# Geospatial Cloud Data Warehouse Optimization Techniques

Ben Scholer - Purdue University
Howell Li - Purdue University

**11:15 - Noon**

## Bio:

**Ben Scholer**: Ben Scholer is a Software Engineer at the Joint Transportation Research Program (JTRP). Scholer holds a BS in Computer Information Technology graduate from Purdue University. Scholer primarily works with connected vehicle data and finding ways to display this data in interactive, real-time web applications for use by state DOTs, namely INDOT.

**Howell Li**: Howell Li is a Research Analyst at the Joint Transportation Research Program (JTRP). Li holds a MS from Purdue University College of Engineering and a BA from New York University in Computer Science. He has over 16 years of experience in front and back-end software development and over 8 years in transportation. Li currently works in connected, electric and autonomous vehicles, highway mobility, and traffic signal performance measures. His focus is on applying web technologies, big data analytics, efficient data structures and computing methods to the research and development of safe and efficient transportation systems.

## Abstract:

In the field of traffic research, connected vehicle (CV) data permits researchers to, without the need for any infrastructure, analyze traffic patterns at scale. The current penetration of CV in Indiana is about 3-5% but is expected to grow significantly within the next decade. This study makes use of cloud data warehousing technology Google BigQuery, to interactively report travel speeds and times on segments of roadway on a web dashboard to agency stakeholders. A barrier to adoption of cloud platforms by public agencies has been the uncertainty of a subscription-based expenditure rather than a one-time capital investment with traditional programs. The contribution of this study is to exercise a popular cloud platform and document expenditures by making queries on over 130 billion geographic data points more efficient and less expensive. By leveraging optimization techniques such as partitioning, clustering, and geohashing, JTRP has been able to extract meaningful information from a dataset in the realm of 10s of terabytes within seconds, most of the time at prices less than $1. This speed is critical to make this real-time information useful to stakeholders such as the Indiana Department of Transportation (INDOT).