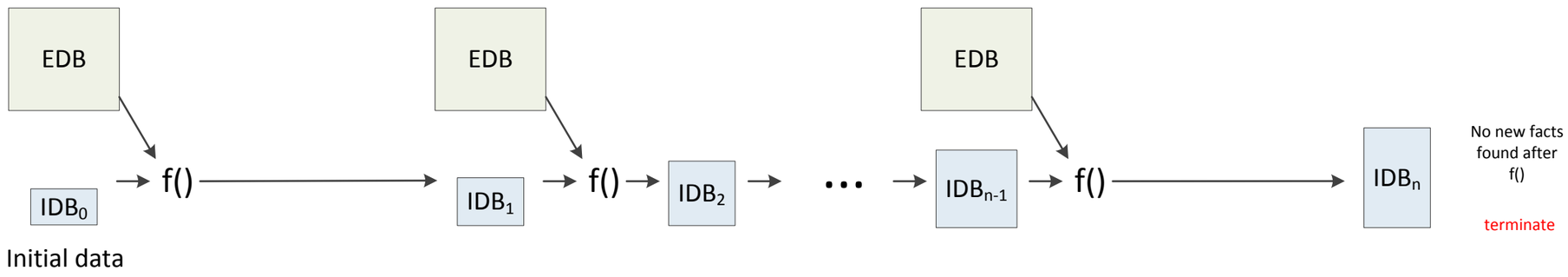
EDB: Extensional database

IDB: Intensional database

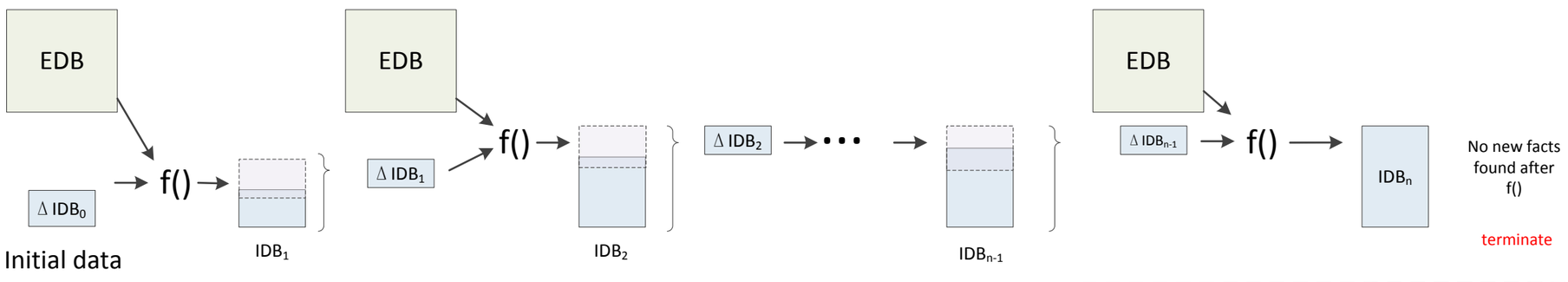
AS: Aggregated state

EDB, IDB: relational data
AS: key-value pairs

model 1:
General recursion
naïve evaluation
monotonic



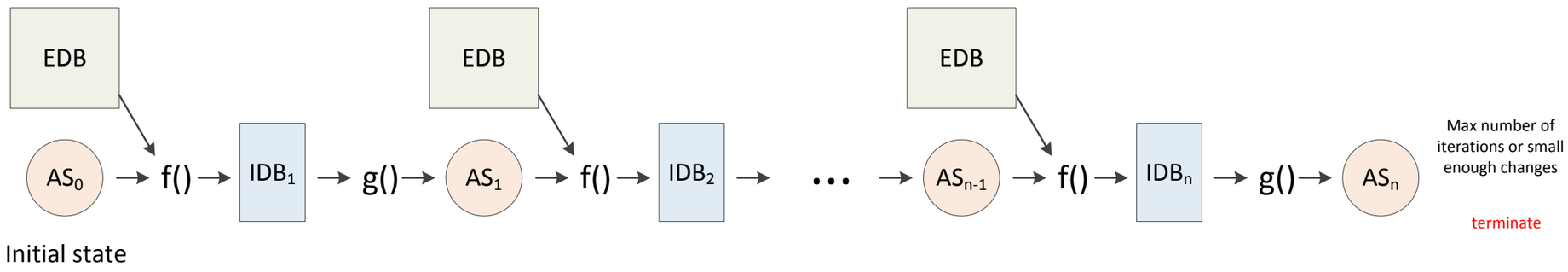
model 2:
general recursion
semi-naïve evaluation
monotonic



model 3:
recursion
with aggregation

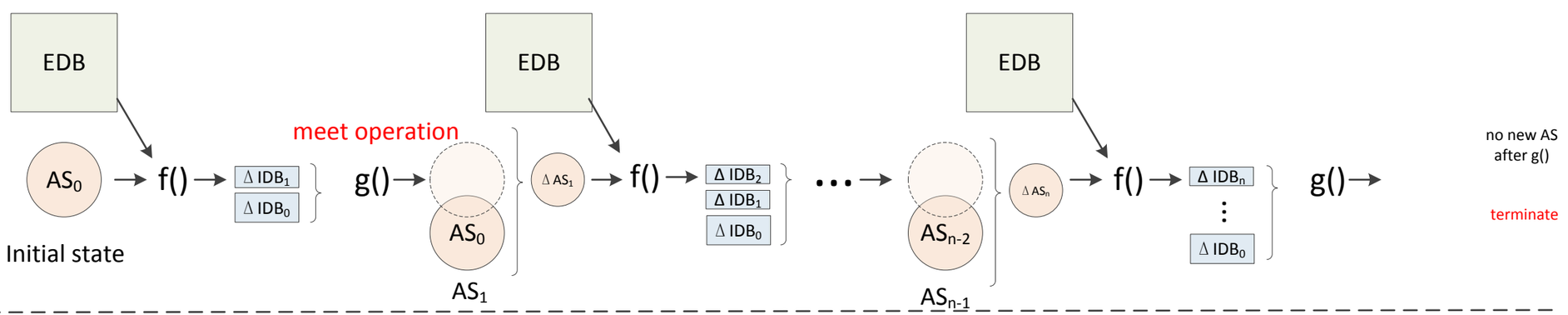
Non-monotonic

Datalography
[ICDB 2016]



model 4:
recursion
with aggregation
semi-naïve evaluation
monotonic

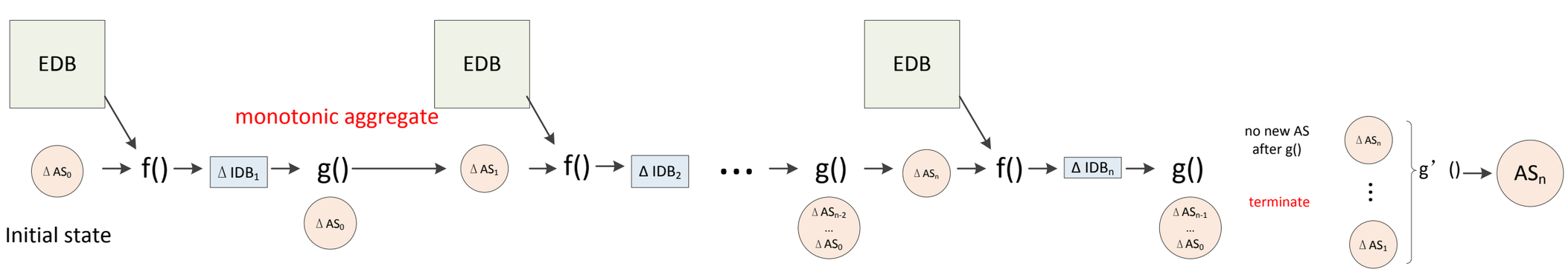
Socialite
(Stanford)
[ICDE 2013]



model 5:
recursion
monotonic aggregation
semi-naïve evaluation
append state

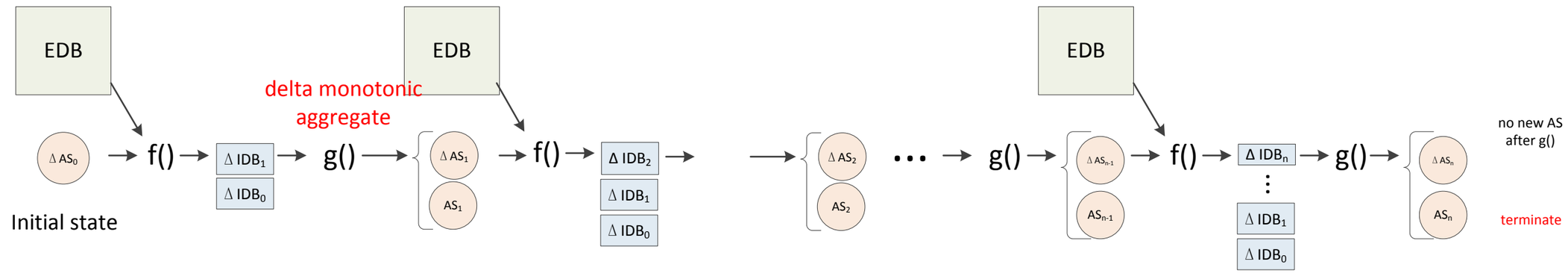
monotonic

DeAL (UCLA)
[VLDB 2013]
[ICDE 2015]



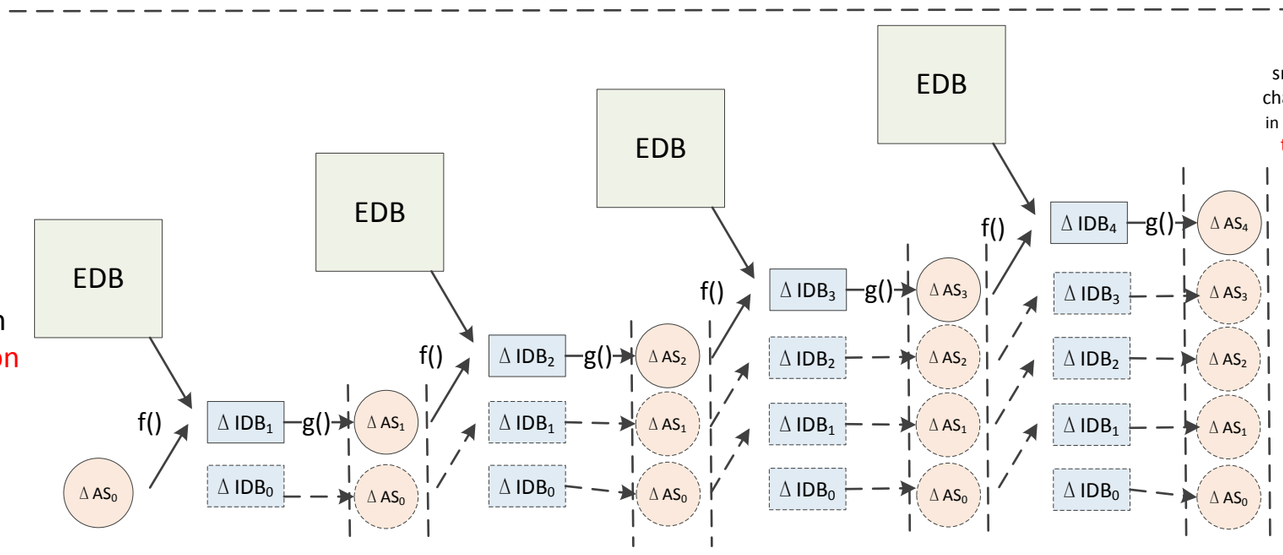
model 6:
recursion
delta monotonic
aggregation
in-place update
semi-naïve evaluation

monotonic



model 7:
recursion
delta monotonic
aggregation
in-place update
semi-naïve evaluation
Asynchronous evluation

monotonic



Conditions:

Monotonicity: $X^* = g(X \cup f \cdot g(X) \cup f \cdot g(f \cdot g(X)) \cup \ldots)$

Associativity: $g(X_1 \cup X_2) = g(g(X_1) \cup g(X_2))$

COmmutativity: $g(X_1 \cup X_2) = g(X_2 \cup X_1)$

Invariance: $f(g(X_1 \cup X_2)) = g(f(X_1) \cup f(X_2))$

Scheduling:

Since monotonic and order independent, schedule the largest Δ AS approaching to X*

Traditional Datalog:
LogicBox, IRIS, ... = model 1 or model 2
SQL recursive CTE = model 2

State-of-the-art Datalog:
DistributedSocialite (Stanford) [VLDB 2013] = model 4 + distributed
Async datalog (UW, Myria) [VLDB 2015] = model 4 + async execution
GLog (HongKong) [ICDE 2014] = model 3 + MapReduce
BigDatalog (UCLA) [SIGMOD 2016] = model 5 + Spark
Datalography (UCSD) [ICBD 2016] = model 3 + datalog + Giraph

Graph Processing:
GraphX, OptIQ, GraphChi, PowerGraph, PowerLyra⋯ = model 3 + sync
GraphLab, Giraph++, GraphUC, GRACE = model 3 + async

Unknown:
BOOM (Berkley) [EuroSys 2010] = model 2 + MapReduce ?
BLOOM (Berkley) [CIDR 2011] = ?