

## **Problem Statement**

Problem Domain: Web Data Integration For a user query over distributed (and possibly duplicated) data sources with unreliable statistics and with unreliable data arrival characteristics – devise a query processor that adapts to real-life changes

## Why Adapt?

- Autonomous & heterogeneous data source => Impacts the query cost function
  Data arrival is unpredictable => impacts static query
- plans
- Data overlap & Redundancy => impacts efficiency of query optimizer



## Tukwila System Components

- Data Schema Catalog
- Metadata view for varying data sources
- Query Reformulator
- Query Optimizer
- Query Execution Engine
- Wrappers



ן ר	lses source mappings
⊐ R t	eformulate user query in terms of he mediated sources
	reate query plan of operators
🗆 D	vivide into fragments
	Atomic units of execution



A plan consists of:

- Fragments
  Pipelined units of execution
  Tree of query operators
- Process each fragment
- □ Decide do we need to re-optimize?















## References

- [1] http://lbd.epfl.ch/e/teaching/SlidesST/adaptive-2005.pdf
- [2] http://wwwlsr.imag.fr/EDBT2002/Other/edbt2002PDF/EDBT2002School-Manolescu.pdf
- [3] <u>http://telegraph.cs.berkeley.edu/techover.html</u>