Optimizing dataflow

Arne Brucksch
About me

Arne Brucksch
Run faster and use less energy!

- Get results faster
- Use less RAM (OoM errors)
- Multi processing
Why and how to do that?
The Start!
Check your workstation, workspace and log files.

- Edit header: FME_PROFILE_RESULT_RESULT_CSV
- Where is the time going?
- Check temporary directory (you want it on a SSD!)
- Run 64bit FME or use 3GB switch on 32bit
FME Transformers
The “good” and the “bad” guys

<table>
<thead>
<tr>
<th>“Good”</th>
<th>“Bad”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Reader</td>
<td>Feature Holder</td>
</tr>
<tr>
<td>Feature Writer</td>
<td>Feature Merger</td>
</tr>
<tr>
<td>Attribute Manager</td>
<td>Overlay</td>
</tr>
<tr>
<td>Attribute Keeper/Remover</td>
<td>Joiner</td>
</tr>
<tr>
<td>Custom Transformers</td>
<td>Readers</td>
</tr>
<tr>
<td></td>
<td>Writers</td>
</tr>
</tbody>
</table>
The Order
# The good and the bad guys

**Good**
- Remove Attributes
- Remove logger and inspector
- Input is Ordered By Group
- Minimize the data to read
- Rethink the order by logic

**Bad**
- Join all with all
- Keep needless attributes
- Overlay all on all
- Writing the huge date at least
Moving processes out of FME
Working with DB!

- Joins, less Feature Merging only the data you want (create views)
- Minimize select use «where» and Published Parameters
- On spatial DB, overlay in DB (use Indexes)
Working with open source or other software!

- Use systemcaller
  (psql - PostGIS)
  (gdal - Raster - Vektor)

...
Multi processes
(with Linked Custom Transformer)
Change from main workbench to Custom Transformer to Linked CT

The Main
- Use Feature readers / writers

Create CT (green)
- Group by (FME_PROCESS_GROUP_BY)
- Parallel: aggressive / extreme
- Feature to read

Create LCT (blue)
- Use PP to modify your LCT in different workspaces
Thank you!
Takk skal du har!
Dankeschön!

If you have any further questions, please don’t hesitate to contact me.
arnbru@norkart.no