FOSS Big Data Storage Solution

Gary Jaffe[1]
Mentors: Michael Ritchson[2], Shaun Smith[2]

California State University, Sacramento[1]
STAR Fellow

National Aeronautics and Space Administration,[2]
Armstrong Flight Research Center, Edwards ,CA

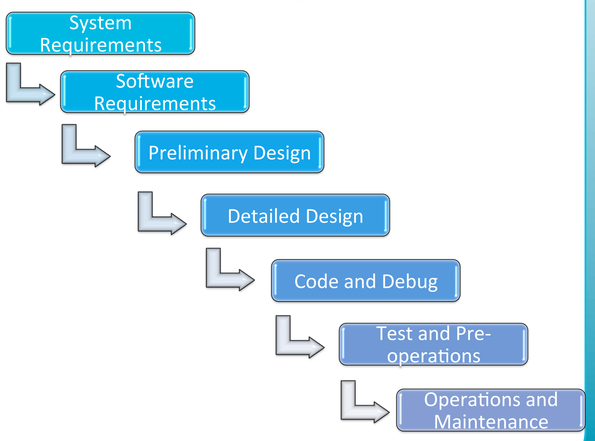
Introduction

NASA projects require a reliable system to store large volumes of data. Accordingly, it is crucial to develop a lightweight, reliable, and scalable database. Current NASA databases bear costly license fees with undesirable speed and flexibility. The purpose of utilizing the AERO Institute as an IT test bed, or "Sandbox," is to design, build, test, and implement software solutions prior to transfer to NASA projects. The Sandbox will design an end-to-end flight data management software solution for Compact Fiber Optic Sensing System (C-FOSS) data collected with the APV3 unmanned vehicle. Cassandra will be validated as a lightweight, open source database capable of managing big data while providing a cluster, fault-tolerant system.

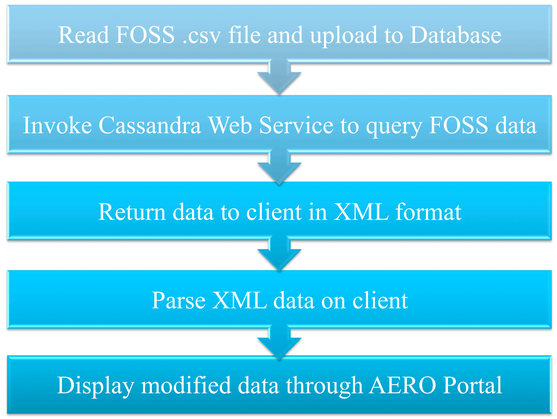
System Requirements

- Store a minimum of 2GB of C-FOSS data in multiple file formats (.csv, .log, .xml, and .jpg)
- Use benchmark tests to verify the speed, flexibility, and reliability of data stored in the Cassandra database
- Create user-friendly interface to query C-FOSS data

Software Development Lifecycle



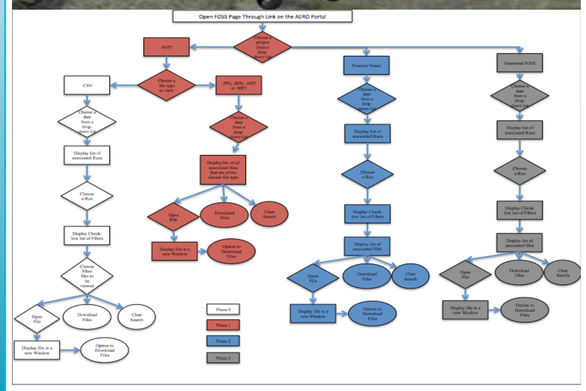
Methodology



Developing The Database

As a technical liaison between the IT team and the C-FOSS scientists, the developer was in charge of gathering and writing system requirements, producing phase goals, and generating an understanding of the projects scope to ensure the databases future launch.

Flight data from this plane is uploaded into Cassandra and accessible via a portal in order to upload, store, and manage it for future use.



This flow-chart represents the capabilities of the front-end user interface through the AERO portal

Software Architecture

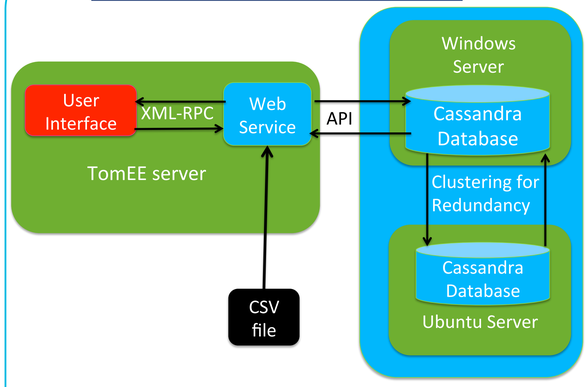
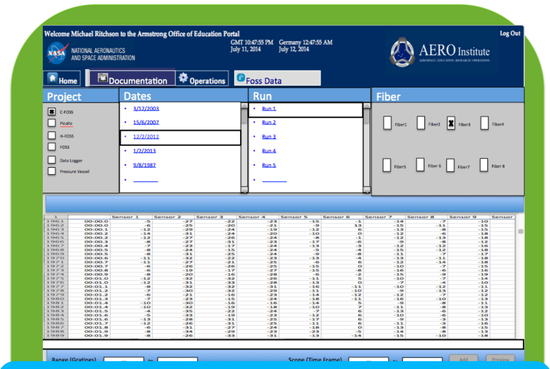


Diagram of flight data management software architecture. Design is loosely-coupled with a lightweight front-end that invokes the Cassandra web service to query C-FOSS data from the Cassandra database. Server clustering enables fault tolerant, redundant data storage.



Next steps include development of a user interface to manage flight data. Existing user interface for document retrieval seen above.

Conclusion

Validation tests prove that uploads are accurate and reliable; a java script found no discrepancies between the original and uploaded data set. Cassandra coupled with the Astyanax API is a viable solution for storing big data.

Future Scope

Next steps include designing, building, and implementing a method for querying and comparing data from C-FOSS, Piccolo, and Arduino simultaneously. The portal will also include an upload tool to automatically load files into the database.

ACKNOWLEDGMENTS

- Michael Ritchson, NASA AFRC
  Software Architect
  Michael.g.ritchson@nasa.gov
- Shaun Smith, NASA AFRC
  Office of Education
  Shaun.smith@nasa.gov
- Ronald Ray, NASA AFRC
  Chief Engineer
  Ronald.j.ray@nasa.gov
- Laura Fobel, NASA AFRC
  Technology Transfer Office
  Laura.j.fobel@nasa.gov
- Karla Shy, NASA AFRC
  Director of Education
  Karla.s.shy@nasa.gov

This material is based upon work supported by the Chevron Corporation, Howard Hughes Medical Institute, the National Marine Sanctuary Foundation, National Science Foundation, and S.D. Bechtel, Jr. Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.
The STAR program is administered by the Cal Poly Center for Excellence in STEM Education (CESAME) on behalf of the California State University.