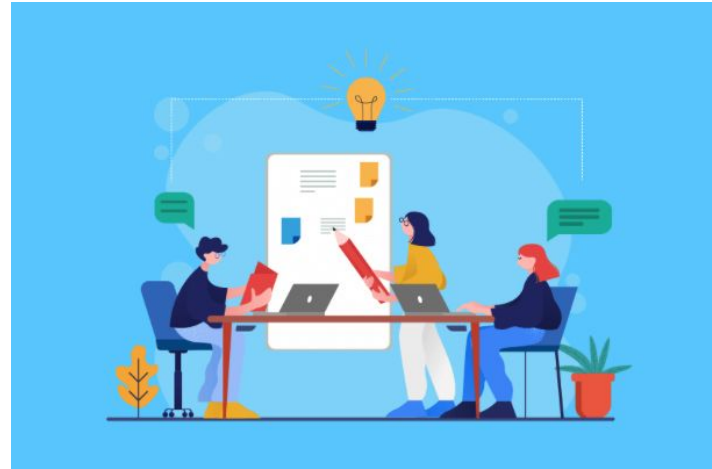# How Vercel Builds Dozens of Metrics from One Heterogenous Table

Thomas Mickley-Doyle

Staff Engineer and Lead, Analytics
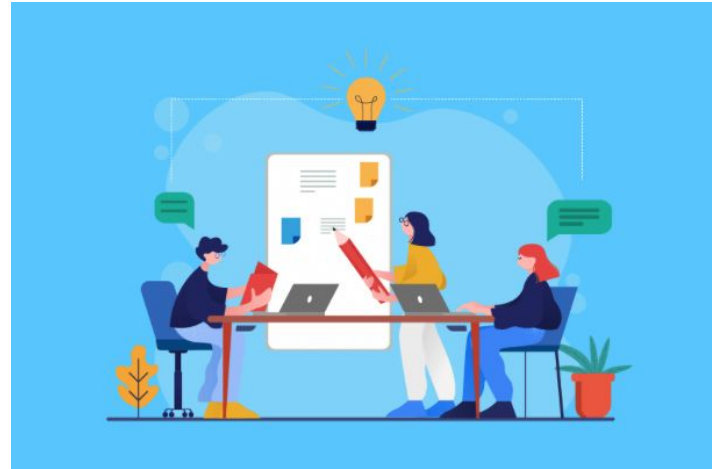
How It Started and How It Is Going

- Data experience
- Starting a data team
- Progress in the first year
- Moving past the first year

How It Started and How It Is Going

- Data experience
- Starting a data team
- Progress in the first year
- Moving past the first year

How It Started and How It Is Going

- Data experience
- Starting a data team
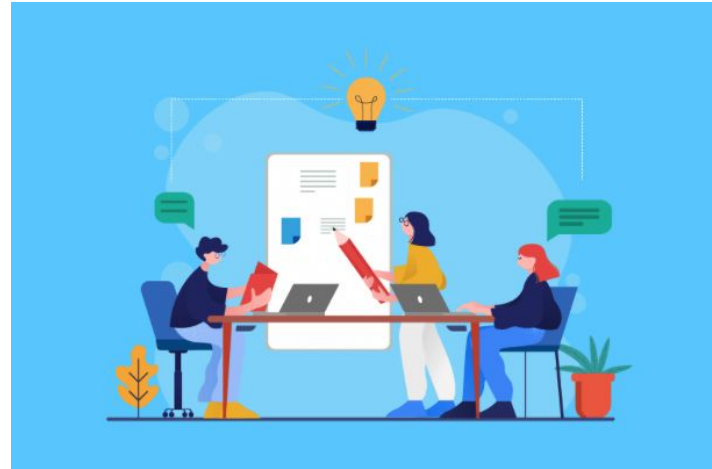- **Progress in the first year**
- Moving past the first year

How It Started and How It Is Going

- Data experience
- Starting a data team
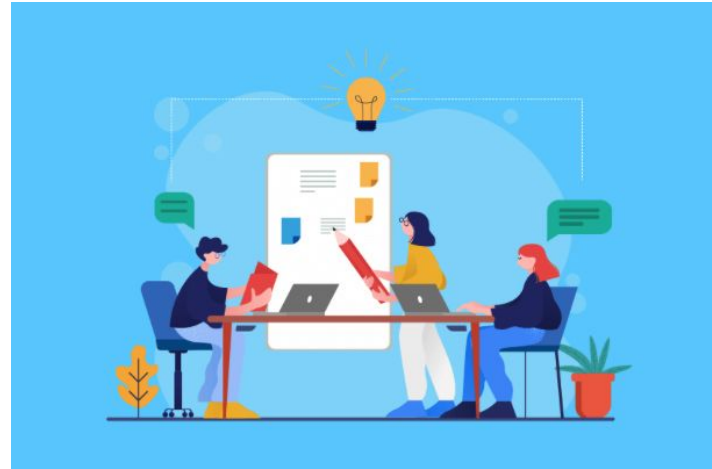- Progress in the first year
- Moving past the first year

Problems

- **Bespoke data pipelines**
- Custom metrics
- Many timeframes
- Random or missing documentation
- Zero SQL standards
- No dashboards

Problems

- Bespoke data pipelines
- **Custom metrics**
- Many timeframes
- Random or missing documentation
- Zero SQL standards
- No dashboards

Problems

- Bespoke data pipelines
- Custom metrics
- **Many timeframes**
- Random or missing documentation
- Zero SQL standards
- No dashboards

Problems

- Bespoke data pipelines
- Custom metrics
- Many timeframes
- **Random or missing documentation**
- Zero SQL standards
- No dashboards

Problems

- Bespoke data pipelines
- Custom metrics
- Many timeframes
- Random or missing documentation
- **Zero SQL standards**
- No dashboards

Problems

- Bespoke data pipelines
- Custom metrics
- Many timeframes
- Random or missing documentation
- Zero SQL standards
- No dashboards

Solutions

- **Standardized and version controlled pipelines**
- Company level metrics everyone can use
- Monthly and weekly timeframes
- Centralized documentation
- All SQL in pipelines requires approval
- Dashboard templates

Solutions

- Standardized and version controlled pipelines
- **Company level metrics everyone can use**
- Monthly and weekly timeframes
- Centralized documentation
- All SQL in pipelines requires approval
- Dashboard templates

Solutions

- Standardized and version controlled pipelines
- Company level metrics everyone can use
- Monthly and weekly timeframes
- Centralized documentation
- All SQL in pipelines requires approval
- Dashboard templates

Solutions

- Standardized and version controlled pipelines
- Company level metrics everyone can use
- Monthly and weekly timeframes
- Centralized documentation
- All SQL in pipelines requires approval
- Dashboard templates

Solutions

- Standardized and version controlled pipelines
- Company level metrics everyone can use
- Monthly and weekly timeframes
- Centralized documentation
- All SQL in pipelines requires approval
- Dashboard templates

Solutions

- Standardized and version controlled pipelines
- Company level metrics everyone can use
- Monthly and weekly timeframes
- Centralized documentation
- All SQL in pipelines requires approval
- Dashboard templates

Audience

Company

Engineering

Product

Design

Operations

Finance

Quarterly planning

Product launches

Board meetings

Many Data Sources into One Table

Dynamo and Cosmos Databases

Kafka Events

AWS Firehose Events

Stripe Logs

Salesforce Tables
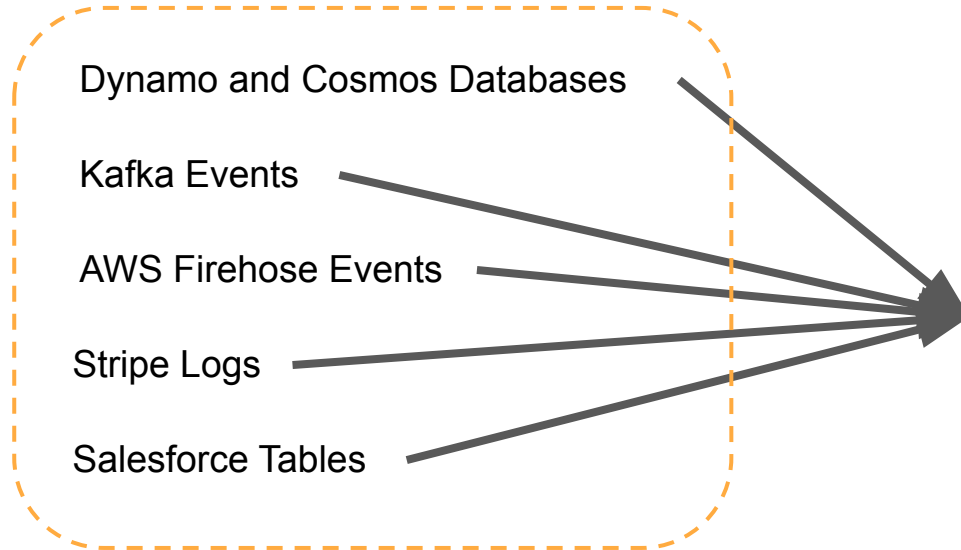
- User ID
- Team ID
- Project ID
- Action
- Product
- Metric Group
- Data source / Code source
- Number of Events
- Date

Metrics Enrollments Table

- Funnel for all metrics
- Versioned schema to prevent unknown errors in metrics
- Structure forces teams to pipe data into a usable format early in the data pipeline
- Adding new metrics does not involve the data team
- Leveraging the different columns, metric hierarchies can be built in the same table

Metrics Enrollments Table

- Funnel for all metrics
- Versioned schema to prevent unknown errors in metrics
- Structure forces teams to pipe data into a usable format early in the data pipeline
- Adding new metrics does not involve the data team
- Leveraging the different columns, metric hierarchies can be built in the same table

Metrics Enrollments Table

- Funnel for all metrics
- Versioned schema to prevent unknown errors in metrics
- Structure forces teams to pipe data into a usable format early in the data pipeline
- Adding new metrics does not involve the data team
- Leveraging the different columns, metric hierarchies can be built in the same table

Metrics Enrollments Table

- Funnel for all metrics
- Versioned schema to prevent unknown errors in metrics
- Structure forces teams to pipe data into a usable format early in the data pipeline
- Adding new metrics does not involve the data team
- Leveraging the different columns, metric hierarchies can be built in the same table

Metrics Enrollments Table

- Funnel for all metrics
- Versioned schema to prevent unknown errors in metrics
- Structure forces teams to pipe data into a usable format early in the data pipeline
- Adding new metrics does not involve the data team
- Leveraging the different columns, metric hierarchies can be built in the same table

Timeframe

- **Daily**
    - Aggregation of events in a given 24-hour period
- **Weekly**
    - Aggregation of events in a rolling 7-day period
        - Examples
            - February 1st to February 7th
            - March 29th to April 4th

- **Monthly**
    - Aggregation of events in a rolling 28-day period
        - Examples
            - February 1st to February 28th
            - March 29th to April 25th

Metrics Enrollments Table

User ID	Team ID	Project ID	Action	Product	Metric Group	Source	# Events	Date
U-100	T-222	P-510	Preview Comment	Experience	Active User	Preview	10	2023-03-01
U-100			Created Account	Growth	New User	Create	1	2023-03-01
U-250	T-222	P-110	Deployed Project	Build	Active User	Deploy	3	2023-03-02
U-425	T-111	P-830	Preview Comment	Experience	Active User	Preview	7	2023-03-02
U-750	T-111	P-830	Deployed Project	Build	Active User	Deploy	2	2023-03-02
U-800	T-777	P-490	Environment Var	Experience	Active User	Security	5	2023-03-05
U-330	T-444	P-610	Created Project	Growth	Active User	Create	1	2023-03-10
U-330	T-444	P-770	Environment Var	Experience	Active User	Security	17	2023-03-10
U-180		P-290	Created Project	Growth	Active User	Create	1	2023-03-11
U-920		P-290	Deployed Project	Build	Active User	Deploy	5	2023-03-11

Metrics Enrollment to Metrics

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

IMPORTANT

Entity ID(s): One more many. Using different combinations of IDs gives deeper views into the metrics

Metrics Enrollment to Metrics: Monthly Active

User ID	Team ID	Project ID	Action	Product	Metric Group	Source	# Events	Date
U-250	T-222	P-110	Deployed Project	Build	Active User	Deploy	3	2023-03-02

March + **U-250** + Deployed Project + Build + **Active User**	**Monthly Active Users**
March + **U-250 + T-222** + Deployed Project + Build + **Active User**	**Monthly Active Team Members**
March + **T-222** + Deployed Project + Build + **Active User**	**Monthly Active Teams**
March + U-250 + **Deployed Project** + Build + Active User	**Monthly Active Developers**
March + U-250 + T-222 + **Deployed Project** + Build + Active User	**Monthly Active Team Developers**
March + **P-110 + Deployed Project** + Build + **Active User**	**Monthly Active Projects**
March + **T-222 + P-110 + Deployed Project** + Build + **Active User**	**Monthly Active Team Projects**

Metrics Enrollment to Metrics: Monthly Active (cont.)

User ID	Team ID	Project ID	Action	Product	Metric Group	Source	# Events	Date
U-425	T-111	P-830	Preview Comment	Experience	Active User	Preview	7	2023-03-02

March + U-425 + Preview Comment + Experience + Active User	**Monthly Active Users**
March + U-250 + T-111 + Preview Comment + Experience + Active User	**Monthly Active Team Members**
March + T-111 + Preview Comment + Experience + Active User	**Monthly Active Teams**
March + U-250 + Preview Comment + Experience + Active User	**Monthly Active Commenters**
March + U-250 + T-111 + Preview Comment + Experience + Active User	**Monthly Active Team Commenters**
March + P-830 + Preview Comment + Experience + Active User	**Monthly Active Projects**
March + T-222 + P-110 + Preview Comment + Experience + Active User	**Monthly Active Team Projects**

Metrics Enrollment to Metrics: Monthly New

User ID	Team ID	Project ID	Action	Product	Metric Group	Source	# Events	Date
U-100			Created Account	Growth	New User	Create	1	2023-03-01

March + U-100 + Created Account + Growth + New User	**Monthly New Users**

One row in Metrics Enrollment might only map to one metric

Metrics Enrollment to Metrics: Notes

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

- **Timeframes are generally Monthly and Weekly**
- Entity IDs can be one ID or a combination of IDs
- Many actions can be linked to a metric and help with deep dive analysis
- Many products can roll up into top level metrics
- Metric groups can be broken down into many timeframes
- The final metric is based on the combination of the inputs

Metrics Enrollment to Metrics: Notes

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

- Timeframes are generally Monthly and Weekly
- Entity IDs can be one ID or a combination of IDs
- Many actions can be linked to a metric and help with deep dive analysis
- Many products can roll up into top level metrics
- Metric groups can be broken down into many timeframes
- The final metric is based on the combination of the inputs

Metrics Enrollment to Metrics: Notes

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

- Timeframes are generally Monthly and Weekly
- Entity IDs can be one ID or a combination of IDs
- Many actions can be linked to a metric and help with deep dive analysis
- Many products can roll up into top level metrics
- Metric groups can be broken down into many timeframes
- The final metric is based on the combination of the inputs

Metrics Enrollment to Metrics: Notes

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

- Timeframes are generally Monthly and Weekly
- Entity IDs can be one ID or a combination of IDs
- Many actions can be linked to a metric and help with deep dive analysis
- Many products can roll up into top level metrics
- Metric groups can be broken down into many timeframes
- The final metric is based on the combination of the inputs

Metrics Enrollment to Metrics: Notes

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

- Timeframes are generally Monthly and Weekly
- Entity IDs can be one ID or a combination of IDs
- Many actions can be linked to a metric and help with deep dive analysis
- Many products can roll up into top level metrics
- Metric groups can be broken down into many timeframes
- The final metric is based on the combination of the inputs

Metrics Enrollment to Metrics: Notes

Timeframe + Entity ID(s)* + Action + Product + Metric Group = Metric

- Timeframes are generally Monthly and Weekly
- Entity IDs can be one ID or a combination of IDs
- Many actions can be linked to a metric and help with deep dive analysis
- Many products can roll up into top level metrics
- Metric groups can be broken down into many timeframes
- The final metric is based on the combination of the inputs

Revisit Before Questions: Metrics Enrollments Table

User ID	Team ID	Project ID	Action	Product	Metric Group	Source	# Events	Date
U-100	T-222	P-510	Preview Comment	Experience	Active User	Preview	10	2023-03-01
U-100			Created Account	Growth	New User	Create	1	2023-03-01
U-250	T-222	P-110	Deployed Project	Build	Active User	Deploy	3	2023-03-02
U-425	T-111	P-830	Preview Comment	Experience	Active User	Preview	7	2023-03-02
U-750	T-111	P-830	Deployed Project	Build	Active User	Deploy	2	2023-03-02
U-800	T-777	P-490	Environment Var	Experience	Active User	Security	5	2023-03-05
U-330	T-444	P-610	Created Project	Growth	Active User	Create	1	2023-03-10
U-330	T-444	P-770	Environment Var	Experience	Active User	Security	17	2023-03-10
U-180		P-290	Created Project	Growth	Active User	Create	1	2023-03-11
U-920		P-290	Deployed Project	Build	Active User	Deploy	5	2023-03-11

Pause for Questions

- Summary of topics
    - No data owners or centralization of SQL logic
    - Target audience
    - Aggregating data sources into one table
    - Metric Enrollments table
    - Timeframes
    - How metrics are calculated from a single source

- Remaining topics if there is time after questions.
    - Segments - subpopulations of users, teams, and projects
    - Value of Data Source Owners

Segments

- **Users**
    - Location
    - Age of customer
    - Number of projects
- **Teams**
    - Number of members
    - Billing plan
    - Number of projects
- **Projects**
    - Age of project
    - Framework
    - Number of contributors

- Each segment can be appended to metrics
    - Examples
        - Monthly New Users by Country
        - Monthly Active Teams by Billing Plan
        - Monthly New Projects by Framework
- The segment data comes from canonical tables
    - Canonical tables are dimension data sources maintained by the Data Team with version controlled SQL in a git repository
    - Canonical tables are versioned by day to allow for metric backfilling and "going back in time" analysis.

Value of Data Sources Owners

- Do not have to track down owners when there is an anomalous event.
- Reduces the amount of questions to the data team related to "who owns the data."
- When there is a code change, the source owners are able to articulate how the change will impact the existing code.
- Before a new data sources is published, it helps engineering teams understand if they should append to an existing data source or support a new source.

Questions