**Final Report References**

**Literature Review**

**Arneil, J. (2012) My Year With Exadata. *Cover Art: UK Oracle User Group Conference 2012*  [online]. ICC Birmingham 3-5 December. e-DBA.**

**Password: CONF2012OUG [Accessed 08 December 2012] Available at:**

# <http://2012.ukoug.org/presdisplayfile.asp?prs\_prsid=7038&filename=UKOUG%5FOCT%5FMy%5Fyear%5Fwith%5FExadata%2Eppt>.

**This source is quite useful as it gives a consumers perspective on Exadata, a summary of exactly what Exadata is, their experiences from the set up, it also goes into detail on the technical specifications of Oracles newly released Exadata X3 which improves greatly on previous versions and finally sheds light on the hardware and software maintenance needed to keep Exadata running. This source is similar to the work of (Kutrovsky 2012) which also goes into detail on Exadata however the source focuses more on the software side for example how the data moves in Exadata while this source is more focused on the new hardware changes with this latest instalment of Exadata. The source therefore provides useful information and does not come directly from Oracle so it is less likely to be biased about Exadata advantages and disadvantages, therefore this source will help me answer my academic question and achieve my aims as supporting material to my research on the emerging database systems, it will be useful to compare them to Oracle's system for handling big data, as well as being incorporated into the report.**

Borkar, R, V., Carey, J, M. and Li, C. (2012). Big data platforms: What's next?. *XRDS: Crossroads*, [online]. **19**(1), pp.44-49 [Accessed 15 February 2013]. Available at:

<http://delivery.acm.org.ezproxy.wlv.ac.uk/10.1145/2340000/2331057/p44-borkar.pdf?ip= 134.220.1.139&acc=ACTIVE%20SERVICE&CFID=281029974&CFTOKEN=14962939&\_\_acm\_\_=1361652571\_a266ed3632371081205a63df4ca9716f>

# This source describes the current status of big data, how it is being used, the history behind it, the source describes several technologies that have emerged to deal with big data such as MapReduce framework, Hadoop, BigTable, dynamo, HBase, Cassandra, Hive, Pig. The source then looks the future and introduces its own system ASTERIX, a big data management system (BDMS), the system is described and is compared to several other big data technologies showing how big data systems have advanced over the years. The source is quite useful as it shows what big data is used for, the technologies that deal with it and a new technology that takes advantage of the developments in big data technology, similar to the work of Tene and Polonetsky (2012), both sources describe big data in detail however this source focuses more on the technology for handling big data while the previously referenced source focuses more on the issues with big data. This source will help me answer my academic question and achieve my aims when it is used to help create my artefact and to make a better decision on its configuration as well as in the final report for when I evaluate the completed artefact.

Bhatewara, A., Waghmare, K. (2012). Improving Network Scalability Using NoSql Database. *International Journal of Advanced Computer Research*, [online]. **2**(6), pp.488-490 [Accessed 21 February 2013] Available at: <http://www.theaccents.org/ijacr/papers/conference/icett2012/87.pdf>

# The source describes technologies more suited to the unstructured nature of big data in this case for use in the cloud such as NoSQL database and Cassandra, it shows the advantages and disadvantages of each and how they compare to a RDBMS. This source is useful as it shows that if higher scalability is needed without upgrading hardware, these technologies are a viable option for this, similar to the work of Stonebraker (2012), they both show technologies that can give the user greater performance when dealing with big data however this source evaluates NoSQL and Cassandra while comparing them to RDBMS which the previously referenced source does not. This source will help me answer my academic question and achieve my aims by being used as reference when creating my artefact and when evaluating it in the final report.

# Brooks, J. (2011). **Oracle** Gives **NoSQL** Market a Boost. *eWeek*, [online]. 28(17), pp.8 [Accessed 27 February 2013] Available at:

# <http://wk6kg9sd8m.search.serialssolutions.com/?ctx\_ver=Z39.88-2004&ctx\_enc=info%3Aofi%2Fenc%3AUTF-8&rfr\_id=info:sid/summon.serialssolutions.com&rft\_val\_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.atitle=Oracle+Releases+NoSQL+Database+Linked+to+2012+Big+Data+Appliance+Launch&rft.jtitle=eWeek&rft.date=2011-10-24&rft.pub=Ziff+Davis+Enterprise&rft.issn=1530-6283&rft.externalDBID=n%2Fa&rft.externalDocID=270709934>

# 

# This source describes the events at Oracle OpenWorld 2011, where Oracle announced its own NoSQl database, this is useful as it provides me with information on another big data technology that may be included into my artefact, Oracle NoSQL is shown to be based upon the Java version of Oracle's BerkeleyDB, as well as being included with services such as supporting scale-out over large numbers of nodes. This source is similar to the work of Henschen, D. (2011), they both give a background of Oracle's NoSQL database and show how it is based upon previous technologies however this source is simply based upon the technology itself and briefly mentions competitor technologies while the previously mentioned source details Oracle's NoSQL database in the context of big data as well as comparing the technology to Cassandra, which this source does not. This source will help me answer my academic question and achieve my aims by being used as reference when creating my artefact and when evaluating it in the final report.

# Clegg, D. (2012) Formulating Your Big Data Strategy: It's Not About Technology - It's About Business Value; It's Not About Data - It's About Insight from Data. *Cover Art: UK Oracle User Group Conference 2012* [online]. ICC Birmingham 3-5 December. Acunu.

**Password: CONF2012OUG**

# [Accessed 08 December 2012] Available at: <http://2012.ukoug.org/presdisplayfile.asp?prs\_prsid=7013&filename=121203UKOUG%2Epdf>.

# This source outlines what the market is using to handle their big data needs, such NoSQL, Hadoop etc and shows exactly what kind of uses they are popular for, such as on-line big data workloads, batch processing. It also shows why RDBMS are not able to handle the issue of big data and how a good analytics program will help to sift through big data to find exactly what is needed. This source is similar to the work of (Howard *et al* 2012) as the source goes into detail on analytics, the uses and benefits therein it also gives a presentation of an analytics program in use unlike this source which provides greater clarity to analytics. This source provides a decent amount of information about the big data solutions currently on the market although it could be considered biased by advertising its own analytics program, however I believe it will still be useful and therefore this source will help me answer my academic question and achieve my aims as supporting material to my research on the emerging database systems that can handle big data and by being included in the report.

Davis, M. A. (2012). NoSQL Equals NoSecurity. *InformationWeek*, [online]. Apr 9, 2012, pp.29-31 [Accessed 26 February 2013] Available at:

<http://search.proquest.com.ezproxy.wlv.ac.uk/docview/1011104685/fulltextPDF?accountid=14685>

The source shows that while technologies like NoSQL and Hadoop are more widely used for their ability to handle big data, it is clear that security is an issue that may have been overlooked. It details how several forms of NoSQL databases are being adopted as well as the use of MapReduce and BigTable to handle big data however security is very lax due to the advancements in such technologies coming from the developers who have so far have not seen fit to add proper security measures to such technologies. It recommends security controls for developers to implement for the database as well as security recommendations for the operating system (OS) and concludes that companies must take more responsibility in terms of security. This source is very useful as it provides a great deal of information on the security issues of big data technologies especially NoSQL databases which is similarly detailed in the work of Tene and Polonetsky (2012). However in contrast to this source the previously referenced source details a background on big data as well as the benefits of big data analysis which this source does not. Therefore this source is very useful as a change from the benefits of big data technologies and it will help me answer my academic question and achieve my aims by using during the creation of my artefact and in the report when I consider the security issues with big data and my artefact.

Henschen, D. (2011). Oracle Releases NoSQL Database, Advances Big Data Plans. [online]. New York: InformationWeek [Accessed 23 February 2013]. Available at:

<http://search.proquest.com.ezproxy.wlv.ac.uk/docview/900341996>

This source details the release of several big data technologies based on the open source Apache Hadoop project, however the source focuses on NoSQL database's more specifically Oracles NoSQL database which it mentions is based on the open source BerkeleyDB and the Sleepycat Software from 2006. This is useful as it provides me with background information on Oracle's NoSQL database as well as comparing this to alternative big data technologies such as Cassandra. This source is similar to the work of Brooks, J. (2011), both sources detail a background of Oracle's NoSQL database and show how it is based upon previous technologies however this source details Oracle's NoSQL database in the context of big data as well as comparing the technology to Cassandra, however the previously referenced source also details the technology itself but not from a big data standpoint and it also mentions competitor technologies such as MongoDB and CouchDb which this source does not. This source will help me answer myacademic question and achieve my aims by being used as reference when creating my artefact and when evaluating it in the final report.

# Howard, P., Sowerby, D. (2012) *The Next BIG Thing?* *Cover Art: UK Oracle User Group Conference 2012* [online]. ICC Birmingham 3-5 December. SolStonePlus.

**Password: CONF2012OUG [Accessed 08 December 2012] Available at: <http://2012.ukoug.org/presdisplayfile.asp?prs\_prsid=7491&filename=SolStonePlus%5F%5FUKOUG%5FConference%5FEndeca%5FPresentation%5F2012%2Epptx>.**

**This source explains what Big data is and how it should be used, it goes on to show how a graphical representation of such data can help to see exactly what you are getting from the data you have collected and stored. This is useful as it provides me with background information on big data and demonstrates how a graphical representation of my artefacts results could help understand the outcome when testing my artefact. This source is similar to the work of Tene, O. and Polonetsky, J.** **(2012) which also provides a detailed background on big data, however it also shows the possible benefits big data can grant with several examples as well as the concerns with big data with several examples which this source does not therefore the previously referenced source gives a more detailed account of big data. The source provides adequate background information on Big data and a demonstration of analytics tools, therefore this is a useful source and it will help me answer my academic question and achieve my aims by furthering my research and will by being included in the report.**

Kutrovsky, C. (2012) Exadata Data warehousing: Leveraging Parallel Query. *Cover Art: UK Oracle User Group Conference 2012* [online]. ICC Birmingham 3-5 December. Pythian.

Password: CONF2012OUG

[Accessed 08 December 2012] Available at:

<http://2012.ukoug.org/presdisplayfile.asp?prs\_prsid=7482&filename=Christo %5FKutrovsky%5F%2D%5FExadata%5FDatawarehousing%5FLeveraging%5FParallel%5FQuery%5F%2D%5F2012%2Epptx>.

The source details the inner workings of Exadata, specifically how the data moves in order to get a result, it describes parallel query concepts including a big sort, group by and join which is useful when dealing with very large tables. There are many useful points in this source that helps to shed light on how Exadata works to give the best result, compared to other sources this source gives the reader a fair amount of information on Exadata similar to the work of Logan and Abbey (2012), which also provides an adequate amount of information on big data and oracles Exadata. This source will help me answer my academic question and achieve my aims by being used in the final report when comparing my artefact to Exadata.

Logan, P., Abbey, M. (2012) Applying Traditional DBA Skills to Oracle Exadata. *Cover Art: UK Oracle User Group Conference 2012* [online]. ICC Birmingham 3-5 December. Pythian.

Password: CONF2012OUG

[Accessed 06 December 2012] Available at:

<http://2012.ukoug.org/presdisplayfile.asp?prs\_prsid=7213&filename=exadata% 5Fukoug12%5Fpost%5Fabbey%5F%2D%5FPaul%5FLogan%2Epptx>.

The source compares and contrasts a relational database management system (RDBMS) to Exadata and how Exadata works to get the most performance possible, backup considerations, the internals of the Exadata machine, functions for compression, performance and storage, this shows what is new with Exadata and how it can help deal with big data. There are useful points in the source which when compared to the work of Howard and Sowerby (2012), which in contrast to this source provides the reader with less information than is expected. This source will help me answer my academic question and achieve my aims by helping tocomparing my artefact against Exadata in the final report.

Mone, G. (2013). Beyond Hadoop. *Communications of the ACM*, [online]. **56**(1), pp.22-24 [Accessed 12 February 2013] Available at:

<http://wk6kg9sd8m.search.serialssolutions.com/?ctx\_ver=Z39.88-2004&ctx\_enc=info%3A ofi%2Fenc%3AUTF-8&rfr\_id=info:sid/summon.serialssolutions.com&rft\_val\_fmt=info:ofi/fm t:kev:mtx:journal&rft.genre=article&rft.atitle=Beyond+Hadoop.%28Apache+Hadoop%27s+continuous+evolution%29%28News+%2F+Technology%29&rft.jtitle=Communications+of+the+ACM&rft.au=Mone%2C+Gregory&rft.date=2013-01-01&rft.pub=Association+for+Computin g+Machinery%2C+Inc&rft.issn=0001-0782&rft.eissn=1557-7317&rft.volume=56&rft.issue=1 &rft.spage=22&rft.externalDBID=n%2Fa&rft.externalDocID=316952108>

# This source shows how Hadoop continues to be used as a leading technology for handling big data, it describes the history and structure of Hadoop as well as mentioning several technologies that add onto or focus more on certain aspects of Hadoop such as, MapR, Pregel, GraphLab, HBase, Cloudant, Dremel for example and gives a description of each. This source is useful as it shows a different side to Hadoop where although it will handle the issue of big data but it alone does not allow for the tools needed to reap the benefits from analysing such data, similar to the work of Borkar et al. (2012), both sources give detailed information on several technologies that handle big data however this source does not go into detail on big data itself while the previously referenced source does and this source focuses on hadoop and its offshoots, while the previously referenced source branches out to several different big data technologies. This source will help me answer my academic question and achieve my aims when I use it as guidance when creating my artefact and in the report when I evaluate the artefact.

Russom, P. (2011) Big Data Analytics. *October*, [online]. **19**(40), pp.1-40 [Accessed 12 March 2013] Available at:

<ftp://129.35.224.12/software/tw/Defining\_Big\_Data\_through\_3V\_v.pdf#page=6>

# This source details the issue of big data from the perspective of analytics, the source gives an introduction to the world of big data analytics, shows the current shape of big data analytics in general such as their adoption, benefits, issues etc, it also shows best practices for big data analytics something which would be invaluable when trying to make use of big data. This source is useful as it details information on both big data and the analytics programs used to discover trends in the data. This is similar to the work of Clegg, D. (2012) as both sources detail analytics in terms of big data however while this one provides the reader with a lot of detail on analytics which the previously referenced source does lack, but it does give a graphical representation of an analytics program which is quite useful to those wishing to make use of big data. This source will help me answer my academic question and achieve my aims when I use it in the report when I evaluate the artefact.

Schmutz, G. (2012) NoSQL Databases for Implementing Data Services – Should I Care? *Cover Art: UK Oracle User Group Conference 2012* [online]. ICC Birmingham 3-5 December. Trivadis.

Password: CONF2012OUG

[Accessed 10 December 2012] Available at:

<http://2012.ukoug.org/presdisplayfile.asp?prs\_prsid=7424&filename=nosql%2Ddatabase%2Dfor%2Dimplementing%2Ddata%2Dservices%2Dshould%2DI%2Dcare%2Epdf>.

The source gives a perspective on NoSQL databases, a background, reasons for using it instead of a RDBMS, a description of different NoSQL databases, a comparison to RDBMS, it shows a different form of programming called Polyglot Programming used for taking advantage of many different languages advantages when tackling different problems and how this can be applied to database systems. This source is similar to the work of Bhatewara and Waghmare (2012). Both sources show in great detail technologies that handle the issue of big data and details why these are used over a traditional RDBMS, this source will help me answer my academic question and achieve my aims when constructing my artefact as well as in the final report.

Stonebraker, M. (2010). SQL databases v. NoSQL databases. *Communications of the ACM*, [online] **53**(4), pp.10-11 [Accessed 22 February 2013] Available at:

<http://mags.acm.org.ezproxy.wlv.ac.uk/communications/201004/?folio=10&CFID=281029974&CFTOKEN=14962939#pg1>

# The source describes the reasons for using a NoSQL database and how to get better performance for online transaction processing (OLTP) databases by removing the 'overhead' or traditional aspects of Atomicity, Consistency, Isolation, Durability (ACID) which is grouped into four parts in OLTP databases, for example; logging, locking, latching and buffer management these points are given descriptions for clarity. Therefore this source is quite useful as it shows the differences between RDBMS and NoSQL databases and how NoSQL allows for greater performance, similar to the work of Schmutz (2012) which also gives its perspective on NoSQL databases and similarly compares it to RDBMS, however this source shows how NoSQL database gives the user greater performance over RDBMS which the previously referenced source does not. This source will help me answer my academic question and achieve my aims by using this source when creating my artefact and in the final report when I evaluate my artefact.

Tene, O., Polonetsky, J. (2012). Privacy in the Age of Big Data: A Time for Big Decisions. *Stanford Law Review Online*, [online] **64**(63), pp.63-69 [Accessed 19 February 2013].

Available at: <http://www.stanfordlawreview.org/online/privacy-paradox/big-data>

The source details a background on big data and how it has relatively recently come into the spotlight, it show the benefits big data can grant such as in the health sector, data can be analysed to attribute health issues to previously unknown causes, big data can also be analysed to predict flu outbreaks or energy demands for example. The concerns with big data are detailed such as protecting privacy and data security and therefore governments and businesses must be aware of this issue. The source therefore is quite useful and raises some issues that come with big data as well as its benefits, similar to the work of Borkar et al. (2012). This source gives the reader a background on big data, benefits etc which is similar to the previously referenced source, however this source also shows the privacy and security concerns with big data which is in contrast to the previously referenced source as it does not mention this.This source will help me answer my academic question and achieve my aims by providing me with background knowledge for the project but also for use in the final report.

Yuan, X., Vega, P., Qadah, Y., Archer, R., Yu, H., Xu, J. (2010). Visualization Tools for Teaching Computer Security. *ACM Transactions on Computing Education*, [online] **9**(4), pp.12-13 [Accessed 29 February 2013]. Available at:

<http://williams.comp.ncat.edu/IA\_visualization\_labs/publications/visualization%20tools-yuan.pdf>

This source details in depth information on the third part authentication service 'Kerberos' which many NoSQL databases support in order to improve on their lack of security, this source is quite useful as it provides me with detailed knowledge on a piece of the security used by many NoSQL databases on the market. This source is similar to the work of Schmutz, G. (2012) as this source details a background on NoSQL databases but also shows the issues specifically security issues with NoSQL databases which this source does not, it mentions Kerberos as a partial solution to this issue but does not go into much detail into the technology itself which this source does. This source will help me answer my academic question and achieve my aims by using this source when creating my artefact and in the final report when I evaluate my artefact.

...............**ADD MORE REFS, check current ref reviews and add more detail where necessary**