{::options parse_block_html="true" /}

On this page

{:.no_toc .hidden-md .hidden-lg}

- TOC {:toc .toc-list-icons .hidden-md .hidden-lg}

GitLab Data Catalog Index

The Data Catalog page indexes Analytics Solutions, Dashboards, Workflows, and Key Terms. Please feel free to contribute to add additional links and resources.

Sisense Dashboards

Sisense is our enterprise standard BI tool.

`IMPORTANT: 2021-09-08 SAFE Dashboard Changes`

On 2021-09-08 Sisense dashboard access will be changing, driven by GitLab's maturation and the SAFE Data Framework. The following section describes the planned access process.

Access to Sisense dashboards is based on job role and goverened by the SAFE Data Access Framework. In Sisense, dashboards are classified into the following spaces:

- the **SAFE Dashboard** space houses all dashboards that meet GitLab's SAFE criteria. SAFE Dashboards are available to GitLab Team Members based on job role or demonstrated need.
- the **GitLab** space is a General Access area which houses all dashboards which do not require SAFE handling. General Access Dashboards are accessible by all GitLab Team Members.

A complete list of Dashboards, including SAFE Dashboards, is available in the GitLab Dashboard Index.

Accessing a GitLab General Access Dashboard

General Access Dashboards are available to all GitLab Team members and no Access Request issue is required. Access instructions are covered in the Sisense overview and development page.

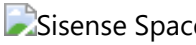
Accessing a SAFE Dashboard

All SAFE Dashboards are stored within the Sisense SAFE Dashboard space and access is granted at the Sisense space level. Access to a single SAFE Dashboard provides access to all SAFE Dashboards. Access to a SAFE Dashboard (and the space) requires:

1. Your immediate manager's approval
2. Departmental VP (or equivalent) approval
3. Approval of the SAFE Dashboard Space Owner defined in the GitLab Dashboard Index

To gain access to SAFE Dashboards:

1. Create an Access Request and list up to three SAFE Dashboards from the GitLab Dashboard Index you require access to. This list of SAFE Dashboards will help approvers understand your needs and intent.
2. Request approval from your immediate manager, your Departmental VP (or equivalent), and the SAFE Space Owner defined in the GitLab Dashboard Index header. Approval is **not needed**, if you have an approval for SAFE Data access in Snowflake, that is not older than 60 days. Skip this step and link to the particular AR.
3. Once the request is approved, tag @gitlab-data/analysts and the Data Team will process the request.
4. After processing is complete you will be able to login to Sisense and access your requested SAFE dashboard and all other dashboards within the SAFE Dashboard Space.

 Sisense Spaces

Please see the Accessing SAFE Data in Snowflake for instructions on to access the SAFE data in Snowflake.

SAFE Data in Google Sheets files can be accessed using the instructions for Accessing a SAFE Dashboard. Please follow those instructions to gain access to SAFE Data in Google Sheets.

`END OF 2021-09-08 SAFE Dashboard Change Announcement`

Data By Subject Area

Marketing

Dashboards

- TD: Marketing Data Mart

- TD: SDR Performance Dashboard

Analytics Projects

- 2021-10-01 SAO Analysis

Sales

Dashboards

- TD: Customer Segmentation
- TD: Sales Funnel
- Manual Usage Data Upload Process

Analytics Projects

- Propensity to Buy

Finance

- TD: Finance ARR

Product

Dashboards

- TD: Product Geolocation
- TD: Pricing Analysis

Data Models and Processes

- Product Usage Data
- SaaS Service Ping
- SaaS Product Events

Analytics Projects

- 2020-12 Product Analytics Offsite

Growth

Dashboards

- XMAU Analysis
- Growth Dashboards

Analytics Projects

- 2021-08 Customer Centric Product Insights
- 2021-08 Experimentation Workshop
- 2021-08 Stage:Secure Adoption and Conversion Analysis
- 2021-08 Namespace Conversion Analysis
- FY22-Q1 Growth Team KPI Review
- 2021-08 SSO Login Deep Dive Analysis
- Growth Insights
- Concluded Growth Experiments

Customer Success

- `Under Construction`

Engineering

- MR Rate

People

- People Metrics Overview
- PTO By Roots (Slack)
- People Key Metrics

- People KPI Deck
- Promotions Report
- Talent Acquisition Metrics
- People Metrics - Data Discovery in Sisense Dashboard

Data Team

- Sisense Usage and Adoption
- Trusted Data Health

Metrics and Terms Index

- Sales Term Glossary
- ARR: Annual Recurring Revenue
- ATR: Available To Renew
- CAC: Customer Aquisition Cost
- LTV: *customer* LifeTime Value
- Namespace
- NDR: Net Dollar Retention
- PQL: Product Qualified Lead
- Product Stage
- SM: Self-Managed aka Self-Hosted
- UPA: Ultimate Parent Account
- xMAU: x Monthly Active Users

Legend

📊 indicates that the solution is operational and is embedded in the handbook.

🚧 indicates that the solution is in a `Work In Progress` and is actively being developed. When using this indicator, an issue should also be linked from this page.

🐴 indicates that the solution is unlikely to be operationalized in the near term.

---

layout: handbook-page-toc title: "Customer Segmentation" description: "Customer segmentation is the process of dividing our customers into groups based on common characteristics so that we can understand who our customers."

On this page

{:.no_toc}

- TOC {:toc}

Customer Segmentation Analysis

Customer segmentation is the process of dividing our customers into groups based on common characteristics so that we can understand who our customers are and provide them with a great customer experience. There are many characteristics that identify our customers including industry, product category, sales segment, delivery, and territory to name a few. The Customer Segmentation Analysis page will provide the information and tools that GitLab team members can use to explore customer data and develop customer insights.

This data solution delivers three Self-Service Data capabilities:

1. Dashboard Users: A Sisense dashboard to visualize ARR and customer count by industry, product category, sales segment, delivery, account owner team, and territory
2. Dashboard Developer: A new Sisense data model containing the complete dimensional model components to build new dashboards and link existing dashboards to customer segmentation data.
3. SQL Developer: An Enterprise Dimensional Model subject area

From a Data Platform perspective, the solution delivers:

1. An extension to the Enterprise Dimensional Model for Customer Segmentation Analysis
2. Testing and data validation extensions to the Data Pipeline Health dashboard
3. ERDs, dbt models, and related platform components

Quick Links

Dashboard User Certification - Customer

Customer Segmentation Dashboard

Segmentation

Dashboard Developer Certification - Customer Segmentation

SQL Certification - Customer Segmentation

Getting started using Sisense Discovery

Self Service Walk-through Video

Data Security Classification

Much of the data within and supporting the Customer Segmentation Dashboard is Orange or Yellow. This includes ORANGE customer metadata from the account, contact data from Salesforce and Zuora and GitLab's Non public financial information, all of which shouldn't be publicly available. Care should be taken when sharing data from this dashboard to ensure that the detail stays within GitLab as an organization and that appropriate approvals are given for any external sharing. In addition, when working with row or record level customer metadata care should always be taken to avoid saving any data on personal devices or laptops. This data should remain in Snowflake and Sisense and should ideally be shared only through those applications unless otherwise approved.

ORANGE

- Description: Customer and Personal data at the row or record level.
- Objects:
  - `dim_billing_accounts`
  - `dim_crm_accounts`
  - `dim_crm_persons`

YELLOW

- Description: GitLab Financial data, which includes aggregations or totals.
- Objects:
  - `dim_subscriptions`
  - `fct_charges`

    - fct_invoice_items
    - fct_mrr

Solution Ownership

- Source System Owner:
    - Salesforce: @jbrennan1
    - Zuora: @andrew_murray
- Source System Subject Matter Expert:
    - Salesforce: @jbrennan1
    - Zuora: @andrew_murray
- Data Team Subject Matter Expert: @paul_armstrong @jeanpeguero @jjstark @iweeks

Key Terms

1. Product Category, Product Tier, Delivery
2. Sales Segment
3. Account Owner Team
4. Territory
5. Customer
6. Industry

Key Metrics, KPIs, and PIs

1. ARR
2. Customer Count

Self-Service Data Solution

Self-Service Dashboard Developer

A great way to get started building charts in Sisense is to watch this 10 minute Data Onboarding Video from Sisense. After you have built your dashboard, you will want to be able to easily find it again. Topics are a great way to organize dashboards in one place and find them easily. You can add a topic by clicking the add to

`topics` icon in the top right of the dashboard. A dashboard can be added to more than one topic that it is relevant for. Some topics include Finance, Marketing, Sales, Product, Engineering, and Growth to name a few.

Self-Service SQL Developer

Key Fields and Business Logic

- Data is sourced from Zuora and Salesforce.
- Parent customers can have more than 1 product; therefore, they can be counted more than once in the product category and delivery dimensions. To get a unique count of total customers, you have to either aggregate the products and delivery into an `ARRAY` and do a `COUNT DISTINCT` of customers or do a `COUNT DISTINCT` of customers without the product category or delivery dimensions included.
- For a charge to be considered recurring, the effective end month must be greater than the effective start month in the data.
- In Zuora, the `effective_end_date` and `effective_end_month` of the charge is the first day or month of the renewal respectively.
- In the monthly ARR calculation, the effective end month indicates when churn would happen and we do not count the effective end month in the ARR calculation. For example, a subscription with `effective start month` = 2020-07-01 and `effective end month` = 2021-07-01 would have its ARR summed from 2020-07-01 through 2021-06-01 for 12 months of ARR.

Entity Relationship Diagrams

Diagram/Entity	Grain	Purpose	Keywords
ARR and Customer Count Analytics ERD	Month, Subscription, Product Category	Provide insights into ARR and Customer Count by various customer dimensions	Parent Customer, Product Category, Delivery, Industry, Account Owner Team, Territory, and Sales Segment
Lead to Cash Overview ERD	All of the below	General overview of all processes for lead to cash	Parent Customer, Product Category, Delivery, Industry, Account Owner Team, Territory, Sales Segment, CRM, Persons, Accounts

Reference SQL

Snippet Library	Description
Customer Segmentation SQL Script	Query to slice ARR and Customer Count by Product Category, Delivery, Industry, Account Owner Team, Territory, and Sales Segment

Snippet Library	Description
Customer Segmentation TY Quarter vs. LY Quarter SQL Script	Query to pull TY versus versus LY ARR and Customer Count by Quarter and slice by Product Category, Delivery, Industry, Account Owner Team, Territory, and Sales Segment

Data Platform Solution

Data Lineage

- Data is sourced from Salesforce.com and Zuora, excluding accounts from manually managed list of zuora excluded accounts
- A complete data lineage can be found at dbt mart_arr lineage chart

DBT Solution

The dbt solution generates a dimensional model from RAW source data. The exceptions are the following fields that are calculated based on business logic implemented within specific dbt models:

field	business logic
product_category	Calculated based of Zuora product_rate_plan_name
delivery	Calculated based of product_category
service_type	Calculated based of product_rate_plan_name
ultimate_parent_account_segment	Calculated based of SFDC ultimate_parent_sales_segment by grouping `Unknown` and `NULL` segments into `SMB`

Trusted Data Solution

See overview at Trusted Data Framework

dbt guide examples for details and examples on implementing further tests

Zuora

- Trusted data dashboard

    - Reporting on all Zuora data tests which include the tdf tag.
- Data pipeline health dashboard
    - Reporting on the row count to highlight any data issues in Zuora account and Subscription data.

EDM Enterprise Dimensional Model Validations

- (WIP) Enterprise Dimensional Model Validation Dashboard
    - Reports on latest Enterprise Dimensional model test and runs

RAW Source Data Pipeline validations

Data Pipeline Health Validations

---

layout: handbook-page-toc title: "WIP Email Marketing Data Mart"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

Email Marketing Data Mart

Over the last year, our team noticed a need for a solution that helped us scale our data requests and our lifecycle marketing tactics.

Enter the email marketing data mart (also known as Email Marketing Database), which integrates several disparate data sources to help us reach our users, customers, and prospects in a more intelligent and scalable way.

In partnership with enterprise applications and marketing teams, we have created a solution that will enable GitLab to more quickly communicate critical customer and user updates, understand the overlap between prospects and users, and enable us to market to them more efficiently.

The goal of this page:

- Help you to use the TD: Email Marketing Data Mart README to generate email campaigns.

- Help you understand the data models used to create the TD: Email Marketing Data Mart.
- `Coming Soon` Assess your understanding by taking a certification most applicable to your role at GitLab.
- And overall help everyone contribute!

Quick Links

README - TD: Email Marketing Database

ERD - TD: Email Marketing Database

Getting Started

To get started we want to make sure you understand:

- Key terms that will explain how we account for the metrics
- The data sources behind the database

Key Terms, Key Fields, and Business Logic

▶ Key Terms

▶ Key Metrics, KPIs, and PIs

▶ Key Fields and Business Logic

Understanding the Data Sources and Data Models

Data Sources:

- Zuora
- Salesforce
- GitLab.com
- Customers DB

Entity Relationship Diagram (ERD):

- Data Lineage
- Example Queries

Additional Resources

- Trusted Data Solution
- EDM Enterprise Dimensional Model Validations
- RAW Source Data Pipeline validations

▶ Data Security Classification
▶ Solution Ownership

---

layout: handbook-page-toc title: "Merge Request (MR) Rate"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

Merge Request (MR) Rate

MR Rate is a measure of productivity and efficiency. The numerator is a collection of merge requests to a set of projects. The denominator is a collection of people. Both are tracked over time (usually monthly).

Additional detail on overall and wider MR Rate can be found here

Quick Links

MR Rate Dashboard

DBT Documentation

Getting Started

The MR Rate Dashboard is built on `rpt_gitlab_employees_merge_request_metrics` and takes into account 2 main filters:

1. `Breakout`: This allows the user to view the dashboard at a company, division, or department level.
2. `Breakout_division_department`: This filter is a child filter and will change depending on the breakout selection.

- `Breakout` = `Company` --> `Breakout_division_department` will give 1 option to be selected `all_company`
- `Breakout` = `Division` --> `Breakout_division_department` will list all the divisions at GitLab
- `Breakout` = `Department`--> `Breakout_division_department` will list all the division-department combinations.

Note: the `Breakout_division_department` filter only allows you to select 1 option as the data is pre-aggregated.

The dashboard itself consists of:

1. The MR Rate Trend
2. A % member of team members with at least 1 MR trend
3. A breakdown of the MR Rate by team members for the previous month
4. A breakdown of which projects are accounted for, and those that are not.

Projects that are part of the product

In the MR Rate and Volume of MR calculations, we consider MRs from projects that contributes to the overall product efforts.

The current list of projects are identified in the gitlab-data/analytics project for the following system databases:

|System Database | File | |GitLab.com |projects_part_of_product.csv | |ops.gitlab.net |projects_part_of_product_ops.csv |

Looking to update the projects list, follow the directions here

Data Model

The underlying data model take into account a lot of data models beforehard, however, we try to make it easier to analyze by bringing this into 1 aggregated data model at the GitLab team member level, and then creating the aggregated level report:

gitlab_employees_merge_requests_xf -- shows all merge requests for GitLab Team Members, and indicates whether the merge request is accounted for in our mr rate calculation with the usage of the `is_part_of_product` field.

In addition, we use gitlab_bamboohr_employee_base to create our base model that breaks down employees by month and organizational attributes (i.e division, department, etc) and how long they are effective.

---

layout: handbook-page-toc title: "Estimation Algorithm"

On this page

{:.no_toc}

- TOC {:toc}

Estimation Algorithm

For the Self-Managed instance that hosts our SaaS GitLab.com product, we receive an real-world data from Usage Pings. For Self-Managed instances, customers have the option to disable Usage Ping and we do not receive usage statistics for these instances. To calculate usage values across all Self-Managed customers, we need to develop an algorithm to predict usage across instances who have disabled their Usage Ping. We call this algorithm our `Estimation Algorithm`.

This section explains our first attempt in these xMAU estimations.

Current Methodology

What we know ?

- The number of Active subscriptions per month
- The number of Active subscriptions that send us a Usage Ping payload per month broken down by version
- The release date (therefore month) of a specific xMAU counter in Usage ping. For example we know that the `merge_requests_users` which is the GMAU of the Create - Source Code Group was released in version 12.9

What we don't know ?

- Number of Active Free Instances

What we can do ?

Let's discuss a real-life example! The counter used as SMAU for Dev:Create stage is `action_monthly_active_users_project_repo`. It was released on version 13.3. To calculate the Estimated SMAU we follow this process step-by-step:

- Calculate recorded GMAU split by delivery

- Calculate per month
  - For October, we know we have 3500 subscriptions that sent us a Usage Ping Payload out of 5000 subscriptions
  - Among them we split between subscriptions that are on a version 13.3 or above and the ones which have 13.2 or below and the ones which don't send us any payload.
  - This SiSense chart saw the split per month
- We get from the chart above a % of subscriptions that send us payloads and that are on 13.4 or above out of all Active Subscriptions
- We then deduce that Estimated SMAU = Recorded SMAU / % on 13.3

Since counters are released on different GitLab version, the estimation is customised form one counter to another.

This number is actually a quite accurate number to estimate paid XMAU but can be used at first for calculating estimated xMAU. Though we can imagine several improvements in order to make this estimator more robust:

- The first improvement would be to use the sum of the seat quantity ordered instead of the number of subscriptions.
- We could also break down the estimator per plan (current problem: a subscription can have 2 instances from 2 different product tiers)
- A major area of improvement is regarding our Core estimation. We currently assume that we have the same patterns between paid and free in terms of version upgrade and usage ping opt-in. While usage ping opt-in rate estimation seems very complicated to improve an critical area of improvement would be around version upgrade and usage data trend for Core Instances

---

layout: handbook-page-toc title: "TD: Snapshot Annual Recurring Revenue (ARR)"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

The goal of this page:

- Help you understand how to navigate through the Snapshot ARR Dashboards.
- Help you understand the data models used to create the Snapshot ARR dashboards.
- And overall help everyone contribute!

Snapshot ARR

ARR in the Zuora data changes on a daily basis as subscriptions are renewed and amended. Therefore, it is necessary to create a snapshot of the data each day so we can have a historical record of ARR in the data warehouse to use for reporting and analysis. The three methodologies we use to Snapshot and report ARR are described below.

Method One:

We use dbt snapshots which implement type-2 Slowly Changing Dimensions over mutable source tables. These Slowly Changing Dimensions (or SCDs) identify how a row in a table changes over time. In Method One, we snapshot the mart_arr table which is the mart used to generate our live ARR metrics. This method produces an exact replica of `mart_arr` each day which is then modeled using date spine techniques. In this method, we do not use the snapshoted raw data and instead snapshot the mart table. We found this has benefits of producing a simple model of the snapshotted mart_arr table that always stays up to date with the columns in the `mart_arr` without having to do full refreshes. This is a great fit for the use case of needing an exact `three ring binder` copy of `mart_arr` that keeps up with the columns and does not need to be fully refreshed. Even if the model is fully refreshed, it would still return the same results all of the time.

Method Two:

We use the Snowflake Clone feature to create a zero-copy clone of `mart_arr`. Method Two satitfies the same use cases as Method One where an exact `three ring binder` copy of `mart_arr` is needed each day. This method serves as a backup to Method One by using a methodlogy that is independent of dbt. These cloned tables will not be the primary tables queried or used for reporting, but will provide redundancy and a fail safe to the dbt snapshots in case there is a performance issue with them on a particular day.

Method Three:

Method Three uses the more traditional bottoms-up snapshoting approach that is detailed in this dbt blog post. We use dbt snapshots which implement type-2 Slowly Changing Dimensions over mutable source tables to snapshot the raw source data. We then use date spinning modeling techniques to build a snapshotted `mart_arr` table from the bottom-up. This model uses a Snapshotted ARR Fact and live CRM Account and Product Details dimensions. This allows us to answer questions about how ARR looks with the current state of the Sales CRM Account Hierarchy and Product Details attributes such as Product Tier and Product Delivery. This method allows for full refreshes and while the total ARR amount on a given day will not change during a full refresh, the slicing of that ARR will change and update according to what the live Sales CRM Account Hierarchy and Product Details dimensions report.

Release Train Cadence:

1. Coming Soon

Maintenance Schedule:

1. As needed, Scheduled Maintenance will be performed every two weeks on Friday, from 9am to 11am EST.

Quick Links

Go To Market Analytics Hub

Getting started using Sisense Discovery

Getting Started

To get started we want to make sure you understand:

- What KPIs/PIs are supported using this dashboard
- Key terms that will explain how we account for the metrics
- The data source behind the dashboard
- To explore further, you can create visual and analysis yourself in Sisense. A great way to start is using the Sisense Discovery tool. Want to get started in Sisense head here.
- To go even deeper, you can explore data in snowflake. The benefit of exploring in Snowflake is you can join to additional information (i.e. other data sources). Additional information on exploring in Snowflake can be found here.

Key Terms, Metrics, KPIs/PIs, and Key Field and Business Logic

- ▶ Key Terms
- ▶ Key Metrics, KPIs, and PIs
- ▶ Key Fields and Business Logic

Understanding the Data Sources and Data Models

The ARR Dashboards and Data models use the data models as seen in the ARR ERD

- ▶ Data Lineage
- ▶ Example Queries

Additional Resources

- ▶ Trusted Data Solution
- ▶ EDM Enterprise Dimensional Model Validations
- ▶ RAW Source Data Pipeline validations
- ▶ Data Security Classification
- ▶ Solution Ownership

---

layout: handbook-page-toc title: "License Utilization Analysis"

On this page

{:.no_toc}

- TOC {:toc}

License Utilization Analysis

Understanding how our customers consumes the licenses they ordered is a very important step towards developing a more complete picture of our customers. That would help product identify potential adoption problems, Sales and TAM teams understand better their customer and identify potential risks of churn or downgrade.

This page is a MVC and needs to be seen as a declaration of intent for the next quarters.

Key Metrics, KPIs and PIs

- License Activity Rate is calculated as Number of Active Users / License ordered.
- License Utilization Rate is calculated as Number of Licenses Used / Licenses Ordered.

Future Analysis needed

We had a preliminary analysis showing the License Activity Rate for both SaaS and Self-Managed. This was our first attempt to chart The Data Team would like to focus more for FY21 Q1 on this analysis and would like to go much deeper in the analysis. Some key areas to focus on would be:

- Try to get the same charts for License Utilization Rate
- Try to find correlation between low License Utilization/Activation Rate and churn/downgrade
- Try to breakdown further the analysis by Product Tier, Deal Size, Industry

Further Work Needed

In order to align this page with our quality standards, we will need to:

- Work on a EDM model and write ERDs
- Work on a L2 solution Page

---

layout: handbook-page-toc title: "Manual Upload of Usage Payload"

On this page

{:.no_toc}

- TOC {:toc}

Activities for the User with Usage Data

- If this is your first time manually uploading usage ping data, you will need access to the `snowflake_imports` GCS bucket. Create an access request. Tag the `@gitlab-data/engineers` group in the access request when access is ready to be provisioned.
- If a folder for the customer does not yet exist in the `snowflake_imports` GCS bucket, create one.
- Place usage ping payload files in the `snowflake_imports` GCS bucket, in that customer's folder: First, click into the folder. Then, press the `UPLOAD FILES` button, select the files, and press upload.
- Organize the file(s), ideally the file(s) will be in json format *i.e companyname_usage_payload_00_00_0000.json*
- Usage ping data captures information that details which features and services are being utilized within a GitLab instance. To learn more please visit this link.

---

layout: handbook-page-toc title: "Group Namespace Conversion Metrics"

On this page

{:.no_toc}

- TOC {:toc}

Group Namespace Conversion Metrics: GitLab.com

Once a new group namespace is created on GitLab.com, it is crucial that the namespace creators and users continue to further engage with GitLab.com by adding members to their namespaces, completing certain stage events, and eventually converting from a free trial to a paid plan. It is also important to know how quickly these events and actions are being taken after the namespace is created on GitLab.com.

Questions answered by the dashboard:

- What percentage of new group namespaces upgrade from a free trial to a paid plan within a certain period?
- What percentage of new group namespaces consist of at least two members within a certain period?

- What percentage of new group namespaces engage with the 'Create' stage within a certain period?
- What percentage of new group namespaces engage with the 'Verify' stage within a certain period?
- Has there been a significant change in number of stage adoptions, conversions, and namespaces being created?

The goal of this page:

- Help you understand how to navigate through our Group Namespace Conversion Metrics
- Help you understand our data models (WIP)
- And overall help everyone contribute!

Quick Links

New Group Namespace Conversion Metrics Dashboard	Group Namespace Cohort Analysis Dashboard Views (WIP)

Data Caveats and Constraints

- Filtering by 'Free' or 'Paid' will reflect the group namespace's current status and does not capture the namespace's account status at the time of the adoption event.
- These reports include group namespaces created by new users and by users on existing paid accounts.
- Data may include namespaces created by spam users.
- The data only reflects the GitLab.com product and does not report on self-managed usage.

Data Security Classification

Much of the data within and supporting the Group Namespace Conversion Dashboard is Orange and should not be made publicly available. Care should be taken when sharing data from this dashboard to ensure that the detail stays within GitLab as an organization and that those appropriate approvals are given for any external sharing. In addition, when working with row or record level customer metadata care should always be taken to avoid saving any data on personal devices or laptops. This data should remain in Snowflake and Sisense and should ideally be shared only through those applications unless otherwise approved.

Solution Ownership

- GitLab.com: `@craig-gomes`

- Data Team Subject Matter Expert: `@dpeterson1` `@kathleentam` `@mpeychet\_` `@iweeks` `@jeanpeguero` `@m_walker` `@pluthra`
- New Group Namespace Create Stage Adoption Rate: `@mkarampalas`
- New Group Namespace Verify Stage Adoption Rate: `@jstava` New Group Namespace with at least two users added: `@s_awezec`
- New Group Namespace Trial to Paid Conversion Rate: `@s_awezec`

Key Terms

- Group Namespace: Top-level namespaces, which can hold subgroups and projects. The data in these KPIs and the related dashboard exclude personal namespaces.
- Namespace Creation Date: Date (week) that new group namespace was created
- Verify Stage: Successful if group namespace registers a ci_pipelines event in the first X days
- Create Stage: Successful if group namespace completes the action_monthly_active_users_project_repo event in the first X days
- Paid Plan: Indicates if namespaces are currently paid or free

Available Dashboard Filters

- Paid Plan: Filter by free plans or paid plans
- First X Days Filter: The conversion and adoption rates for these KPIs within X days of the group namespace's creation. Current options include first 3, 7, 14, 30, 60, 90, and 180 days
- Aggregation: Aggregates the namespace creation date by a selected date interval (day, week, month, quarter, year)
- Date Range: Filter charts to only show group namespaces created within the last X days.

Key Metrics, KPIs, and PIs

New Group Namespace Trial to Paid Conversion Rate

New Group Namespace Create Stage Adoption Rate

New Group Namespace Verify Stage Adoption Rate

New Group Namespace with at least two users added

Key Fields and Business Logic

- Data is primarily sourced from Snowflake.

Snippets

Snippets are used to create a string of SQL code that can be reused in different charts. For more information, visit the Sisense SQL Snippets page.

Reference SQL

Snippet Library	Description
denominator_group_namespaces_on_trial	This snippet is used as the denominator for the New Group Namespace Trial to Paid Conversion Rate KPI and calculates the total number of group namespaces created that began a trial within the first X days.
denominator_all_group_namespaces	Query is used as the denominator for the remaining three KPIs listed above. The denominator calculates the number of group namespaces that were created. Unlike the denominator_group_namespaces_on_trial snippet, this snippet does not exclude new group namespaces that haven't started a trial.

Entity Relationship Diagrams

Diagram/Entity	Grain	Purpose	Keywords
Group Namespace Conversion Metrics ERD	Date, Location, Plan Type	Provide insight into new namespace adoption of key stage events and actions	Stage Adoption, Invite Members, New Groups, Paid Conversion, New Namespaces, GitLab.com

Data Platform Solution

Data Lineage

`Coming Soon`

DBT Solution

`Coming Soon`

Trusted Data Solution

- Trusted Data Framework

```
Coming Soon
```

EDM Enterprise Dimensional Model Validations

- Enterprise Dimensional Model Validation Dashboard

```
Coming Soon
```

RAW Source Data Pipeline validations

```
Coming Soon
```

CustomerDB

- Trusted Data Dashboard
    - Reporting on all CustomerDB data tests which include the tdf tag.

```
Coming Soon
```

---

layout: handbook-page-toc title: "People Analytics Overview"

On this page

{:.no_toc .hidden-md .hidden-lg}

TOC {:toc .hidden-md .hidden-lg}

{::options parse_block_html="true" /}

People Analytics

The study of people at work! Human resource departments are flipping their approach to organizational behavior and instead of using judgements or opinions to make decisions are making it fact based with the power of data!

The Benefits of Data in the People Space

- Build a more streamlined talent acquisition process that helps GitLab build a strong and diverse team, as well as provides candidates going through the process a well-crafted experience.
- Drive teams to constantly be improving the experience of GitLab team members, and we mean it when we say "Everyone can contribute." We want individuals to thrive in their careers here and using engagement surveys and KPIs to drive our team goal we put people at the forefront.
- Tie in the other data! Whether it be sales data, engineering data, etc it is all important to us! We want to understand how we can help team members grow in their journey and help leaders understand the impact of their organizations.

Generic Rules and Guidance

- The People Group should be able to do their daily work within the operational software (BambooHR, Greenhouse, etc.).
- People data in the warehouse should be for reporting general People information “up and out” in the organization.
- Always know the roles and users that have access to the data from “cradle to grave” to understand the risk.
- Only what is needed for reporting should be brought into the data warehouse and leave other potentially sensitive data in the operational tools.
- Anonymize sensitive data that is used in metic calculations and reporting to reduce risk when possible.

People Data Sources

BambooHR

HR management system.

Greenhouse

Recruiting and Applicant Management System

PTO By Roots

A slack application that captures team member time off

---

layout: handbook-page-toc title: "People KPI Deck Report"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

People KPI Deck Report

This report takes into account the metrics accounted for in the People Group monthly KPI meeting.

The report can be found in this google report.

Underlying Data:

Documentation for the `rpt_people_kpis` data model can be found here

This captures data for the past 12 months.

---

layout: handbook-page-toc title: "Promotions Report"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

Promotions Report

This report is created for the Total Rewards team.

The reports can be found in this google folder.

Updating the Promotions Report

1. Update the `Promotions Last 12 Months` tab. A. Head to snowflake (login via Okta) B. Run the following code:

```
SELECT *
FROM "PREP"."SENSITIVE"."BAMBOOHR_PROMOTIONS_XF"
WHERE promotion_month BETWEEN DATEADD(year,-1, DATE_TRUNC(month, CURRENT_DATE())) AND DATEADD(month,-1,
DATE_TRUNC(month, CURRENT_DATE()))
ORDER by promotion_date
```

C. Download the resutls and place in the `Promotions Last 12 Months` tab

2. Update the `Budget Spend` tab A. In snowflake, running the following code:

```
SELECT *
FROM "PREP"."SENSITIVE"."BAMBOOHR_BUDGET_VS_ACTUAL"
ORDER BY fiscal_year, fiscal_quarter, division
```

B. Download the results and and place in the Budget Spend tab

3. Refresh the budget vs actual pivot Note: to update the budget totals go here: People Budget Sheetload. this data gets brought in to update the overall report

4. Update the promo budget when refreshed here: sheetload.budget_people

---

layout: handbook-page-toc title: "Slack Dashboard"

Slack Dashboard

This dashboard is used to monitor the percent of messages that arenot DMSa as related to the the Chief of Staff Team to the CEO KPIs. This dashboard can help us understand how to increase the use of public channels and be handbook-first.

The dashboard can be found here

###Updating Data

1. Head to https://gitlab.slack.com/stats
2. Change `Date Range` to all time and export data
3. Use the following code to clean the data (this can be done through R Studio) - https://gitlab.com/gitlab-data/analytics/-/blob/master/transform/general/clean_slack_data.R
4. Once you have the cleaned data heads to sheetload.slack_stats(https://docs.google.com/spreadsheets/d/15a2PVvSs7K_C-EsGq2hKLUYNywxPmE74Ldx-YbXqN-w/edit#gid=673732546) and replace data in current tab with new data

---

layout: handbook-page-toc title: "Talent Acquisition Metrics"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

Talent Acquisition Key Metrics

GitLab is a focused on supporting our team members, and just like other business stakeholders our Talent Acquisition Team within People Group is data focused to constantly improving the hiring process to help teams internally hire the right team members, and also making sure candidates that go through the process have a good experience.

Many of the Talent Acquisition KPIs can be found in our KPI Index.

The Talent Acquisition Metrics Dashboard takes into account our KPIs, in addition to other metrics.

Quick Links

Talent Acquisition Metrics Dashboard

Model Documentation

Using the Dashboard People Metrics Dashboard

The talent acquisition metrics dashboard allows users to come in and get a quick insight of where GitLab stands for the last month, the previous fiscal quarter end, and identifies how the company has been trending for the selected date range (defaulted to past 12 months).

The dashboard allows users to filter on:

- Talent Acquisition Division
- Talent Acquisition Department (a division must be selected to go to the department level)
- Talent Source
- Talent Source Name (a talent source must be selected)
- Talent Acquisition Team (i.e. engineering talent acquisition team, the sales talent acquisition team, etc)
- Talent Acquisition Team Recruiter (dependent on talent acquisition team filter)
- Candidate Coordinator
- Is Prospect (to identify if the candidate is a potential prospect that we have reached out to versus a candidate where they have started the hiring process by applying through other sources)
- Sourcing Team Flag
- Sourcing Team Sourcer (dependent on `sourcing team flag set` = `sourcing team`)

Some key tips while using this dashboard:

1. If you see the metric name in blue there is an accompanying dashboard that will allow you to explore the data deeper. Click on the link to head there!

Talent Acquisition Metrics Data Model

This dashboard is built off `greenhouse_stage_analysis`. The source data comes from Greenhouse.

This data model provides an overview of each candidate that has gone through the hiring process and what stages they have hit in addition to key attributes (i.e. source, recruiter, division the job is for, etc)

The underlying data to create this model is accessible to team members with access to `greenhouse` role.

In Sisense you'll see the underlying query displayed as:

```
  WITH talent acquisition_data AS (

    SELECT *
    FROM legacy.greenhouse_stage_analysis
    WHERE [division_modified=Talent Acquisition__Division]
      AND [department_name=talent acquisition_department]
      AND [source_type=Talent_Source]
      AND [sourcing_team=Sourcing_Team_Flag]
      AND [sourcer_name=Sourcing_Team_sourcer]
      AND [talent acquisition_team=talent acquisition_team]
      AND [candidate_recruiter=Talent Acquisition_Team_Recruiter]
      AND [source_name=Talent Acquisition_Source_Name]
      AND [candidate_coordinator=candidate_coordinator]
      AND [is_prospect=Is_Prospect]


  [talent acquisition_metrics] --- this uses a sisense snippet that identifies all the key metrics i.e. calcuating out which
candidates are considered an applicant, what the application to reviewed rate is, what the application to hire rate is, etc.

  The entire snippet can be found [here]
(https://app.periscopedata.com/app/gitlab/snippet/recruiting_metrics/55081acd27f44d7fb1706d47f44b5ae8/edit)

  SELECT prospected
  FROM metrics
  WHERE month_stage_entered_on = DATE_TRUNC(month, CURRENT_DATE())
```

---

layout: handbook-page-toc title: "Team Member Separations"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

Team Member Separations Report

This report identifies all separated team members. Note: This data is accessible only in snowflake to users with the bamboohr_sensitive_separation role in roles.yml. This data is not accessible in Sisense.

````
Data can be access by running:

```
SELECT *
FROM "PREP"."SENSITIVE"."BAMBOOHR_SEPARATIONS"
```
````

Getting Access

To get access to this report you will need to create an access request to get access to:

1. Snowflake
2. Access to `bamboohr_sensitive_separation` role.

This access request should be reviewed by the people team - Total Rewards Manager or Senior Director, People Success.

---

layout: handbook-page-toc title: "PTO by Roots"

On this page

{:.no_toc .hidden-md .hidden-lg}

- TOC {:toc .toc-list-icons .hidden-md .hidden-lg}

PTO by Roots (Slack)

Goal

GitLab team members use slack to log time off. Tying this data to other data sources can be really helpful in helping us to understand:

- What % of our team members have not taken time off
- Help plan for Family & Friends Day
- Better understand upcoming capacity to plan milestones and OKRs
- Measure impact on performance metrics

As a more specific example, we use PTO data to help understand fluctuations in the Development Department Narrow MR Rate, an indicator of how productive our team members are. We encourage our team members to take the time off to recharge, and by considering time off we are able to explain perceived drops or increases in MR rate over time. For example, an increase in time off from the previous month may explain a drop in the narrow MR rate this month.

Please do not use this data to evaluate the specific amount of time or reason why individual team members are absent. Instead, if you have any concerns about an individual's attendance, you should notify your People Business Partner.

Data Process

The data we capture:

- At the data level, we care about the day that was taken off, not about why. Hence, in our data we remove all reasons for time off in the extract layer or what is available in snowflake.
- The data is then aggregated up to be used for the various data solutions. For example, for MR Rate we show the KPI at a month, division, department level. The time off and the team members information is not made it accessible in Sisense.

Access to this time off data is limited to data team (Parul Luthra, Data Engineer- to be determined). If you are looking to get access, an access request form must be submitted and approved by the Slack PTO admin, and the Senior Director, People Success.

---

layout: handbook-page-toc title: "People Metrics"

People Key Metrics

GitLab is a focused on supporting our team members, and just like other business stakeholders our People Group is data focused when it comes to creating an environment that allows team members to focus on the contribution they have at GitLab.

Many of the People KPIs can be found in our KPI Index.

The People Metrics Dashboard takes into account our KPIs, in addition to other metrics that help drive people data decisions.

Quick Links

People Metrics Dashboard | People Metrics Overview | Model Documentation

People Data Discovery Dashboard | Demo of Data Model | Video Demo of Data Model and Dashboard

Using the Dashboard People Metrics Dashboard

The people metrics dashboard allows users to come in and get a quick insight of where GitLab stands for the last month, the previous fiscal quarter end, and identifies how the company has been trending for the selected date range (defaulted to past 12 months).

The dashboard can be filtered down to 3 levels:

Dashboard View	Filter 1	Filter 2
GitLab Overall	`Breakout = Company`	`Breakout_Division_Department = All Company`
GitLab Overall	`Breakout = Division`	`Breakout_Division_Department = Filter on the applicable division`
GitLab Overall	`Breakout = Department`	`Breakout_Division_Department = Filter on the applicable division-department`

**Note: This only allows you select 1 division or 1 department at a time.

Some key tips while using this dashboard:

1. If you see the metric name in blue there is an accompanying dashboard that will allow you to explore the data deeper. Click on the link to head there!
2. If you need to see everything aggregated in a report --> head to the People Metrics Overview - Current Month. This is accessible within the dashboard.
    - This works in the same way as the drilldown report. For example, filtering on `People_Metrics_View` = 'Division` will show the breakout of all kpis in a report with divisions across the top and GitLab Overall at the end to provide a quick comparision at the division level.
        - This is great for questions like which division has high attrition, or which divisions are we behind on hiring for?
    - By filtering on `People_Metrics_View` = 'Department` it will allow the viewer to dive deeper into the above questions, and determine if there is an issue at the department level.

People Metrics Data Model

The majority of the dashboard is built off `bamboohr_rpt_headcount_aggregation`.

This data model aggregates all relevant people metrics at 4 levels on a monthly basis:

1. `kpi_breakout` -- this shows all metrics aggregated up to the company level
2. `division_breakout`
3. `department_breakout`
4. `eeoc_breakout` -- this shows metrics aggregated to understand diversity metrics (i.e gender, ethnicity, region)

The underlying data to create this model is accessible to team members with access to `bamboohr_analytics` role and then for more sensistive data (i.e. eeoc data and compensation) `bamboohr` role.

Create Your Own Report:

Looking to create a report yourself, check out the demo video that walks through how to use the `discovery` feature in Sisense. This allows users to drag and drop fields.

A preview of this feature can also be found on the People Data Discovery Feature Dashboard

---

layout: handbook-page-toc title: "Pricing Analysis"

On this page

- TOC {:toc}

{::options parse_block_html="true" /}

Pricing Analysis

Pricing is the process of analyzing the value customers receive from GitLab at specific price points along with the profitability of those price points. The analysis also includes understanding how these prices affect the overall business and determining what the optimal price points are for customers and GitLab's profitability. The Pricing Analysis page will provide the information and tools that GitLab team members can use to explore our current pricing strategy and develop insights to optimize it.

The goal of this page:

- Help you understand how to navigate through the Pricing - Customer Discounts Dashboard
- Help you understand the data models used to create the Pricing - Customer Discounts Dashboard.
- Have you asess your understanding by taking a certification most applicable to your role at GitLab.
    - To learn more about how to use the dashboard, take the Dashboard user certification.
    - To learn more about developing Sisense dashboards, take the Dashboard Developer certification
- And overall help everyone contribute!

Quick Links

Pricing - Customer Discounts Dashboard

PnP Test and Research Dashboard

Pricing - Customer Overview Dashboard

Getting started using Sisense Discovery

Getting Started

To get started we want to make sure you understand:

- What KPIs/PIs are supported using this dashboard
- Key terms that will explain how we account for the metrics
- The data source behind the dashboard
- To explore further, you can create visual and analysis yourself in Sisense. A great way to start is using the Sisense Discovery tool. Want to get started in Sisense head here.
- To go even deeper, you can explore data in snowflake. The benefit of exploring in Snowflake is you can join to additional information (i.e. other data sources). Additional information on exploring in Snowflake can be found here.

More of a visual learner? Make sure you watch this Pricing Analysis Solution Training

Key Terms, Metrics, KPIs/PIs, and Key Field and Business Logic

- ▶ Key Terms
- ▶ Key Metrics, KPIs, and PIs
- ▶ Key Fields and Business Logic

Understanding the Data Sources and Data Models

This dashboard is created off the `base_pricing_discount` snippet in Sisense. Sisense snippets and views allow us to reference a simple string of sql in as many charts as we want. This helps to take the guess work out for an end user in understanding logic behind calculations like ARR, or number of seats and also the guesswork out of understanding which joins are possible.

To create your own dashboards off this model you'll simply be able to type the following in your query in Sisense:

> [base_pricing_discount]

`Base_pricing_discount` is built off `mart_discount_arr`, which provides insights into discounts and reseller deals by various customer dimensions (i.e. parent customer, product category, delivery, account owner team, reseller/not, subscription, and accounts).

This mart takes into account the data models as seen in the Entity Relationship Diagram (ERD):

▶ Data Lineage

▶ Example Query