**DMETL Term Project**

**“** **Working with XML”**

Project Report by

**Pratiksha Mete(H21017)**

**Pooja Bornarkar(H21015)**

**Sampada Petkar(H21023)**

**Gauri Deshmukh(H21011)**

**Latika Sawant (H21014)**

**Priyanka Sargam(H21020)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. **Using "tXMLMap" to read XML**

Objective:

* The objective was to read a XML file and convert it to a flat file(Excel file)
* It Shows how we can convert XML record saved in a file into a format that is readable by "tXMLMap" and how can we read and process the data in XML record.

Components used:

1. tFileInputXML : XML file

2. tXMLMap : Joins.

3. tLogRow : display flat(table) Output on console

4. tFileOutputExcel

Working:

- The "tFileInputXML" component uses "XPATH" to convert the input XML into a JAVA Document object, so that an XML tree can be created by the "tXMLMap" component.

- By pointing to "tXMLMap" at our XML file , we were able to import the XML structure from the file and we were able to manipulate the fields individually.

- Once the document tree has been defined, elements can then be copied from the output to the input as with a normal "tMap."

Result:

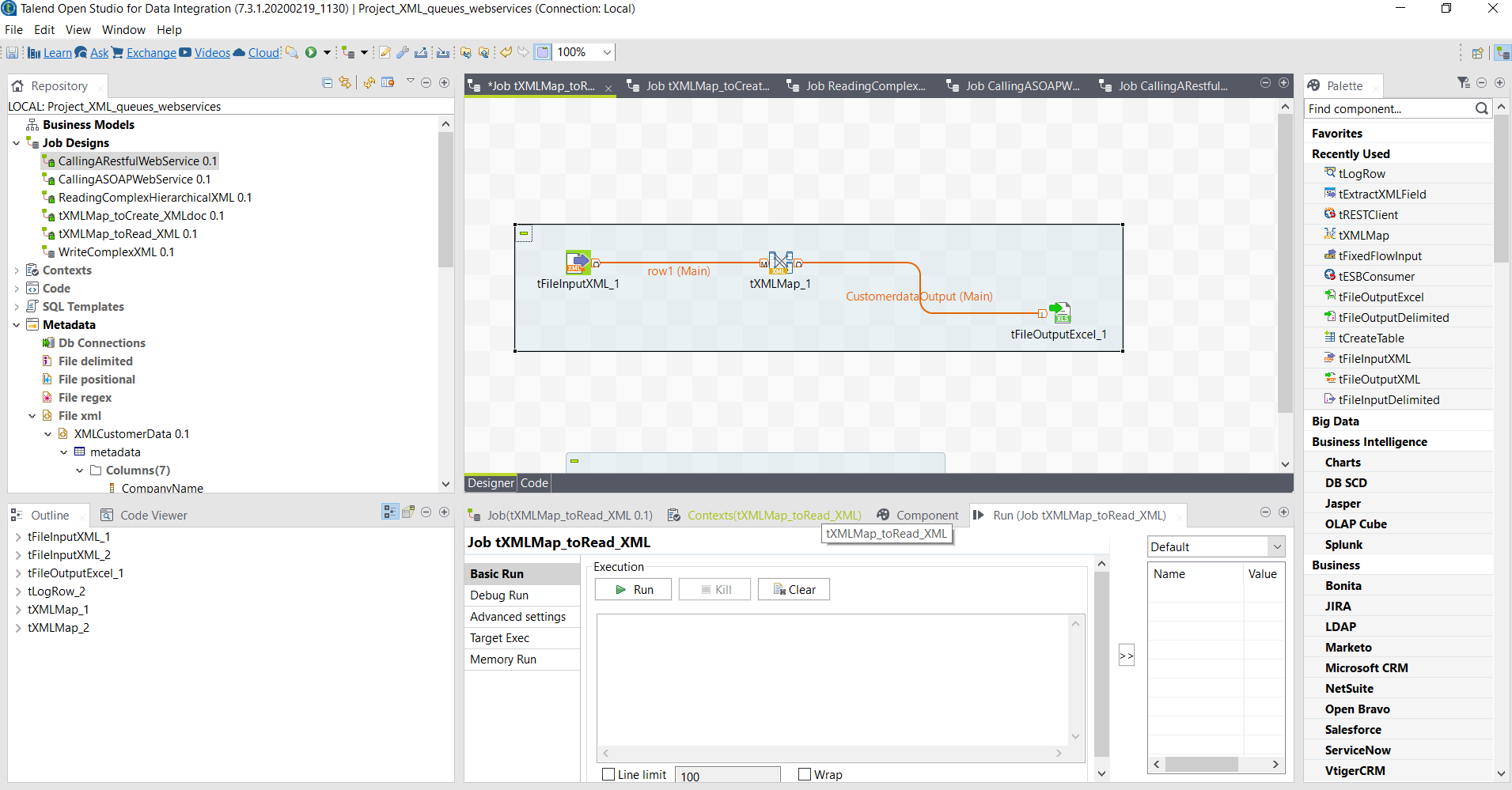
Fig.1 Using "tXMLMap" to read XML



Fig.2 XML input file

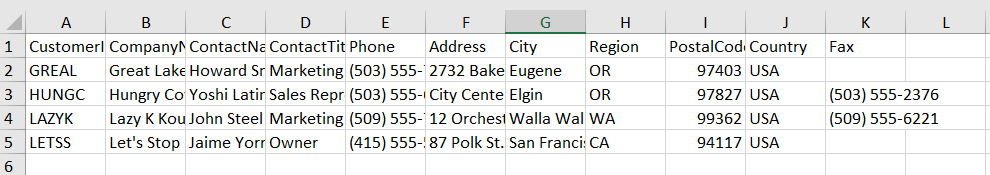


Fig.3 Excel Output file

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Using tXMLMap component to create XML Document**

Objective:

* The objective was to read from a flat file and convert it to a XML document for output file.

- Reading from a flat file and converting it to a XML document for output file.

Components used:

1. tFileInputDelimited

2. tXMLMap

3. tFileOutputXML

Working:

- Defining the output type of "Document" allows us to define an XMLformat within the "tXMLMap" component into which we can then map our input data.

- the "tFileOutputXML" component by default will create an XML structure from a normal schema. However, it is possible to force it to handle a document.

Results:

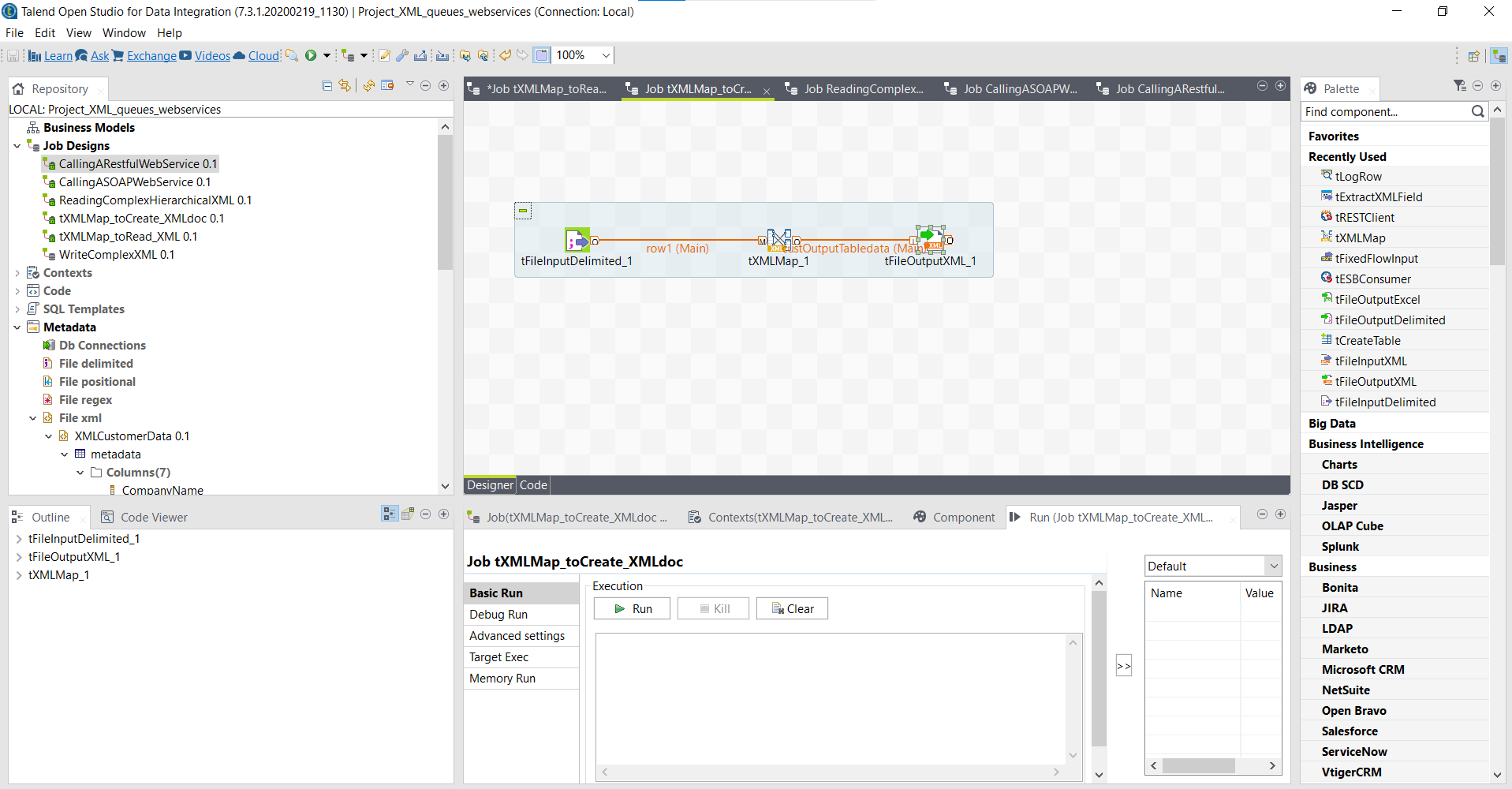


Fig. 4 Using tXMLMap component to create XML Document

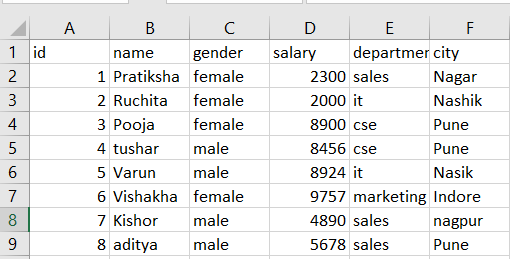


Fig.5 Input Excel File



Fig.6 XML output file

1. **Reading Complex Hierarchical XML**

Objective:

* The objective was to read XML document and convert it to a flat file.

- It shows how can we deconstruct a more complex XML record into individual sets of data while ensuring that the hierarchical relationships between the data are not lost.

Components used:

1. tFileInputXML

2. tLogRow

Working:

- The XML schema component allows us to map data from the XML structure into a flattend Talend schema easily, ready for use in the downstream components.

- First, we defined an XML schema to extract just the "customer" fields.

- Then, we did the same for the "order" fields, and also extracted the key for the "customer", which is "customerId".

This will ensure that for each set of order data of which there are two, the "customerID" is present.

- We Repeated this process again for the "orderItem" fields, remembering to extract the "customer" and "order" fields.

- Finally we dragged the schemas to the canvas, linked them and added "tLogRow" outputs.

Results:

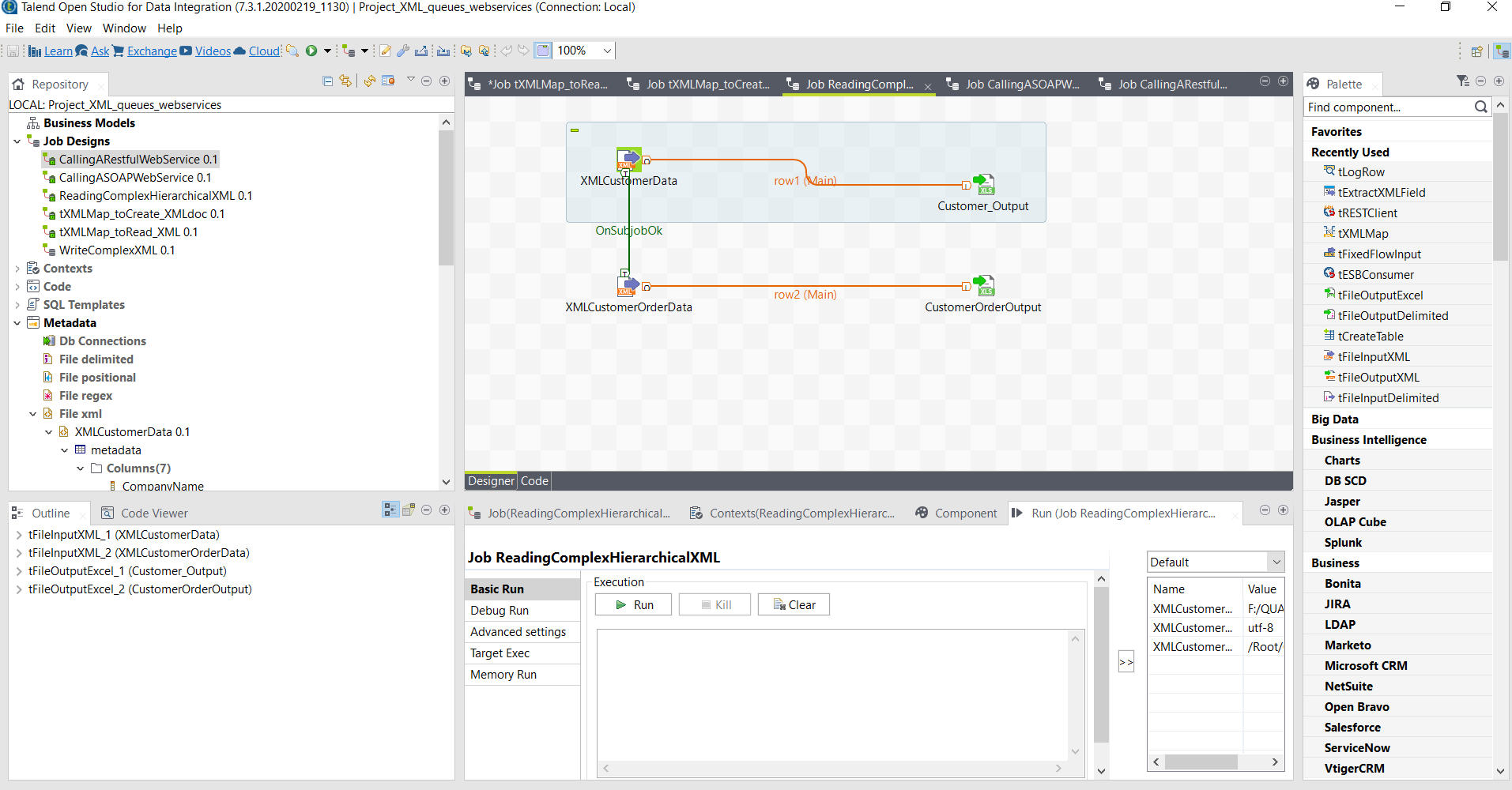


Fig. 7 Reading Complex Hierarchical XML

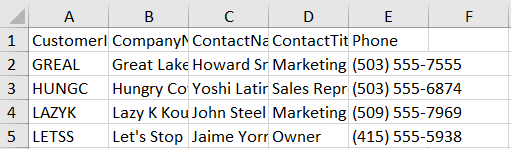


Fig.8 Excel Output File1

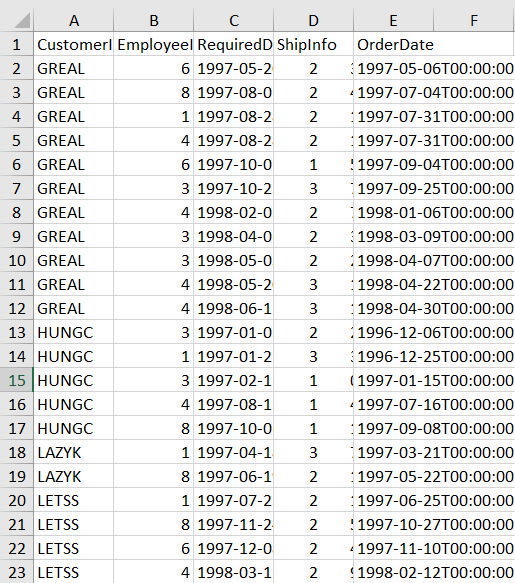


Fig.9 Excel Output File2