1.Thrfit.

Apache Thrift is a framework for implementing RPC in services, with cross-language support(supports multiple language)

i.e.,we can implement a function in Java, host it on a server and then remotely call it from Python or

any other language.

In simple terms , thrift eliminates the code written for the communication between the webserver and the services

as it is having its own Interface Definition Language(IDL) it will satisfy the request by the client to know the functions and

paramters exposed by the service.

2. REST

REST-Representational State Transfer.

REST uses HTTP(Standard protocol for web applications) for communication therefore it is suitable for heterogenous systems.

Internally both REST and Thrift use the common HTable-based client API to access the tables.

Comparing both REST and Thrift , Thrift is more lightweight and faster.

3.bulk import in Hbase

The bulk loading feature of HBase provides a way to insert data into the table more easily and faster compared to

other techniques.

4.work flow and an oozie workflow

A workflow is a business automation process in which certain group of work is scheduled to run automatically.

Oozie is a workflow scheduler that schedules a job as a DAG(Directed Acyclic Graph) consisting of operations/Script/commands etc.,

to run in the flow specified in the workflow.xml file.

It is a server based workflow engine.

5.tracking an Oozie job.

Oozie job can be tracked using job tracker in job.properties.

6.jobs scheduled with Oozie.

The 2 types of jobs that can be scheduled with oozie are

Oozie Workflow jobs are Directed Acyclical Graphs (DAG), specifying a sequence of actions to execute in a Directed flow.

Oozie Coordinator jobs are recurrent Oozie Workflow jobs that are triggered by time and data availability.

7.oozie co-ordinator and its uses

Oozie Coordinator is a collection of predicates (conditional statements based on timefrequency and data availability)

and actions (i.e. Hadoop Map/Reduce jobs, HDFS, Hadoop Streaming, Pig, Java and Oozie sub-workflow).

Actions are recurrent workflow jobs invoked each time predicate returns true.

The coordinator file conatains the path to the workflow.xml file and the frequency and time/condition by which

the control should be transferred to the workflow.