

T-Systems Principal Architect  
Christoph G. Jung (l.) and Erik Redl,  
Head of the BigAnalytics Department  
of the Telekom subsidiary



# Digital harvesters.

Although state-of-the-art technology is no longer able to cope with the analysis of huge volumes of data, experts are still talking about "huge data". A patent-pending T-Systems development is now finding the "digital truffles" hidden in vast expanses of data in record time.



COPY — Sven Hansel

Every eight weeks on Thursday at 9 p.m., an icon of German television in 1987: the car test on the program “Telemotor” by broadcaster ZDF. Fast-paced music, dynamic camera work, and quick cuts phase the brand-new Audi 90 2.3 into the picture. The car roars through deep, artificially created puddles on the test track. The camera stops abruptly, and the hard-hitting tester starts in: “In practice, consistent aerodynamics also have disadvantages. Turbulent water heavily soils the body of the Audi and rain or snow falls onto the seats when the doors are opened,” the speaker says in a serious and sonorous

voice. No question about it... at that time, it was a statement with high information value for all viewers, but more or less “measured” on the basis of experience, intuition and pure manual work. Today, in contrast, test drives are the real force behind the carefully planned propagation of mountains of data.

The futuristic autonomous vehicles which premium manufacturers are currently developing are becoming more and more complex because all the components communicate with one another and are networked for external access. HD cameras with panoramic view, distance sensors, radar devices, emission sensors, internal

# 10,000 channels

capture data from the car's advanced sensors, including not just traffic signs and passers-by, but also the driver's own pupil movements to counteract inattention or fatigue.

microphones: All of these record signals, providing important insights into the quality of the advanced driving functions for the prototype pre-production tests. "These vehicles deliver one to three terabytes of specially coded data per hour," says Christoph G. Jung, principal architect at T-Systems, describing the changing times – and thus a new challenge for all digitized industrial sectors.

For the companies concerned, it is almost irrelevant whether the number of devices and sensors interconnected by the IoT will be 50 billion or 60 billion by 2020. The immense challenge lies in what their measurement and control units generate in terms of big data, and its subsequent evaluation in real time. This is basically a sort of hunt for digital truffles. If, so to speak, the harvest time of the precious raw data is the commodity, then the valuable commodity must be extracted in the shortest possible time and made palatable; otherwise, the contained information becomes obsolete. In the automotive industry, there are several hundred vehicles that professional test drivers push to the limits around the world and around the clock in multi-shift test track operations – always looking for abnormalities, and always focused on discovering any safety issues in the "thinking" ECU software as early as possible – scrupulously accurate and down to the last bit and byte. Speed, consumption, engine and transmission data, radar scans: Up to 10,000 channels capture data from the car's advanced sensors, including not just traffic signs and passers-by, but also the driver's own pupil movements to counteract inattention or fatigue. During the journey, all this information is logged using a sort of "black box" on modern, shock-resistant, solid-state disks, which, at the end of the working day, "only" actually needs to be output at the depot and fed into the evaluation software – actually.

After all, with the bandwidths required worldwide for data processing, every company reaches its limits with today's on-board resources (4G networks, WLAN, VPN, and host computers). It does not take long to reach the multi-digit petabyte range, which is why one speaks today of "huge data". However, engineers need to be able to evaluate the captured signals in just a few hours to fix critical errors and prepare the next important tests while the data is fresh. It can be compared to a huge field with a considerable number of truffles of the highest quality hiding in the soil, but there are far too few harvest workers laboring for the star chef.

With conventional technology, automotive engineers are often condemned to a test of patience over several days and hypothetical games of thought, as the reading and analysis of the resulting data mountains take far too long – a prohibitive cost and factor that is not insignificant for industry.

This is because reading data is a special technological challenge.

Unlike texts, for example, this "signal data" so far can only be compressed poorly and interpreted efficiently. Software developers know this: The files of a large book, for example, can be broken down, figuratively speaking, into more manageable excerpts. One computer then scans the first half, another takes care of the second in the meantime, and the results of both analyses are then combined. The job is completed twice as fast with two computers; with a whole stack of computers (cluster), the result is available after just a few seconds.

However, such a procedure has not been able to be applied in automotive development so far. "When recording machine signals, fixed character sets are not used, as is the case with texts, but rather variable, situation-dependent codes. For example, if a car changes to a higher speed range, then certain channel groups related to the engine will have to be sampled more often," explains Jung. Translated to a cookbook, this would mean that each of the international recipes would have been written in its native language (and also its proper writing system!), such as German, Spanish, Russian, Greek, Chinese, etc. This is difficult for classic data compression methods. "But I was not prepared to settle for this and I thought more about it," reports Jung.

His groundbreaking invention – patented by Telekom – overcomes two obstacles. First, it cracks the supposedly unpredictable data formats and brings them into logically related technical pieces (called chunks). These are put "in the cradle" of the computer system as a kind of second foreign language. And secondly, the solution – a "transcoder" similar to an MP3 converter in modern audio equipment – ensures rapid and compressed storage, even in the cloud. When an engine runs faster, the temperature or oil pressure does not always suddenly change faster. The resulting software-based signal processing ("big data signal processing") takes advantage of this fact and can thus operate without loss of information on a fraction of the original data, but at the same



**Hardware-in-the-loop – based on ever more powerful microprocessors, HiL systems will increasingly make ideal laboratory test environments possible.**



time on every computer core of a provided cluster. The speed achieved in practice is 40 times higher than in previous methods; the stored data amount shrinks depending on the measured channels to up to 10 percent of the original volume.

“Unlike what private MP3 users are used to, we can also return the data exactly to its original form. So, if an engineer wants to investigate a detected anomaly in detail and needs the corresponding partial detail in full, then this is possible at any time with our huge data method,” reveals Jung.

In addition to the original signals, derived channels that a simulation computer artificially generates can also be faded into this microscope function. The new type of signal processing is flexible enough to handle simulated test runs. After all, if a supplier changes the software of one of its ECUs only minimally, automobile manufacturers would ultimately be forced to repeat the entire test drive on the road – an elaborate affair, because up to nearly 100,000 miles of test track is the industry standard. Instead, the manufacturers install the updated control units in a simulator, which merely plays the recorded signals of the test vehicle to the relevant control unit and then records its changed reaction (“hardware-in-the-loop”). The T-Systems development fits right into this profitable and possibly repeated harvesting loop.

The huge data turbocharger was so well received among users that the T-Systems inventors are already planning their next coup. Consequently, endurance tests – long-term tests on rough desert terrain, tropical humidity or Arctic cold – generate huge amounts of data, especially

On their test drives of up to nearly 100,000 miles, the “prototypes” of the automotive industry produce data volumes in the multi-digit petabyte range.

in inhospitable areas, where even a future 5G supply would reach its limits. Which is why Jung is working with his colleagues on a mobile cluster, a sort of transportable, air-conditioned mini-computer center, which can also be shipped to and operated in the most remote locations in the world. “That’s where our invention comes in; the compressed results of the analyses are sent to the company headquarters of the carmaker quickly,” reports Erik Redl, head of the BigAnalytics department at T-Systems.

The principles of a successful data harvest are so fundamental that they can be seamlessly transferred to other industries. In rail transportation, for example, for detecting malfunctions, defects, or structural defects in the track bed by means of cameras mounted under the trains. “We have now also been able to transfer our solution to multi-hour video files and have them evaluated by hundreds of computer cores simultaneously,” says Jung of the progressive development. The result: similar time gains as the carmakers.

For this reason, well-known big data experts like Dr. Bange foresee great potential for such inventions. “They cover all industries where machines are involved. They cover individual production. The healthcare industry and complex products in any case – digitization makes new procedures such as this imperative,” says the CEO of Würzburg BARC GmbH. Big data will have to be filtered at the source in the future, novel reduction methods are important, “and, ultimately, EDGE computing will certainly play an increasingly important role,” explains Dr. Bange.

In short, the digital harvester is already an indispensable tool for successful data analysis in the automotive industry as soon as the dimension of data far exceeds the scope of big data. Certainly, in the future it will ignite the turbo across industries.

**“Ultimately,  
EDGE computing  
will certainly  
play an increas-  
ingly important  
role.”**

**CARSTEN BANGE,**  
Managing Director of  
Würzburg-based BARC-GmbH



C-Jung@t-systems.com (Christoph G. Jung)



[www.t-systems.com/solutions/big-data-analytics](http://www.t-systems.com/solutions/big-data-analytics)

[www.t-systems.com/whitepaper/big-data-automotive](http://www.t-systems.com/whitepaper/big-data-automotive)