Cloud Data Analytics from Excel

Roger Barga, Dennis Gannon, Jared Jackson, Jaliya Ekanayake, Wei Lu, and Mohamed Fathalla,
eXtreme Computing Group (XCG)

Data analytics as a service on Windows Azure, bringing the power of the cloud to the laptop

- Extensible collection of data analytics and machine learning algorithms running in the cloud;
- Customizable data analytics ribbon.
- Invoke models, perform analytics and visualization to gain insight from data.
- Sample and manipulate extremely large data collections in the cloud;
- Share data and results in the cloud, w/ annotations to facilitate data discovery and reuse;

Researchers use a familiar tool (Excel), familiar but differentiated.

Making sense of Big Data.

Large corporations, small and med sized businesses, and research groups are swimming in a rising sea of Big Data. Many are flailing because Big Data is difficult store, process, and analyze using conventional systems and methods.

“"The speed of business these days and the amount of data that we are now swimming in means we need to have new ways and new techniques of getting at the data, finding out what's in there, and figuring out how we deal with it." — Bud Albers of Disney

Why now? Key factors are driving demand

- Increase in data volume
- Increasing availability of cost-effective storage and compute
- Increasing complexity of analytical types (image, recognition, behavioral analysis…)
- Changing analysis model (voice and handwriting recognition, crowd sourcing…)
- Computer savvy C-level executives, making business decisions using data

Data Sources

- Excel
- Proprietary data sources
- Enterprise data

Data Ingest

- Sample
- Clean
- Compress
- Register

Refine

- Transform
- Aggregations
- Partition, shard, and replicate

Analyze

- Cluster
- Outlier Detection
- Predict
- Classify

Visualization

- Excel
- Pivot Visualization
- Egress to legacy systems

Identify Hidden Associations

Automatically identify associations hidden in data. Each association is fully specified by its strength, the number of records affected, and its preconditions.

Transform Your Data

Create, rename, delete, or convert fields. Filter, sample, or rank rows. Scramble values, compare values, apply complex formulas and dozens of built-in functions.

Visualize Correlations

Filtering, area selection, and automatic group detection makes it possible to easily identify groups of highly correlated fields, or fields with little influence on others.

Forecast Time Series

Train a model using all available data, allowing it to detect and learn patterns such as seasonal variations, underlying general trends, or user activity patterns.

Discover Similarities

Records are projected onto a two-dimensional map, in a way which attempts to project similar records to close by cells.

Categorize Records

Automatically discover broad categories of records using cluster detection algorithms.

Detect Anomalies

Users can visualize abnormal records and their neighbors, for example to address data quality or fraud issues.