

SANKHYA Translation Framework: A Framework for Model-driven Data Integration and Translation

Murali Desikan,

Gopi Kumar Bulusu

Sankhya Technologies Private Limited

Contents

- Data Integration – Brief Introduction
- Introduction to STF
- STF Features
- STF Editions
- STF Architecture
- STF Components
- SANKHYA Translation Modeling Language
- Usage Scenario
- Comparative Technologies
- Summary

Data Integration – Brief Introduction

■ Data Integration

– Involves:

- » data collection from multiple sources (files, databases, applications)
- » data transformation
- » presentation

– Issues:

- » heterogeneous formats
- » distributed data sources
- » legacy applications

Introduction to STF

- Novel framework for building dynamic, model-driven data integration systems and parsers
- Powerful modeling language for describing:
 - document schema
 - language grammars
 - translation and transformation rules
- Applications:
 - document processing
 - protocol conversion
 - text/database to XML conversion
 - server page processing
 - parsers and translation tools

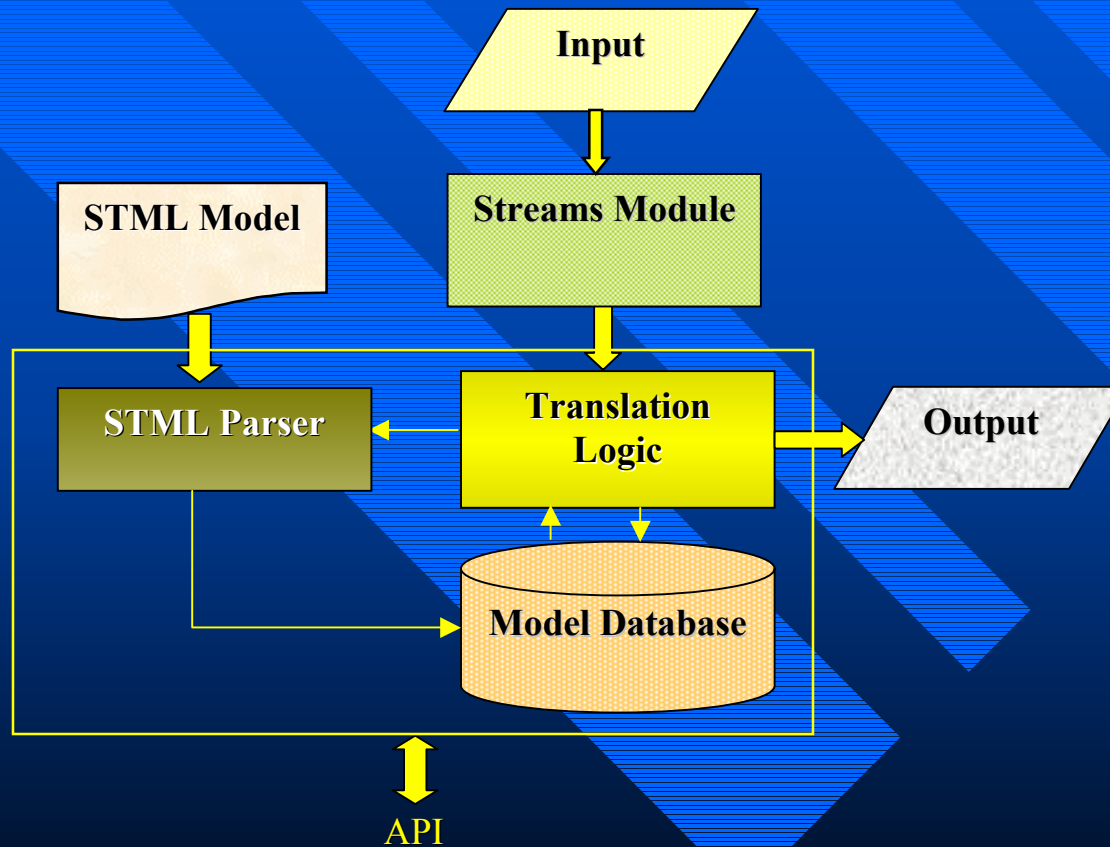
STF Features

- Powerful modeling language for describing document format, message format and language grammars
- Supports complex transformation between multiple formats
- Supports reversible transformations
- Pluggable streams for processing input from multiple sources of data (file, URL, database)

STF Editions

- Command-line utility for simple translation and data integration needs
- CORBA-based server for distributed client-server applications
- C++ API for complex translation and transformation requirements

STF Architecture



STF Components

- Sankhya Translation Modeling Language (STML)
- STML Parser
- Model Database
- Translation Logic
- Streams Module

SANKHYA Translation Modeling Language

- Hierarchical information modeling
- Multi-representational modeling
- Inheritance and aggregation of model elements
- Support for different data types (word, string, symbol, any)
- Support for attributes

STML Elements

- Unified representation of nodes of a hierarchy
- Element Types
 - STML Root
 - » Represents the root node of an hierarchy
 - STML Node
 - » Represents any internal node in an hierarchy
 - STML Leaf
 - » Represents a leaf node of an hierarchy
 - STML Sequence
 - » Represents a sequence of root, node, leaf elements
 - STML Union
 - » Represents a set of alternatives for an element

STML Elements (Contd.)

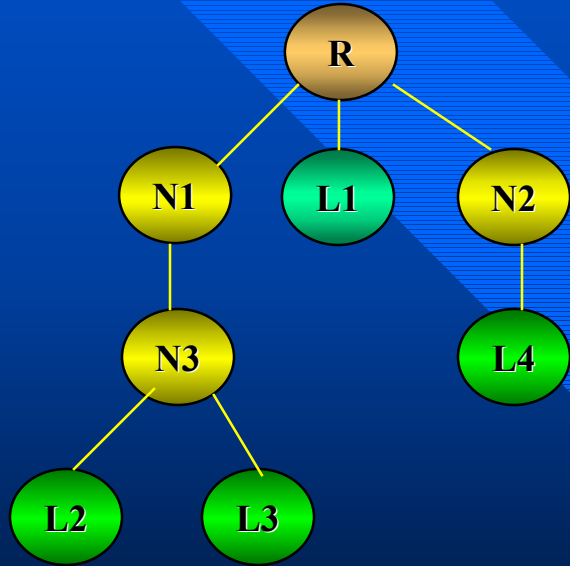
- STML Value
 - Place holder for data
 - Constant data or reference to other elements, symbols
- STML Representation
 - List of STML Values representing a node in a particular format
- STML Types
 - Word, Any, Symbol

STML Actions

- Attached to STML elements and represent some external action to be performed on nodes
- Action triggered when element matches during translation/transformation
- Sample Actions:
 - Update symbol table
 - Evaluate expressions

Modeling Using STML

■ Sample Hierarchy



```

STMLModel M {
  STMLLeaf L1 {
    rep1 = { ... }; rep2 = {...};
  };
  // L2, L3 L4 not shown
  STMLNode N2 {
    L4 l4;
    rep1 = {l4}; rep2 = {...};
  };
  STMLNode N3 {
    L2 l2; L3 l3;
    rep1 = {l2, l3}; rep2 = {...};
  };
  STMLNode N1 {
    N3 n3;
    rep1 = {n3}; rep2 = {...};
  };
  STMLRoot R {
    N1 n1; L1 l1; N2 n2;
    rep1 = {n1, l1, n2}; rep2 = {...};
  };
};
  
```

STF Components (Contd.)

- STML Parser
 - Syntax and semantic analysis of STML Model
 - Creates internal representation of the model
- Model Database
 - Repository for parsed model information
- Translation Logic
 - Translates/Transforms input to different representations based on given STML Model

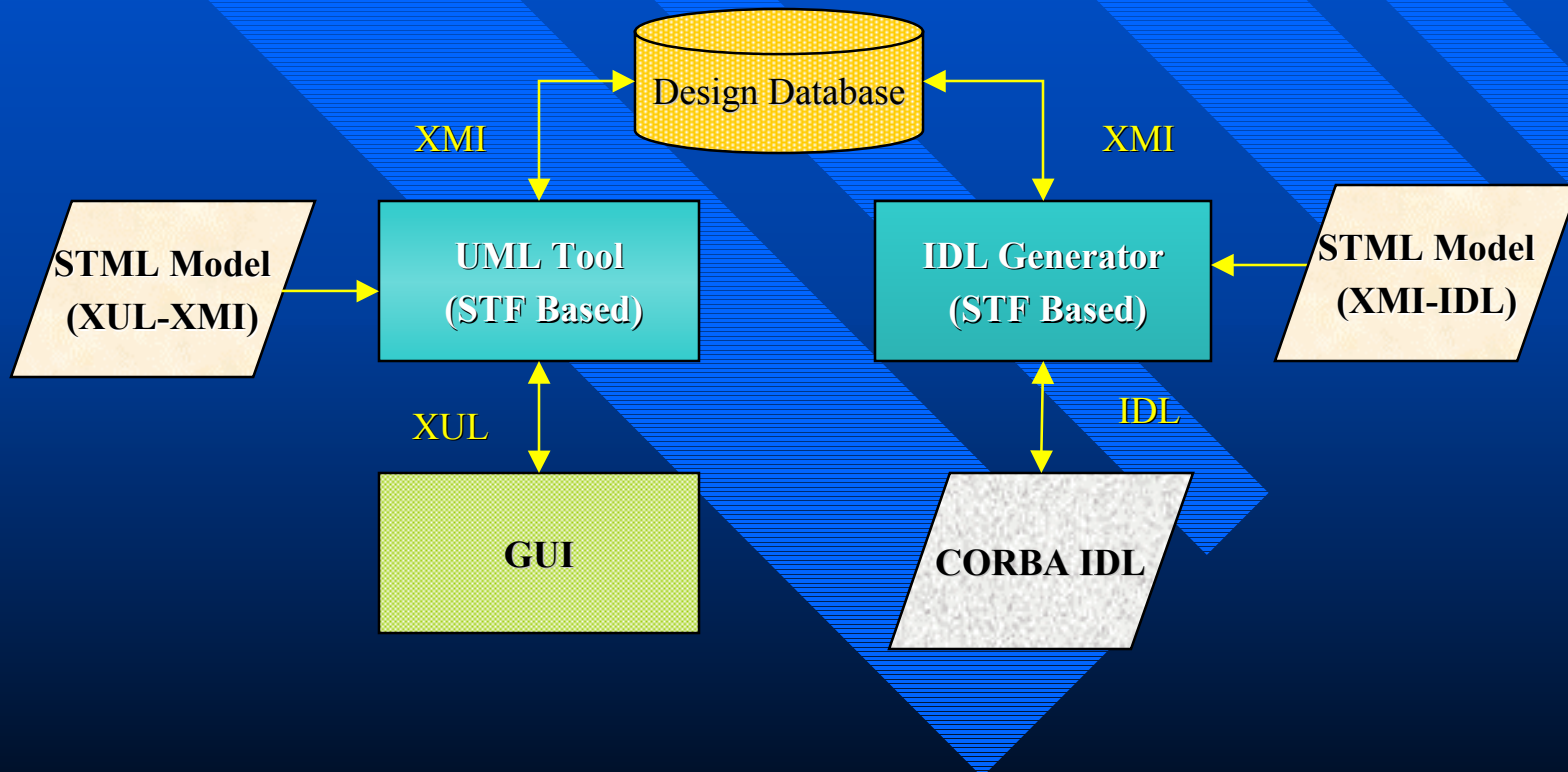
STF Components (Contd.)

■ Streams Module

- Pluggable module for obtaining input from multiple sources
- File stream, string stream, ODBC stream
- New streams can be implemented as DLLs and registered with the framework (e.g., ftp stream, http stream etc)

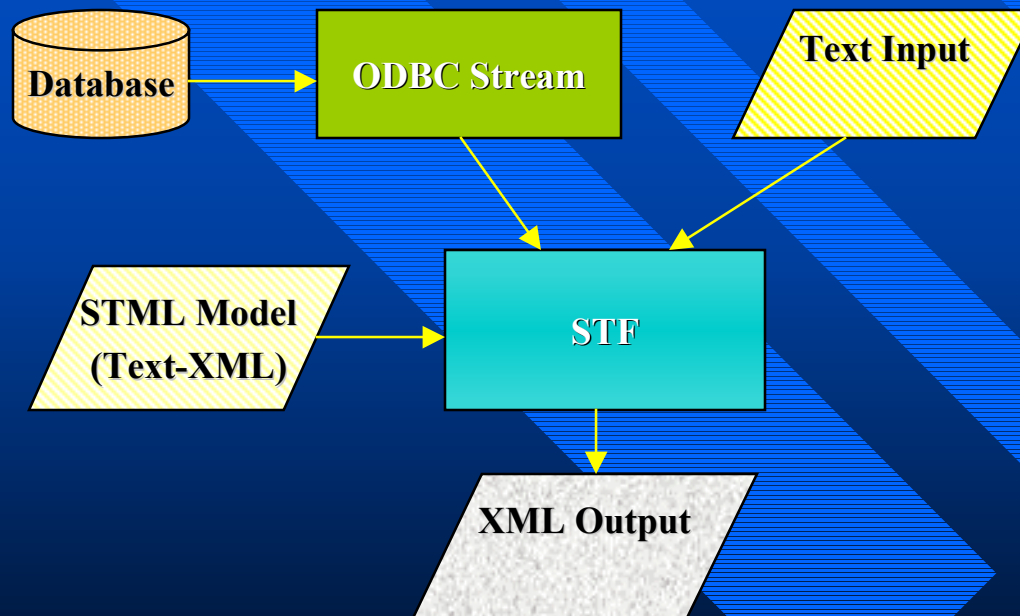
STF Usage Scenario

■ Tool Integration



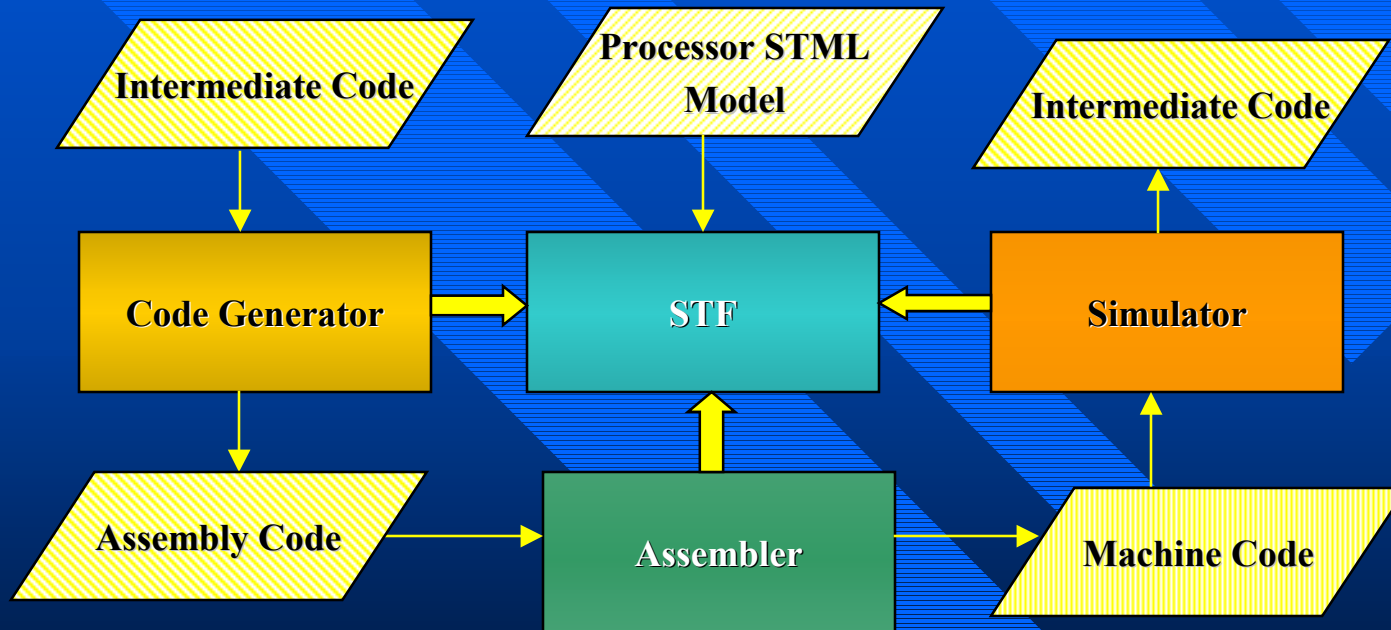
STF Usage Scenario

■ Text/Database to XML Conversion



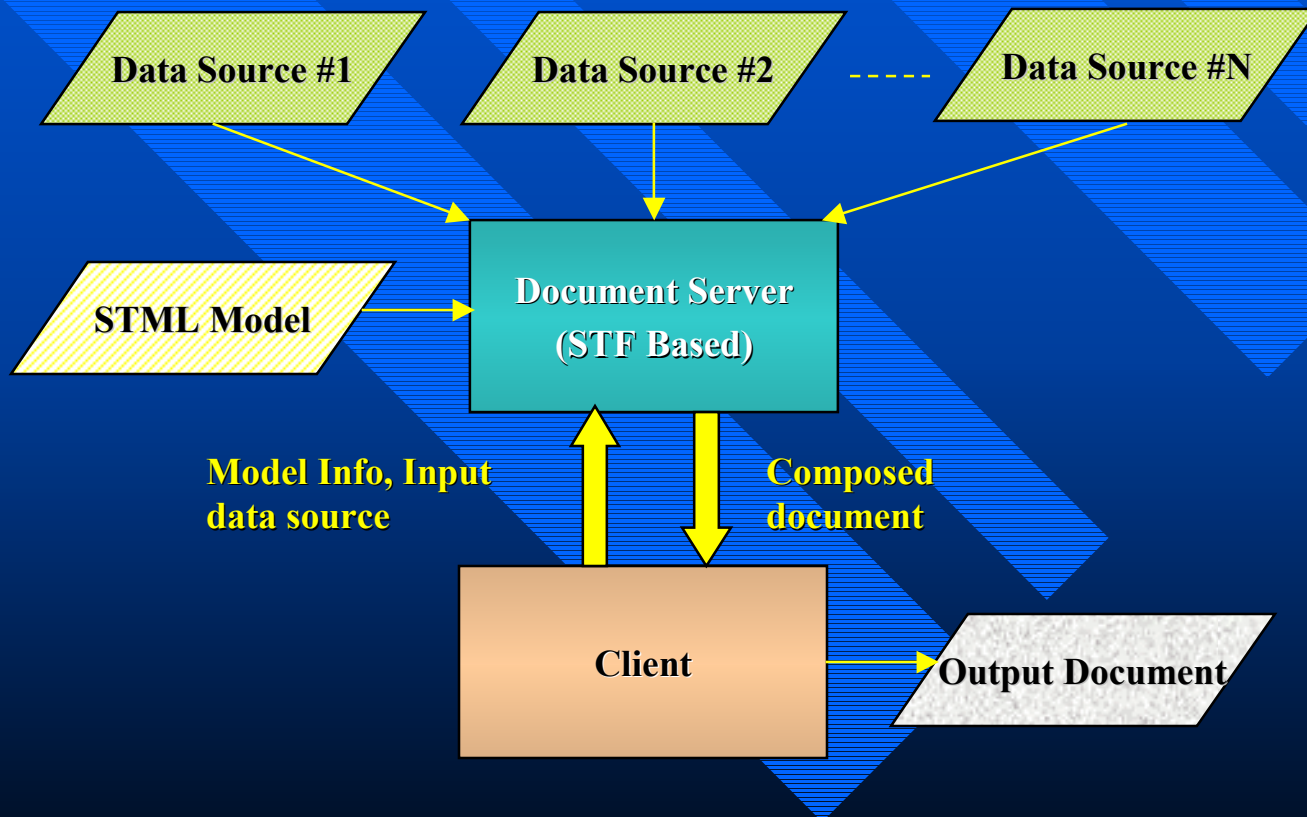
STF Usage Scenario

■ Program Translation Tools



STF Usage Scenario

■ Model-driven Document Servers



Comparative Technologies

- Data Modeling: XML
 - Represent data in multiple formats in a single model
- Transformations: XSLT
 - Transform between multiple formats
 - Execute external actions
 - Support for multiple sources of data
- Parsing: Lex, Yacc
 - Grammar specification external to tool
 - Support for attributes
 - Inherit grammar elements and attributes
 - Top-down and bottom-up parsing support

Summary

- Framework for building model-driven data integration and translation tools
- Modeling language for multi-representational modeling
- Pluggable streams for obtaining input from multiple data sources
- Build data conversion tools, program translation tools, document servers

Thank You

Sankhya Technologies Private Limited

<http://www.sankhya.com>

Corporate Office:

Sankhya Technologies India Operations
#30-15-58, Third Floor, Silver Willow,
Dabagardens, Visakhapatnam 530020,
INDIA

Ph: +91 891 554 2666
Fax: +91 891 554 2665

Registered Office:

Sankhya Technologies Private Limited
#13/2, Third Floor, "JAYASHREE"
Jayalakshmiipuram First Street,
Nungambakkam, Chennai 600 034,
INDIA

Ph: +91 44 2822 7358
Fax: +91 44 2822 7357