Actian Analytics™

Make Big Data Analytics Faster and Easier

Extract useful information from modern datasets using Actian Analytics – a combination of high-performance analytics and in-database advanced mathematical, statistical, data mining and industry-specific functions.

Boost Analytic Performance

In-database analytics can greatly improve performance and efficiency by moving the processing out to the data, rather than moving the data to the processing. You can run more complex analyses on more data in a fraction of the time required by traditional solutions.

Expand the Use of Advanced Analytics

Leverage off-the-shelf or custom functions directly through SQL-based queries against all available data, regardless of scale. Use multiple advanced analytic libraries and dozens of dimensions in an ad-hoc process and deliver valuable competitive insights.

Advanced Analytics Libraries

Gain access to more than 700 in-database analytics functions. Basic analytic functions are included with Actian Analytics Database - Matrix. Actian also offers advanced and vertical market analytic packages. Create your own custom analytics, store them in an Actian library, and make them available for use by others in your organization.

Base Mathematic/Statistical Packages

Mathematic Functions

- Trigonometric Functions – Sine, Cosine, Tangent, Cotangent, Secant, Cosecant
- Inverse Trigonometric Functions – Sine Inverse, Cosine Inverse, Tangent Inverse, Cotangent Inverse, Secant Inverse, Cosecant Inverse
- Hyperbolic Trigonometric Functions – Sine Hyperbolic, Cosine Hyperbolic, Tangent Hyperbolic, Cotangent Hyperbolic, Secant Hyperbolic, Cosecant Hyperbolic
- Inverse Hyperbolic Trigonometric Functions – Sine Hyperbolic Inverse, Cosine Hyperbolic Inverse, Tangent Hyperbolic Inverse, Cotangent Hyperbolic Inverse, Secant Hyperbolic Inverse, Cosecant Hyperbolic Inverse
- Greatest Common Divisor (GCD) & Least Common Multiple (LCM)
- Conversion of Value – Absolute, Modulo, Floor, Ceiling, Round, Truncate, Degrees, Radians, nth root, Sign
- Exponential & Logarithm
- Gamma Function and Factorial of non-integral values
- Matrix Algebra – Product¹, Transpose, Inverse², Determinant
- Area Under Curve – Trapezoidal, Simpson’s 2/3rd, Simpson’s 3/8th
- Interpolation – Linear interpolation, cubic spline, Nelson Siegel Method³
Statistical Functions

- Arithmetic / Geometric / Harmonic Mean
- Median
- Percentile
- Mode
- Square of Deviation from Mean
- Variance
- Standard Deviation
- Skewness
- Kurtosis
- Covariance
- Rank Covariance
- Correlation
- Rank Correlation (Spearman’s Correlation)
- Sum Product
- Weighted Average
- Weighted Variance
- Weighted Standard Deviation
- Count
- Count Missing
- Minimum
- Maximum
- Rank
- Percent Rank Correlation
- Distance Measures – Euclidean, Manhattan, Mahalanobis
- Cross Tab Evaluation – Chi-squared, Phi Coefficient, Cramer’s V, Contin-gency Coefficient, Cohen’s kappa
- Hypothesis Testing – Student’s t-test, F-test, Binomial test, Wilcoxon Signed rank test, Chi-squared test, Mann Whitney test, Kolmogorov-Smirnov test, One-way ANOVA, Anderson-Darling test

Advanced Mathematic/Statistical Packages

Univariate/Multivariate

For each distribution, the following functionality is available:
- Calculation of probability and cumulative probability distribution
- Calculation of mean, median, mode, standard deviation, skewness and kurtosis
- Calculation of the inverse of a cumulative distribution
- Monte Carlo simulations: Generation of random numbers that follow the underlying distribution

Univariate Distribution

- Beta
- Bradford
- Burr
- Cauchy
- Chi
- Chi-square
- Cosine
- Erlang
- Exponential
- ExtremeLB
- Fisk
- Folded Normal
- Gamma
- Generalized Logistic
- Gumbel
- Half Normal
- Hyperbolic Secant
- Inverse Normal
- Laplace
- Logistic
- LogNormal
- Maxwell
- Nakagami
- Normal
- Pareto
- Power
- Rayleigh
- Reciprocal
- Semicircular
- Student’s T
- Triangular
- Uniform
- Weibull

Data Mining

Principal Component Analysis
- Includes formation of correlation matrix, calculation of Eigen Values & Eigen Vectors, and framing the orthogonal matrix

Clustering Methods
- K Medoids, Fuzzy K Means
- Hierarchical Clustering – Hierarchical K Means, Hierarchical O – Cluster

Other Supervised Learning Methods
- Linear Discriminant Analysis
- Flexible Discriminant Analysis
- Mixture Discriminant Analysis
- Decision Trees
- Naïve Bayes Classifier

Linear Regression
- Calculation of regression coefficients (B’s), standard error of coefficient, t Statistic, p Value, confidence interval
- Goodness of fit measures — F statistic, R-squared, adjusted R-squared, variance inflation factor (VIF), residuals and tests for heteroschedasticity, Variable subset selection

Logistic Regression & Probit Model
- Calculation of regression coefficients (B’s), standard error of coefficient estimate, chi-square, p Value
- Goodness of fit measures — concordance/discordance, Gini coefficient, false positives, false negatives, variance inflation factor (VIF)
- Variable subset selection
Financial Services Package

Corporate Finance

Valuation of Cash Flows
- Present Value Analysis — NPV, IRR, periodic payment
- Estimation of discount rate using CAPM — levered and un-levered Beta

Firm Valuation
- Stock valuation using Dividend Growth model
- Asset valuation based on equity price (Merton model)

Options and Derivatives

Equity Derivatives
- Risk-neutral valuation of calls and puts using Black Scholes with and without dividends
- Greeks — Delta, Theta, Gamma, Vega, Rho with and without dividends

Currency Derivatives
- Currency Derivatives, Cross-currency Derivatives and Quantos

Interest Rate Derivatives
- Models for short rate i.e., Vasicek, Cox, Ingresoll and Ross, Ho & Lee, Hull & White, etc.
- Models for forward rate i.e., HJM, LMM

Fixed Income / Bond Math

- Price
- Coupon dates — first coupon date, all coupon dates
- Computation of yield measures for fixed-rate bonds (Current yield, Yield to maturity, Yield to first call, Yield to first par call date, etc)
- Discount Margin
- Duration, Modified Duration, McCaulay Duration
- DV01, Convexity
- Computation of Forward Rates and bond valuation using forward rates
- Spreads — Zero-Volatility Spread, Option-adjusted Spread (OAS), Option Cost, Nominal Spread
- Spot rate using bootstrapping
- Historical and Implied Yield volatility

Time Series Analysis

- Autoregression ARCH(m)
- Moving Average EWMA
- GARCH(p,q) Model
- Regime Switch Model
- Holt Winters

Actian transforms big data into business value for any organization – not just the privileged few. Our next generation Actian Analytics Platform software delivers extreme performance, scalability, and agility on off-the-shelf hardware, overcoming key technical and economic barriers to broad adoption of big data, delivering Big Data for the Rest of Us™. Actian also makes Hadoop enterprise-grade by providing high-performance ELT, visual design and SQL analytics on Hadoop without the need for MapReduce skills.
1. We have parallel implementation of matrix product using Strassen’s algorithm which is $O(n^{2.708})$ instead of $O(n^3)$.

2. We perform matrix inversion using Singular Value Decomposition, which results in much better speed. Singular Value Decomposition is done using our proprietary implementation of Cuppen’s method for finding Eigen Values and Eigen Vectors.

3. Nelson Siegel method can be used for fitting curves with parsimonious data. The disadvantage with this method is that it is an extremely slow converging algorithm. We have a proprietary implementation on Nelson Siegel method, which includes an intelligent grid search and accelerated iteration.

4. The Archimedian Copulas have closed form solutions and Monte Carlo simulations are not required to find the area under the curve. Such areas can be calculated using closed form solutions. For the other copulas, simulations are required to calculate the area under the curve.

5. Eigen Values and Eigen Vectors are calculated using a proprietary implementation of Cuppen’s method. This method can be easily parallelized and is much faster than the traditional QR algorithm.

6. For clustering algorithms, we have a proprietary algorithm that is used to determine the appropriate number of clusters required.

7. Variable subset selection is done using a proprietary implementation of a leaps and bound algorithm published by Furnival & Wilkinson, 1974. The original algorithm is infeasible when the number of variables is larger than say, 40. We improvised on the original research and our algorithm can be used for a much larger number of variables.