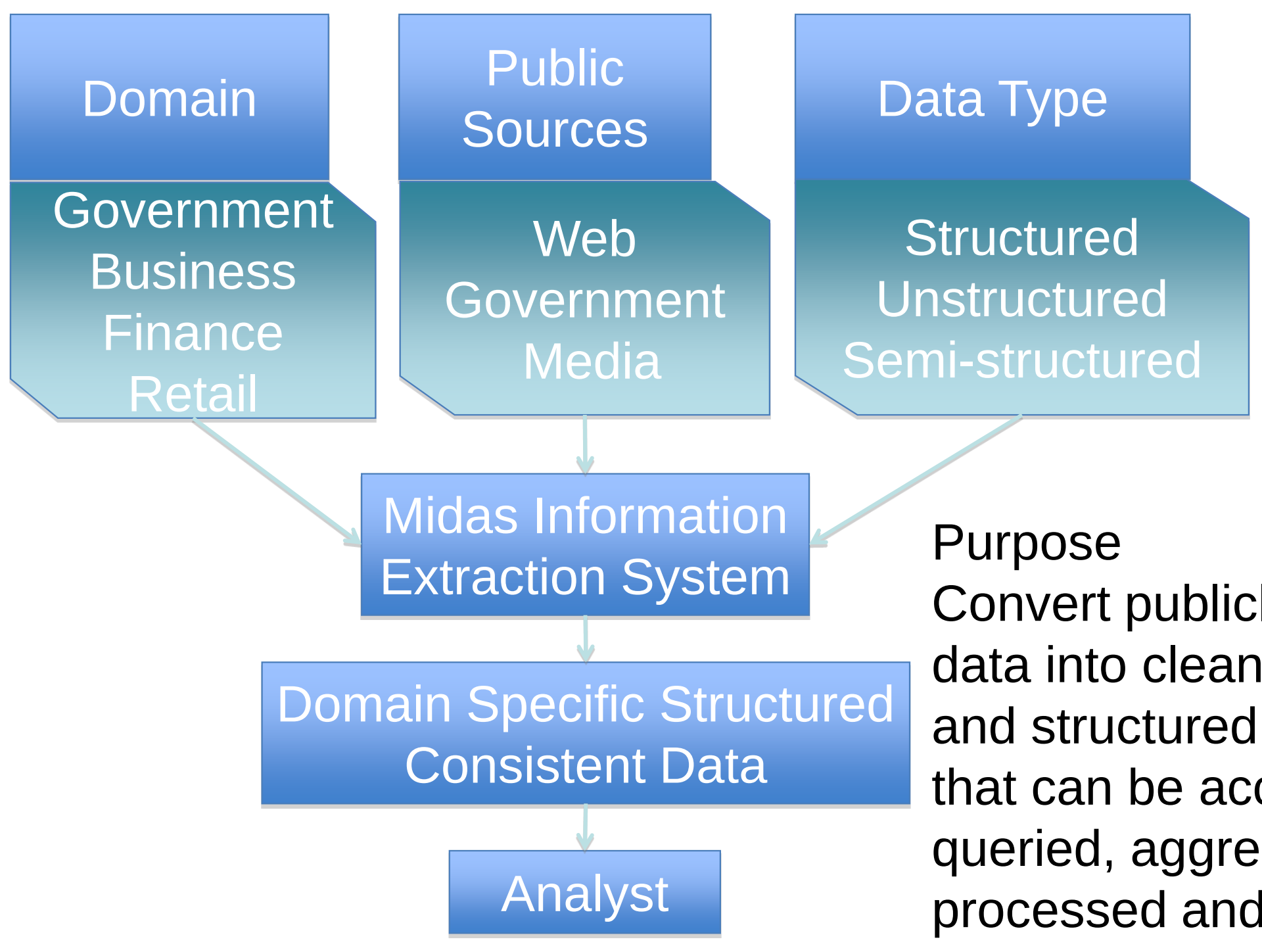
Partnership for International Research and Education
A Global Living Laboratory for Cyberinfrastructure Application Enablement

SEC File Extraction and Financial Table Analysis for Midas

Student: Jonatan Gonzalez, Undergraduate Student, Florida International University
Research Advisor: Professor Naphtali Rishe, Florida International University
CI-PIRE Partner Advisors: Dr. Howard Ho, IBM Research - Almaden, Dr. Ullas Nambiar, IBM IRL India

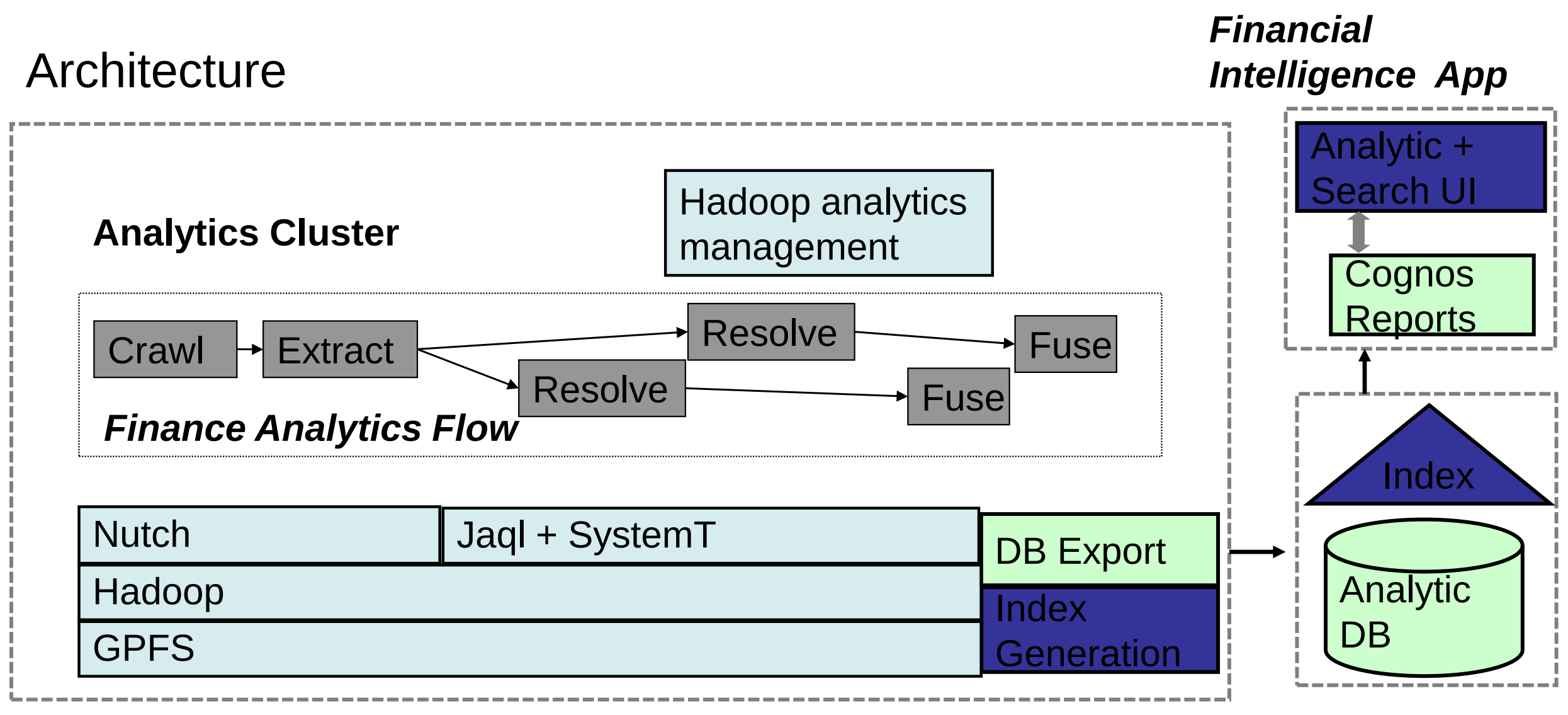
I. Research Overview and Outcome

Midas Overview



Purpose
Convert publicly available data into clean, normalized, and structured temporal data that can be accurately queried, aggregated, processed and analyzed

Architecture



Tools

JAQL is used to perform SQL like queries in a Hadoop MapReduce environment

Hadoop and MapRedue are used for scalability and processing the large amount of data required

Java is used in situations where Jaql is not flexible enough to perform the required actions.

Eclipse is the development environment used for the java functions and accessing the CVS repository

Work Done

IBM Almaden Research Lab

Background
The Midas project uses publicly available data as input and converts it to a uniform format to be analyzed. The government provides information about bank call reports. These filings are provided by the Federal Financial Institutions Examination Council (FFIEC). Originally the files were downloaded manually ever quarter. My task was to create a web crawler in java to access the web service provided by the FFIEC to access the files in the repository.

Solution
The FFIEC webpage provided a C# client to access the repository. The Java solution however was not so simple. In order to access the web service java needs Axis2. Axis2 provides the functionality to automatically create the java files required to access the web service. Much time was spent figuring out what was the proper policy file was due to the username token and password setup required to access the repository.
The web crawler uses multiple threads to expedite the download process. The crawler however is limited by the FFIEC web service to 2500 downloads per hour. The web crawler automatically handles when the limit is reached by waiting until the next hour. It provides functionality to download the bulk data from any quarter. As per request it can download the only the latest call reports for specified banks. After all the data is downloaded it is unzipped to a user specified location organized by quarter.

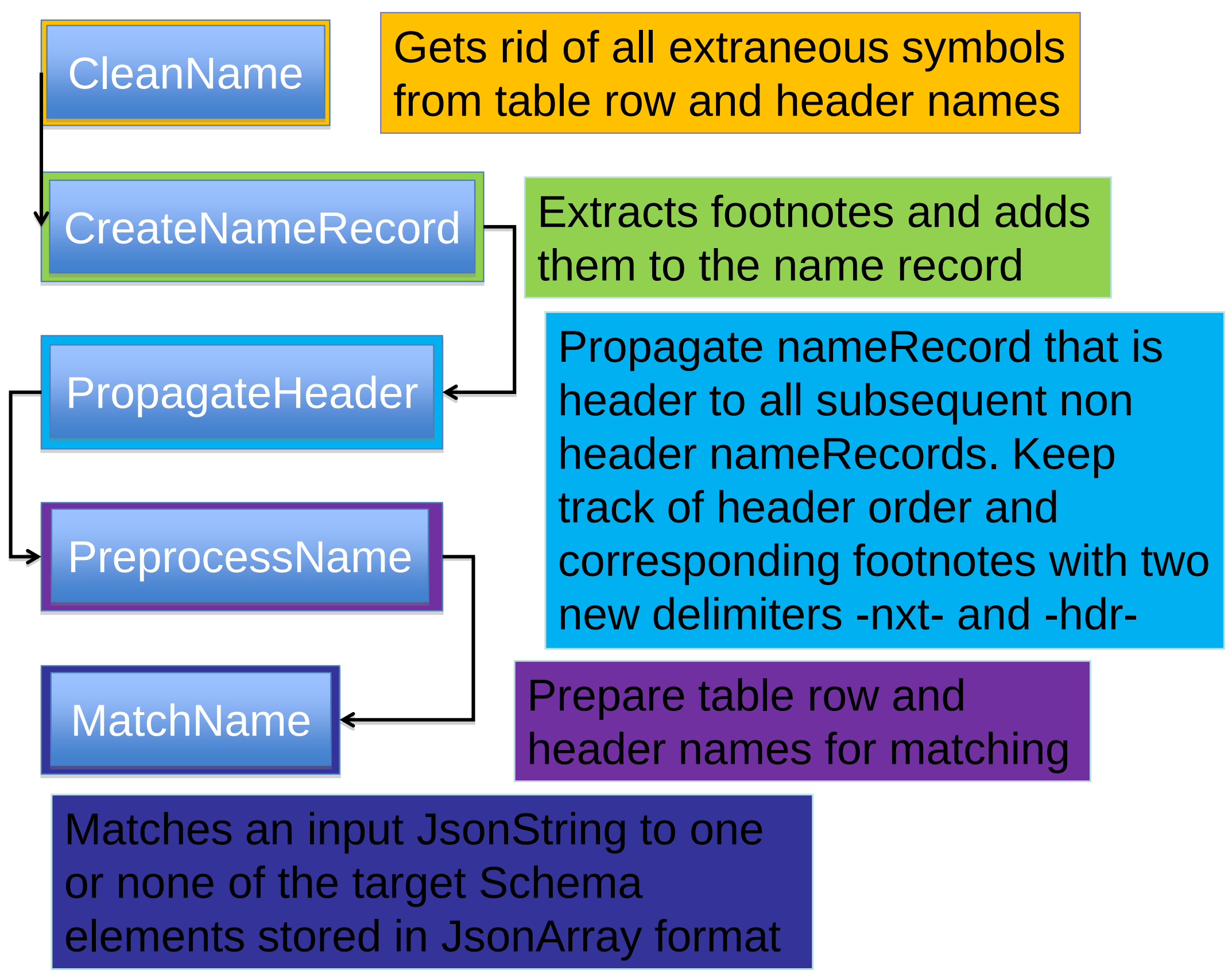
When I completed this task I worked on becoming familiar with Jaql in order to assist Melita with her part of the project.

IBM India Research Lab

Background
The Midas project's main focus is to provide structured data for the average user. Once data is transformed into a uniform format it can be easily analyzed by computer to retrieve valuable information.

Solution
Using Jaql scripts financial tables are manipulated from their unstructured/semi-structured formats to a uniform format that was created to normalize these tables. There were numerous situations however where Jaql was not flexible enough to provide the functionality required. The most prominent example of this is the propagate header function which needed to extract headers and propagate them to each subsequent row name.

Java Functions



CleanName: Gets rid of all extraneous symbols from table row and header names

CreateNameRecord: Extracts footnotes and adds them to the name record

PropagateHeader: Propagate nameRecord that is header to all subsequent non header nameRecords. Keep track of header order and corresponding footnotes with two new delimiters -nxt- and -hdr-

PreprocessName: Prepare table row and header names for matching

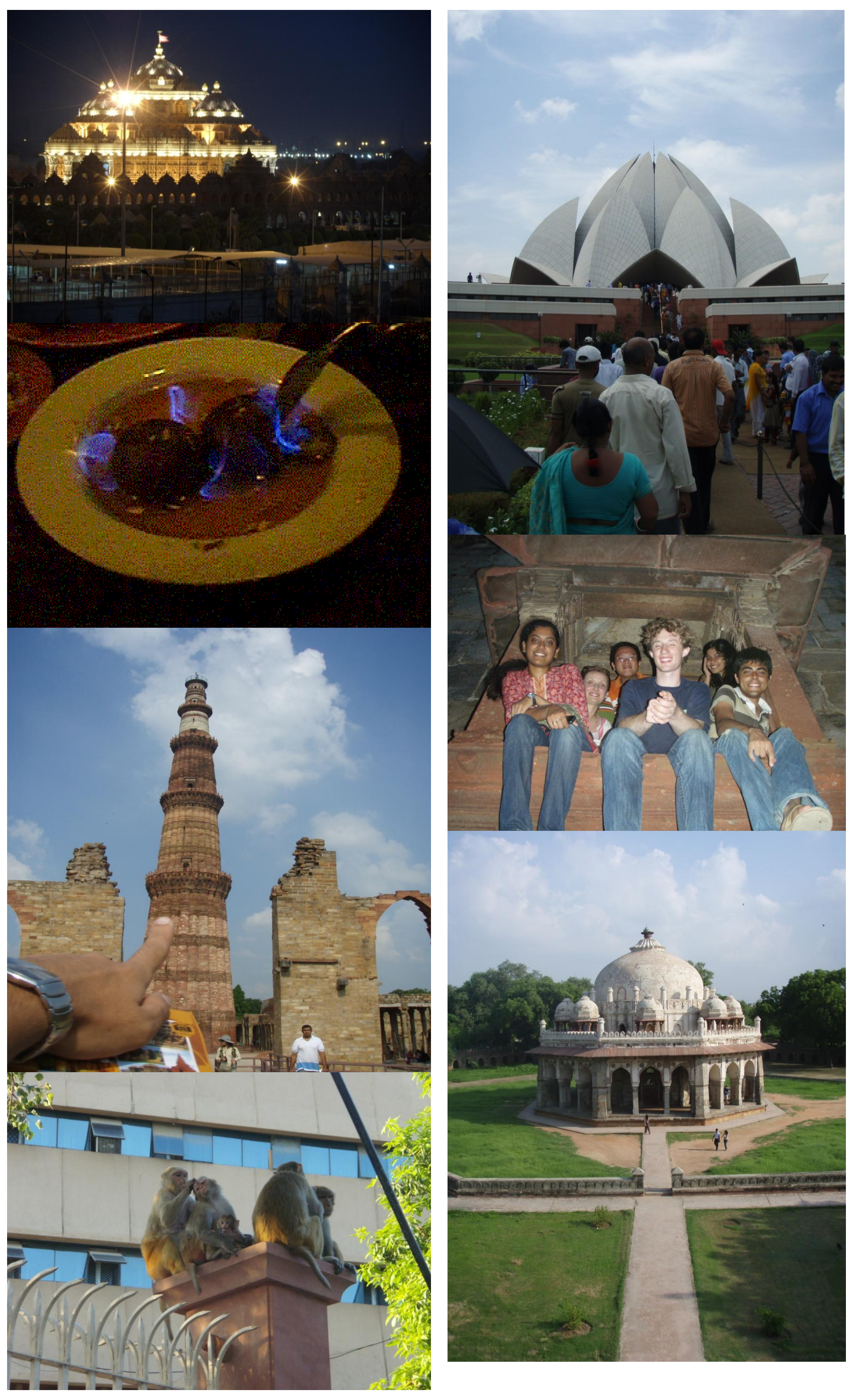
MatchName: Matches an input JsonString to one or none of the target Schema elements stored in JsonArray format

II. International Experience

Lab Friends

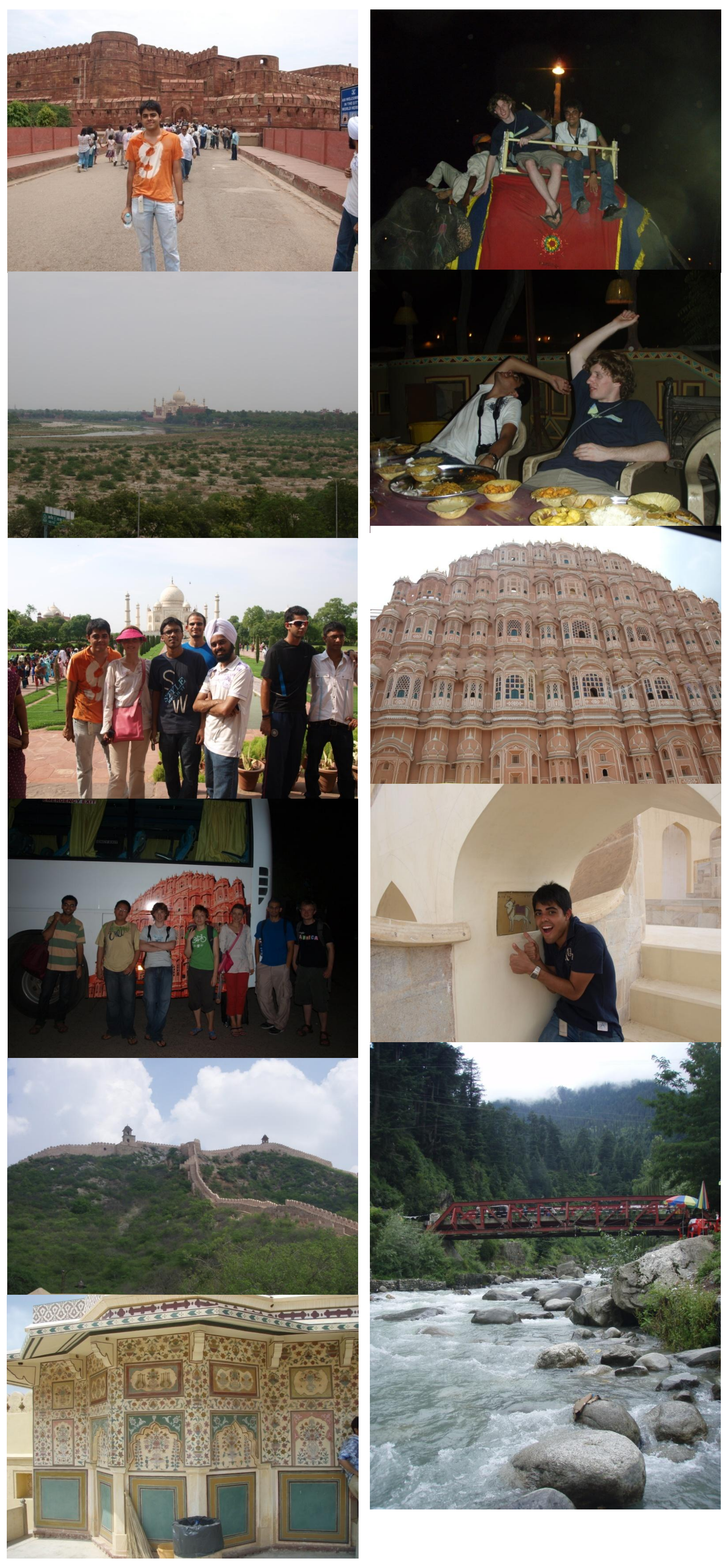
Many great connections were made with amazing people. Its incredible how you can find like minded people half way across the world.

Delhi



We stayed in Green Park, New Delhi in a guest house called Exotica. The city was filled with history and old monuments. It took a few weekends to see a portion of what Delhi had to offer.

India



Though resistant at first India had won me over with its immense beauty and amazing people. It was an experience to help ground me as a person and appreciate the opportunities I have in my country.

III. Acknowledgement

The material presented in this poster is based upon the work supported by the National Science Foundation under Grant No. OISE-0730065. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.