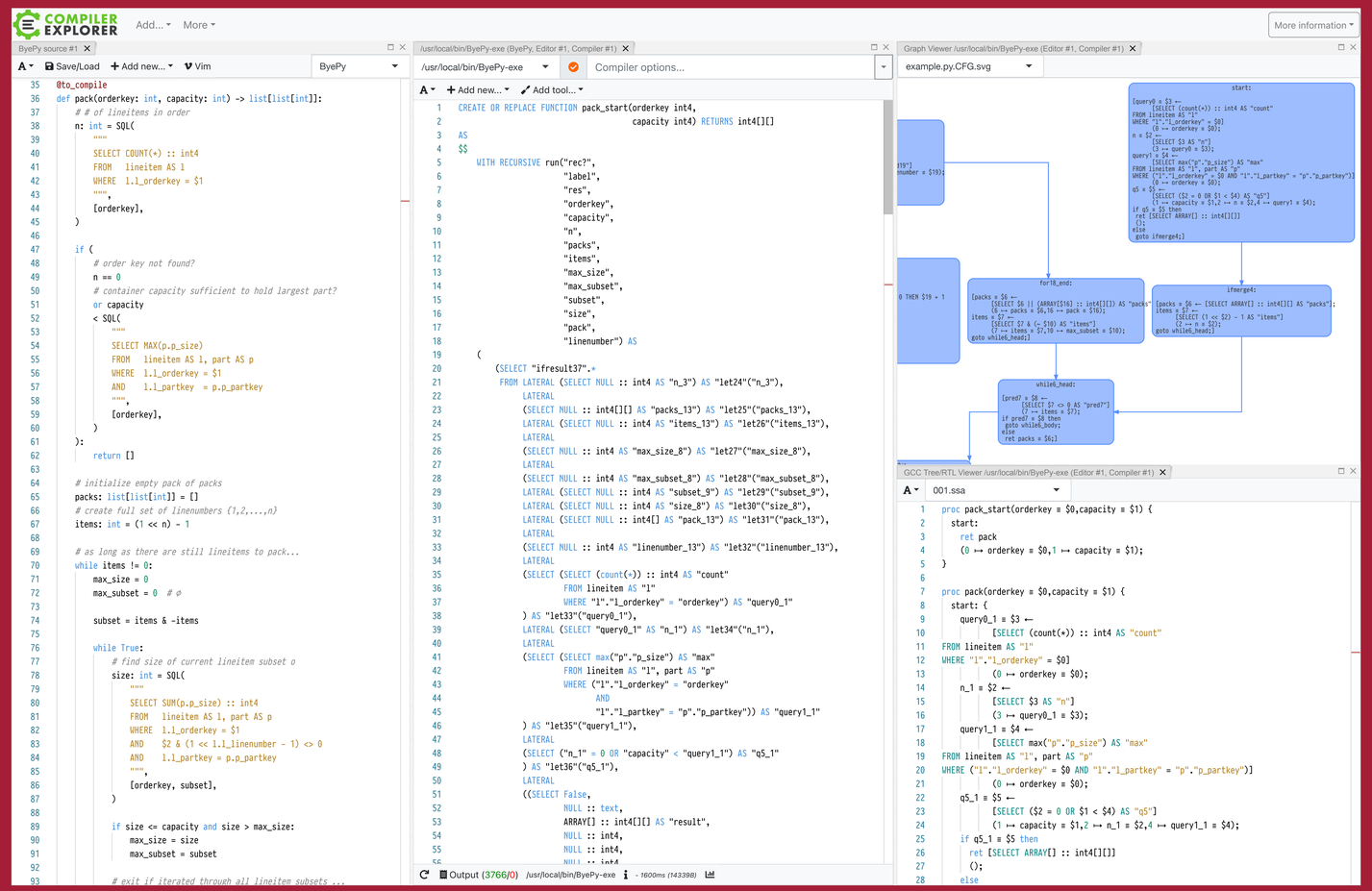
Tim Fischer
Denis Hirn
Torsten Grust

[tim.fischer,denis.hirn,torsten.grust]@uni-tuebingen.de

Snakes on a Plan — ByePy

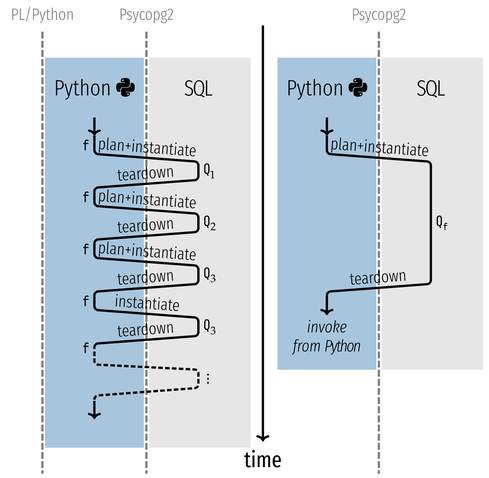
Bye, Python! — How we compile UDFs with **complex control flow** into **one recursive SQL CTE.**



Python Subset Understood by ByePy

- looping and iteration (`while`, `for v in range/array`)
- flow control statements (`continue`, `break`, `return`),
- conditional statements (`if`, `elif`, `else`) as well as expressions (e_1 `if` e_2 `else` e_3),
- variable assignment ($v = e$, v `+=` e, $v[e_1] = e_2$) and reference,
- lists ($[e_1, \ldots, e_n]$), indexed access and slicing ($e_1[e_2]$, $e_1[e_2 : e_3]$), stateful list methods (e`.pop`, e`.append`, e`.extend`),
- a large range of builtin operators and functions (`+`, `%`, `**`, `&`, `~`, `<<`, `<=`, `==`, `and`, …, `len`, `max`, `ceil`, `sqrt`, `coalesce`, …),
- embedded read-only queries (`SQL(`q`,[`e_1`,…,`e_n`])`)

Interplay of the Python Interpreter and SQL Engine



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

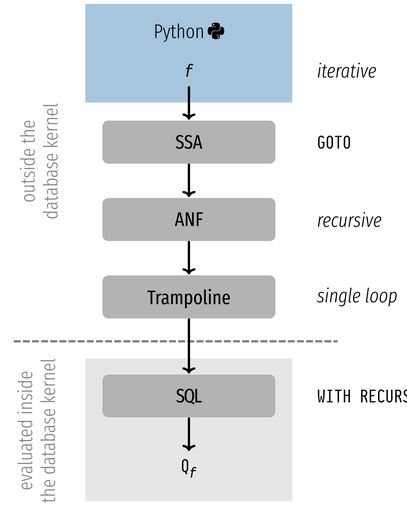
Try the demo
https://apfel-db.cs.uni-tuebingen.de

Download the full paper
https://db.cs.uni-tuebingen.de/staticfiles/publications/snakes-on-a-plan.pdf

Compiling UDF f: Intermediate Program Forms

1. Bring f into **SSA form** in which iterative as well as conditional control flow is *exclusively* expressed in terms of **`GOTO`**,
2. translate the resulting *graph of SSA blocks* into a bundle of **tail-recursive functions in ANF**,
3. form a central **trampoline** function which dispatches to the functions in the bundle, then loops back to itself, and
4. **inline** the functions into the trampoline, after which the **recursive CTE** Q_f can be read off this final intermediate form.

Compiler Stages



A Collection of Compiled Python UDFs

UDF	Description	Speedup
force	n-body simulation (Barnes-Hut quad tree)	5.7×
march	track border of 2D object (Marching Squares)	10.0×
margin	buy/sell TPC-H orders to maximize margin	4.7×
markov	Markov-chain based robot control	3.6×
packing	pack TPC-H lineitems tightly into containers	16.0×
savings	optimize supply chain of a TPC-H order	31.0×
vm	execute program on a simple virtual machine	24.0×

ACM SIGMOD PODS 2022
Philadelphia, PA, USA