



# Javlin, a.s.

Den s průmyslovými partnery FI 10. 12. 2009

December 9, 2009

- Founded in 2005
- Global HQ in Prague, Offices in Washington, DC and UK
- ~45 Employees
- Developer of the CloverETL Solution
- Javlin solutions for mass data processing include
  - Data Quality
  - Master Data Management
  - Data Integration
  - Data Warehousing

# A few of Javlin's Worldwide Customers



PRICEWATERHOUSECOOPERS 

T-Mobile



Initiate 



accenture

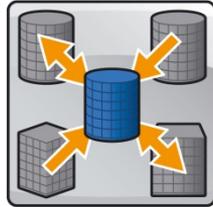


# What is CloverETL



- Dominate your data : huge batches or lots of small ad-hoc jobs
  - ▶ Read from one or more data sources of almost any type
  - ▶ Transform data in almost any way imaginable
  - ▶ Write data out to any number of data targets of almost any type
- Common Approach : Programming or heavy DBA work
- Rapid Setup. Very fast execution. Scalable.
- Runs on PC, Mac OS X, Linux, Unix, AS/400
- GUI Driven

# CloverETL - Dominate Your Data



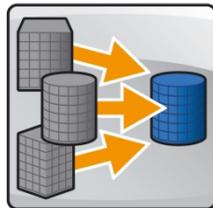
Master Data Management

Master Data Management



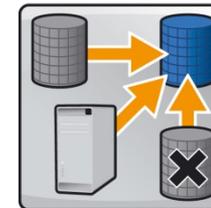
Data Migration

Data Migration



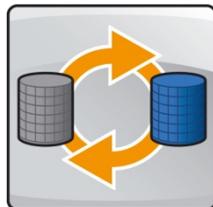
Data Warehouse

Data Warehouse



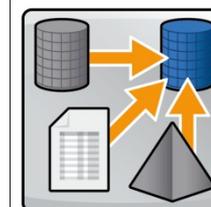
Data Consolidation

Data Consolidation



Data Synchronization

Data Synchronization



Data Integration

Data Integration

# CloverETL - Continuous Product Line



- CloverETL Desktop Designer



- CloverETL Server



- CloverETL Cluster



# Clover Designer - Quick Demo



The screenshot displays the Clover Designer interface within an Eclipse SDK. The main workspace shows a data flow graph for a project named 'SOAP\_EXAMPLE.grf'. The graph starts with an 'XML Reader' component, which feeds into two parallel 'Join Location with...' components. These lead to 'Join time with lay...' and 'Join temp value wi...' components, both followed by 'Simple Copy' components. The 'Simple Copy' components feed into 'Dedup' components. One 'Dedup' component feeds into a 'Create synthetic k...' component, which then feeds into another 'Dedup' component. The 'Dedup' components feed into 'Join Location' and 'Join time with temp' components. The 'Join Location' component feeds into a 'Simple Copy' component, which then feeds into an 'Aggregate' component. The 'Aggregate' component feeds into two 'XLS Data Writer' components. The interface also includes a Navigator, Outline, Properties, Console, Problems, Clover - Regex Tester, Clover - Graph tracking, and Clover - Log/Debug panels.

| Name                       | Status      | CPU ... | CPU | Avg bytes | Avg rows | Total bytes | Total rows |
|----------------------------|-------------|---------|-----|-----------|----------|-------------|------------|
| Phase # 0 (12 s, 8.525 KB) |             |         |     |           |          |             |            |
| AGGREGATED (0%)            | FINISHED... | 0 s     | 0%  |           |          |             |            |
| DEDUP0 (0%)                | FINISHED... | 0 s     | 0%  |           |          |             |            |
| DEDUP1 (0%)                | FINISHED... | 0 s     | 0%  |           |          |             |            |
| EXT_HASH_JOIN0 (0%)        | FINISHED... | 0 s     | 0%  |           |          |             |            |
| EXT_HASH_JOIN1 (0%)        | FINISHED... | 0 s     | 0%  |           |          |             |            |

# CloverETL - Server



 **CloverETL**

License for CloverETL Server 2.8  
Demo license for CloverETL server 2.8  
Server time: Tue Dec 8 13:44:23 2009

Tester  
en  
logout  
Page complete

[Monitoring](#) [Executions history](#) [Sandboxes](#) [Users](#) [Groups](#) [Launch Services](#) [Scheduling](#) [Tasks history](#)

[Graph event listeners](#) [File event listeners](#)

**Filter**

run ID:  graph:

from date:  to date:

status:  executed by:

| run ID | Node   | sandbox / graph  | executed by | status      | started               |                        |
|--------|--------|--|-------------|-------------|-----------------------|------------------------|
| 24     | node01 | <a href="#">default/graph/graphMountainsCSV.grf</a>          | clover      | FINISHED_OK | 2009-11-10 02:51:39.0 | <a href="#">detail</a> |
| 23     | node01 | <a href="#">playground2/graph/graphAggregateUnsorted.grf</a> | clover      | FINISHED_OK | 2009-11-09 09:22:19.0 | <a href="#">detail</a> |
| 22     | node01 | <a href="#">playground2/graph/graphAggregateSorted.grf</a>   | clover      | FINISHED_OK | 2009-11-09 09:20:04.0 | <a href="#">detail</a> |
| 21     | node01 | <a href="#">playground2/graph/graphAggregateSorted.grf</a>   | clover      | FINISHED_OK | 2009-11-09 09:18:39.0 | <a href="#">detail</a> |
| 20     | node01 | <a href="#">default/graph/graphMountainsCSV.grf</a>          | clover      | FINISHED_OK | 2009-11-09 08:15:31.0 | <a href="#">detail</a> |

# CloverETL Cluster



## Increased processing power & throughput

- ▶ Parallel execution of transformations over multiple servers / nodes
- ▶ Load balancing based on individual nodes utilization

## Increased fault tolerance

- ▶ Fail over in case of problems with particular node(s)
- ▶ Data replication

## Increased flexibility

- ▶ Cluster can be dynamically reconfigured by adding or removing nodes on the fly

Large quantity of  
data Loaded



Processed in  
parallel in Cluster



Written out in  
parallel



# Cloud Computing



- Amazon EC2, Voxel, Rackspace, MS Azure / Google App Engine, Salesforce
- Rent as many servers as you like by the minute / hour / month
- Highly configurable and rapidly deployable
  - ▶ Server CPU / RAM / Hard Disk setup fully configurable
  - ▶ Offline storage (costs extra / GB)
  - ▶ Data snapshot capability and Machine Images
  - ▶ Provider server farms and storage available West Coast, East Coast, Europe (Amazon)
  - ▶ Content Delivery Network
  - ▶ Autoscaling, Monitoring, Load-Balancing
  - ▶ Custom and automated deployments

## The Washington Post

- 20,000 pages of non-searchable images released on paper
- Needed to convert to text format using OCR within 24 hour news cycle
- Data had to be scanned, formatted, image cleaned, resized and OCRed
- 200 Amazon EC2 server instances used for 9 hours
- Total cost \$144.62

# CloverETL Cluster - In the Cloud

---



- Clover in the Sky with Diamonds!
- CloverETL Server Cluster will often run on In-House Servers
- But can be deployed very rapidly on Cloud servers
- Will offer substantial cost savings when load is controlled and predictable

# CloverETL Cluster In the Cloud



- Processing of 6 million records of data
  - Calculates discount offered and date reformat
  - sort all 6 million records
  - Aggregate total Price, Discount and Quantity
  - Calculate shipping period
  - Sort by Ship Time
  - Write out slowest 100 orders to be fulfilled

| <b>Notebook</b> | <b>2 Nodes</b> | <b>3 Nodes</b> | <b>4 Nodes</b> |
|-----------------|----------------|----------------|----------------|
| 150 seconds     | 60 seconds     | 47 seconds     | 31 seconds     |



# Cloud Computing - Yearly Costs



| Utilization       | Cloud Cost (Amazon)   | In-House Servers      |
|-------------------|-----------------------|-----------------------|
| 5%                | \$340                 | \$8'700               |
| 10%               | \$680                 | \$8'700               |
| <b><u>25%</u></b> | <b><u>\$1'700</u></b> | <b><u>\$8'700</u></b> |
| 50%               | \$3'400               | \$8'700               |

# CloverETL Cluster Scalability

---



- Ditch versus hole
- Excellent scalability for data that can be “chunked”
- Each Node reads a separate “chunk” of the source data
- Even better performance can be realized if source data can be pre-split by the source application

# Who are we looking for?

---



- **Java/C programmers**
  - Challenge-oriented programming
  - Rapidly developing family of products
  - Dbs, UNIX, data management experience is an advantage
  - English

# Who are we looking for?

---



- **Consultants**
  - You have the knowledge
  - We offer the experience
  - Work abroad (international customers)
  - Problem solving skills
  - General knowledge in IT
  - Soft skills (customer oriented work)
  - Willingness to accept challenges and learn
  - English