HawkEye Configuration Guide

Release Version 3.0.0

July, 2020

Document Version History

Version Number	Original / Modified	Authored / Modified By	Date
1.0	Original	Documentation Team	July 07, 2020

Release Version

Project	Release Date	Release Version
HawkEye Configuration Guide	July 07, 2020	V 3.0.0
HawkEye Configuration Guide	September, 2020	V 3.2.0
HawkEye Configuration Guide	October 07, 2020	V 3.3.0

About this Document

This document is intended to provide an overview of EkStep's HawkEye project and details to set up and configure its infrastructure.

In Scope

- Brief overview of the HawkEye project
    - Infrastructure components
    - Actors, their roles and responsibilities
- Prerequisites
- Workflow
- One-time setup
- Configuring reports regularly using Apache Superset

Out-of-Scope

- Detailed report generation architecture
- API documentation (linked)
- Druid data model (linked)
- Report request template (linked)
- Dashboard views
- Sample reports
- Druid analytics database
- Portal configuration and publishing
- Telemetry datastore

Intended Audience

The information in the document is primarily meant for:

- Report creators
- Report reviewers

Glossary and Abbreviations

Term/Abbreviation	Description/Expansion
PMU	Project Management Unit
MHRD	Ministry of Human Resources Development

Overview

The Hawk Eye project enables reporting infrastructure that packages:

- Generation of actionable datasets from Druid
- Configuration of visualisations using Superset
- Publishing the charts as reports using Portal with additional functionality to add a summary, export and download .

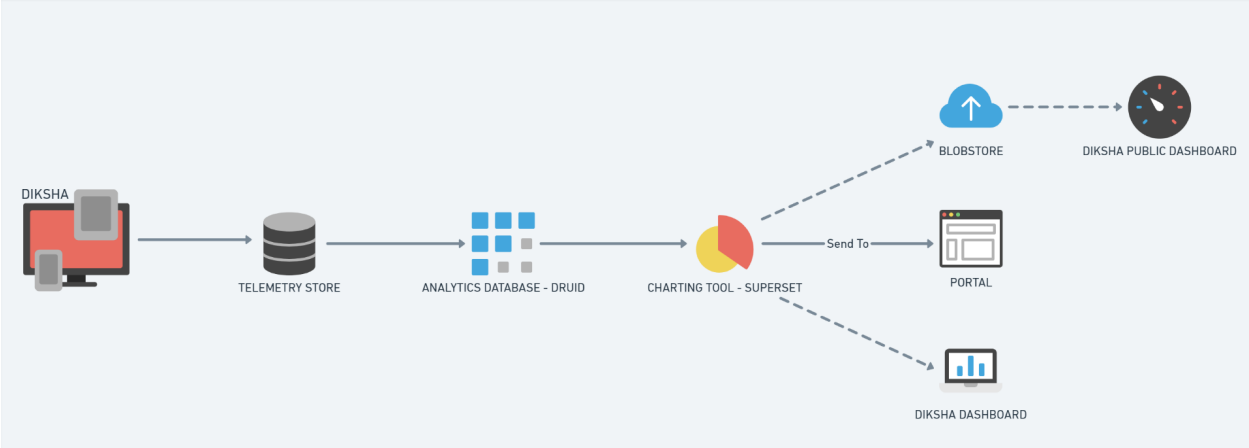
The generated reports allow the State PMUs, State outreach program administrators, users from the MHRD and others to view and understand DIKSHA usage in the context of their programs in and across states. They enable the States and the MHRD to drive targeted outcomes and demonstrate concrete progress made towards achieving stated goals.

Infrastructure Components

The HawkEye infrastructure mainly consists of:

- Druid - A high-performance, real-time, flexible data exploration analytics database
- Apache Superset - A modern, enterprise-ready business intelligence web application. HawkEye uses it as a library to visualise the druid data
- Portal - The easy-to-configure report publishing dashboard with real-time visibility into critical data and essential resources

The information flow among the components is briefly depicted in the following diagram:



Actors Involved

Team	Role	Responsibility
State PMU	Report requestor	Requests for a report in the prescribed template
Report Team	Report creator	● Validates the report request ● Configures the report in Apache Superset ● Validates the telemetry variables ● Generates the draft report ● Sends the report configuration and draft for a review
Report Team	Report reviewer	● Validates report logic ● Validates report configuration ● Validates the chart configuration against the request template ● Verifies Druid queries ● Publishes the report on Portal as a Draft ● Shares the report's system-generated link with the State PMU SPOC/ Domain Expert via email
State PMU / Domain Expert	Report admin	● Opens the report link received via email ● Verifies the draft report received on Portal ● Adds the report summary and actions to be taken on Portal ● Publishes the live report on the DIKSHA public portal, post-approval
Report Team	Report viewer	● Views the published report ● Actionizes tasks based on report recommendations

Prerequisites

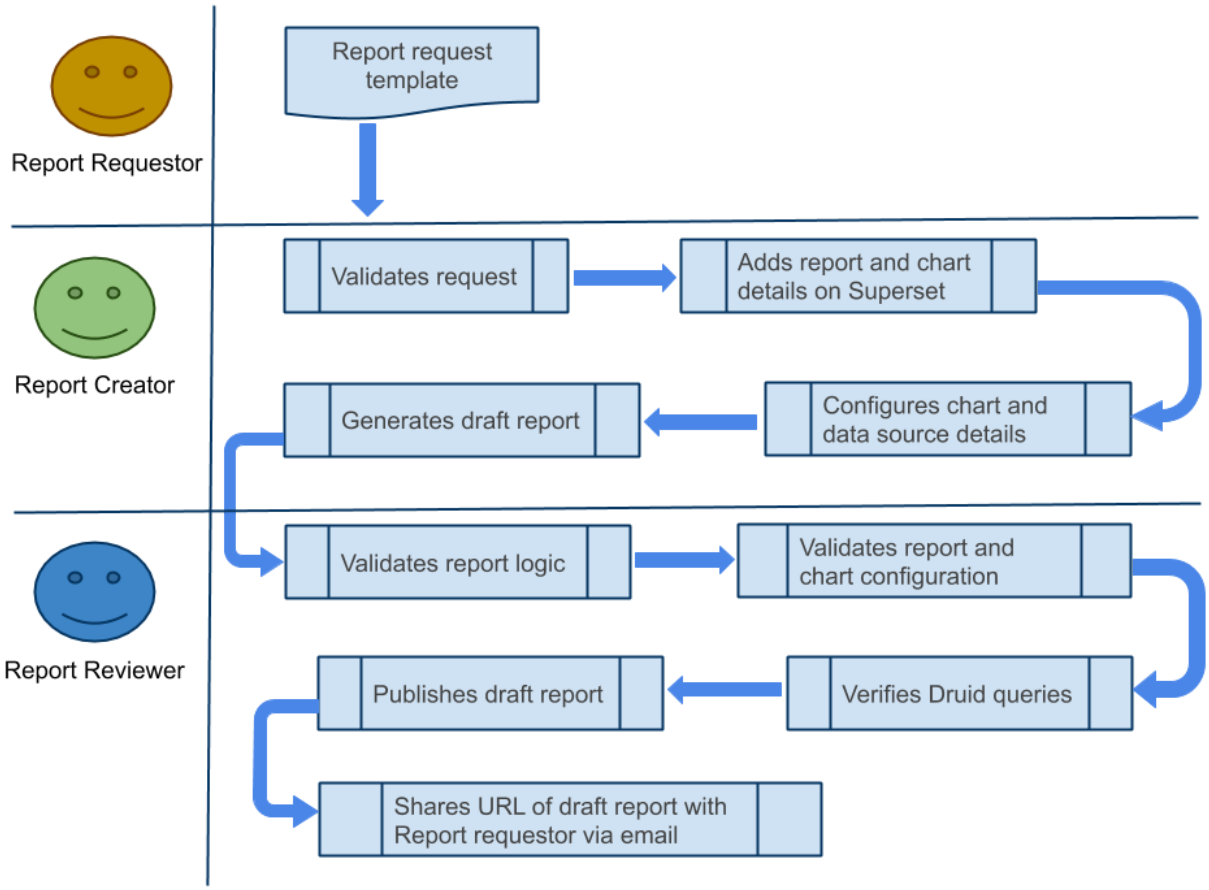
Before you start creating or reviewing reports on Apache Superset, ensure that you have the following permissions:

- VPN connection - To get an approval for VPN access, send a mail to noc@optit.in with the following details:

Field	Description
Environment	The name of the environment that you want to access through VPN. **Valid values:** Prod, Pre-prod
User Name	Specify your username, for example**;** Ram12
Email address	Specify your email address, for example; Ram12@xyz.org
Role	Specify the role(s) using which you will access Apache Superset. **Valid values:** Creator, Reviewer
Team/State	Specify the name of your team or State, for example; Product, Karnataka

- Access to Apache Superset (URL http://10.2.13.4:8088/)
- Credentials to log in to Apache Superset
- Basic understanding of Superset tool and chart.js

Workflow



Report APIs

HawkEye uses report APIs to:

- Perform create, update, read, delete, publish, and retire operations on the reports
- Add summaries to charts and reports
- Process Druid queries

Apache Superset uses these APIs appropriately to generate the requested reports.

Note: *For details, refer to [Report API Documentation](#).*

Configuration

Configuration for HawkEye can be separated into:

- One-time setup
- Regular report configuration

One-time Setup

- Install all infrastructure components
- Ensure that the telemetry store, Druid, Apache Superset and Portal are connected

Repeated Report Configuration

The reports team gets report requests regularly from domain experts or program managers to monitor usage and progress of their programs. To generate the requested report, the report creator and/or report reviewer can:

- Configure parameters on Apache Superset
- Check and validate the Report APIs
- Verify Druid queries

Report creators and reviewers configure charts and queries on Apache Superset. A report can contain multiple charts. Domain experts or State program managers add report summaries before publishing the reports. Charts can also be added to published reports.

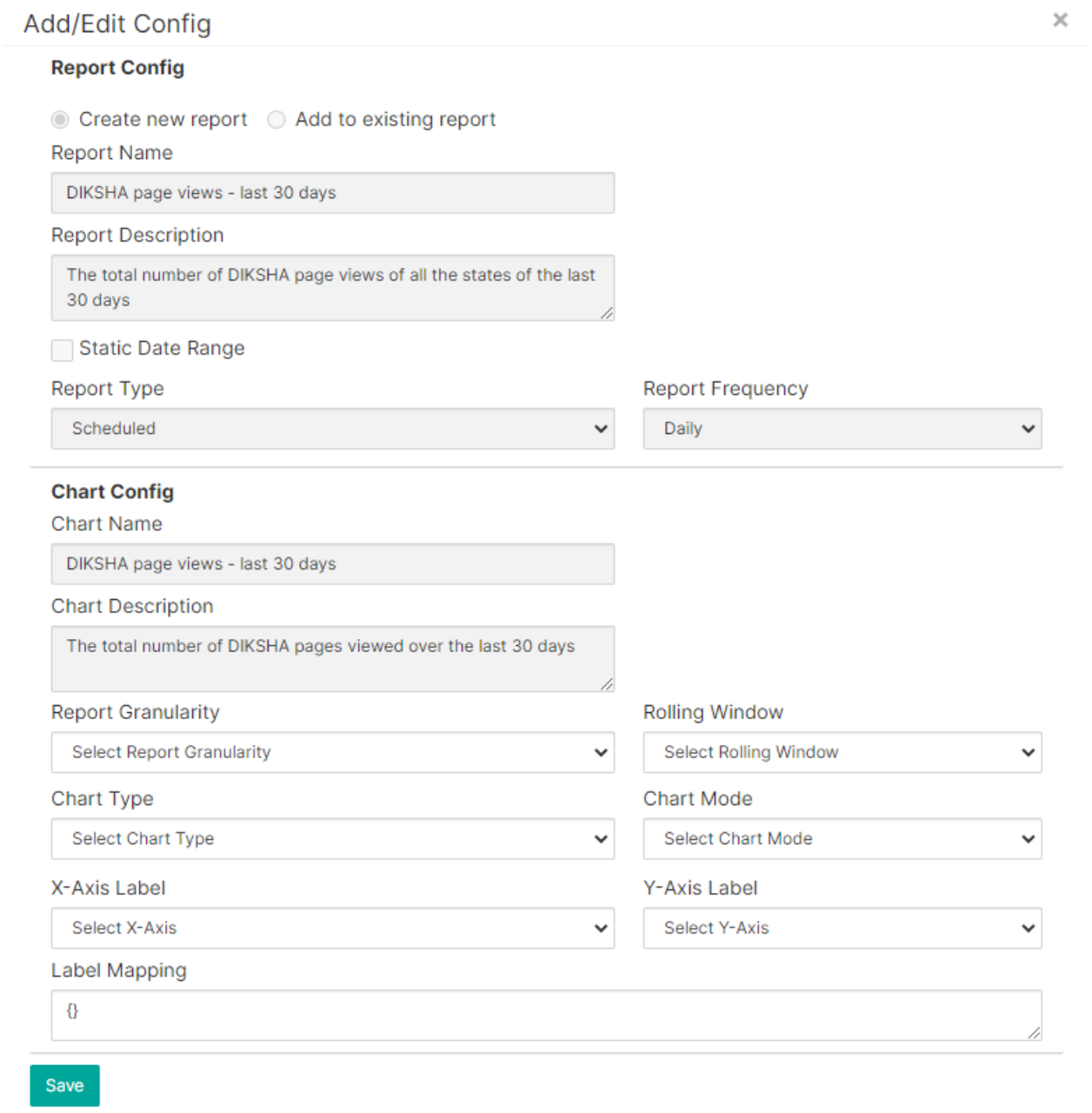
Configuration on Apache Superset

Use the following instructions to update configuration parameters and report details on Apache Superset.

Edit Configuration

1. Click **Edit config** on the landing page

2. The ***Add/Edit Config*** screen is displayed. You can specify report and chart details on this screen.

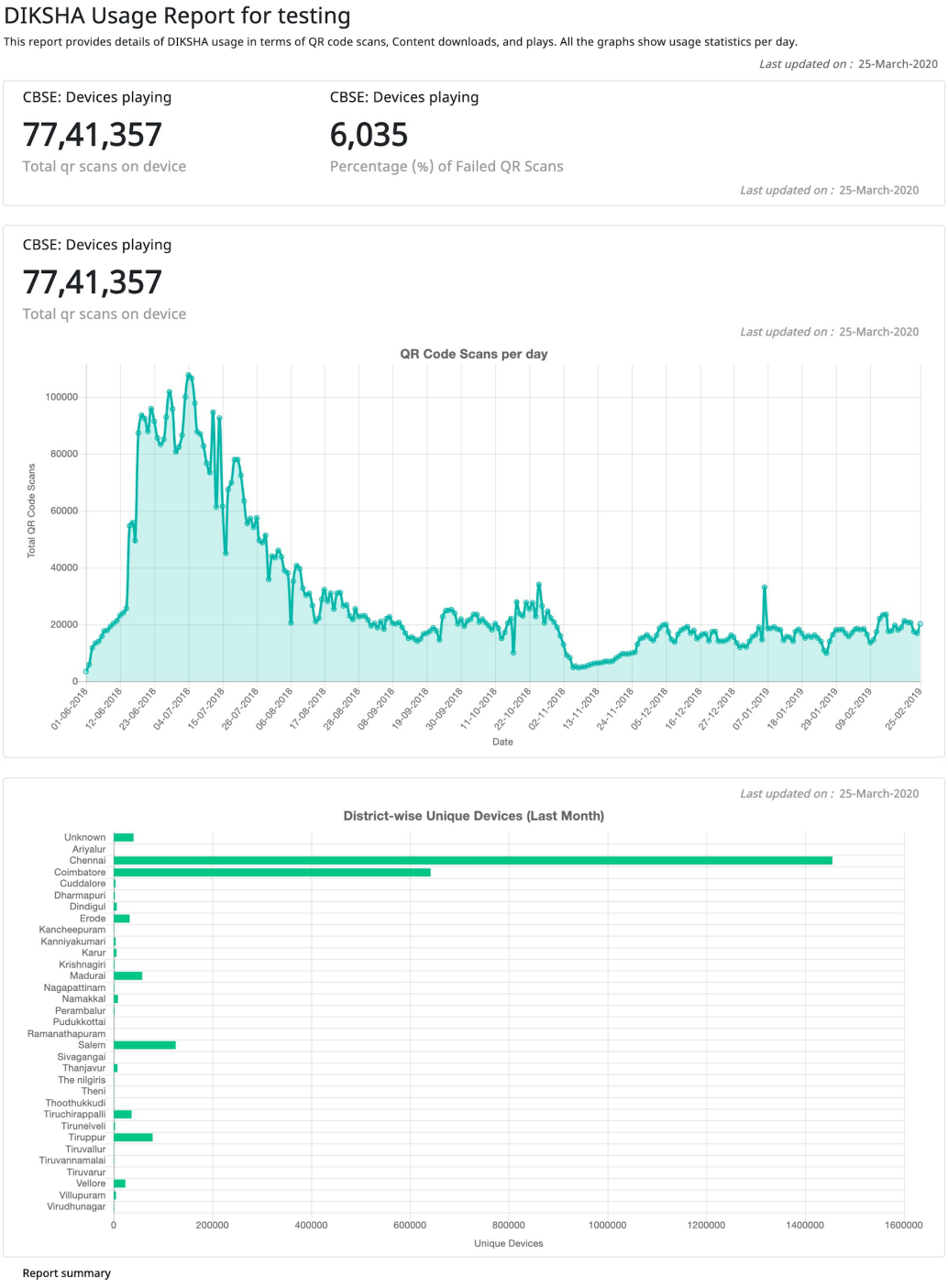


Screen Details

Parameter	Description
Report Config ***Note:*** *This section allows you to specify basic report details. A Report comprises multiple charts. A report is what is published onto the portal.*	
Create new report/Add to existing report	Select whether you want to create a new report or modify parameters of an existing report
Report Name	Specify the report name. ● **For New report:** Enter the report name ● **ForExisting report:** Select the report name from the dropdown
Report Description	Enter a description for the report.
Static Date Range	Select this check box to take the date range from the Superset tool query panel. If this field is not selected, the reports are generated based on the values provided in the **Rolling Window** field.
Report Type	Select if the report should be generated once or at a scheduled frequency for a particular duration. **Valid values:** One-time, Scheduled
Report Frequency	Select the frequency at which the report should be generated. The report data will be updated at the frequency mentioned. **Valid values:** Daily, Weekly, Monthly
Chart Config ***Note:*** *This section allows you to specify details for charts to be added in the report. A report can have multiple charts. Charts can be added to published reports as well*	
Chart Name	Specify the chart name. A report can have multiple charts. Specify the name for each chart separately. To add names for multiple charts, you need to: ● Save the configuration done for the first chart ● Select Add to existing report in the Report config section ● Select the report name from the Report Name drop down ● All other report details will be displayed ● Re-enter all chart details in this section
Chart Description	Enter the description for each chart.

Report Granularity	Select how detailed the data analysis should be for the report. The generated report splits the data based on the time range mentioned in this field. **Valid values:** day, week, month, all
Rolling Window	Select duration range for which the report should be generated. The report will show data only for the window selected in this field. **Valid values:** Last day, Last 7 days, Last 10 days, Last 30 days
Chart Type	Select the chart type. **Valid values:** Line, Horizontal bar, Vertical bar, Pie When you select the chart type as ***Pie*** the following fields appear ● **Show percentage**: Pie chart values with accumulated percentage ● **Show top n records**: The selected number of the pie chart is shown from the list of records
Chart Mode	Select the mode in which the report should be generated. **Valid values:** ● **Add** - to add the generated report to the previous report ● **Replace** - to replace the previous report with the current report
X-Axis Label	The label values that should be shown on the X-Axis of the report.
Y-Axis Label	The label values that should be shown on the Y-Axis of the report.
Label Mapping	Enter label mapping JSON for druid query processor and labels to be added in chart and legend in the chart map/ the field in the Superset, field names are coming from database in metrics. for example**;** {"date": "Date", "total_count": "total_content_plays_on_portal", "total_content_plays_on_portal": "Page Views", "legend": "Plays"}

3. Update the fields on the ***Add/Edit Configuration*** screen.
4. Click **Save**. The report is created on the Superset tool.



A report with multiple chart is generated

APIs use these configuration details to generate reports.
Note: *For details on APIs refer to the section, Report APIs.*

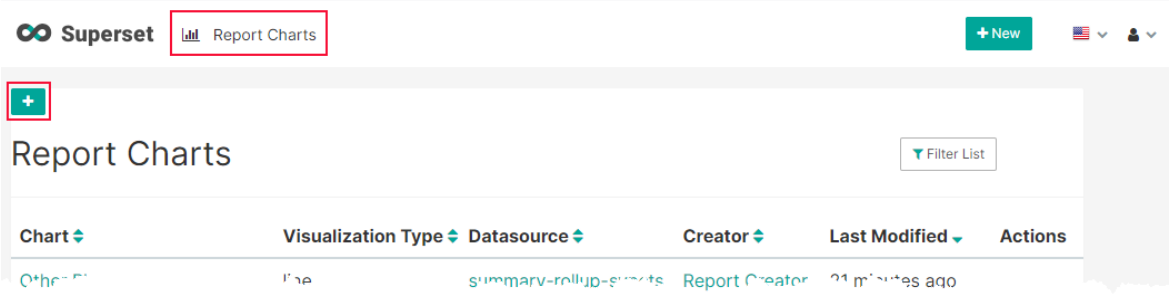
Chart Type Configuration

Depending on the requirements, each report may contain multiple charts. Apache Superset has the following in-built chart designs that can be included in the report:

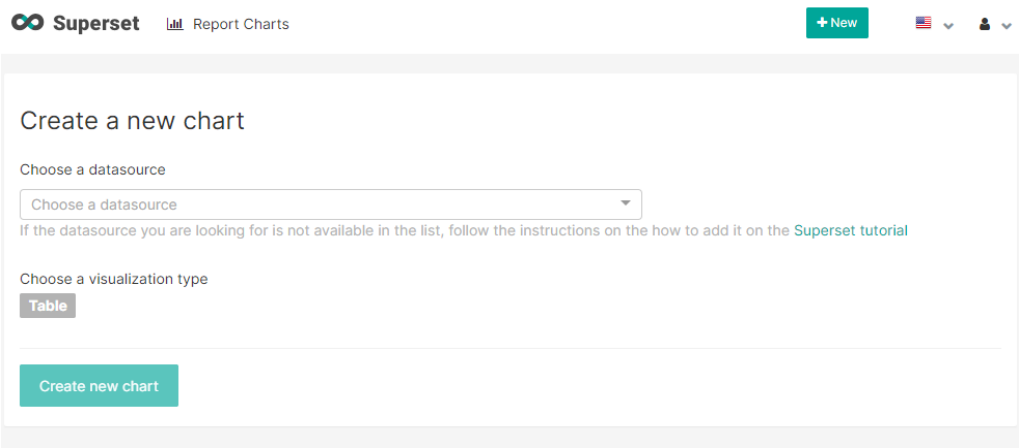
- Line
- Bar (horizontal and vertical)
- Pie
- Stacked bar

To create a new chart

1. Click **Report Charts** on the landing page.
2. The ***Report Charts*** screen is displayed.



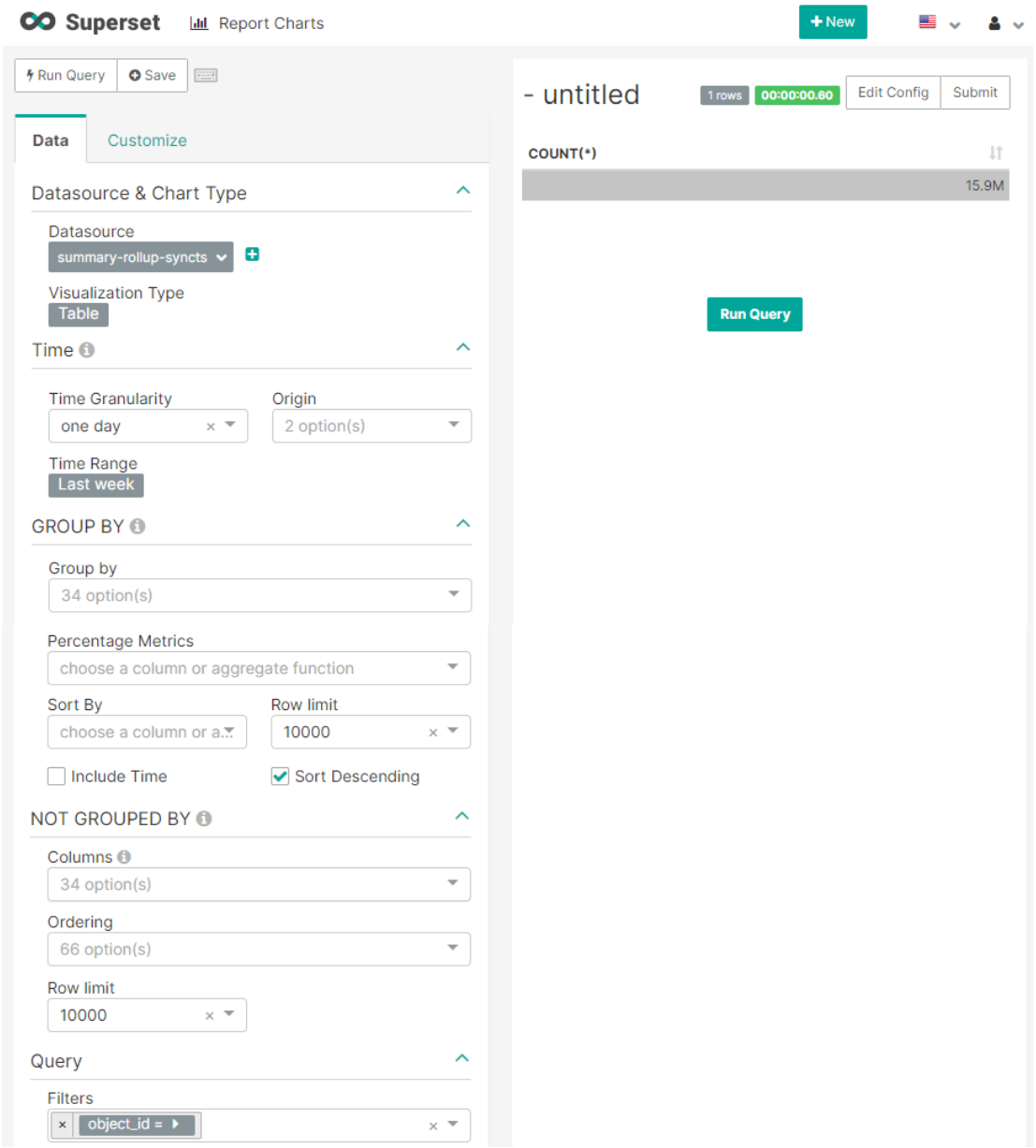
3. Click **+**
4. The ***Create a new chart*** screen is displayed.



Screen details

Parameter	Description
Choose a datasource	Select the datasource. Datasources include both databases and druid clusters.
Choose a visualization type	Select the format in which set of information is represented.
Create new chart	Click to create a new chart.

5. Update fields on the **Data** tab.


Superset
Report Charts
+ New
Run Query
Save
Data
Customize
Datasource & Chart Type
Datasource
summary-rollup-syncts
Visualization Type
Table
Time
Time Granularity
one day
Origin
2 option(s)
Time Range
Last week
GROUP BY
Group by
34 option(s)
Percentage Metrics
choose a column or aggregate function
Sort By
choose a column or a...
Row limit
10000
Include Time
Sort Descending
NOT GROUPED BY
Columns
34 option(s)
Ordering
66 option(s)
Row limit
10000
Query
Filters
object_id =
- untitled
1 rows
00:00:00.60
Edit Config
Submit
COUNT(*)
15.9M
Run Query

Screen Details

The Data tab allows you to build the Druid queries for the report that you are configuring.

Note: *The fields customised for Hawkeye in Superset is explained here. For further details on Superset tool, refer to the Apache Superset Tutorial*

Parameter	Description
Datasource	The datasource selected in the initial screen is displayed here. Datasources include both databases and druid clusters.
Visualization type	The format in which you want the information to be represented as selected in the initial screen is displayed here.
Time: *The parameters related to the time period of the report.*	
Time Granularity	Select how detailed the time-frame for the data should be, for example; daily, hourly, weekly, etc.
Origin	Select the source from where the telemetry data should be picked up.
Time Range	The duration of the data being represented in the chart.
Group By or Series: *The parameters by which you want to group your data.*	
Percentage Metrics	Specify the metrics to be added in the visualization.
Sort by	Select the column on which to sort the data.
Row limit	Select the maximum number of rows to be grouped.
Filters	Select the saved filter.

Report Configurations

To customise the reports further add these configurations in the Create Report API. Use the values in the following table to add specifications.

Parameter	Description
DataSource ({id: string, path: string} [])	Use the Data sources that are used within the report. ***Note:*** *For details, refer DataSources*
reportLevelDataSourceId (string)	Use an ID from the dataSource array.
Charts (chart[])	Defines all charts for the report. They are rendered in order as mentioned. ***Note:*** *For properties, refer to the* ***Chart configuration*** *table.*
Table (table[])	Defines all the tables to be shown in the report.
downloadUrl (String)	Specify the URL of the file in the Azure blob storage, if you have only one file to download. The path has a format that is similar to the Datasource path.
Files (File[])	Specify the configurations for multiple files, if multiple files need to be downloaded.

Chart Configuration

Update the following details in:

- Create Report API to specify the chart details
- Update Report API to append existing configuration

Parameter	Description
datasets ({data, label, stack, hidden, stack }[])	Use this when you need to hardcode the data.
datasets ({dataExpr: string, label: string, hidden: boolean, stack: string}[])	Use this to fetch data from the table. dataExpr should be same as the column name in the table

labels (Label[])	Provide X-axis labels, mandatory for charts like line, bar and radar. For polarArea, pie and doughnut the label appears on hover-over of the mouse. The label is either a single string, or it may be an array of string[] with each multi-line label where each array element is on a new line
labelsExpr (String)	Use this to fetch X-axis labels from datasets. labelsExpr value should match the column name
chartType (ChartType)	Use this to indicate the type of charts for example: line, bar, radar, pie, polarArea, doughnut
options (ChartOptions)	Specify the chart options, refer Chart.js documentation
colors (Color[])	Charts will use a default and/or random colors, unless stated Colors can be replaced using the colors attribute, if not specified, colors can be generated and applied randomly If the color field is not specified in the config, then the colors are generated and applied randomly.
legend: (boolean = false)	If true show legend below the chart, otherwise not be shown
bigNumbers ({header: string, footer: string, dataExpr: string}[]**)**	Defines a big number at the top of the chart. dataExpr - column name
iframeConfig ({ sourceUrl: string})	Use this if you want to show an external chart like India heap map within the report
Filters ({controlType: 'multi-select | 'select | 'date'" , displayName: string, reference: string}[])	If any filters need to be defined for the report use this field. controlType: type of element (multi-select, select, date), displayName: display message on the filter Reference: column name whose value should be used for the filter

Druid Data Model

Superset has a native connector to Druid and a majority of Druid's features are accessible through Superset.

Note: *To know more about telemetry events in Druid, refer to the Druid Data Model.*