IT-Capacity Analysis and Forecasting with KNIME and R

Markus Schmid
T-Systems International GmbH

KNIME UGM  Zurich, 2014-02-12
AGENDA

T-Systems
Capacity Management: Scope and Challenges
Capacity Reporting with KNIME: Architecture
Real-Life examples (KNIME/R/BIRT)
  • Resource level
  • Service-Level
Forecast-Approach
Lessons learned
Summary
ABOUT T-SYSTEMS

- **T-Systems International:**
  - present in more than 20 countries worldwide
  - 52,000 employees in total, about 23,000 in Germany
- **T-Systems provides IT Services for the Deutsche Telekom Group as well as for external customers**
  - **Telekom-IT:**
    - T-Systems division with focus on
      - Applications development & operation
      - IT support for complex business processes
  - for the Customer Deutsche Telekom
IT CAPACITY MANAGEMENT: SCOPE

Balancing of Costs and Capacity

„As small as possible, still as big as necessary“

- Scalability
- Capacity
- Performance
- Costs

12.02.2014 IT-Capacity Analysis and Forecasting with KNIME and R / Dr. Markus Schmid
Balancing of Costs and Capacity
„As small as possible, still as big as necessary“

Scope:
- IT-Capacity (primarily logical and physical server infrastructure, storage)
  - Initial sizing for new projects
  - Capacity monitoring and forecasting for systems in operation

Scalability
Capacity
Performance
Costs
IT CAPACITY MANAGEMENT: SCOPE

Balancing of Costs and Capacity
„As small as possible, still as big as necessary“

- Scope:
  - IT-Capacity
    (primarily logical and physical server infrastructure, storage)
    - Initial sizing for new projects
    - Capacity monitoring and forecasting for systems in operation

- Non-Scope:
  - Staff
  - Desktop systems
Purpose of IT infrastructure: Support of business processes

- technical monitoring depicts load that is typically caused by business activities
- In a telecommunications company typically complex process chains that involve a number of
  - business support systems (BSS)
  - operations support systems (OSS)

Business development has a direct impact on system load

- provisioning of additional capacity depends on underlying platform
  (classic servers, virtualization, cloud-environments)

Evaluation of business forecasts is essential for balanced capacity provisioning
CAPACITY REPORTING & FORECASTING

Capacity reporting

- Did things work out as planned?
- Are there long-term trends to react to?
  - Avoidance of capacity problems and incidents

Capacity forecasting

- Evaluate the impact of business forecasts to IT infrastructure
- Challenging in large-scale deployments
  - Permanent change in
    - processes
    - applications and interfaces
    - technical infrastructure
ARCHITECTURAL OVERVIEW
ARCHITECTURAL OVERVIEW

Asset data (CMDB)

Technical monitoring data

Capacity Warehouse
ARCHITECTURAL OVERVIEW

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Asset data (CMDB)

Technical monitoring data

Capacity Warehouse
ARCHITECTURAL OVERVIEW

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Asset data (CMDB)

Technical monitoring data

Capacity Warehouse

KNIME WebPortal

KNIME Server
ARCHITECTURAL OVERVIEW

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Asset data (CMDB)

Technical monitoring data

Capacity Warehouse

AdHoc analysis & specialized reports

KNIME WebPortal

KNIME Server
ARCHITECTURAL OVERVIEW

- AdHoc analysis & specialized reports
  - KNIME WebPortal
  - KNIME Worker
  - KNIME Server

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Asset data (CMDB)

Capacity Warehouse

Technical monitoring data
ARCHITECTURAL OVERVIEW

AdHoc analysis & specialized reports

KNIME WebPortal

KNIME Worker

KNIME Worker

KNIME Worker

KNIME Server

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Preprocessed data

Capacity Warehouse

Asset data (CMDB)

Technical monitoring data

DB access (JDBC)
ARCHITECTURAL OVERVIEW

- AdHoc analysis & specialized reports

KNIME WebPortal

KNIME Server

- Service monitoring data & forecasts
  - service invocations
  - concurrent users
  - product sales numbers
  - ...  

- Asset data (CMDB)

- Technical monitoring data

Preprocessed data

Capacity Warehouse

GNU R with

- service invocations
- concurrent users
- product sales numbers

DB access (JDBC)

With extension packages
ARCHITECTURAL OVERVIEW

AdHoc analysis & specialized reports

Preprocessed data

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Asset data (CMDB)

Technical monitoring data

Capacity Warehouse

KNIME WebPortal

KNIME Server

DB access (JDBC)

GNU R with extension packages

IT-Capacity Analysis and Forecasting with KNIME and R / Dr. Markus Schmid – internal – 12.02.2014

T·Systems
**ARCHITECTURAL OVERVIEW**

Service monitoring data & forecasts
- service invocations
- concurrent users
- product sales numbers
- ...

Asset data (CMDB)

Technical monitoring data

Preprocessed data

Capacity Warehouse

KNIME WebPortal

WebService Interface

KNIME Worker

KNIME Server

DB access (JDBC)

GNU R with extension packages

AdHoc analysis & specialized reports

automated generation of recurring standard reports (PDF)
SOME NUMBERS...

Recurring standard capacity reporting (per application)

• About 250 Reports per month
• One PDF-Report per application
  • From 30 to about 250 pages
  • File size between 2 and 25 MB

KNIME JDBC access to capacity warehouse

• Fine-grained data for the last 2-3 years
• Total: about 4.7 TB of data

KNIME workflow for standard report

• Consists of 2,468 nodes (and growing)
• Overhead due to preprocessing and formatting of data, error handling
• Worker Instance uses up to 10GB of main memory
PER APPLICATION OVERVIEW:
SERVER CPU-LOAD HEATMAP (MO-FR 08-18:00)
PER APPLICATION OVERVIEW:
SERVER CPU-LOAD HEATMAP (MO-FR 08-18:00)
# MAX CPU PER DAY (.95 PERCENTILE) (24HRS)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SERVER0001</td>
<td>18</td>
<td>16</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>SERVER0002</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>SERVER0004</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>SERVER0008</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>SERVER0009</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>SERVER0010</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0011</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>SERVER0012</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0013</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0014</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0015</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0016</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0017</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>SERVER0018</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
</tbody>
</table>

Max. CPU Usage (.95 percentile)

- no data
- 0-10 %
- 10-20 %
- 20-30 %
- 30-40 %
- 40-50 %
- 50-60 %
- 60-70 %
- 70-80 %
- 80-90 %
- 90-100 %
## TECHNICAL CAPACITY RATING - OVERVIEW

<table>
<thead>
<tr>
<th>Host</th>
<th>Component</th>
<th>CPU Rating</th>
<th>CPU Trend</th>
<th>Memory / Swap Rating</th>
<th>Memory / Swap Trend</th>
<th>Storage Rating</th>
<th>Storage Trend</th>
<th>I/O Rating</th>
<th>I/O Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERVER0001</td>
<td>Web Server</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0002</td>
<td>MQ</td>
<td>18</td>
<td>→</td>
<td>20</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0003</td>
<td>Web Server</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0004</td>
<td>MQ</td>
<td>17</td>
<td>→</td>
<td>29</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0005</td>
<td>Stage A AppSrv</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0006</td>
<td>Stage B AppSrv</td>
<td>24</td>
<td>→</td>
<td>3</td>
<td>↑</td>
<td></td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>SERVER0007</td>
<td>Stage A AppSrv</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0008</td>
<td>Stage B AppSrv</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0009</td>
<td>Stage A AppSrv, Stage B AppSrv</td>
<td>29</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0010</td>
<td>Stage A AppSrv</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0011</td>
<td>Stage B AppSrv</td>
<td>24</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0012</td>
<td>Stage A AppSrv</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0013</td>
<td>Stage B AppSrv</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0014</td>
<td>Stage A AppSrv, Stage B AppSrv</td>
<td>29</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0015</td>
<td>Database Server</td>
<td>31</td>
<td>→</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERVER0016</td>
<td>Database Server</td>
<td>3</td>
<td>↑</td>
<td>31</td>
<td>→</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SERVER0014 : CPU LOAD

AIX 6.1, IBM, 9179-MHC PowerPC_POWER7

Reason: (RunQueue > Threshold) > 119 Min. (for 29 days)
<table>
<thead>
<tr>
<th>File system</th>
<th>Reason</th>
<th>Min free</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.175.36.207:/opt/WebSphere/install_sourcen, 52 G</td>
<td>Allocated space is constantly below 20% and growth rate is near 0 (31 days)</td>
<td>3 G</td>
<td>☹️ 31</td>
</tr>
<tr>
<td>/dev/chrootlv, 41 G</td>
<td>Allocated space is constantly below 20% and growth rate is near 0 (31 days)</td>
<td>11 G</td>
<td>☺️ 29</td>
</tr>
<tr>
<td>/dev/optwebs4, 41 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>23 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/varwebs4, 27 G</td>
<td>Allocated space is constantly below 20% and growth rate is near 0 (31 days)</td>
<td>25 G</td>
<td>☹️ 31</td>
</tr>
<tr>
<td>/dev/exportlv, 21 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>8 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/cognoslv, 11 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>3 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/optoralv, 11 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>2 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/hd2, 6 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>0 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/tqslv, 4 G</td>
<td>Allocated space is constantly below 20% and growth rate is near 0 (31 days)</td>
<td>2 G</td>
<td>☹️ 31</td>
</tr>
<tr>
<td>/dev/ITM6_lv, 3 G</td>
<td>Allocated space is constantly below 20% and growth rate is near 0 (31 days)</td>
<td>1 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/hd10opt, 3 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>1 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/opencv_lv, 3 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>1 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/hd4, 2 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>1 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/hd3, 2 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>1 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/hd1, 2 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>0 G</td>
<td>☺️ 31</td>
</tr>
<tr>
<td>/dev/hd9var, 2 G</td>
<td>Number of days &lt; 30 until filesystem reaches 100% capacity (8 days)</td>
<td>0 G</td>
<td>☺️ 31</td>
</tr>
</tbody>
</table>
BUSINESS DATA CONTROL CHART
SERVICE_X: TREND OF SERVICE INVOCATIONS PER DAY
BUSINESS DATA CONTROL CHART
SERVICE_X: TREND OF SERVICE INVOCATIONS PER DAY

Number of Service invocations per day

BUSINESS DATA CONTROL CHART
SERVICE_X: TREND OF SERVICE INVOCATIONS PER DAY
BUSINESS DATA CONTROL CHART
SERVICE_X: TREND OF SERVICE INVOCATIONS PER DAY

Number of Service invocations per day

Linear trend (reporting period)

Linear trend (long term)

Average number of historical service invocations per day of week +/- 3x std. deviation

IT-Capacity Analysis and Forecasting with KNIME and R / Dr. Markus Schmid
BUSINESS DATA CONTROL CHART
SERVICE_X: TREND OF SERVICE INVOCATIONS PER DAY

- Number of Service invocations per day
- Linear trend (reporting period)
- Linear trend (long term)
- Average number of historical service invocations per day of week +/- 3x std. deviation
- Anomaly in number of service invocations
Challenges:

- Constantly changing, large-scale environment: Use of fine-grained modeling techniques is not adequate
- Point of measurement for business data often unknown
  - Possibly delay in measurements, unknown impact on quality, ...
  - Mapping to IT infrastructure is challenging

Determine correlation between historical business data and technical monitoring data (e.g. CPU load)

- Multi-step process:
  - Mapping of business data to technical metrics (e.g. service calls)
  - Mapping of service calls to resource load
  - Resource mapping based on CMDB asset data
  - Complexity: Quality of data source, granularity of data; resources may be active or inactive, changes in resource capacity, ...
  - Make sure, that all relevant load factors have been taken into account

Use regression techniques to forecast infrastructure load based on business forecast

- Possibly use historical information (e.g. on forecast quality) to improve predictions
FORECASTING OF RESOURCE LOAD

- Calibrate model (hist. period A)
- Store model in DB
- Forecast
- Verify forecast (hist. period B)
BUSINESS DATA / CPU LOAD: STATISTICAL ANALYSIS

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimate</strong></td>
<td>Estimate</td>
</tr>
<tr>
<td>StdErr</td>
<td>StdErr</td>
</tr>
<tr>
<td>95% LCL</td>
<td>95% LCL</td>
</tr>
<tr>
<td>95% UCL</td>
<td>95% UCL</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
</tr>
<tr>
<td>Pr(&gt;</td>
<td>t</td>
</tr>
<tr>
<td>19,442</td>
<td>19,442</td>
</tr>
<tr>
<td>0,8104</td>
<td>0,0015</td>
</tr>
<tr>
<td>17,8514</td>
<td>0</td>
</tr>
<tr>
<td>21,0326</td>
<td>0,0016</td>
</tr>
<tr>
<td>23,9918</td>
<td>38,625</td>
</tr>
<tr>
<td>0,6427</td>
<td>0</td>
</tr>
</tbody>
</table>

IT-Capacity Analysis and Forecasting with KNIME and R / Dr. Markus Schmid – internal – 12.02.2014
VISUALISATION OF CORRELATION BETWEEN REFERENCE SERVICES AND OTHER SERVICES
CORRELATION BETWEEN REFERENCE SERVICES AND OTHER SERVICES

Histogram of Residuals

Histogram of Standardized Residuals

Residuals vs Fitted

Normal Q-Q

Scale-Location

Residuals vs Leverage

IT-Capacity Analysis and Forecasting with KNIME and R / Dr. Markus Schmid
Prognose auf Basis Modell-Kalibrierungszeitraum 01.11.2013 bis 31.12.2013

Prognostic based on model calibration period from 01.11.2013 to 31.12.2013
VERIFICATION OF CPU FORECAST: SERVER_B

Prognose auf Basis Modell-Kalibrierungszeitraum 01.11.2013 bis 31.12.2013
LESSONS LEARNED:
LARGE SCALE WORKFLOW DESIGN

KNIME/BIRT + R is a powerful tool combination for statistical data analysis and graphical presentation.

Processing of large data sets is easily possible.

While KNIME scales well across multiple CPUs, BIRT only uses a single core.

KNIME allows easy transition from ad-hoc analysis to provisioning of automated, recurring tasks.

Designing and testing of large workflows is a complex task.

- In-workflow documentation of functionality is essential!
- Use a proper build chain with development, testing and production environment (Also helpful for testing upgrades)
LESSONS LEARNED: LARGE SCALE WORKFLOW DESIGN

Always make sure, your input data is stored with the correct column type (int vs. double problem)

- If unsure, enforce conversion before loop-ends and after invocation of R nodes

Document your column types, names and order of data to report nodes and make sure, they don’t change (Otherwise BIRT may silently delete some of your scripts)

If things start to slow down: Check the heap memory requirements of your workflows

- 75%-Threshold on Pool “PS Old Gen” causes GC

Server-based execution stops on some errors you don’t notice when testing in KNIME desktop (e.g. unconnected nodes):

- Check the logs
  - ..., but debugging is still hard with thousands of nodes

Decrease the debug-level in production: this significantly speeds things up
Thank you...

Any Questions?