DryadLINQ: A System for General-Purpose Distributed Data-Parallel Computing Using a High-Level Language

Yu et al. 2008

Presentation for R202

Data Centric Systems and Networking
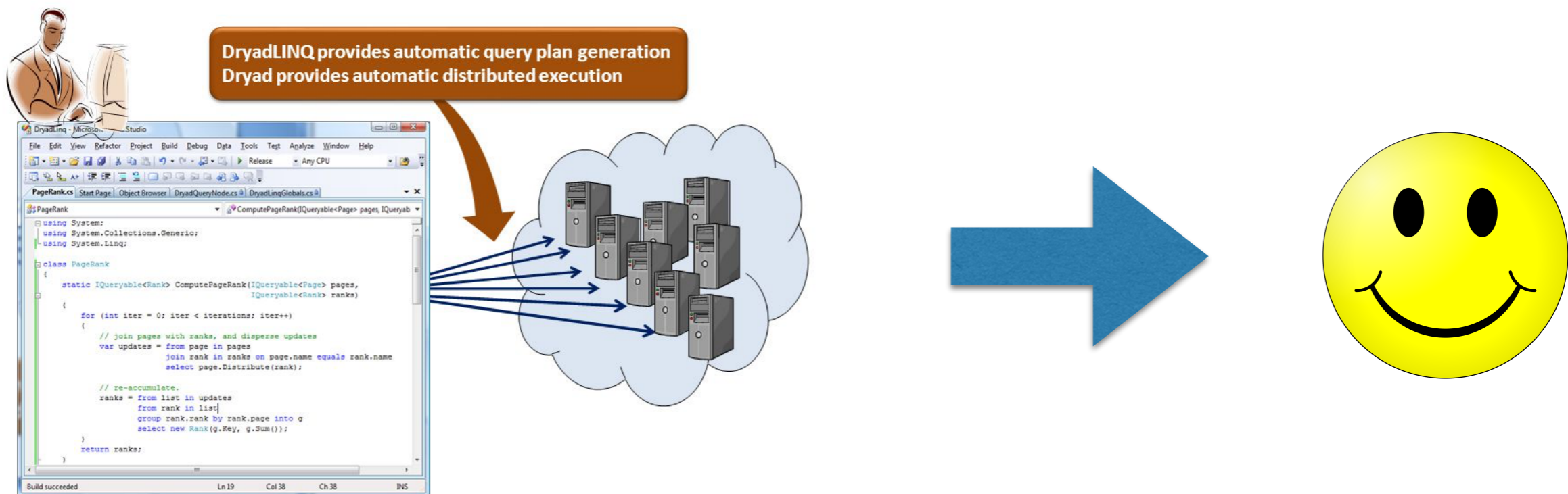
James Trever

20 October 2015

Overview

- Set of language extensions - LINQ Expressions

- Built on .NET

- Generalises SQL, MapReduce and Dryad

Motivation

- Make computation on large sets of data easier for developers

- SQL is to limited for large data sets

What else was going on

- Language

- Systems

- Storage

How have things changed?

- DryadLINQ and Dryad are longer supported
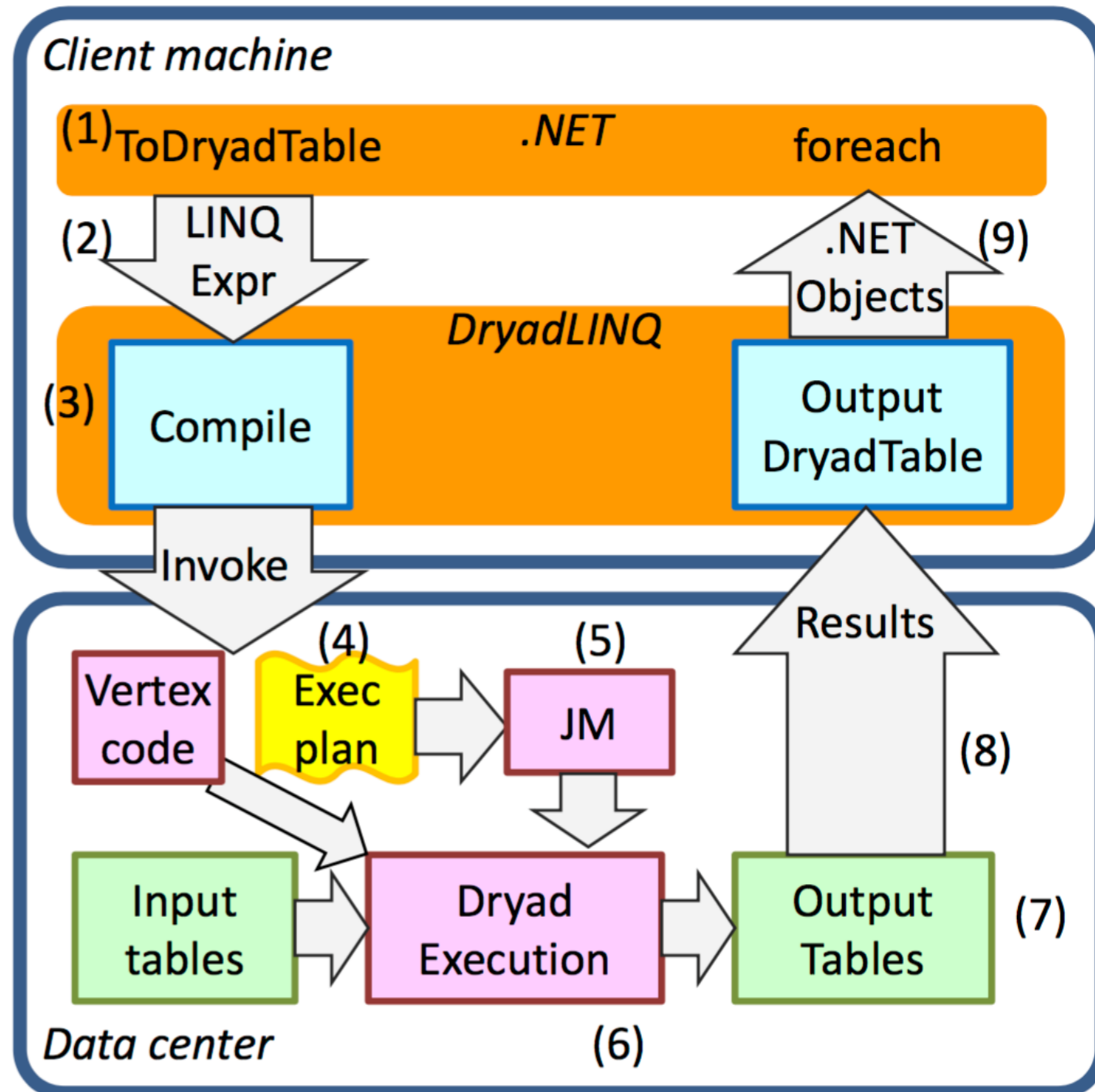
- Apache Spark is growing

Key Ideas

- Building on .NET allows use of other high level languages

- Expressive data model of strongly typed .NET objects

- Compiler deals with the scheduling, distribution and fault tolerance

What did they actually do

- They implemented the high level language extensions

- Implemented the compiler with static optimisations

How it works

Results

- Good feedback from users

- Apparently good results for their testing

- Minimised the amount of code needed

Critical Evaluation

- Good Idea
- Multiple tests run on the system
- Not had to touch Dryad
- .NET Language Support
- Accomplished Goals
- Easy Debugging

- No performance Debugging
- No real comparative tests