

Sensor Data Analysis for a Large Mining Major

Authors : Sudipto Shanker Dasgupta

Abstract : One of the largest mining companies wanted to look at health analytics for their driverless trucks. These trucks were the key to their supply chain logistics. The automated trucks had multi-level sub-assemblies which would send out sensor information. The use case that was worked on was to capture the sensor signal from the truck subcomponents and analyze the health of the trucks from repair and replacement purview. Open source software was used to stream the data into a clustered Hadoop setup in Amazon Web Services cloud and Apache Spark SQL was used to analyze the data. All of this was achieved through a 10 node amazon 32 core, 64 GB RAM setup real-time analytics was achieved on '300 million records'. To check the scalability of the system, the cluster was increased to 100 node setup. This talk will highlight how Open Source software was used to achieve the above use case and the insights on the high data throughput on a cloud set up.

Keywords : streaming analytics, data science, big data, Hadoop, high throughput, sensor data

Conference Title : ICDE 2015 : International Conference on Data Engineering

Conference Location : Chicago, United States

Conference Dates : October 08-09, 2015