THRIFT IN FACEBOOK

1. **Introduction**. Thrift is as software library and a set of code generation tool which was developed at the Facebook Office at Palo Alto, California, to expedite development and implementation of scalable and efficient backend services. The primary goal of thrift is to enable efficient and reliable communication across programming languages by abstracting the portions of each language that tend to require the most customization, into a common library that is implemented in each language. This is done by allowing the users to define the data types and service interfaces in a common Interface Definition Logic File (IDL File) which is supposed to be language neutral file and it generates all the necessary code to build Remote Procedure Calls to clients and servers. This report explains the design choices and implementation level details and also tries to demonstrate a sample Thrift Service.

2. **Concept**. The whole concept of Thrift stemmed out from the fact that a new direction was required to tackle the resource demands problems for many of Facebook's on-site applications, which couldn't be addressed by staying within the LAMP framework. LAMP is the acronym for Linux, MySQL, Apache and PHP. When Facebook was being laboriously designed, it was done from ground up using this LAMP framework. By 2006 Facebook was widely accepted all over the world as the social networking site and consequently its network traffic also grew giving rise to the need for scaling its network structure for many of its onsite applications like, search, ad selection, delivery and event logging. Scaling these operations to match the resource demands was not possible within the LAMP framework. In their implementation of creating many of these services like search and event logging, various programming languages had been selected to optimize for the right combination of performance, ease and speed of development, availability of existing libraries etc.

3. **Thrift Design Features**. The primary idea behind Thrift is that it consists of a language neutral stack, which is implemented across various programming languages and an associated code generation engine which transforms a simple interface and data definition language into client and server remote procedure call libraries. Thrift is designed to be as simple as possible for the developers who can define all the necessary data structures to be as simple as possible for the developers who can define all the necessary data structures and interfaces for a complex service in a single short file. This file is called as Thrift Interface Definition Logic File or Thrift IDL File. The developers identified some important features while evaluating the technical challenges of cross language interactions in a networked environment.

4. **Types.** A Common type system should exist across all the programming languages without requiring the need for the developers to write their own serialization code. Serialization is the process of transforming an object of one type to another. For example, if a programmer has written an application implementing a strongly typed STL map for a python dictionary. Neither programmer should be forced to write any code below the application layer. Dictionary is a data type in Python which allows sequencing a collection of items or elements using keys. It is very similar to 'Associative Arrays'.

5. **Transport**. Each language must have a common interface to bidirectional raw data transport. Consider a scenario where there are 2 servers in which, one is deployed in Java and the other one is deployed in python. So a typical service written in Java should be able to send the raw data from that service to a common interface which will be understood by the other server which is running on Python and vice-versa. The same application code should be able to run against TCP Stream Sockets, raw data in memory or files on disk.

6. **Versioning**. For the services to be robust they must evolve from their present version. They should incorporate new features and in order to do this the data types involved in the service should provide a mechanism to add or delete fields of an object or alter the arguments list of a function without any interruption in service. This is called versioning.

7. **Processors**. Processors are the ones which process the data streams and accomplish Remote Procedure Calls. Thrift allows programmers to develop completely using thrift's native data type rather than using any wrapper objects or special dynamic types. It also does not require the developer to write any serialization code for transport. The developer is given the freedom to logically annotate their data structures in Thrift Interface Definition Logic File (IDL File), with minimal amount of extra information necessary to tell the code generator how to safely transport the objects across languages.

8. **Facebook Thrift Services**. Thrift has been employed in a large number of applications at Facebook, including search, logging, mobile, ads and the developer platform. Two specific usages are discussed below :-

   (a) **Search**. Thrift is used as the underlying protocol and transport layer for the Facebook Search service. The multi-language code generation is well suited for search because it allows for application development in an efficient server side language (C++) and allows the Facebook PHP-based web application to make calls to the search service using Thrift PHP libraries. Thrift has allowed the search team to leverage each language for its strengths and to develop code at a rapid pace.

   (b) **Logging**. The Thrift T-File Transport functionality is used for structured logging. Each service function definition along with its parameters can be considered to be a structured log entry identified by the function name. This log can then be used for a variety of purposes, including online and offline processing, stats aggregation and as a redo log. Thrift has enabled Facebook to build scalable backend services efficiently by enabling engineers to divide and conquer. We have found that the marginal performance cost incurred by an extra layer of software abstraction is far eclipsed by the gains in developer efficiency and systems reliability.

9. Finally, Thrift has been added to Apache Software Foundation as the Apache Thrift Project, making it open source framework for cross-language services implementation.