



Dagster Deep Dive

Dagster Running Dagster

How We Use Compass

2/17/2026

Speakers



Alex Noonan

Developer Advocate



Nick Roach

Data Engineer



Anil Maharjan

Data Analyst

Table of contents

- 01 Intro
- 02 How Do We Use AI Analytics
- 03 Do We Still Need BI?
- 04 Data Modeling
- 05 Managing Context
- 06 Dagster+ Integration Storytime
- 07 Demo
- 08 Q&A

“Just use AI for that.”

Move aside “Synergy.”

Move aside “Let’s offline this.”

There is a new empty corporate platitude in town.

- Goals of Today:
 - In a practical sense, how do we actually use AI Analytics?
 - If I were a LinkedIn Influencer, what hot take would I have on BI?
 - How has the use of AI impacted data modeling?
 - Demo something interesting

Simple vs Complex

How Do We Use AI Analytics?

Answering Simple One Offs

Questions defined in your business tables can be answered instantly.

This might not seem like a big deal, but consider what happens every time an ad-hoc question comes in...

The Context Switch Tax

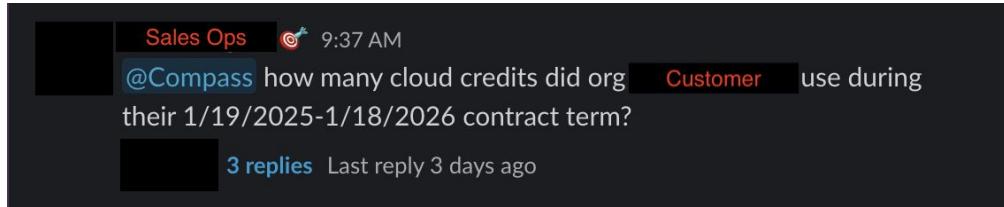
- Stop current deep work
- Switch context to answer
- Lose flow state
- Struggle to get back on track

Self-Service with Compass

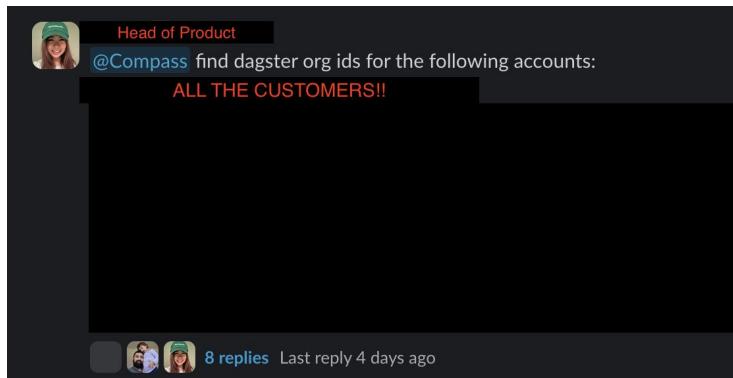
- Force people to find their own answers
- No analyst interruption
- If data isn't there: Auto-create a data request
- If context isn't there: Auto-create a PR

If it's less than 5 clicks in a BI tool or does not require multiple CTEs, AI should handle it.

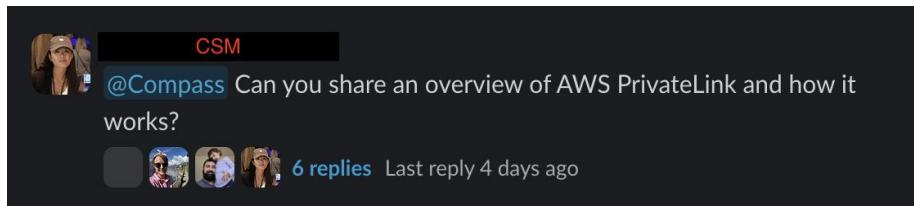
Examples



Sales Ops asking about product usage that would normally go through a Dashboard or an ad hoc report in Salesforce



Product looking up all of those pesky UUIDs that are always important but always a pain



CSM mixing Web Search and then combining it with information about a customer



Anil Maharjan 10:44 AM

@Compass - staging - can you write a sql query that shows the number of anonymous_ids by day that hit 2 different domains in the Compass web data



7 replies Last reply 1 month ago



Me. Forgetting how to write SQL since the bots took over.

Starting Point for Deep Analysis

The Old Way

Hack around in SQL for a while, trying different queries, exploring tables, hoping to stumble onto something useful.

Hours spent before any real analysis begins.

Now

Ask a handful of questions to quickly understand the landscape and identify which direction your analysis needs to go.

Get oriented in minutes, not hours.

With complex analysis, knowing what path to go down is often the hardest part. Find your direction before you commit to a specific approach.



Account Executive PM

@Compass can you help me get up to speed on an account I recently started working with? The account is Customer I can thread specific questions I'm looking to have answered

114 replies

Last reply 6 days ago



Anil Maharjan 11:00 PM

@Compass - can you use the token_usage event in compass events and map out token usage over the last four weeks. break out by token type

19 replies

Why am I doing this at 11PM? Who knows

Account Executive getting up to speed on a new account.

Get the directional answer for your stakeholder to get them through their next meeting and then dig in further.



Anil Maharjan Jan 27th at 1:43 PM

@Compass - staging - how much new ARR is generated by customers who have an upsell closed won opportunity within 12 months of their first New Business opportunity?

Sizing up whether it is worthwhile to proceed with a larger analysis.

Increasing Data Capacity



Guide Data Work

Compass gives you a guide on what new data models you need and any new data engineering involved.



Fast Track the Research

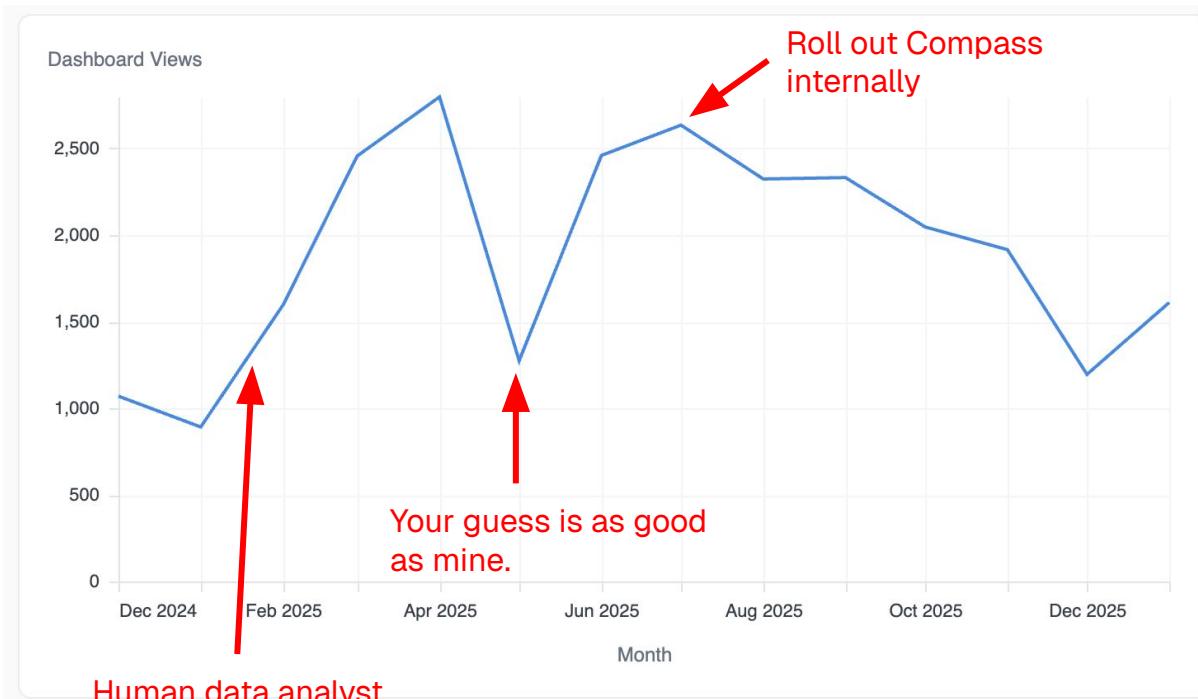
Get past the “hope for the best” queries and get into the actual analysis that makes your team valuable.

Reveal the questions people are actually asking, helping you prioritize data investments.

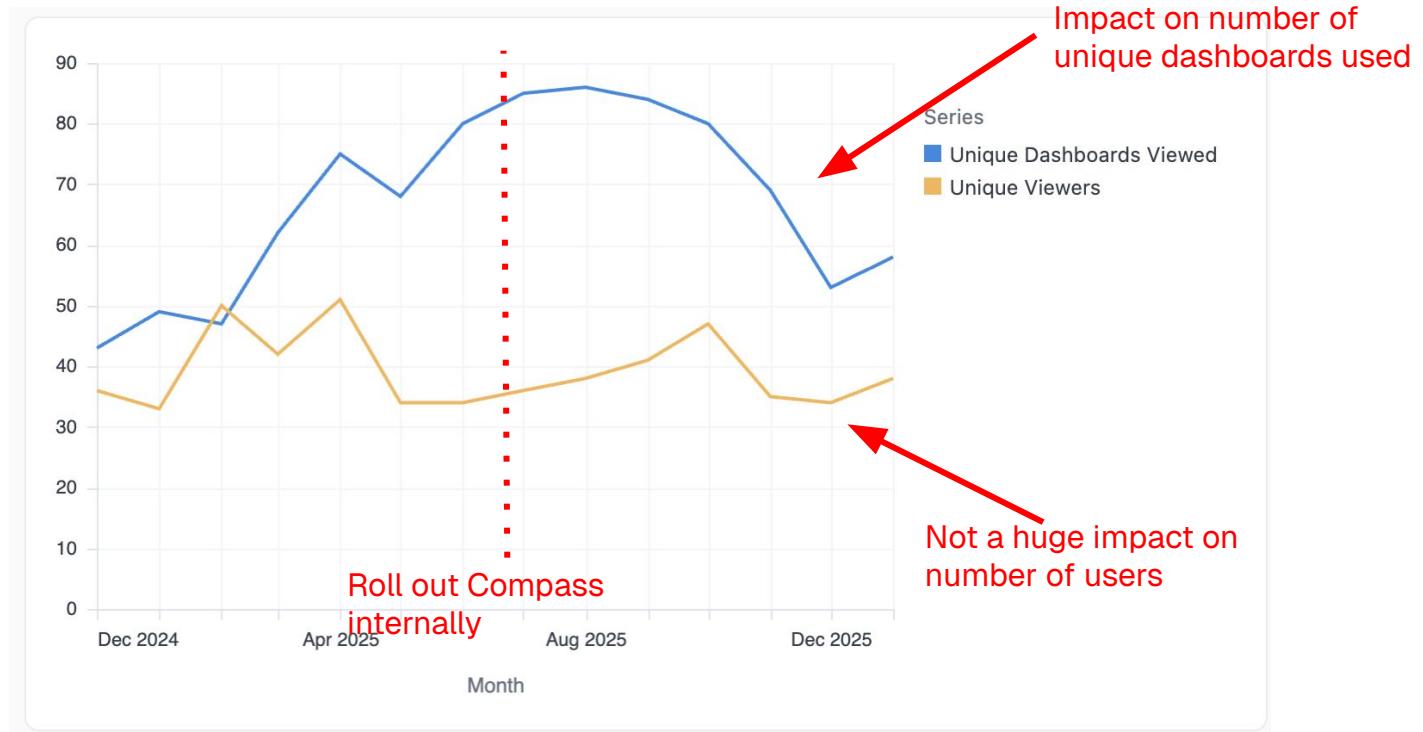
It depends

Does BI Still Have a Place?

Dashboard Views over Time



Unique Users and Dashboards



BI has a place in our stack... for now

- BI is still essential for:
 - Showing metrics that need to be consistently reviewed (e.g. weekly team review)
 - Heavily regulated KPIs (e.g. financial)
- Fewer one-off dashboards that never get reused

Best Practices for Compass

Data Modeling

Our Three Principles for AI Data Modeling



Traditional BI constraints and best practices don't apply. Don't be afraid to design for AI comprehension, not just visual dashboards

1

The Richer the Better

AI can handle summarizing unstructured call transcripts and chat logs far better than an analyst.

2

Minimize Overlap

Reduce overlap between tables to give Compass an easier path. Give the AI the fewest possible paths to the right answer.

3

Go Wide and Nest

It's okay to have really wide tables, AI handles it well. Don't be afraid to nest reasonably sized data to avoid

Best Practices for Compass

Managing Context

Data Documentation

- Context is king... blah blah blah
- Compass generates docs based on scanning your table
 - Updates weekly
- Pro Tips:
 - Persist your documentation
 - Add checks for missing docs in CI/CD
 - Add a command for your AI agent

The screenshot displays a data documentation interface with two main sections: a table summary and a CI/CD pipeline.

SALES_MEETING_TRANSCRIPTS Table Summary

- Keywords:** Sales calls, meeting transcripts, Gong, call recordings, sales pipeline, opportunity tracking, conversation analysis, call metrics, participant engagement, sentiment analysis, account management, deal progression, Zoom/Teams/Google Meet
- Overall Dataset Characteristics:** Total Rows: 9,581 completed sales meeting transcripts
- Data Quality Observations:**
 - Excellent data completeness for core call identifiers (0% nulls on CALL_CONVERSATION_ID, CALL_STATUS)
 - Significant null values (10-16%) for meeting metadata fields (end datetime, duration) - some calls lack complete recording information
 - 100% null columns present (PROVIDER_UNIQUE_ID, YEAR1_AMOUNT) indicating eit
 - Call spotlight analysis has 5-13% null rates, suggesting not all calls receive Gong AI i
 - All calls marked as "COMPLETED" status
 - Dataset spans multiple years with recent data through 2025
- Notable Patterns:**
 - Conference calls dominate (majority of CALL_DIRECTION values)
 - Sentiment scores normally distributed around 0 (-0.85 to +0.73 range)
 - Significant variation in call duration (0-5418 seconds) and participant engagement r

Data Pipeline CI/CD

on: pull_request

- ✓ check_paths
- ✓ Manage Databases for PR (PURINA)
- ✓ Manage Databases for PR (DWH_REPORTING)
- ✓ Manage Databases for PR (DWH_SCIENCE)
- ✓ Manage Databases for PR (COMPASS_DAGSTER_PLUS_TABLES)
- ✓ Manage Databases for PR (DWH_OPERATIONS)
- ✓ Check for Untested dbt Models (DWH_REPORTING)
- ✓ Check for dbt Models Without Descriptions (DWH_REPORTING)**

Context Management

- Context is king
- Data team can approve, deny, or modify the context before it is merged

Governance

[Github Context Repo](#)

| Status | Request type | Title | Updated | Context | Controls |
|----------------------|----------------|--|-----------|-------------------|---|
| Open | Context update | DATASET MONITORING: Update DWH_REPORTING.BUSINESS.COMP... | 2/4/2026 | 🔗 | ✓ ✗ |
| Open | Context update | DATASET MONITORING: Update DWH_REPORTING.PRODUCT.COMP... | 2/2/2026 | 🔗 | ✓ ✗ |
| Open | Context update | CONTEXT: Concurrency metrics [REDACTED] | 2/2/2026 | 🔗 | ✓ ✗ |
| Open | Context update | CONTEXT: [REDACTED] concurrent slots post-migration, not ... | 1/28/2026 | 🔗 | ✓ ✗ |

20251023_New-ARR-definition-expanded-Q3-2025-to-include-New-Business-Upsell-Renewals-vs-Q1-Q2-FY25-which-only-counted-New-Business-for-targets.yaml

csbot CONTEXT: New ARR definition expanded Q3 2025 to include New Business,... 39105c4 · 4 months ago

8 lines (8 loc) · 557 Bytes

[Code](#) [Blame](#)

```
topic: New ARR definition change starting Q3 2025
incorrect_understanding: New ARR targets are consistently measured as New Business
only across all quarters of FY25
correct_understanding: Starting Q3 2025, the definition of New ARR changed to include
New ARR derived from New Business, Upsell, and Renewals. Prior quarters (Q1 and
Q2 FY25) used New Business only, but Q3 targets include all opportunity types.
search_keywords: New ARR, Q3 2025, definition change, New Business, Upsell, Renewals,
FY25, targets, opportunity types, quarterly metrics
```

Dagster+ Integration in Action

Story Time

The Pipeline Problem



40-50%

Failure Rate

Postgres → Snowflake

Pipeline

A critical data pipeline was failing nearly half the time, causing reliability issues and data delays.

The question: What was causing these failures?

Compass Finds the Root Cause

 **Compass - staging APP**

 **compass insight:**

`DagsterExecutionStepExecutionError` may be the most significant bottleneck, affecting 146 runs over the past 2 weeks and causing 51+ assets to fail repeatedly in `compass_analytics_hourly_job`

- The root cause may be a 502 error from the Dagster Cloud artifact service (https://element1.agent.dagster.cloud/gen_artifact_get) when loading component definitions
- `compass_analytics_hourly_job` experienced 32 affected reruns from this error, with most recent failures on Dec 2 @ 23:04 UTC
- Generic `Failure` errors come in second (99 affected runs), often related to Hightouch connector setup issues like "Connector 'disclose_cigarettes' cannot be synced as it has not been setup"
- `DagsterExecutionInterruptedError` (42 runs) suggests some jobs are being interrupted during execution, possibly due to resource constraints or timeout policies
- The error pattern shows relative stability over the 2-week window, though `compass_analytics_hourly_job` needs immediate attention

details in thread 

 **Anil Maharjan**

Could you tell me a bit more about the errors associated with: `compass_analytics_hourly_job`

detailed breakdown of `compass_analytics_hourly_job` errors

summary overview

the job has had **53 affected runs** over the past 2 weeks, with issues concentrated in 2 main areas:

- `prod_replication step` - 32 runs (60% of failures)
- `fivetran_google_sheets_compass_customer_id_exclusion step` - 15 runs (28% of failures)
- `other dbt & replication steps` - 6 runs (12% of failures)

Finding: Tables were being written while simultaneously being dropped

The Fix & Results

root cause analysis

based on snowflake documentation and the error patterns, this is likely:

session isolation issue + race condition:

- snowflake temporary tables are session-scoped (visible only to the session that created them)
- sling may be creating temp tables in one session and trying to merge from another
- when the source session drops the temp table, the merge session can't find it

when it fails on organizations:

- sling successfully reads ~240-250 rows from postgres
- sling creates a temporary staging table (ORGANIZATIONS_TMP) in snowflake
- sling attempts to merge this temp table into the production COMPASS.ORGANIZATIONS table
- **one of two things happens:**
 - the temp table is dropped before the merge completes (session timeout/race condition)
 - the temp table is created in one session but referenced from another (snowflake session isolation issue)
 - record counts don't match due to duplicate detection or filtering logic

Result



Fix deployed — pipeline now works great!

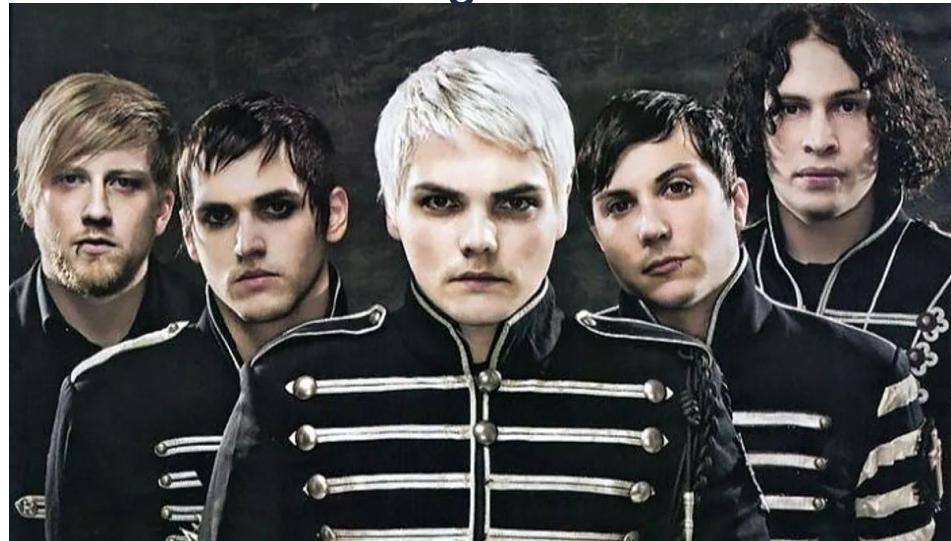
Demo



**For those of you born in a year
starting with a 1**



**For those of you born in a year
starting with a 2**



Key Takeaways

Simple Questions

Self-service eliminates context switching. Missing data?
Auto-creates requests.

Deep Analysis

Your starting point. Find direction before hacking SQL for hours.

Data Modeling

Nested columns, wide tables, minimize overlap



Compass increases your data team's capacity by revealing what people actually need — and helping them get answers faster.

Questions? → compass.dagster.io

Q&A

Check Out Compass



Compass & Dagster+



Stay Connected

