

**List of Topics for DHBW Seminar Work papers in the area of Data Warehouse / Business Intelligence (without Points)**

- **Team Size/Effort/Pages:** group work (2 members); ~10-12 hours/~10-15 pages
- **Language/Deadline/Details:** English/22.12.2023/ Examination Info - Seminar Work
- **Evaluation:** Dr. Hermann Völlinger (send to [hermann.voellinger@gmail.com](mailto:hermann.voellinger@gmail.com))
- **References:** As a source of information and further references to the respective themes, it is recommended to refer to the instructions in the corresponding slides of the lecture.

No	Topic	Details	Students/Points (max=100)
DW01	<b>Investigate the BI-Data Trends in 2023</b>	Prepare/show the results of the e-book “BI_Daten_Trends_2023”. Compare Moodle: <a href="https://elearning.dhbw-stuttgart.de/moodle/pluginfile.php/573359/mod_folder/content/0/BI-Data-Trends-2023_DE.pdf">https://elearning.dhbw-stuttgart.de/moodle/pluginfile.php/573359/mod_folder/content/0/BI-Data-Trends-2023_DE.pdf</a> How can DWH & BI help to overcome the current problems (i.e. food supply shortages, global climate crisis, etc.) and build also the basics for more digitalization and Artificial Intelligence (AI) solutions? Examine 10 data trends to support these requirements.	: X
DW02	<b>Investigate the catchwords: DWH, BI and CRM</b>	Investigate the catchwords. Information sources are newspaper or magazine articles or books (see literature list). Show also trends or new development in these areas, which are defined by the catchwords (project reports are also possible): 1. Data Warehousing (DWH) 2. Business Intelligence (BI) 3. Customer Relationship Management (CRM)	: X
DW03	<b>Compare three Data Catalogue Tools</b>	Select 3 of the Data Catalogue (DC) tools from the two “Market Study - DC” slides and prepare a report (SW paper) about the functionality of these tools. Information source is the internet. See also links in the “Market Study –DC” slides. See the directory “Supporting Material” in the Moodle: Kurs <a href="#">DW 21E</a> ).	: X
DW04	<b>Compare DWH Architectures</b>	Compare the three DWH architectures (DW only, DM only and DW & DM) in the next slide. List the advantages and disadvantages and give a detailed explanation for it. Find also a fourth possible architecture (hint: ‘virtual’ DWH). Solution Hint: Use the criteria-table of the lecture.	: X
DW05	<b>DWH for Basel II/III and RFID</b>	Theme: Give a definition and impact of new trends on Data Warehouse: 1. Basel II / Basel III 2. RFID Look also for examples of current projects in Germany. Information sources are newspaper or magazine articles or internet.	: X
DW06	<b>Referential Integrity (RI)</b>	Explain what is “Referential Integrity” (RI) in a Database? Sub-Questions: • What means RI in a Data Warehouse? • Should one have RI in a DWH or not? (collect pro and cons) Find explanations and arguments in DWH forums or articles about this theme in the internet or in the literature.	: X
DW07	<b>Compare STAR-with SNOWFLAKE-model</b>	Compare MDDM Model schemas STAR and SNOWFLAKE. See in IBM Redbook ‘Data Modelling Techniques for DWH‘ (see DWH lecture) chapter 6.4.4. Build a list of advantages and disadvantages for each of these concepts, in the form of a table (same as in the lecture).	: X

DW08	<b>Discover- and Prepare-Tools</b>	In the lecture we have seen the steps - “Discover” “Prepare” and “Transform” - for a successful data population strategy. Please investigate for the first two steps examples of two tools for each step (see lecture). Show details like functionality, price/costs, special features, strong features, weak points, etc. You can use the examples of the lecture or show new tools, which you found in the internet or you know from your current business.	: <b>X</b>
DW09	<b>3 ETL Tools</b>	Show the highlights & build a Strengthens/Weakness Diagram for the following 3 ETL Tools. Use the information from the internet: 1. Informatica – PowerCenter --→ <a href="http://www.informatica.com">www.informatica.com</a> 2. IBM - Infosphere Inform. Server - DataStage ---→ <a href="https://www.ibm.com/us-en/marketplace/datastage?loc=de-de">https://www.ibm.com/us-en/marketplace/datastage?loc=de-de</a> 3. Oracle – Warehouse Builder (OWB) --→ <a href="https://docs.oracle.com/cd/B28359_01/owb.111/b31278/concept_overview.htm#WBDOD10100">https://docs.oracle.com/cd/B28359_01/owb.111/b31278/concept_overview.htm#WBDOD10100</a> . Show the three tools in competition to each other.	: <b>X</b>
DW10	<b>3 Reporting Tools</b>	Show the highlights and build a Strengthens/Weakness Diagram for the following 3 Reporting Tools. Use the information from the internet: 1. MicroStrategy --→ <a href="http://www.MicroStrategy.com">www.MicroStrategy.com</a> 2. BusinessObjects ---→ <a href="http://www.BusinessObjects.com">www.BusinessObjects.com</a> 3. Cognos ---→ <a href="http://www.Cognos.com">www.Cognos.com</a> Show the three tools in competition to each other.	<b>X</b>
DW11	<b>Compare three leading DM/ML tools</b>	Search for the actual “Gartner Quadrant” of DS/ML (DM) tools (see lecture). Give detail descriptions of three of the leading tools in the quadrant. See the link in the lecture.	: <b>X</b>
DW12	<b>Evaluate the Technology of UseCase “Semantic Search”</b>	Evaluate and find the underlying technology, which is used in “UseCase – Semantic Search: Predictive Basket with Fact-Finder”. See: <a href="https://youtu.be/vSWLafBdHus">https://youtu.be/vSWLafBdHus</a> Found also more information about Computer Linguistic (NLP) and AI within Fact-Finder under: <a href="https://www.fact-finder.com/library">https://www.fact-finder.com/library</a>	: <b>X</b>
DW13	<b>K-means Clusters of IRIS Dataset</b>	The Iris dataset contains the data for 50 flowers from each of the 3 species - Setosa, Versicolor and Virginica. <a href="http://www.lac.inpe.br/~rafael.santos/Docs/CAP394/WholeStory-Iris.html">http://www.lac.inpe.br/~rafael.santos/Docs/CAP394/WholeStory-Iris.html</a> . The Develop a Python program by using the Scikit-learn library can be seen under: <a href="https://github.com/bhattbhavesh91/k_means_iris_dataset/blob/master/K_in_K_means_Clustering.ipynb">https://github.com/bhattbhavesh91/k_means_iris_dataset/blob/master/K_in_K_means_Clustering.ipynb</a> Describe the solution steps of the python program.	: <b>X</b>
DW14	<b>Advanced Analytics (AA) vs. Artificial Intelligence (AI)</b>	Look for example on the blog: <a href="https://seleritysas.com/blog/2019/05/17/data-science-and-data-analytics-what-is-the-difference">https://seleritysas.com/blog/2019/05/17/data-science-and-data-analytics-what-is-the-difference</a> Give a short summary of this blog. If necessary you can also use additional information from the internet. What are the main statements? What are the similarities and what are the differences.	: <b>X</b>
DW15	<b>Run &amp;compare 3 KNIME Data Mining solutions</b>	Run and compare the three KNIME-Basics workflow of exercise E9.5. Give technical explanations to the solution steps (use given workflows). See the KNIME documentation from the lecture.	: <b>X</b>
DW16	<b>Image-Classif. with MNIST Data using KNIME</b>	Rebuild the complete KNIME Workflow (use given workflow) of Exercise E10.4 for Image-Classification and give technical explanations to the solution steps. Use the information from the KNIME documentation. Formulate and explain your own insights.	: <b>X</b>