound from openvpn server

Nginx proxy & Load Balancer:

- create a Nginx server
- configure it to connect to nginx on port 80, 443

Domain Name:

- create a domain name
- configure CNAME of load balancer to domain name

SSL:

- create a certificate from any certificate provider
- attach it to nginx load balancer

Firewall:

- port 80, 443 inbound from 0.0.0.0/0

13.5 Public Subnet

Public subnet will contain two instances, one is for OpenVPN and another is for Nginx, which will act as a reverse proxy. It is used to provide connectivity with the private subnet.

The steps involved to create the public subnet

- Create a subnet in that same VPC where the private subnet was created with a Routing Table attached to the Internet Gateway.
- Create the first instance with OpenVPN server and configure it to connect with the private subnet.

- Create the second instance with Nginx to connect with the cQube server which is in the private subnet.

VM2: Openvpn server

- create an instance with openvpn
- create users to access the instances and for cQube admin pages

firewall:

- port 22 inbound to 1 public ip (admin's)

VM3: Nginx server

- create an instance with ubuntu 18.04
- configure the nginx to connect Load balancer and cQube server
- configuration file will be provided

firewall:

- port 80, 443 inbound to Load balancer
- port 22 inbound to openvpn server

Section 2 - Process Steps

Step 1 - Server procurement

- 3 Virtual Machines needs to be procured along with the above mentioned configuration
- 500 GB storage disk

VM1:

- 8 core cpu, 32 gb ram and 1 tb storage

Firewall ports to be opened:

- port 4200, 3000, 8000 inbound from nginx
- port 4201, 3001, 9000, 8080, 8096, 22 inbound from openvpn server

VM2:

- 2 core cpu, 2 gb ram and 10GB storage
- create an instance with openvpn
- create users to access the instances and for cQube admin pages

firewall:

- port 22 inbound to 1 public ip (admin's)

VM3:

- 2 core cpu, 2 gb ram and 10GB storage

- create an instance with ubuntu 18.04
- configure the nginx to connect Load balancer and cQube server
- configuration file will be provided

firewall:

- port 80, 443 inbound to Load balancer
- port 22 inbound to openvpn server

Step 2 - Subdomain setup

- Subdomain needs to be created
- Map NGINX server IP to subdomain

Step 3 - Login to VM1 and Clone Git repository

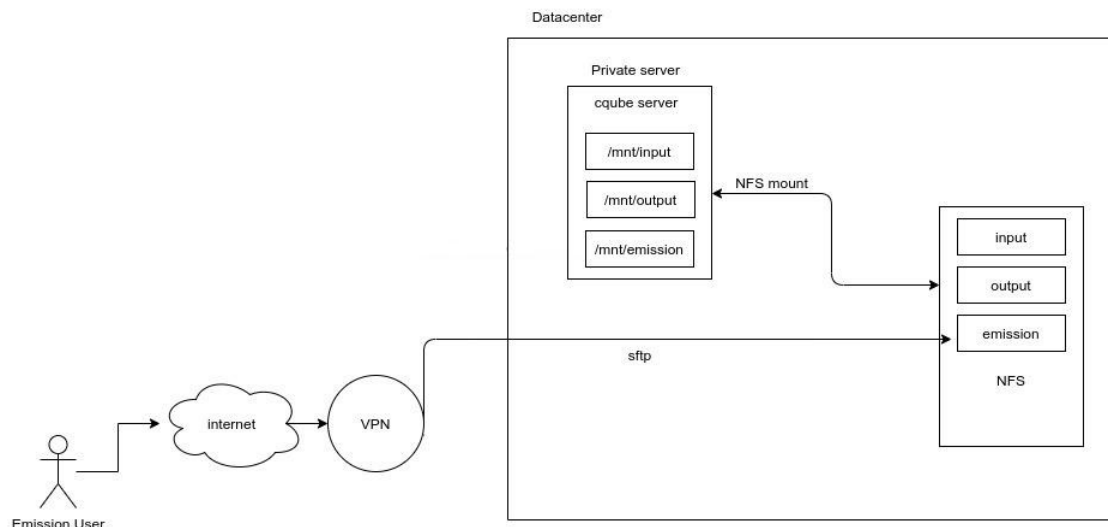
- Login to VM1 and clone cQube git repository
- Fill the configuration details as per git readme file
- Start the installation
- Once installation is done verify on UI

Step 4 - Data emission and next process

- Emit data for viz

13.6 Data emission process

The below image will describe the emission process for a private data center.



- The center should have the NFS For the emission of the data.
- The emission user should emit the data file using sftp.
- If the user can emit the data files through SFTP through the VPN, if the user is emitting the files from an external location.

14. Reports

As earlier said cQube is a “decision making tool”.

In explanation, cQube’s every report in Analytics is made up of dimensions and metrics. cQube reports provide good predictive insight and anticipates upcoming correlations at different components of the data and activities of the educational system. These reports help cQube admins to make appropriate decisions based on the insights.

14.1 Landing page dashboard

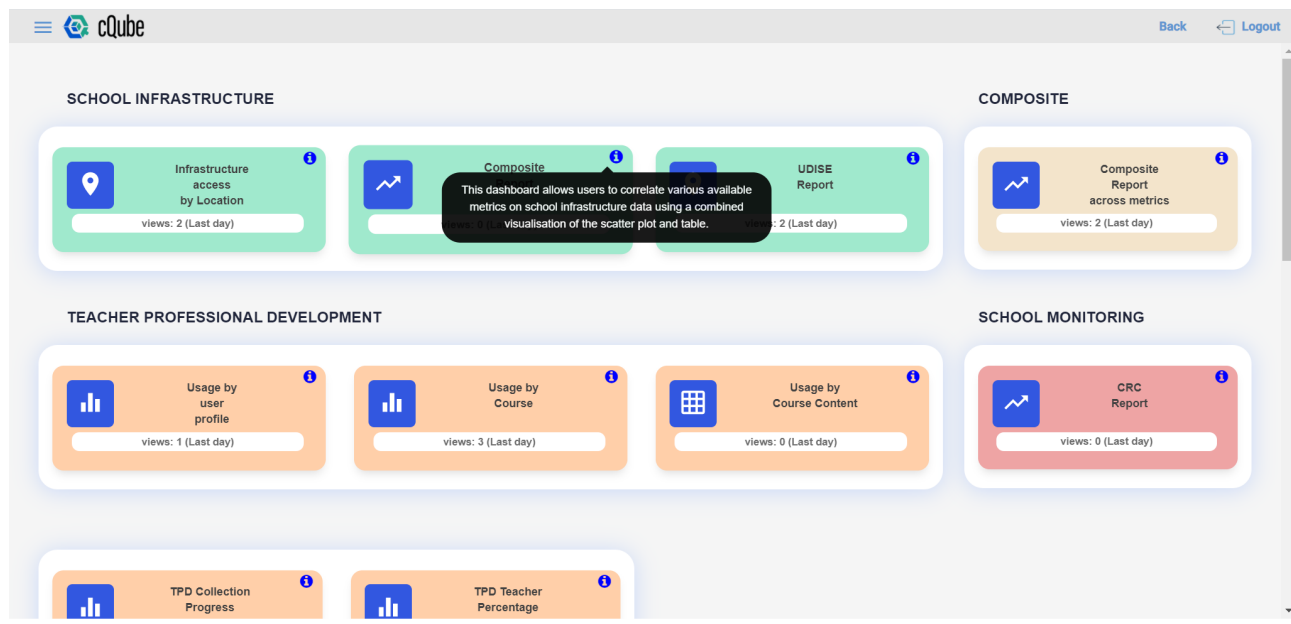
cQube dashboard will contain the shortcut icons of all the reports. Telemetry details will be displayed on each icon.

The number of views will be displayed for the last one day, 7 days and last 30 days with a 15 seconds time delay.

Each icon will have an ‘i’ symbol at the right corner of the icon which will display the description of the report with whoever mouse.

The dashboard will be shown as like below,

Figure – 8: cQube Dashboard - Landing page



For the visualization of the metrics cQube uses the below types of reports.

1. Map based visualization
2. Scatter plot visualization

3. Stacked bar report
4. Table format visualization
5. Bar chart visualization

More details of the visualization feature wise details are explained below,

14.2 Map based visualization

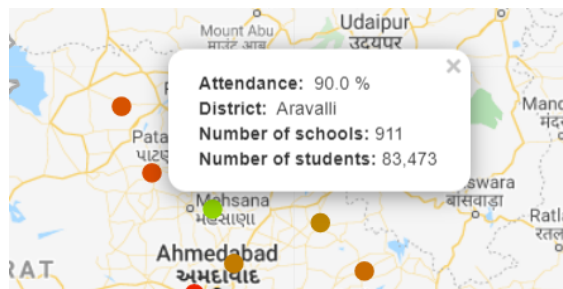
Landing page of the Map report will be represented for the whole state by having the colour gradient dots on the different districts of the state. Colour gradient will be declared from Red-Green where Red represents the poor performance and the Green represents the good performance.

All the dots will be displayed at the center of the screen. Zoom functionality will be applicable only for the school level representation that too only up to the state dimensional level only.

Maps should be dragged up to state specific. Entire India will not be able to see at a time.

Tooltip: On mouse hover of the dot, The user will be able to see the tooltip which can give the results in formation of the metric calculations. Along with the results the tooltips are containing the data like id and the location name and location details also.

Figure – 9: Map report tooltip



Blocks button: Block button will be a clickable function and it will display the information and the metrics results with the color gradient dots, tooltip information of the total blocks of the state.

Clusters button: Clusters button will be a clickable function and it will display the information and the metrics results with the color gradient dots, tooltip information of the total clusters of the state.

Schools button: Schools button will be a clickable function and it will display the information and the metrics results with the color gradient dots, tooltip information of the total Schools of the state.

Drill down functionality: Users will be redirected to the next hierarchical level of the locations by clicking on the dots.

For example: If the user clicks on some district's dot, then the visualization will be populated by all the blocks of the districts.

If the user clicks on some block's dot, then the visualization will be populated by all the clusters of the districts.

Drop down functionality: Drop down functionality will be experienced by the selection from the "Choose a District", "Choose a Block" and "Choose a cluster" drop down box values.

Time series: By selection of the time represented dropdown values users will get the time series information. Time series will be changed according to the data sources.

- Student attendance report has the Year and Month time series selection.
- Semester report is having the Semester wise time series selection.
- Infrastructure report does not have the time series. It represents the whole information of the infrastructure data.
- CRC report, Teacher attendance, student attendance, PAT map report, UDISE is having the Previous day, last 7 days & 30 days time selection.

Absolute & Relative toggle button: The Absolute and Relative toggle button in the footer will represent the dots color in the absolute and relative calculation explained below.

Absolute color will be calculated in the range of 1 to 100. The statically derived 50 gradient colors will be distributed between Red and Green colors. Each color will be assigned with 2 numbers interval

Relative color will be calculated by taking the difference between the maximum & Minimum values of the data and applying the statically derived 50 gradient colors will be distributed between Red and Green colors. Each color will be assigned the mean value of the difference.

Download Functionality: The download icon provides the download functionality of the report metrics based on the selection of the drop down values. The report names are named as below

```
{report_name}_{time_period/region_specification}_{district_id/block_id/cluster_id/school_id}_{date}.csv  
time_period -->{last_1_day,last_7_days,last_30_days,overall,month_year}  
Region_specification -->{blocks_of_district,clusters_of_block,schools_of_cluster}  
date & time -->{DD-MM-YYYY}
```

14.2.1 Map exception reports

The functionalities of the exception reports will be the same as the map reports except the color representation of the dots.

The color of the dots will be represented in the reverse order of the normal report. The colors will be applied from Red to Green gradient for the maximum to minimum value.

14.3 Scatter plot visualization

The landing page of the scatter plot report will be represented for the whole state by having the several normalised metrics using the raw data and you can read more about them.

A key feature of this dashboard is its ability to Zoom In and Out at various administrative levels. The administrative levels include District, Block and Cluster.

This has been done to provide relevant insights at the appropriate administrative level. In addition to visualising data, the dashboard also gives you the ability to download the data at various administrative levels.

This feature has been enabled to provide freedom to power users to derive additional insights that may not be captured in this dashboard. You can download the data using the dropdown option on the top right corner.

Table representation: The table in the scatter plot will have the different insights as columns of the table which represent the performances of different educational areas.

X-Axis & Y-Axis: these are the dropdown boxes which contain the columns of the above mentioned table. By selecting the x-axis value and the y-axis value the dot will be displayed at the cross point of the x & y results.

Drop down functionality: Drop down functionality will be experienced by the selection from the “Choose a District”, “Choose a Block” and “Choose a cluster” drop down box values.

Download Functionality: The download icon provides the download functionality of the report metrics based on the selection of the drop down values.

14.4 Stacked bar report

The landing page of the scatter plot report will be represented for the whole state by having the several normalised metrics in the bars with the combination of multiple items. The results will be represented in the x & y axis representation.

Example: Different subjects will be represented in a single bar between the X-Axis & Y-Axis.

Drop down functionality: Drop down functionality will be experienced by the selection from the “Choose a District” drop down box values.

Time series: By selection of the time represented dropdown values users will get the time series information. Time series will be changed according to the data sources. Diksha is having the Previous day, last 7 days & 30 days time selection.

Download Functionality: The download icon provides the download functionality of the report metrics based on the selection of the drop down values.

14.5 Table format visualization

Diksha is using the table format representation. The landing page of the table format visualization is having the metrics in combination with the 'Textbook content plays' and the 'Course content plays'. This report is having the pagination and a total of 1000 records will be shown in each page.

Each cell of the table will represent the color gradient based on the performance. The color gradient is applied same as heat map and 10 different colors between red to green

Content Selection: Content selection is the dropdown box which will represent the overall content, Textbook content and course content. By selecting the value table will be filtered as per the selected value.

Drop down functionality: Drop down functionality will be experienced by the selection from the "Choose a District" drop down box values.

Time series: By selection of the time represented dropdown values users will get the time series information. Time series will be changed according to the data sources. Diksha is having the Previous day, last 7 days & 30 days time selection.

Download Functionality: The download icon provides the download functionality of the report metrics based on the selection of the drop down values.

14.6 Bar Chart visualization

Diksha is using the bar chart visualization. The landing page of the bar chart visualization is having the metrics of in combination with the 'Textbook content plays' and the 'Course content plays' in the form of horizontal bars.

Content Selection: Content selection is the dropdown box which will represent the overall content, Textbook content and course content. By selecting the value table will be filtered as per the selected value.

Choose collection Name: By choosing the Collection name from the dropdown the bars will represent the selected results.

Drop down functionality: Drop down functionality will be experienced by the selection from the "Choose a District" drop down box values.

Time series: By selection of the time represented dropdown values users will get the time series information. Time series will be changed according to the data sources. Diksha is having the Previous day, last 7 days & 30 days time selection.

Download Functionality: The download icon provides the download functionality of the report metrics based on the selection of the drop down values.

14.7 Heat Map Report visualization

Periodic Assessment Test (PAT) and Diksha Teacher professional development (TPD) are using the Heat map visualization. This report contains the average performance of the district, block, cluster and school wise users.

X-axis contains the location names and Y-axis contains the consolidated information of the course, exam date, grade and subject details.

For Diksha TPD Course Progress Report & Diksha TPD Teachers Percentage Report, A drop down box was added for selection of multiple courses. All the courses will be available in the dropdown box and users will be able to select multiple courses from the dropdown box. When the user selects the causes and clicks on the submit button the results will be displayed according to the selection.

Heat map should have the pagination functionality. Records will be divided into 40 records per page. Each page will display the 40 records with vertical scroll bar operation.

The color indicators were added to the heat map for the understanding of the color differences of various performance ranges.

Heatmap contains various filters as described below.

Choose view by: This filter contains the Question ID & Indicator. By the selection of Choose view the report should show the Y-axis changes the representation to Question id and Indicator name.

Heat map shows the results by applying the colour gradient will be declared from Red-Green where Red represents the poor performance and the Green represents the good performance.

14.8 Progress card Report

Progress card report is the report where all the report highlights are displayed at one page. This report should contain the information of Student attendance, Semester, PAT, CRC, UDISE and School infrastructure details.

Progress card index report contains various filters as described below.

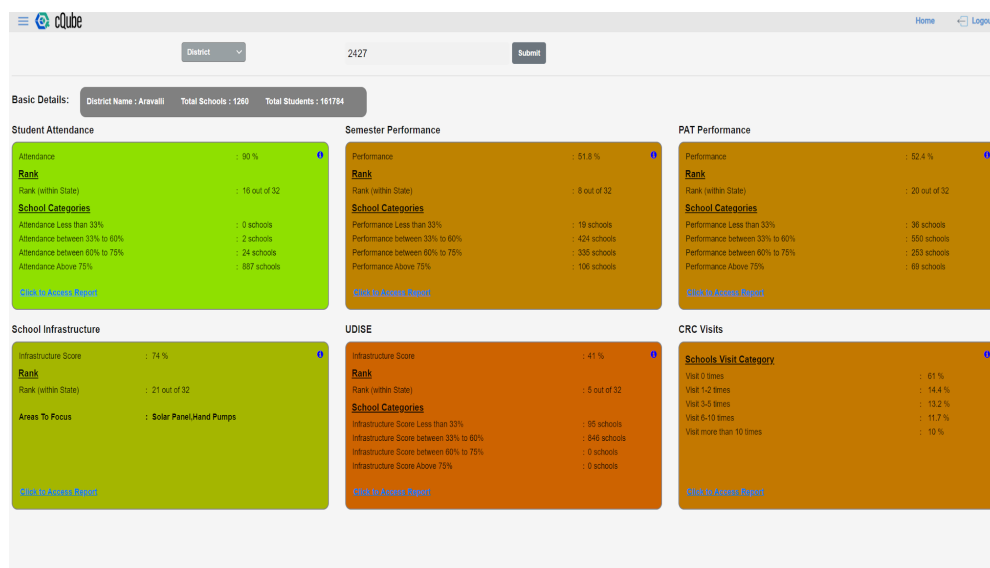
Choose Level: This filter contains the text of District, Block, Cluster and School options. Users may select one of the options and have to enter either name (OR) ID of the selected level.

By clicking on the submit button the data for the selected place will be displayed.

Info icon has been placed for each report data box which will display the complete details of the metrics of the selected area.

A hyperlink will be provided at the bottom of each report data box which will be redirected to the actual report and shows the selected place at the actual report. Each map report will be displayed with the link “Click here to access Progress card” to access the Progress card index report at the right which will redirect the user to the Progress card with selected options.

The report will be looks like below,



14.9 Trends Chart

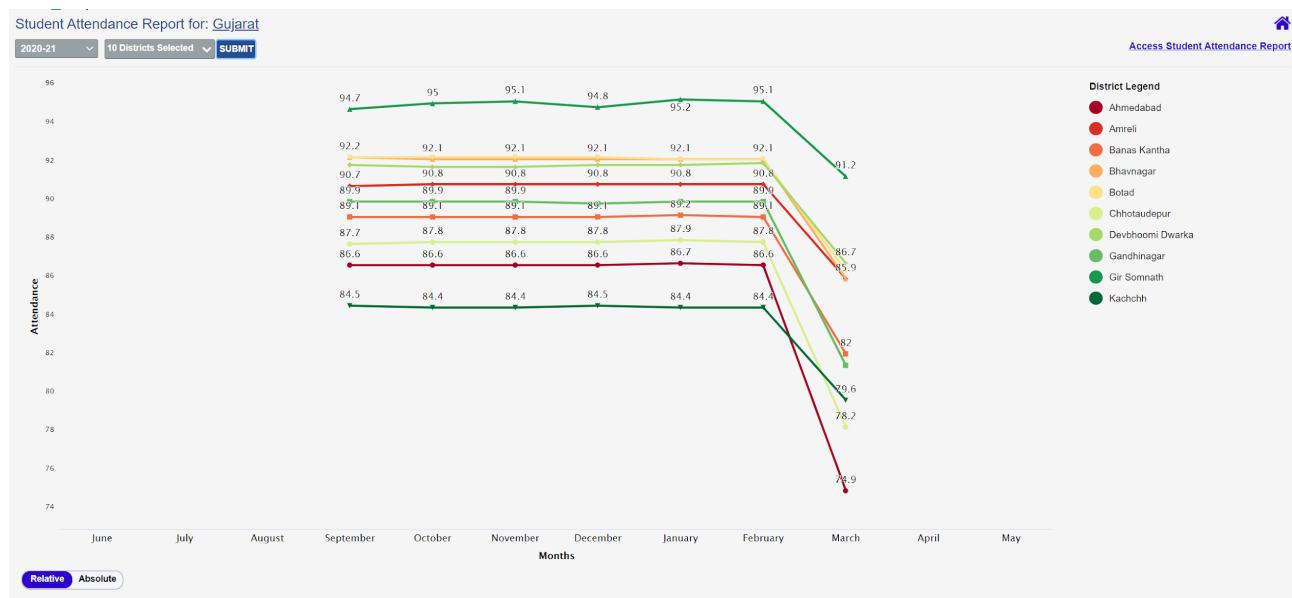
Trends Chart is introduced to cQube to observe how the schools are performing in comparison to the previous months. It will give the user to understand the performance of schools and how it is trending month by month.

Trends chart should be applied to each data source. As a part of cQube 1.12 release applied the trends chart for student attendance and It will be applicable to the rest of the data sources in the further releases. A hyperlink will be provided in the report as [Click on Trends](#)

By clicking on the hyperlink user will be redirected to the trends chart. Below are the trends chart functionalities

- By default trends chart will load the full state information on the home page
- At a time user can select up to 10 inner hierarchical level selections such as Districts, Blocks and clusters and user has to click the submit button

- If the user selects only one option and clicks on the submit button, Then the next level of hierarchical locations will be displayed.
- Users will be able to select the academic year from the “Choose year” select box.
- Each option will be represented with different color coding and the code will be given as a legend at the right side of the screen
- X- Axis should be the performance and the Y-axis should be the months of the academic year.
- Users will be able to see the Relative & Absolute markers on Y-Axis by selecting the switch at the left bottom of the report.



Data source wise reports Visualizations are as below

Student Attendance Report: Map visualization, Heat Map

Student Attendance Exception Report: Exception Map visualization

Teacher Attendance Report: Map visualization

Teacher Attendance Exception Report: Exception Map visualization

Semester Report: Map visualization

CRC Report: Scatter plot visualization

School Infrastructure Reports: Map Visualization and Scatter plot visualization

Diksha: Stacked bar, Bar chart and table type visualizations

UDISE: Map visualization

PAT: Map visualization, Heatmap, Table type visualizations

PAT Exception Report: Exception Map visualization

Composite Report : Scatter plot visualization

Telemetry : Map visualization

(i) cQube common Issues and Resolutions

Please refer the below document for understanding the most common issues and its resolutions were added in the below documents.

https://github.com/project-sunbird/cQube_Workflow/blob/release-3.0/documents/Solution%20Doc%20-%20Error%20Log.pdf

(ii) List of NIFI processor groups & UI Components

Please refer to the document below for understanding the NIFI processor groups and the UI Reports and its descriptions.

https://github.com/project-sunbird/cQube_Workflow/blob/release-3.0/documents/Nifi%20Processor%20Group%20%26%20UI%20Repost%20List.pdf