Institution: University of Derby		
Unit of Assessment: 11		
Title of case study: Distributed Data-Stream Analytics		
Period when the underpinning research was undertaken: March 2011–August 2019		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Professor Ashiq Anjum	Director of the Data Science Research Centre, Professor of Distributed Systems	March 2011–February 2020
Professor Nikos Antonopoulos	Pro Vice-Chancellor Research & Innovation, Dean of College of Engineering & Technology, Professor & Head of School of Computing & Mathematics	March 2009–January 2019
Professor Richard Hill	Head of Department, Electronics, Computing & Mathematics, Professor of Intelligent Systems	January 2010–January 2017
Period when the claimed impact occurred: 1 August 2013–31 December 2020		
Is this case study continued from a case study submitted in 2014? No		

1. Summary of the impact (indicative maximum 100 words)

Distributed Data-Stream Analytics (DDSA) is the application of deep learning to networks of sensors and systems that gather and process data in real-time. The research of Professor Anjum and his colleagues at the University of Derby (UoD) has combined Cloud and Edge Computing with a partitioned Convolutional Neural Networks (CNN) model to several applications including digital content and virtual reality development for engineering, high energy physics data processing and rail asset management. The DDSA solution reduces the network bandwidth requirements and platform costs while maximising the utilisation of the available resources enabling the network to respond to tasking and queries in near real-time.

Evidence of the Derby Group's impact can be found in:

1. The processing of high energy physics experimental results at ***CERN***.
2. The management of rail infrastructure with ***Fishbone Solutions***.
3. The processing of real-time digital twins at ***Bloc Digital***.

In each case the impact of the research opened up new opportunities for the partner organisations in terms of the reach of their research (CERN) or development of new products and services (Fishbone and Bloc Digital).

2. Underpinning research (indicative maximum 500 words)

The creation of a feedforward based deep learning model capable of optimising results while operating in a distributed resource constrained environment is central to DDSA.

The deep learning model is based on the biologically inspired CNN, a class of neural networks able to efficiently process visual imagery. The underpinning research carried out by the Derby Group includes:

1. Cloud-based deep-learning for data analytics [3.1, 3.2]
In deep learning for video-analytics, the video-stream is passed through an analysis pipeline that filters low-value (uninteresting) data using a CNN model. This approach demonstrated the ability to firstly move the processing of video stream to cloud-based resources [3.1], and secondly, successfully classify the video stream through the detection of objects of interest [3.2]. Initial applications extended to vehicle detection and facial recognition.

2. Distributed deep learning [3.3, 3.4]
When there is a requirement for real-time video analytics, cloud-based resources are not always sufficient to process large amounts of data in the time required. The creation of a CNN feedforward model that can be decomposed and distributed to the available resources at the edge of the network increases the available computer power to perform real-time video analytics throughout the network while minimising bandwidth requirements. The approach is based on intelligently decomposing the CNN layers among the available edge, in-transit and cloud resources. The first few layers are computed on the edge resource, remaining layers are computed on the in-transit nodes, finally, cloud processes the output layer to provide an overall result.

3. Data Quality Monitoring [3.5]
The automatic analysis of complex datasets in real-time is a problem for a wide range of applications, such as Data Quality Monitoring of high-energy nuclear and particle physics experiments at the Large Hadron Collider (LHC). The learning of multiple consecutive tasks with continual identification of new features from streamed data is considered a challenging task, known as lifelong learning. Our approach tackles this problem by combining fast analysis on the data-stream from a shallow network, with parallel deep networks trained to adapt to new features in the data incrementally, using transfer learning techniques. The proposed model addresses the continuous analysis of high-energy physics data quality, making it available in the O2 framework in the A Large Ion Collider Experiment (ALICE) to be used in the upcoming data-taking period.

3. References to the research (indicative maximum of six references)

University of Derby researchers are indicated by black, underlined text:

[3.1] A. Anjum, T. Abdullah, M. F. Tariq, Y. Baltaci and N. Antonopoulos, 'Video Stream Analysis in Clouds: An Object Detection and Classification Framework for High Performance Video Analytics,' in *IEEE Transactions on Cloud Computing*, vol. 7, no. 4, pp. 1152-1167, 1 Oct.-Dec. 2019, doi: 10.1109/TCC.2016.2517653. https://ieeexplore.ieee.org/document/7381631

[3.2] M.U. Yaseen, A. Anjum, O. Rana, R. Hill, 'Cloud-based scalable object detection and classification in video streams', *Future Generation Computer Systems*, Volume 80, March 2018, pp. 286-298. https://www.sciencedirect.com/science/article/pii/S0167739X17301929

[3.3] M. Ali, A. Anjum, M.U. Yaseen, A.R. Zamani, D. Balouek-Thomert, O. Rana, M. Parashar, 'Edge Enhanced Deep Learning System for Large-Scale Video Stream Analytics,' *2018 IEEE 2nd International Conference on Fog and Edge Computing (ICFEC)*, Washington, DC, 2018, pp. 1-10, doi: 10.1109/CFEC.2018.8358733. https://ieeexplore.ieee.org/abstract/document/8358733

[3.4] M. Ali, A. Anjum, O. Rana, A. R. Zamani, D. Balouek-Thomert and M. Parashar, 'RES: Real-time Video Stream Analytics using Edge Enhanced Clouds,' in *IEEE Transactions on Cloud Computing*, May 2020, doi: 10.1109/TCC.2020.2991748. https://ieeexplore.ieee.org/document/9084281

[3.5] M. U. Yaseen, A. Anjum, O. Rana and N. Antonopoulos, 'Deep Learning Hyper-Parameter Optimization for Video Analytics in Clouds,' in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 253-264, January 2019, doi: 10.1109/TSMC.2018.2840341. https://ieeexplore.ieee.org/abstract/document/8386680

All papers were peer reviewed in line with IEEE guidelines.

Funding
Science and Technology Facilities Council (STFC) Nuclear Physics Consolidated Grant, ST/P005047/1 GBP18,020 (October 2017–September 2021), used to support work with **CERN**.

Innovate UK Knowledge Transfer Partnership grants with **Bloc Digital** GBP133,477 – number 511773 (November 2019–January 2022) and GBP259,388 – number 511413 (February 2019–April 2021). Figures are for portion of grant which came to UoD.

European Regional Development Fund (ERDF) grant, through D2N2 Local Enterprise Partnership, with **Fishbone Solutions** EUR9,080 (October 2017–January 2018). Figure is the contribution to the associate funding from the 'Enabling Innovation' Knowledge Exchange for Information (KEI) funding scheme.

4. Details of the impact (indicative maximum 750 words)

Professor Anjum and his team's research in DDSA has made the extraction of information from large data sets practical in real-time. The impact of their work in the wider world can be measured in two ways, technically and economically. DDSA was the essential component for several success stories for multinational organisations including CERN as well as companies such as Bloc Digital and Fishbone Solutions.

High energy physics: Professor Anjum and his team's research work in high performance analytics has led to collaborative projects with the **European Organization for Nuclear Research (CERN)** – one of the world's largest scientific research centres, established by international treaty. The UoD team have been collaborating with different teams at CERN, supporting seven people (directly or indirectly) for over a decade. The team focussed their more recent work on ALICE which investigates the state of matter produced a few microseconds after the Big Bang.

LHC Data Analytics (October 2017–June 2020) [5.1]: Collaborating with **ALICE** to investigate scalable machine learning and statistical models for large amounts of data (over 200 petabytes) that the LHC experiments are producing. Our work in scheduling and resource management has produced significant advances for LHC Data Analytics. The exploitation of concepts and software produced has contributed to industrial grade projects for large-scale data analysis and optimal use of infrastructure. A data intensive and network-aware scheduling approach and its use of machine learning algorithms to optimally plan workflows before they are considered for scheduling, has impacted on the way applications manage and use distributed resources.

The Deputy Coordinator of the Detector Control System states: *"Professor Anjum and his team's work on Distributed Data Stream Analytics (DDSA) has had a significant impact on our thought processes and the design of the ALICE next generation data acquisition system"* [5.1].

The improved ALICE data acquisition system will be used operationally in the experiment, when the Large Hadron Collider resumes operation in 2021/2. Initial testing has shown an improvement in the data acquisition rate from 2.5GB/s to 4TB/s. The final data output will increase by a factor of 10,000, greatly increasing the reach of this research. The Deputy Coordinator stated that it will: "…*allow us [ALICE Collaboration] to move from a reactive to proactive approach. This will give us a fundamental improvement in our understanding of the behaviour of the quark-gluon plasma"* [5.1].

Data analytics for rail innovation: The Derby Group have been working with rail companies to exploit the benefits for rail infrastructures and services of the Internet of Things (IoT) and High-Performance Data Analytics by setting up a Data Analytics Laboratory providing local SMEs with easy access to rail data and a range of analytics services to extract value from that data. Applying DDSA to their data analytics ecosystem has delivered the following impact:

DDSA for predictive maintenance (August 2017–July 2020) [5.2]. **Fishbone Solutions** partnered with the University of Derby to exploit data science innovations for predictive maintenance in rail. This has led to increased operational efficiencies by offering predictive maintenance regimes and reducing the time it takes to halt operations when trains are in depots for maintenance. A testimonial from the Director at Fishbone corroborates that the Distributed Data Streaming Analytics research of Professor Anjum and the University of Derby was core to the creation of a system that automated the process of converting a proprietary format On-Train Monitoring Recorder (OTMR) files into an open-source format [5.2]. This has allowed Fishbone to bring to market a new product family of equipment-health monitoring solutions.

Data processing for real-time visualisation: With the advent of affordable augmented and virtual reality hardware for industrial applications, Bloc Digital has been at the forefront of the market, developing solutions for companies such as Rolls-Royce, GSK and Siemens. Augmented/Virtual Reality (AR/VR) solutions require specialised digital twin models to be created that are optimised to allow refresh rates of between 90-120Hz to prevent latency issues and subsequent nausea in the user.

DDSA for data visualisation in VR (November 2018–July 2020) [5.3, 5.4]: We worked with Bloc Digital to produce real time visualisation of engineering models in VR. The approaches used to produce high performance visualisations are underpinned by research led by Professor Anjum and his team, applying DDSA to process and visualise data in VR in real time. The latter part of work with KTP associates was supervised by Dr Bo Yuan and Dr Leonardo Stella, current members of academic staff, and this work has continued.

Examining Figure 1 below, a comparison of a traditionally developed and DDSA developed real-time digital model is shown. The digital twin on the right was developed from a master CAD model over 65 man days costing GBP16,000. The digital twin on the left was generated from a master CAD (non-real-time) model. The CAD model was decomposed into separate components, processed in parallel using a CCN and then recombined to create the finished product. The DDSA processed digital twin was created in 6 days at a total cost of less than GBP10 and achieved a 98% accuracy to the original CAD model and was identical to the human eye [5.3].

Figure 1: A comparison of traditionally (right) and DDSA developed (left) real-time digital twins.

From a market perspective, this has allowed premium industrial AR and VR solutions to drop in cost by an order of magnitude. Although the marketplace has been disrupted by COVID-19, there has been a significant up-turn in business interest as a consequence [5.3].

5. Sources to corroborate the impact (indicative maximum of 10 references)

[5.1] Letter from the Deputy Coordinator of the Detector Control System, ALICE Experiment at the LHC, CERN, stating the impact of DDSA on their HEP research (August 2020).
[5.2] Letter from Fishbone Director stating the impact of DDSA on their work (August 2020).
[5.3] Letter from Bloc Digital Director stating the impact of DDSA on their work (August 2020).
[5.4] Bloc Digital website quoting the company's Director: https://bloc-digital.com/blocnews/growth-and-innovation-drives-bloc-digital-move-to-university-of-derbys-enterprise-centre/ (Accessed: 14 December 2020).