

Sparksheet - Transforming Spreadsheets into Spark Data Frames

Oscar Castañeda-Villagrán
Universidad del Valle de Guatemala



About

- Researcher at Universidad del Valle de Guatemala.
- Research Interests:
 - Program Transformation,
 - Programming Education Research,
 - Online Learning to Rank.

Prototyping ...

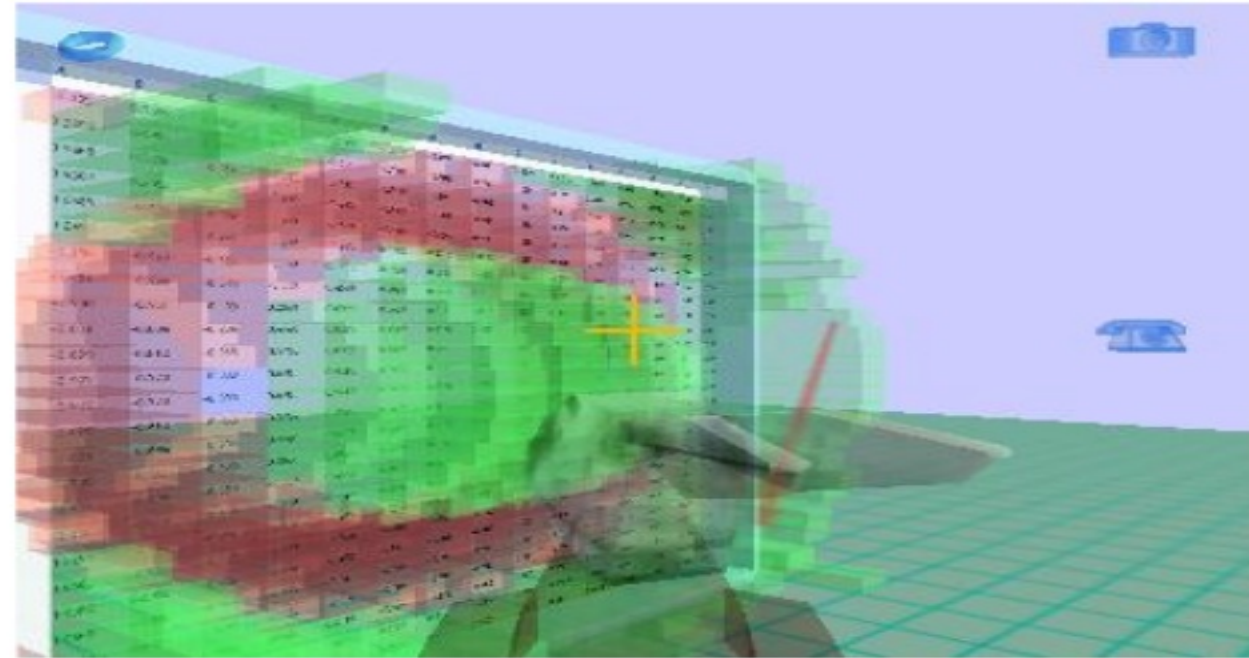


<http://bit.ly/2eSGmwy>

Prototyping Spark programs with ...



<http://bit.ly/2eSGmwy>



<http://bit.ly/2eSGmwy>

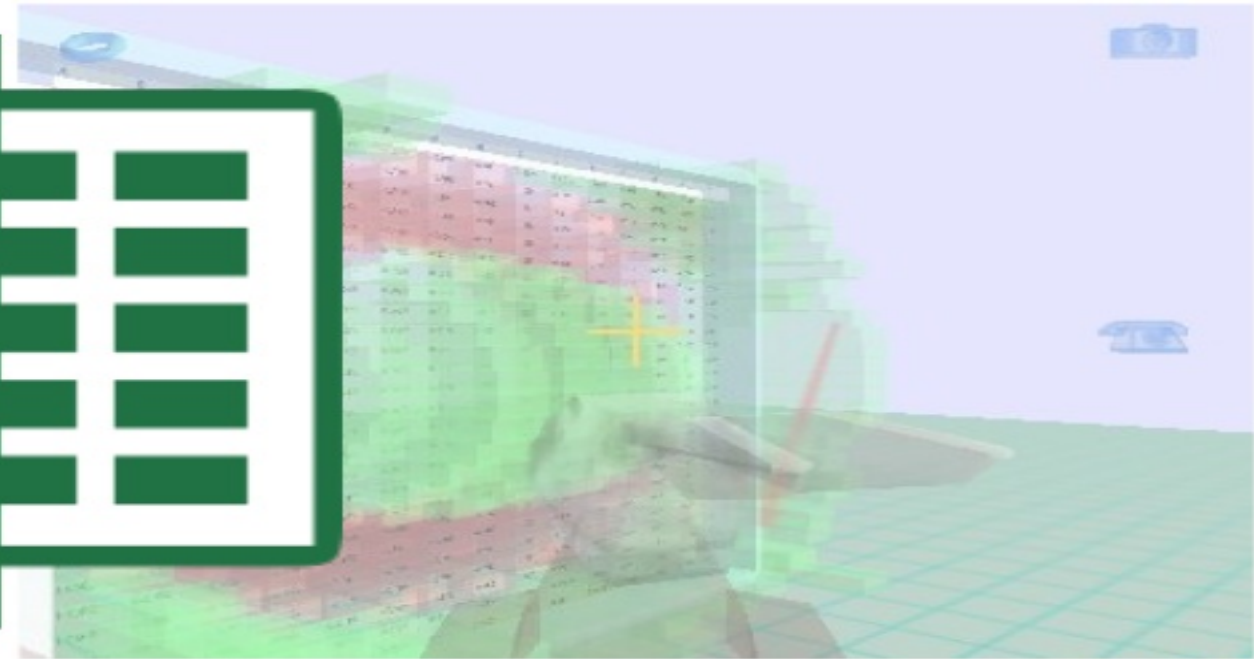
Prototyping Spark programs with Excel



<http://bit.ly/2eSGmY>



<http://bit.ly/2eG1768>



<http://bit.ly/2eG1768>

Agenda

- Problem Statement and Motivation
 - Architecture
- Program Transformation
 - Pipeline
 - Code-to-Code Transformation
- Parsing Excel Formulas
 - Grammar
 - Parse Tree
 - XLParse
- Excel as a DSL
 - Generating Code
- Demo
- Q&A

Disclaimer(s)

- Ongoing research ...
- We will focus on how to create a Program Transformation Pipeline.

Problem Statement

Spark programs can be prototyped in Excel but **manually translating** Excel formulas to Spark programs is tedious and error-prone.

Motivation

- “***Straight path***” between column-oriented Excel “programs” and Spark programs that make use of the DataFrame API.
- But, **manually translating** Excel formulas to Spark is tedious and error-prone.
- **What if:** Excel compiler?

Problem Statement

Given that column-oriented Excel applications can be manually translated to Spark programs, ...

... find a way to automate translation of Excel formulas so that data pipelines can be prototyped in Excel ...

... and Scala/Python code generated to run in Spark.

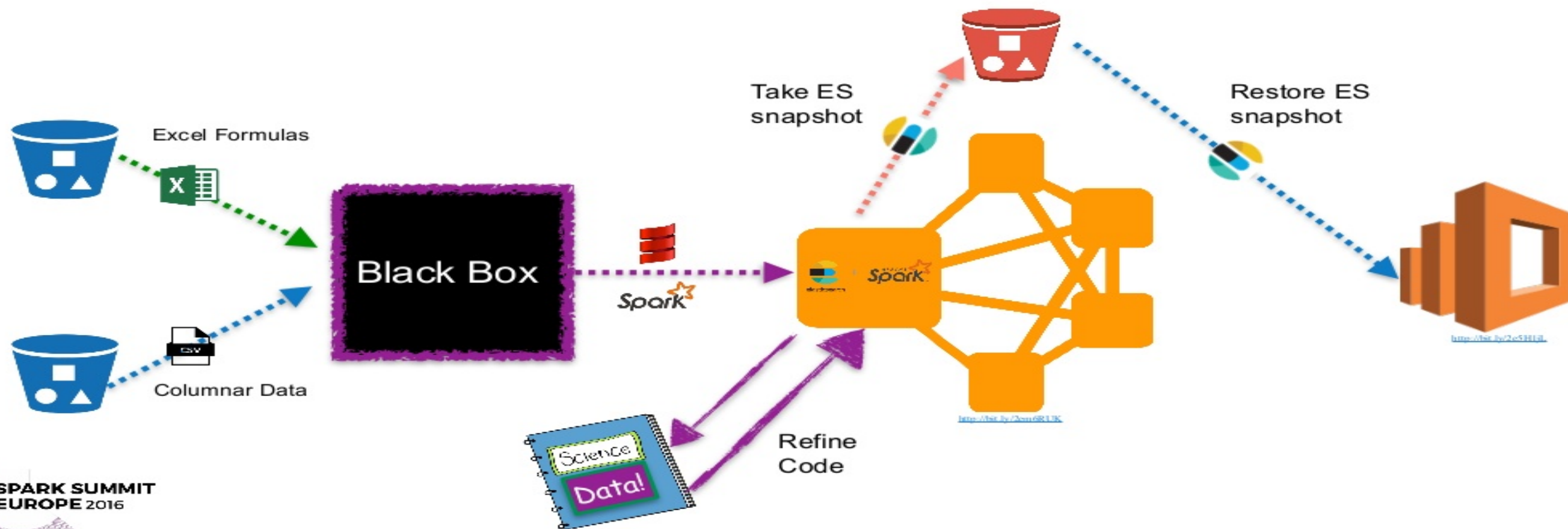
Motivation

Automatically translate
Excel Formulas to ...

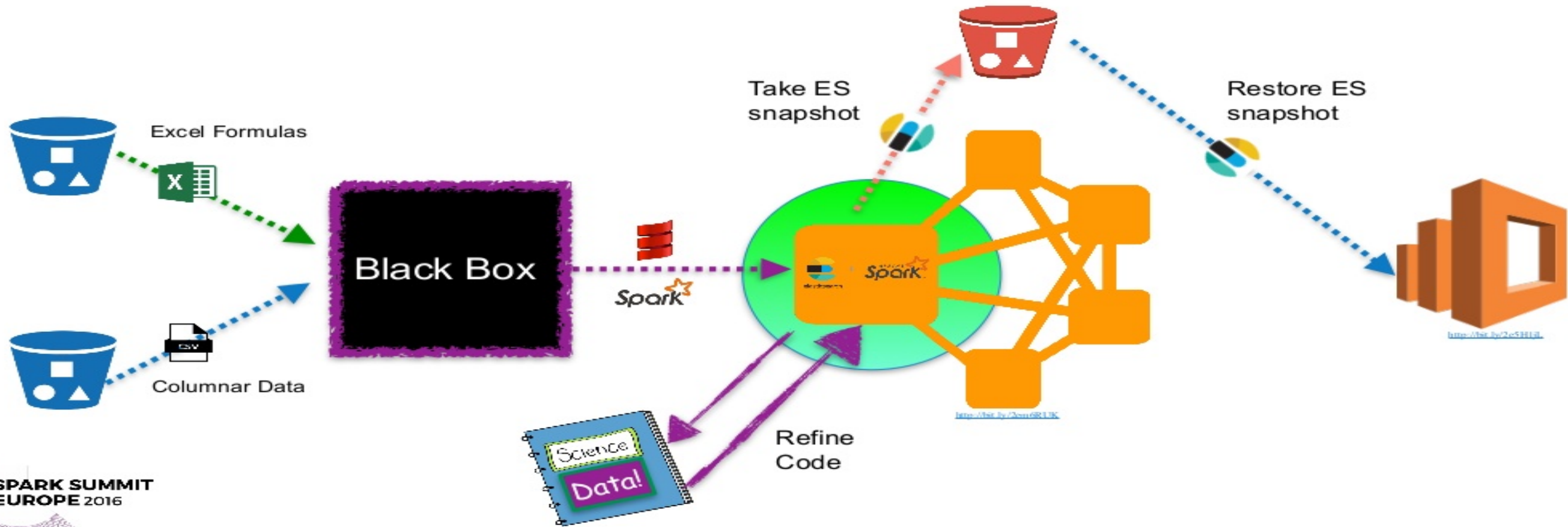
Spark.



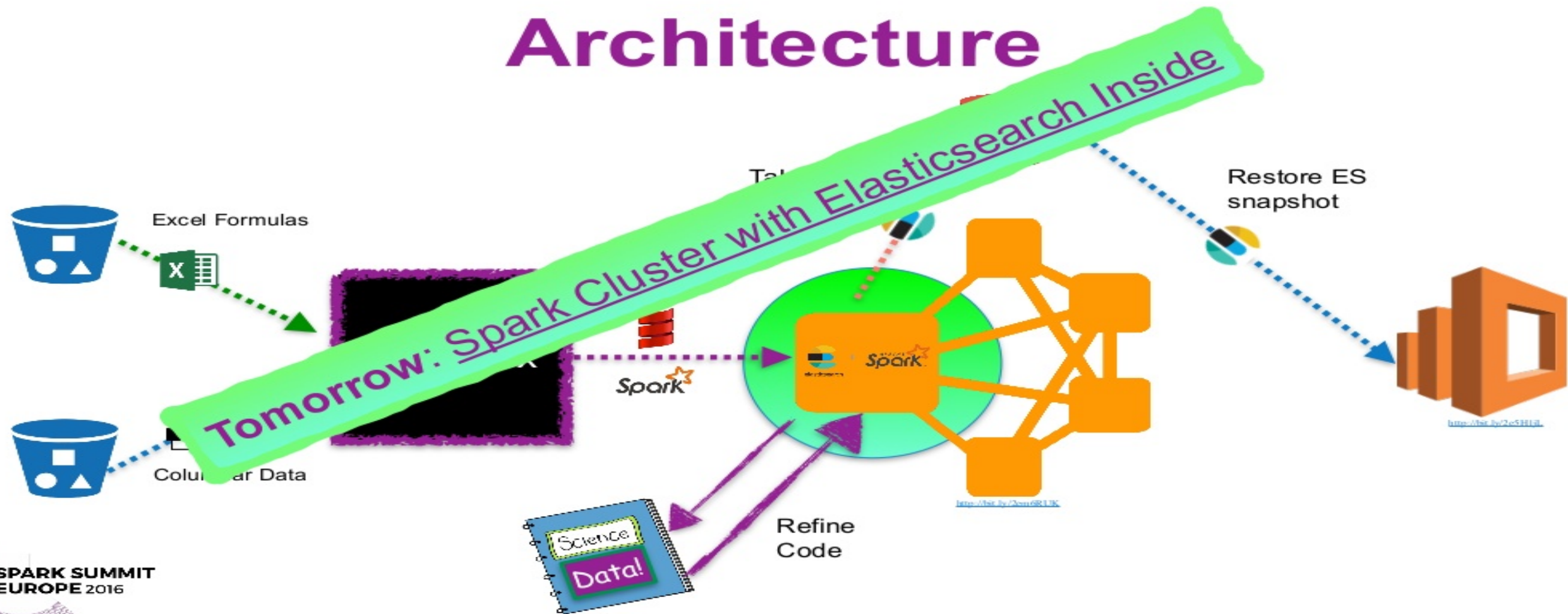
Architecture



Architecture



Architecture

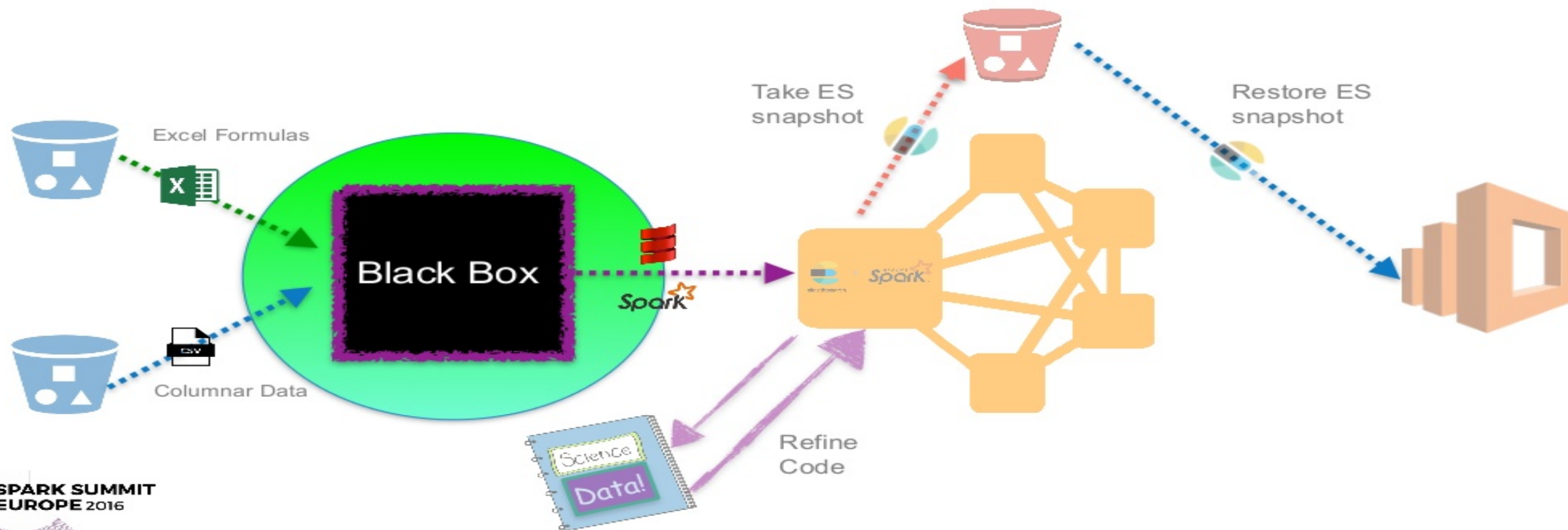


Program Transformation

Program Transformation

“A **program transformation** is any operation that takes a **computer program** and generates another program.”

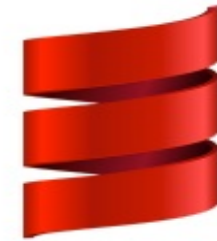
Architecture



Program Transformation Pipeline



<http://bit.ly/2e0TZA8>



<http://bit.ly/2e0ab44>

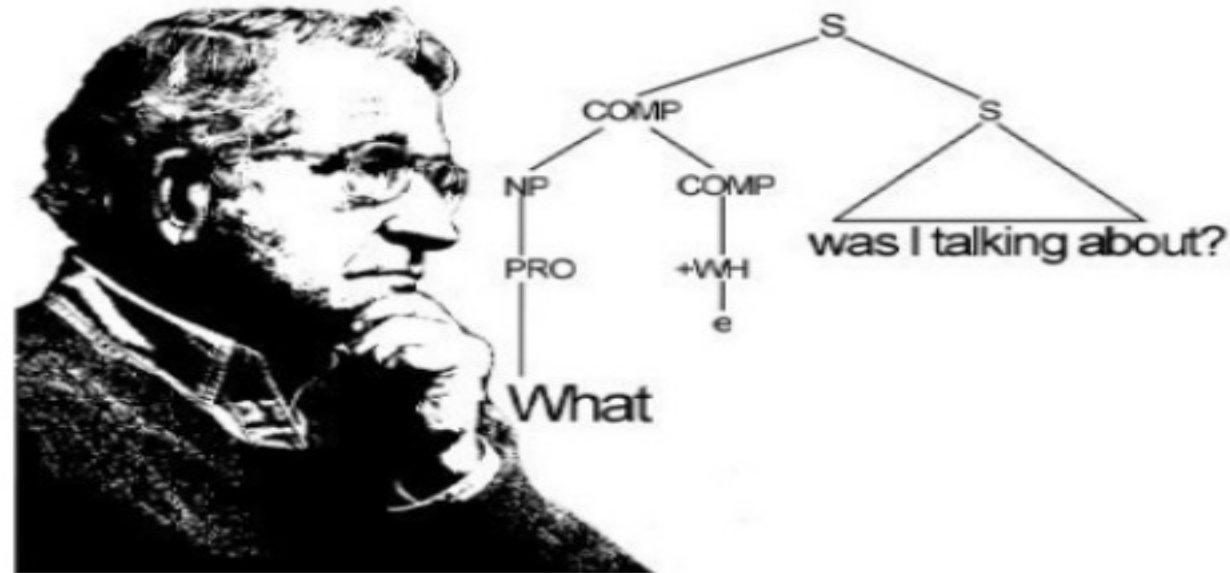


<http://bit.ly/2e0DeEg>

Code-to-Code Transformation

“The input to the code generator typically consists of a parse tree or an abstract syntax tree.”

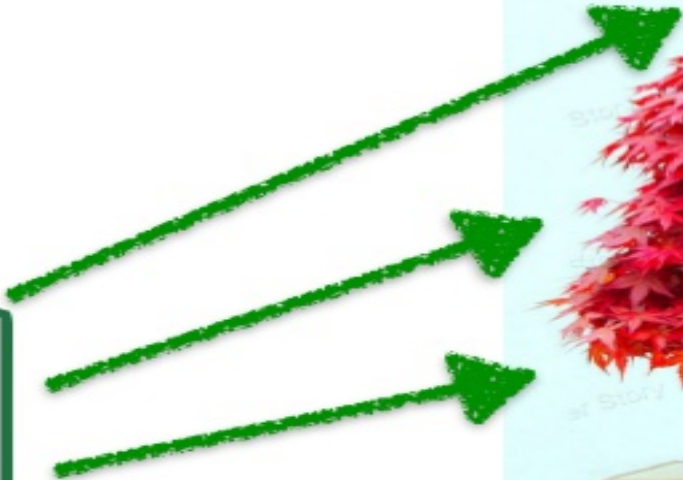
We need a Grammar!



<http://bit.ly/2dH0ybE>

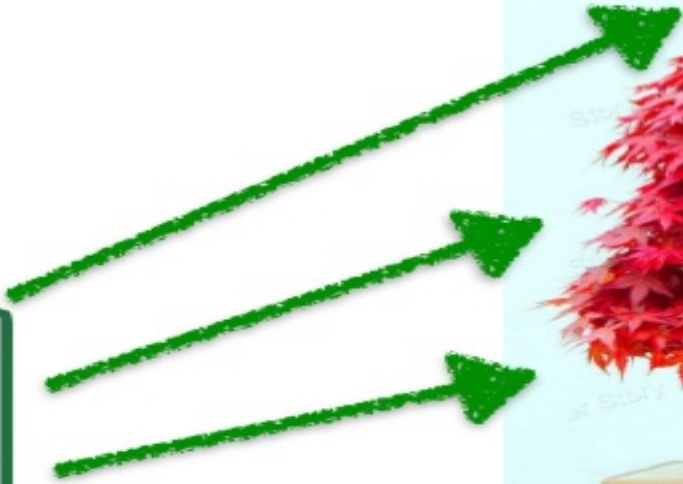
We need a Parse Tree!

Parse Excel Formulas



We need a Parse Tree!

Parse Excel Formulas

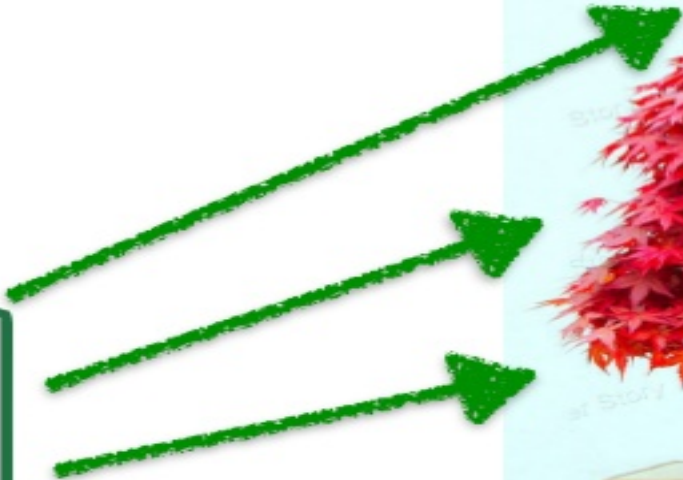


<http://bit.ly/2e0TZAS>

SPARK SUMMIT
EUROPE 2016

We need a Parse Tree!

Parse Excel Formulas



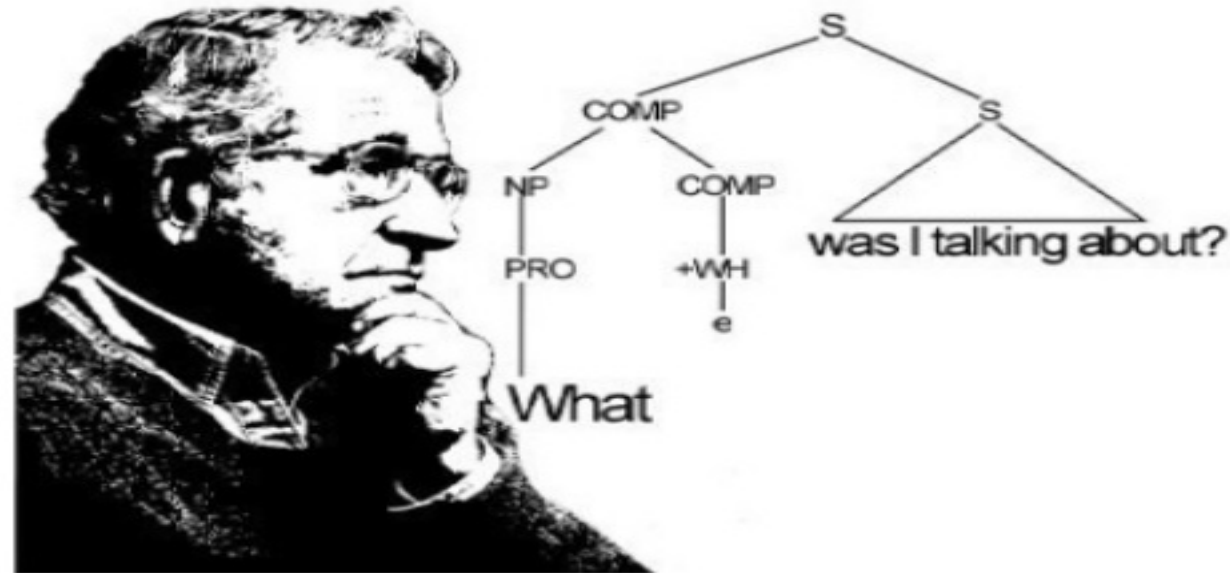
Generate Scala Code



<http://bit.ly/2e0TZAS>

SPARK SUMMIT
EUROPE 2016

We need a Grammar!



<http://bit.ly/2dH0ybE>

XLParser

"If I have seen further, it is by standing upon the shoulders of giants"

— Sir Isaac Newton

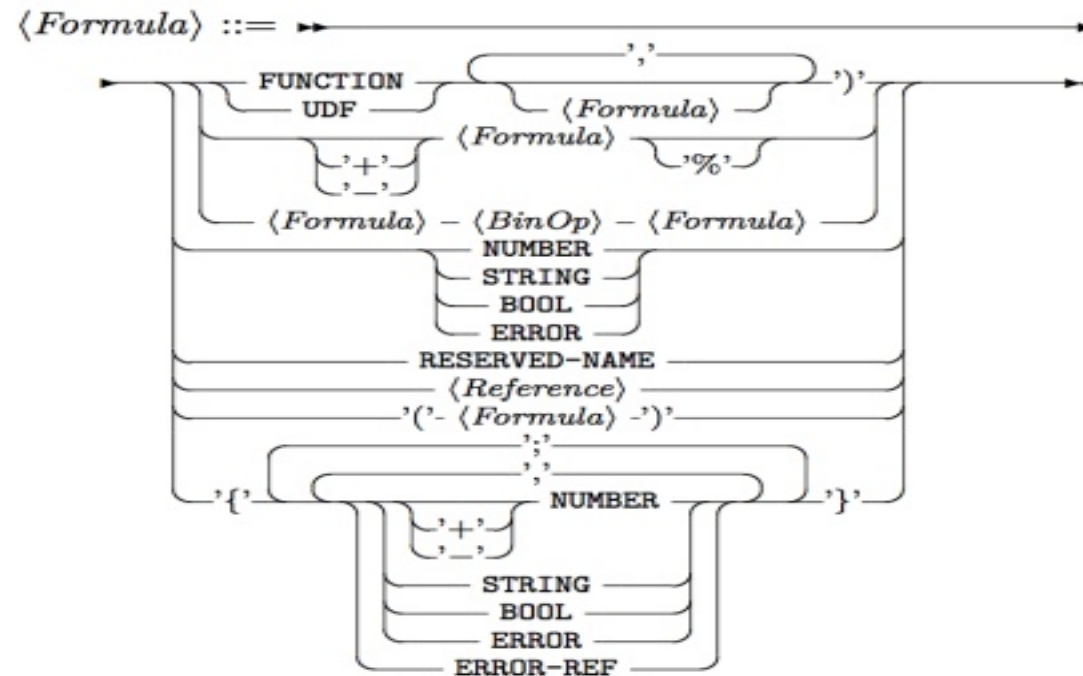
A Grammar for Spreadsheet Formulas Evaluated on Two Large Datasets

Efthimia Aivaloglou, David Hoepelman, Felienne Hermans
Software Engineering Research Group
Delft University of Technology
Mekelweg 4, 2628 CD Delft, the Netherlands
e.aivaloglou@tudelft.nl, d.j.hoepelman@student.tudelft.nl, f.f.j.hermans@tudelft.nl

XLParser

"If I have seen further, it is by standing upon the shoulders of giants"

— Sir Isaac Newton



A Grammar for Spreadsheet Formulas Evaluated on Two Large Datasets – Efthimia Aivaloglou, David Hoepelman & Felienne Hermans, Proceedings of SCAM '15

XLParser

"If I have seen further, it is by standing upon the shoulders of giants"

— Sir Isaac Newton

Excel Formula

SUM(A,C)



<http://xlparser.perfectxl.nl/demo>

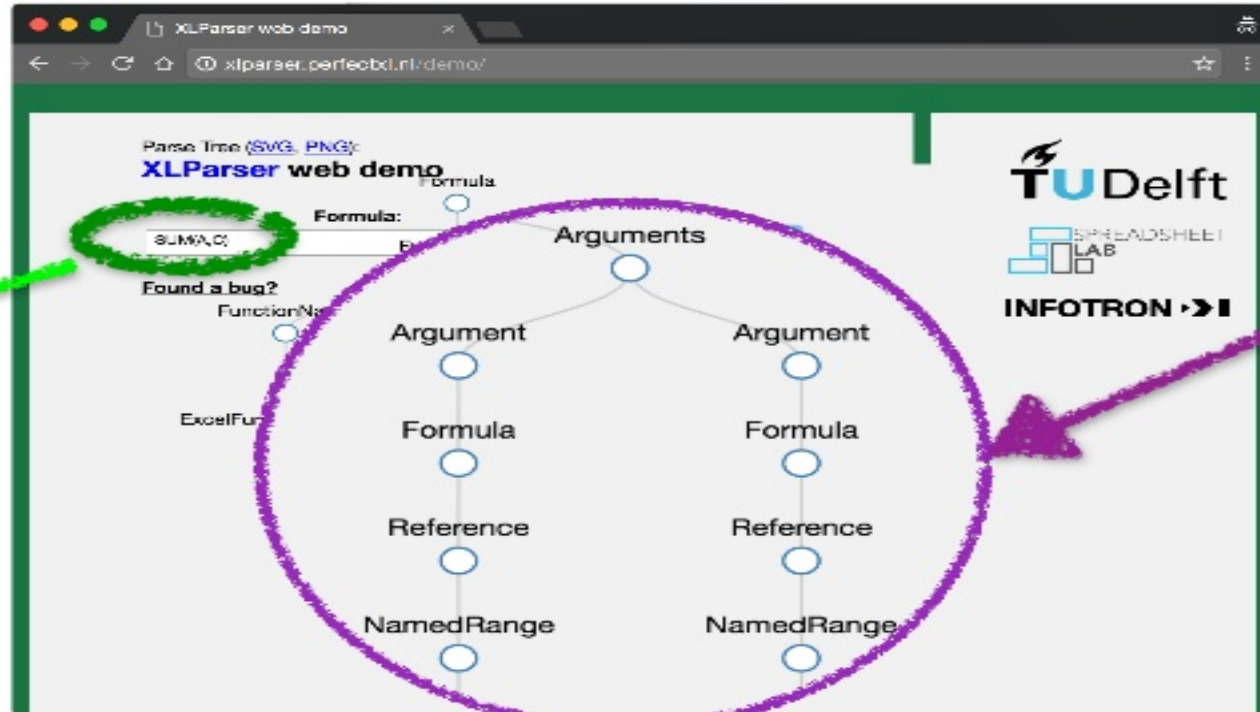
XLParser

"If I have seen further, it is by standing upon the shoulders of giants"

— Sir Isaac Newton

Excel Formula

SUM(A,C)



Parse Tree!

<http://xlparser.perfectxl.nl/demo>