**Assignment-14.6**

1. **Explain what is Thrift.**

* Thrift is an [interface definition language](https://en.wikipedia.org/wiki/Interface_definition_language) and [binary communication protocol](https://en.wikipedia.org/wiki/Binary_protocol)that is used to define and create [services](https://en.wikipedia.org/wiki/Service_(systems_architecture)) for numerous languages.
* It is used as a [remote procedure call](https://en.wikipedia.org/wiki/Remote_procedure_call) (RPC) framework and was developed at [Facebook](https://en.wikipedia.org/wiki/Facebook" \o "Facebook) for scalable cross-language services development.
* It combines a software stack with a code generation engine to build [cross-platform](https://en.wikipedia.org/wiki/Cross-platform) services that can connect applications written in a variety of languages and frameworks.
* In this language you define the functions and their parameters.And then, use Thrift compiler to generate corresponding code for any language of your choice.

1. **Explain what is REST.**

* Representational state transfer (REST) or RESTful [Web services](https://en.wikipedia.org/wiki/Web_service) are one way of providing interoperability between computer systems on the [Internet](https://en.wikipedia.org/wiki/Internet).
* REST-compliant Web services allow requesting systems to access and manipulate textual representations of [Web resources](https://en.wikipedia.org/wiki/Web_resource) using a uniform and predefined set of [stateless](https://en.wikipedia.org/wiki/Stateless_protocol) operations. Other forms of Web service exist, which expose their own arbitrary sets of operations such as [WSDL](https://en.wikipedia.org/wiki/WSDL) and [SOAP](https://en.wikipedia.org/wiki/SOAP).
* "Web resources" were first defined on the [World Wide Web](https://en.wikipedia.org/wiki/World_Wide_Web) as documents or files identified by their [URLs](https://en.wikipedia.org/wiki/URL), but today they have a much more generic and abstract definition encompassing every thing or entity that can be identified, named, addressed or handled, in any way whatsoever, on the Web.
* In a RESTful Web service, requests made to a resource's [URI](https://en.wikipedia.org/wiki/URI) will elicit a response that may be in [XML](https://en.wikipedia.org/wiki/XML), [HTML](https://en.wikipedia.org/wiki/HTML), [JSON](https://en.wikipedia.org/wiki/JSON) or some other defined format.

1. **What is bulk import in Hbase?**

* HBase includes several methods of loading data into tables.
* The bulk load feature uses a Map Reduce job to output table data in HBase's internal data format, and then directly loads the generated Store Files into a running cluster. Using bulk load will use less CPU and network resources than simply using the HBase API.
* The HBase bulk load process consists of two main steps,
* **Preparing data via a MapReduce job**
* The first step of a bulk load is to generate HBase data files (StoreFiles) from a Map Reduce job usingHFileOutputFormat. This output format writes out data in HBase's internal storage format so that they can be later loaded very efficiently into the cluster.
* **Completing the data load**
* After the data has been prepared usingHFileOutputFormat, it is loaded into the cluster usingcompletebulkload. This command line tool iterates through the prepared data files, and for each one determines the region the file belongs to. It then contacts the appropriate Region Server which adopts the HFile, moving it into its storage directory and making the data available to clients.

1. **What is meant by a work flow and what is an oozie workflow.**

* A workflow consists of an orchestrated and repeatable pattern of business activity enabled by the systematic organization of resources into processes that transform materials, provide services, or process information.
* It can be depicted as a sequence of operations, declared as work of a person or group, or one or more simple or complex mechanisms.
* Apache Oozie is a real time scheduler and workflow engine that blends well with large production environments.It is a server based workflow engine.
* To run oozie workflows, two files are needed namely ,
* workflow.xml (stored in HDFS)
* job.properties (stored in local)

1. **How can you track a Oozie job.**

* Oozie job can be tracked using job tracker in job.properties which will be stored in the local.

1. **What kind of jobs can be scheduled with Oozie.**

There are two basic types of jobs that can be scheduled using Oozie,

* Oozie Workflow jobs are Directed Acyclical Graphs (DAGs), specifying a sequence of actions to execute. The Workflow job has to wait.
* Oozie Coordinator jobs are recurrent Oozie Workflow jobs that are triggered by time and data availability.

1. **What is meant by oozie co-ordinator and how it is useful.**

* The Oozie Coordinator system allows the user to define and execute recurrent and interdependent workflow jobs (data application pipelines).Real world data application pipelines have to account for reprocessing, late processing, catchup, partial processing, monitoring, notification and SLAs.
* Coordinator Action: A coordinator action is a workflow job that is started when a set of conditions are met.
* Coordinator Application: A coordinator application defines the conditions under which coordinator actions should be createdand when the actions can be started. The coordinator application also defines a start and an end time. Normally, coordinator applications are parameterized and is written in XML.
* Coordinator Job: A coordinator job is an executable instance of a coordination definition. A job submission is done by submitting a job configuration that resolves all parameters in the application definition.
* Data pipeline: A data pipeline is a connected set of coordinator applications that consume and produce interdependent data sets.
* Coordinator Definition Language: The language used to describe data sets and coordinator applications.
* Coordinator Engine: A system that executes coordinator jobs.