Cassandra GSoD 2019 application

Overview

Google is sponsoring a program called Season of Docs. Here is their description taken from their site:

> The goals of Season of Docs are to give technical writers an opportunity to gain experience in contributing to open source projects, and to give open source projects an opportunity to engage the technical writing community.
>
> During the program, technical writers spend a few months working closely with an open source community. They bring their technical writing expertise to the project's documentation, and at the same time learn about open source and new technologies.
>
> The open source projects work with the technical writers to improve the project's documentation and processes. Together they may choose to build a new documentation set, or redesign the existing docs, or improve and document the open source community's contribution procedures and onboarding experience.
>
> Together, we raise public awareness of open source docs, of technical writing, and of how we can work together to the benefit of the global open source community.

The goal of this page is to put together a successful application for Apache Cassandra so we could get a Technical Writer in the community to improve the documentation in Cassandra. Program details are here.

Volunteers

The following people have volunteered to help out with this project. If you'd like to help out, please feel free to **add** your name. Alternatively, if things have changed on your end, please cross out your name.

Note: Please subscribe to the Season of Docs Announce mailing list for updates.

Sr.	Name	Email
1	Dinesh Joshi	djoshi@apache.org
2	Nate McCall	zznate.m@gmail.com
3	Chris Lohfink	clohfink@apache.org
4	Ben Slater	ben.slater@instaclustr.com
5	Rahul Singh	rahul.xavier.singh@gmail.com
6	Aaron Ploetz	aaronploetz@gmail.com
7	Horia Mocioi	a.mocioi@gmail.com
8	Stefan Miklosovic	stefan.miklosovic@instaclustr.com
9	Laxmikant Upadhyay	laxmikant.hcl@gmail.com

Open Source Organization Administrators

TBD. The program requires each Open Source organization to have at least 2 administrators.

Application

Here are some tips to create a successful application. **Timeline & Deadlines are available here.** Mentoring organizations (us) can start submitting applications as soon as April 2 but the final deadline to get the application in is **April 23, 2019.**

We need to get together the following -

- ☑ Project Ideas
- ☑ Provide details about working with Technical Writers and Documentation
- ☐ Previous experience working in similar programs such as Google Summer of Code, etc.

☐   A link to any material we publish about our participation in GSoD such as a blog post.

Potential Project Ideas

Please feel free to add project ideas below.

1. ~~Update Cassandra documentation for 4.0 release~~
2. ~~Operator's Handbook for Cassandra~~
3. ~~Improving CQL documentation~~
4. ~~Detailed nodetool documentation for operators~~
5. ~~Get rid of all TODOs on Cassandra Website (docs)~~
6. ~~Common Application Architecture patterns using Cassandra~~
7. ~~Document Cassandra Architecture 2.x, 3.x, 3.0.x, 4.x in detail~~

Project Descriptions

1 - Apache Cassandra 4.0 Documentation Update

Project Name: Apache Cassandra 4.0 Documentation Update

Project Description: Review and update existing Apache Cassandra documentation to ensure that it is aligned with the forthcoming Apache Cassandra 4.0 release. Our proposed approach to this would be to first put out a call to the cassandra-dev mailing list for Cassandra contributors to highlight areas of the document which are in need of updates. Secondly, we would put together a structured review of the existing documentation to identify potential gaps. For example: in the structured review, the technical writer could compare the Cassandra configuration file documentation (http://cassandra.apache.org/doc/latest/configuration/cassandra_config_file.html) with the configuration options actually contained in the source (https://github.com/apache/cassandra/blob/trunk/src/java/org/apache/cassandra/config/Config.java).

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org

2 - Cassandra Operator's Handbook

Project Name: Apache Cassandra Operational Handbook

Project Description: Apache Cassandra is a distributed database that is seeking a qualified technical writer to create an extensive operator's handbook. This handbook would cover topics from Installation to Tuning & Troubleshooting. There are several top level topics which we would like to cover, including but not limited to:

1. Installation
2. Configuring
3. Securing
4. Monitoring
5. Tuning
6. Safely restarting
7. Adding, removing, & replacing nodes
8. Upgrading Cassandra
9. Backing up & Restoring
10. Expanding, shrinking & replacing
11. Troubleshooting
12. Miscellaneous tools & utilities

There may be additional topics to be added later. This handbook would also cover prior versions of Cassandra (2.x, 3.x, 3.11.x, etc) highlighting the differences between them. We estimate that each topic might take somewhere around 1-5 pages to cover the topic. The project will also include adding sample code, commands and illustrations.

Note: The scope of this project is undoubtedly large. We will work with you, the technical writer, to ensure that it is scoped appropriately.

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org

3 - Improving CQL Documentation

Project Name: Apache Cassandra CQL Documentation

Project Description: Apache Cassandra is a distributed database that is seeking a qualified technical writer to create documentation for Cassandra's Query Language (CQL). CQL, like SQL, is a query language that allows users to interact with Cassandra. Similarly, CQL provides syntactic structures such as DDL (Data Definition Language), DML (Data Modification Language) and DCL (Data Control Language). CQL also has well defined grammar and syntax which may vary with the Cassandra version. While Apache Cassandra 4.0 will be the newest version, the documentation should also provide support for the earlier versions of Cassandra 2.x, 3.x, and 3.11.x. There are approximately 40 CQL commands, with several options available for each. The document should contain example usages for each CQL command, as well as how its use varies amongst the different versions. This could theoretically cover also some core schema design principles and best practices, as well as how to design tables in order to achieve optimal performance and usability.

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org

4 - Detailed Nodetool Documentation for Operators

Project Name: Apache Cassandra Nodetool Documentation for Operators

Project Description: Apache Cassandra is a distributed database that is seeking a qualified technical writer to create documentation for Cassandra's nodetool. nodetool is a command line tool which Cassandra operators will often use for their day-to-day operations. It is one of the mechanisms by which an operator interacts with a node in the cluster. Therefore, it is a critical aspect of building and managing a Cassandra cluster. This documentation should support Cassandra 2.x, 3.x, 3.11.x, 4.x. There are approximately 80 nodetool commands, with several options available for each. The document should also contain examples for each nodetool command, as well as how it varies amongst the different versions. It would be very nice to have not only the "how" behind a particular command, but also "why" these commands would be used. The operator should know what each command is for and under which circumstances one should use them. It may be useful for the commands to be grouped by category (ex: commands for grouping SSTables, et al). There could also be some definition around frequent scenarios and workflows to highlight using various nodetool commands to identify, debug and resolve issues within a Cassandra cluster, in terms of every-day usage.

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org

5 - Resolve TODOs in Apache Cassandra Documentation

Project Name: Resolve TODOs in Apache Cassandra Documentation

Project Description: The Apache Cassandra documentation contains a number of sections highlighted as "TODO" (or tasks yet to be completed). These are:

- Operating Cassandra / Backups
- Operating Cassandra / Bulk Loading
- CQL / Data Definitions (limited review points)
- Data Modelling
- Dynamo / Gossip, Failure Detection and Token Ring
- Architecture/Guarantees
- Operating Cassandra/Hints
- Architecture/Overview (this is proposed as a separate project)
- Operating Cassandra/Read Repair

Many members of the community have produced material to cover these topics (including public blog posts, Stack Overflow posts, etc). The most productive approach to completing these sections may be to request good references nd material from the mailing lists and synthesize into a concise document of the Apache Cassandra documentation set.

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org

6 - Common Application Architecture patterns using Cassandra

Project Name: Common Application Architecture patterns using Cassandra

Project Description:

Apache Cassandra is a Distributed Database which is very versatile and can be applied to various problem domains. If applied correctly, it can scale linearly for the user. This documentation project will define the common architectural patterns which are successfully applied to various domains, such as time series data (ex: Financial Stock Data, Weather Data) and query-based data models. There are many patterns and best practices which could be documented as a part of this project. At a minimum, we should try to document the most common use-cases, as well as calling out architectural modifications for the various versions of Cassandra. This project can also include use cases around integration with other technologies such as tools like Apache Solr and ElasticSearch (for search and analytics). The scope is open and can be defined at a later point.

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org

7 - Document Cassandra Architecture 2.x, 3.x, 3.0.x, 4.x in detail

Project Name: Document Cassandra Architecture 2.x, 3.x, 3.0.x, 4.x in detail

Project Description:

Apache Cassandra is a Distributed Database which is very versatile. There are various stable versions of the database which are available for use. This documentation project will try to produce detailed documentation of Cassandra's internal architecture for all stable versions. This is critical for new developers, as well as experienced operators to understand how the database functions "under the hood." This will also help new developers gain a better understanding of the internals, thereby improving the ease and quality of their contributions. There are various aspects which can be covered here including, but not limited to:

1. General Cassandra Architecture
2. Code Oganization
3. Storage Engine
4. Internode Communication
5. Native Protocol
6. Secondary Indexes
7. Materialized Views
8. Gossip

The documentation should describe the overall architecture, relevant classes, and differences in the various versions of Cassandra.

Note: The scope of this project is undoubtedly large. We will work with you, the technical writer, to ensure that it is scoped appropriately.

Related material:

- Existing published documentation: http://cassandra.apache.org/doc/latest/
- Documentation source (RST format): https://github.com/apache/cassandra/tree/trunk/doc
- Project source: https://github.com/apache/cassandra
- Developer mailing list archive: https://lists.apache.org/list.html?dev@cassandra.apache.org
- User mailing list archive: https://lists.apache.org/list.html?user@cassandra.apache.org