

Pre-aggregations Workshop



Pavel Tiunov



Ryan Pei



Community Code of Conduct

- We want to foster an open and welcoming environment where everyone feels they belong in the Cube.js community.
- The full text of our Code of Conduct is available at https://github.com/cube-js/cube.js/blob/master/CODE_OF_CONDUCT.md
- Any instances of inappropriate/unacceptable behavior can be reported to conduct@cube.dev.

Some quick notes

- If you have any questions during the workshop, please feel free to type them in the “Q&A”.
- We will be using Cube Cloud during the “Getting started” section.
- Recording of today’s workshop will be posted on the Cube Dev YouTube channel.
- All attendees will receive a post-event survey and we’d appreciate your feedback to help us with future events.
 - In the survey, you can also provide your address so we can mail your pair of Cube.js socks 😊

What we will discuss today

- Why use pre-aggregations?
- Getting started
- Tips & tricks
- Resources
- Q&A

Why Cube.js pre-aggregations?

In the world of big data

Applications' backend datasets are reaching enormous scales...

...and yet, to remain competitive, your applications must be **fast, responsive** and **cost-effective** at presenting information to users.

What is a pre-aggregation?

Condensed versions of source data that are partitioned, stored, and refreshed in a way that optimizes for speedy and efficient queries

By far, the most common type of pre-aggregation are **rollups**.

You can think of the “data condensing” mechanism of rollups as analogous to the **GROUP BY** function in SQL.

Example of pre-aggregation data

Raw Data

| Order ID | Order Time | Product | Quantity |
|----------|-------------------------|-------------|----------|
| 1 | July 20, 2021 10:00 AM | Sugar Cubes | 2 |
| 2 | July 22, 2021 08:00 AM | Ice Cubes | 4 |
| 3 | July 26, 2021 03:00 PM | Ice Cubes | 2 |
| 4 | July 28, 2021 12:00 AM | Sugar Cubes | 1 |
| 5 | August 3, 2021 12:00 PM | Ice Cubes | 4 |
| 6 | August 5, 2021 09:00 AM | Ice Cubes | 1 |

Row Count

~1M (5 years of data)

⚡ Pre-Aggregated Data

| Order Time | Product | Quantity |
|------------|-------------|----------|
| July | Ice Cubes | 6 |
| July | Sugar Cubes | 3 |
| August | Ice Cubes | 5 |

Row Count

~120

Before Pre-aggregations

Timeline of a typical analytics project, before pre-aggregations:

Day 1: Customers complain your reporting/analytics features are slow

Day 2: Team decides to add a caching layer and/or reorganize backend data

Day 96: Many debates over cache invalidation tactics and scaling your data warehouse follow... project is deprioritized

Pre-Aggregations to the rescue!

Benefits of pre-aggregations:

- Speed up your queries
- Reduce your data processing costs
- Save your engineering team days/weeks of headaches

And why is Cube well-suited to provide and manage pre-aggregations?

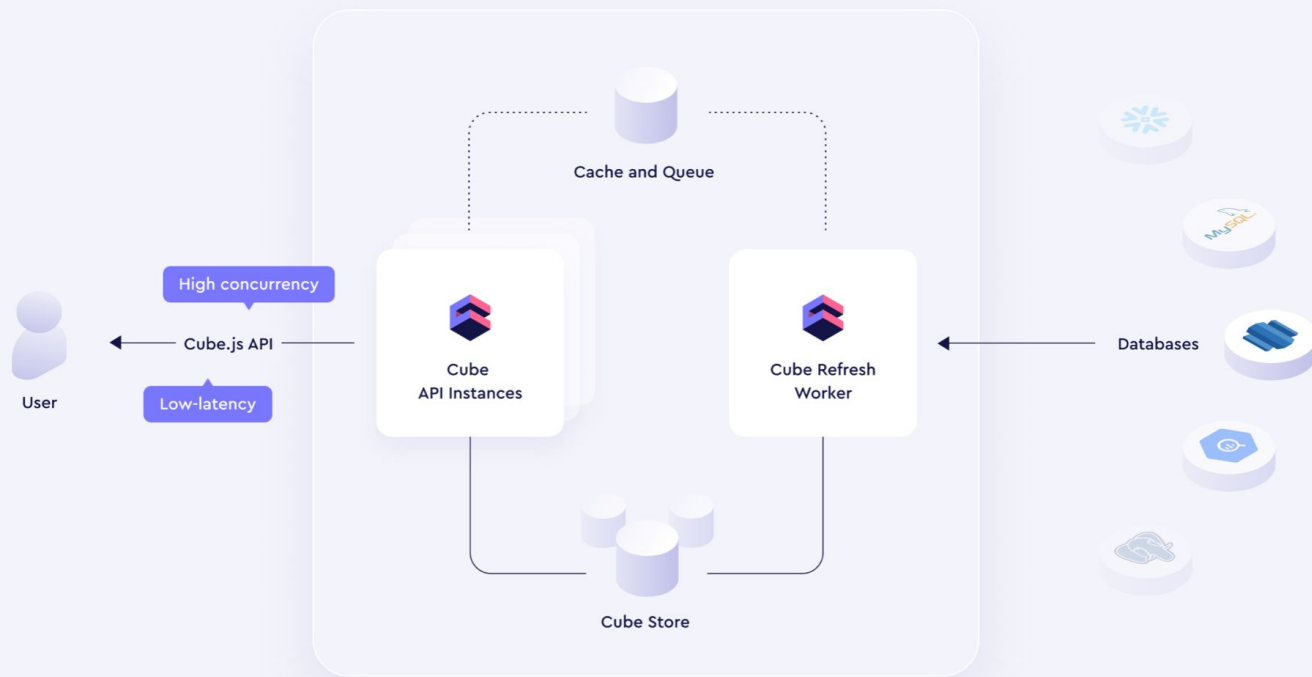
Two components are key: **the API and Cube Store**.

The API makes pre-aggregations easy for developers to create and use.

And Cube Store is the best at storing and serving pre-aggregated data. Its designed to store multi-billion row rollup tables.

Serving time doesn't depend on the pre-aggregation complexity.

Pre-Aggregations behind the scenes



Getting started with pre-aggregations

Let's now login to Cube Cloud!

Tips & tricks

FAQ

- When does it make most sense to use pre-aggregations?
 - **count**, **sum**, **max**, **min**, **countDistinctApprox**
 - **avg** and other calculated measures are also doable, but more complicated
- How do I increase the chances of queries hitting my pre-aggregations?
 - More dimensions, more measures → bigger pre-aggregation → more queries it serves (but that comes with a tradeoff of slightly longer query times)

Production checklist

- ✓ Cube Store
- ✓ Dedicated Refresh Worker
- ✓ Dedicated Redis
- ✓ Batching and/or exporting data
- ✓ Pre-aggregation build/refreshes may need tuning

Resources

To learn more about pre-aggregations

- Documentation: <https://cube.dev/docs/caching>
- Blog post:
<https://cube.dev/blog/high-performance-data-analytics-with-cubejs-pre-aggregations/>
- Community support:
 - Discourse: <https://forum.cube.dev/>
 - Slack: <http://cube-js.slack.com/>



