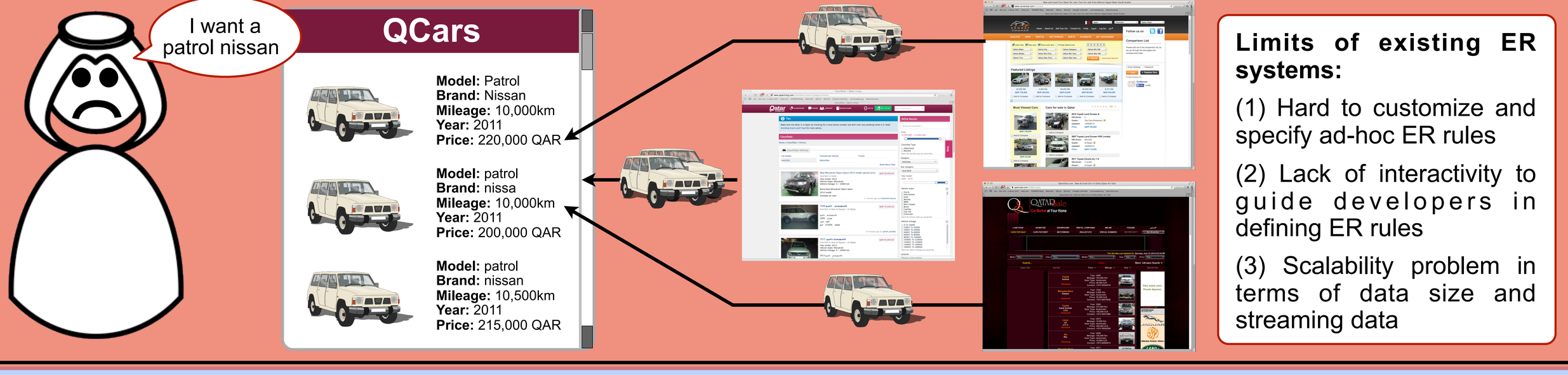
NADEEF/ER: INTERACTIVE ENTITY RESOLUTION

Ahmed Elmagarmid[1], Ihab Ilyas[2], Mourad Ouzzani[1], Jorge Quiané[1], Nan Tang[1], Si Yin[1]

[1] Qatar Computing Research Institute [2] University of Waterloo

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو في مؤسسة قطر



Limits of existing ER systems:

(1) Hard to customize and specify ad-hoc ER rules

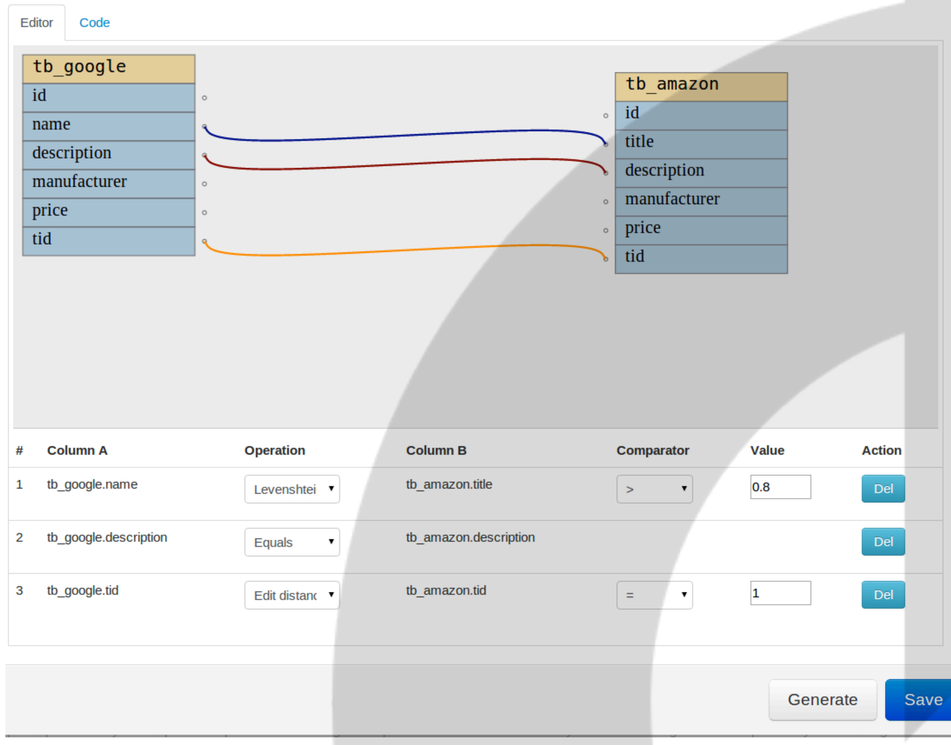
(2) Lack of interactivity to guide developers in defining ER rules

(3) Scalability problem in terms of data size and streaming data

NADEEF/ER

Easy specification

Rich graphical interface for rule specification

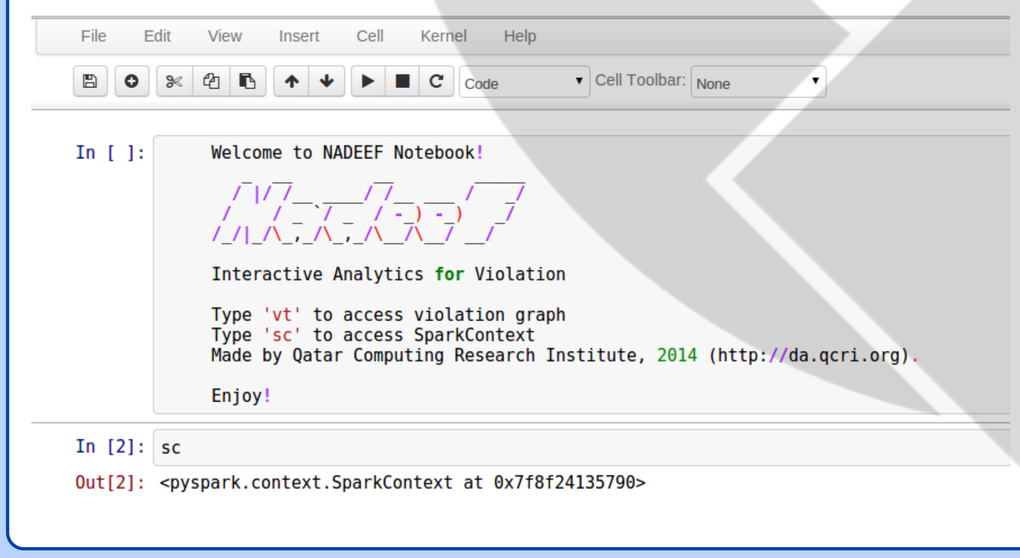


Extensibility

Customizable ER process

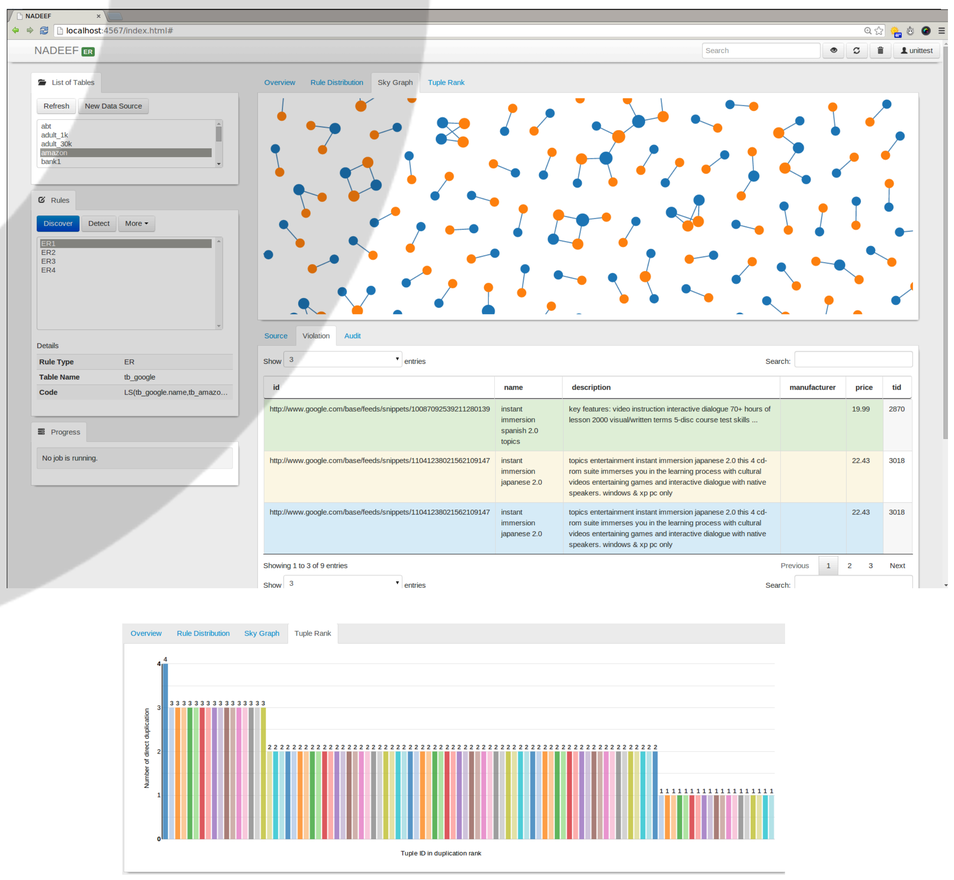
Scalable Ad-hoc Analytics

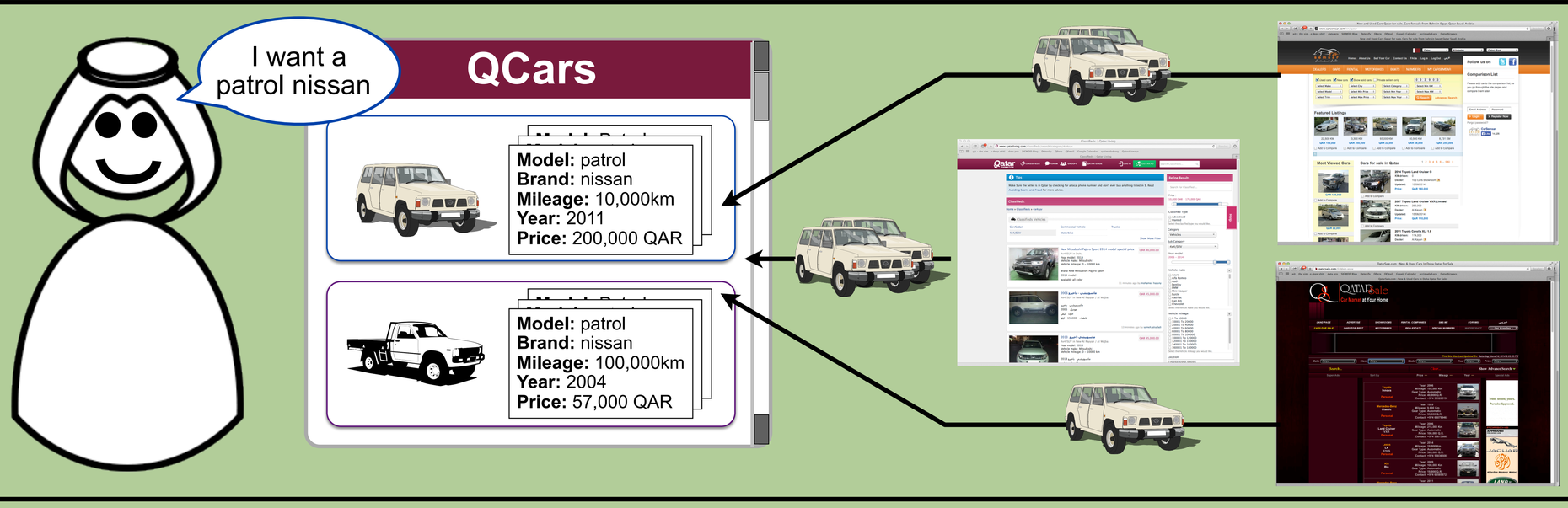
Spark, Python, and R support



Interactivity

Better understanding of duplicates





NADEEF/ER Advantages:

(1) Generality

(2) Ease-of-use

(3) Data repair integration

(4) Scalability