Abstract

***What is the problem statement of the topic*. The number of IoT applications in various industries and cases cannot be separated from the results of data processing generated from installed nodes. Looking at the characteristics of the data generated, IoT data can be categorized as big data and requires big data solutions to precess it. Commonly used architecture of big data is to send all the resulting data to central server and process it there. However with the large number of nodes installed and the volume of data points generated, processing all data at the central server would be costly in terms of processing and transport from node to server.**

***Why is the topic interesting*. As the development and complexity of IoT system grow, data that needs to be process will also grow. If all data processing need to be handled at central server, all the data needs to be send which will cause high cost on processing and network.**

***What is the solution*. With the development of hardware capabilities on nodes, processing data on the node before it is sent to the server becomes possible. To reduce data sent to server, node can do part of data processing. This approach is called distributed IoT. An application which can take advantage of this approach is road traffic data processing. The road traffic data can be processed using complex event processing and context aware k-means clustering to detect bad traffic. With distributed IoT approach, the two processing can be divided between nodes and central server. Complex event processing can be processed on nodes while the clustering needs to happen in central server. This will reduce the data that needs to be sent to server and the latency on the detection. If the application is using central server approach, both processing needs to happen at central server.**

***Main result.* Comparison between 2 approaches can be done by applying the traffic data processing application on both approaches. By applying using both approaches, throughput on central server and latency from data produced until it is processed by complex event processing can be measured.**
Keywords : distributed IoT, context aware k-means clustering, complex event processing

---