Season of Docs Case Study Example

Create the GeomScale Reference Documentation

GeomScale is a research and development project that delivers open source code for state-of-the-art algorithms at the intersection of *data science*, *optimization*, *geometric*, and *statistical computing*. The current focus of GeomScale is scalable algorithms for sampling from high-dimensional distributions, integration, convex optimization, and their applications. One of our ambitions is to fill the gap between theory and practice by turning state-of-the-art theoretical tools in geometry and optimization to state-of-the-art implementations. We believe that towards this goal, we will deliver various innovative solutions in a variety of application fields, like finance, computational biology, and statistics that will extend the limits of contemporary computational tools. GeomScale aims to serve as a building block for an international, interdisciplinary, and open community in high dimensional geometrical and statistical computing.

The main development is currently performed in **volesti**, a generic open source C++ library, with R and Python interfaces, for high-dimensional sampling, volume approximation, and copula estimation for financial modelling. In particular, the current implementation scales up to hundred or thousand dimensions, depending on the problem. It is the most efficient software package for sampling and volume computation to date. It is faster by orders of magnitude compared to packages that solve the same problems in several cases. It can compute challenging multivariate integrals and approximate optimal solutions in optimization problems.

It has already found important applications in systems biology, for [analyzing large metabolic networks](#) (e.g., the latest human network), and in FinTech for detecting shock events and evaluating portfolios performance in stock markets with thousands of assets. Other application areas include AI and in particular approximate weighted model integration and data-driven power systems in control.

Website: https://geomscale.github.io

Problem Statement/Proposal Abstract

GeomScale develops scientific and research oriented software, therefore, detailed and well written documentation is an important requirement to reach the communities, the users, the practitioners, and the researchers it may concern.

GeomScale's software can solve several complex and high dimensional problems efficiently in various fields; so our aim now is to create the essential tools to make it well-known and easily accessible across open source communities.

The main bottleneck for onboarding new contributors to GeomScale is the nature of the project that requires knowledge from various fields of advanced applied mathematics and theoretical computer science. The creation of a complete and detailed documentation will be a valuable tool towards overcoming that burden. Therefore, solid documentation is a stepping stone to grow our organisation to become the reference open source software in geometric and statistical computing in high dimensions.

https://geomscale.github.io/GSoD-application/

Project Description

Creating the proposal

GeomScale heard about the Season of Docs program from the participation in Google Summer of Code program. The board discussed the program at their meeting, and agreed to ask some core contributors of GeomScale for their feedback. The feedback was very positive and then the board members drafted a proposal and asked for more feedback from contributors to end up in the final version of the proposal.

Budget

None of GeomScale members or contributors have worked solely as technical writers but they have participated in projects with documentation writing. The location of funding was successful and we didn't face any issues ending spending more or less than the grant award.

Participants

The core team working on this project was:

- Vissarion Fisikopoulos (GeomScale board, GSoD coordinator)
- Apostolos Chalkis (GeomScale board, admin help)
- Elias Tsigaridas (GeomScale board, subject-matter expert)
- Alex Manochis (technical writer)

We have collaborated with Alex Manochis in the past. Alex is a contributor of volesti library. This is very important for our project since the technical background in both computational mathematics and templated generic C++ programming is a rare skill. We understood this while examining technical writers' proposals, where Alex was clearly the best candidate.

Timeline

In the beginning of the GSoD, the GeomScale members together with the technical writer worked on creating the infrastructure to host the volesti documentation (read the docs, Sphynx, doxygen). Then the technical writer wrote the C++ documentation and the GeomScale members reviewed the result.

The final schedule is as follows:

Stage	Completed By
Review docs audit	17 May
Friction log use case	1 June
Review friction logs with Vissarion and Apostolos, answer queries	5 June
Doc infrastructure development https://github.com/GeomScale/volesti/pull/232	6 Jun
Final review of doc infrastructure	1 Jul
First draft of docs	28 June
Draft reviewed by Vissarion, Apostolos and Elias	28 July
Most of GeomScale members was on vacation Aug 1-20	--
Begin testing of new docs in community	21 August
Testing feedback incorporated	10 September
Copyedit and proofread of new docs	17 September
Final version of docs (github PR https://github.com/GeomScale/volesti/pull/253)	17 November
Merge docs PR and official documentation launched	18 November
This case study created	19 November
Case study submitted	30 November

Results

The documentation of volesti is created and several legacy documentation items have been updated.

We have created the infrastructure to host the volesti documentation (read the docs, Sphynx, doxygen).

Also github actions for testing that the documentation does not break after a commit or PR are added. The actions can be seen as a guide to build the docs by the users.

Metrics and analysis

GeomScale receives an average of 15 pull requests a quarter to add a new feature or optimization or propose a bug fix. The majority of these pull requests (>80%) are from previous contributors. We believe that this improved documentation will result in more pull requests and more pull requests from new contributors.

We will track three metrics: (a) number of new feature pull requests, (b) number of pull requests from new contributors, (c) standard metrics (number of views, downloads, web traffic to GeomScale site, time on page), monthly after the documentation is published. We will also track the number of contributors who have made more than three contributions overall, starting quarterly after the documentation is published.

In our proposal we stated that we would consider the project successful if, after publication of the new documentation at least three of the following hold:

- The number of new feature pull requests increases by 10%
- The number of pull requests from new contributors increases by 15%
- The number of contributors who have made >2 contributions increases by 5% (beginning the quarter after the documentation is published)
- The standard metrics increased by 10% on average
- The number of forks and stars in our repository increased by 10%

Note that here for the scope of this case study we consider the publication of the infrastructure i.e. 1 July 2022. After that point 19 new PRs have been submitted, 13 have been merged now, 8 were about new features all by new contributors. Thus we consider the project **successful**.

In the next few months we expect to see the effect of the publication of the full documentation.

Summary

To sum up, our project is considered successful! We achieved our documentation deliverables and our metrics seem to be in line with our goals.

Our experience was very positive since we now have basic documentation to help new users start and use our project more effectively.

Moreover, the whole process helped our technical writer to get engaged even more into our project and the GeomScale members to gain valuable experience with writing documentation.

We strongly encourage other projects to invest time in building their documentation.

We are looking forward to continue towards building a complete documentation as described in our proposal https://geomscale.github.io/GSoD-application/#the-problem i.e. write *tutorials*, *how-to guides* and *explanations*.