Sigi is at the front-edge of using state-of-the-art cloud-based geocomputational platforms. We have extensive experience using cloud infrastructure for planetary scale spatial analyses on large datasets of high-resolution satellite imagery. We develop interactive web-based applications on top of our cloud infrastructure to enable on-the-fly custom exploration of spatial data.

**DATA MINING AND MACHINE LEARNING**
- Artificial Neural Networks
- Bayesian Networks
- Decision Trees
- Naive Bayes
- Bagging and Boosting
- Clustering (e.g., K-means, ISODATA)
- Classification (e.g., K-nearest Neighbors, Random Forest)
- Simulated Annealing
- Support Vector Machines
- Natural Language Processing
- Topic Modelling (e.g., Latent Dirichlet Allocation)

**PROGRAMMING LANGUAGES**
Programming Languages: Java, Javascript, Clojure, Clojurescript, SQL, Bash, GnuPlot, Org-Babel, C, C++, C# .NET, Node.js, Django, Python, R

**WEB BASED GIS DATASYSTEMS**
We have expertise in building either custom-hosted or cloud-hosted open source web applications with HTML5, CSS3, Python, Java, Clojure, Javascript, and Clojurescript. Embedded web maps may be included using the open source OpenLayers or Leaflet Javascript libraries. Proprietary map layers may be integrated using Google Maps API.

**Skills:** GIS programming (w/ PostGIS, GDAL/OGR, GRASS, or QGIS Python) Server management (OS configuration, software installation and upgrades, network security protocols, load monitoring, shell scripting)

GeoServer and Web Map Services (WMS, WCS, WFS)

**GEOSPATIAL ANALYTICS**
Ate add content

**OTHER**
- Web development (front end and back end), Database programming (SQL and/or NoSQL DBs), Command line application development, Statistical Programming (e.g., R, NumPy, Incanter)
- Literate Programming (Emacs Org-mode, Python PWeave, NoWeb, or other Tangle+Weave facility)
- Programming Paradigms: Imperative, Object-Oriented, Functional, Logic

**STAFF CONTACTS**

Gary Johnson
gjohnson@sig-gis.com

David Saah
dsaah@sig-gis.com
Software Development

BIG DATA & DISTRIBUTED COMPUTING
- Using reliable, scalable, and distributed computing frameworks like Hadoop and Spark, SIG can design and deliver services, applications and solutions requiring distributed processing of large data sets across clusters of computers using simple programming models. SIG Solutions are designed to scale up from single servers to thousands of machines, each offering local computation and storage.
- Using a High-Availability framework of compute clusters, SIG solutions do not rely solely on hardware to handle failures but in addition can detect and handle failures at the application layer to deliver a highly-available computational environment on top of a cluster of computers, each of which may be prone to failures.
- SIG integrates Distributed Computing Frameworks with Application Development to implement large scale machine learning and statistical algorithms in pipelines, including:
  - Summary statistics, correlations, stratified sampling, hypothesis testing, random data generation
  - Classification and regression: support vector machines, logistic regression, linear regression, decision trees, naive Bayes classification
  - Collaborative filtering techniques including alternating least squares (ALS)
  - Cluster analysis methods including k-means, and latent Dirichlet allocation (LDA)
  - Dimensionality reduction techniques such as singular value decomposition (SVD), and principal component analysis (PCA)
  - Feature extraction and transformation functions
  - Optimization algorithms such as stochastic gradient descent, limited-memory BFGS (L-BFGS)

CUSTOM SOFTWARE DEVELOPMENT FOR:
- Wildfire Simulation Modelling (GridFire)
- Property Level Fire Risk Assessments (IWAP)
- Human-Driven Satellite Imagery Classification (Mapcha & Collect Earth Online)
- Automated Satellite Imagery Classification (EcoDash and SurfaceWaterTool)
- Interactive Web Mapping of Curated Datasets (Indian Land Trust Data Portal)