An Oracle White Paper
October 2014

Agile Advanced Analytic Platform (A3P)

Applied to National Security

ORACLE

Disclaimer

This White Paper includes data that shall not be disclosed outside the Government and shall not be duplicated, used, or disclosed—in whole or in part—for any purpose other than to provide information to the Government. This restriction does not limit the Government's right to use information contained in this data if it is lawfully obtained from another source without restriction.

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Executive Overview

Advanced analytics is a key factor in producing good Intelligence. However, analytics means many things to many people. Unfortunately, the overloaded use of the word "analytics" creates not only ambiguity in conversation but also incomplete and inefficient solution architectures. Advanced analytics utilizes data of different types, from different sources and requires different algorithmic processing. Valued Intelligence results from the correlations and insights amongst this data and the inter-relationships that exist from different data sources.

Frequently, different data types are managed in different data stores. This causes information fragmentation and hinders analysis. Additionally, there are many inefficiencies in the steps that have to be performed to do analysis. These inefficiencies result in significant delays in mission execution that are unnecessary given the advancements with today's technology.

What is needed is an efficient and robust data management and data analysis platform that addresses the most common workhorse analytic technologies and techniques used by the Intelligence Community today. The desired solution needs to support rapid, low cost, iterative and adaptive analytics. By empowering analysts with a self-service platform, users can perform rapid tests and evaluations on the data thereby decreasing the time and costs to derive analytical value.

The platform should be complete with support for traditional and cutting edge data management capabilities and data analysis techniques yet simple and fully maintain consistent security, manageability, scalability and interoperability. Ultimately this platform represents the compilation of many systems working together to perform the necessary analytic functions. The entire 'system of systems' needs to be professionally designed, optimized, certified and supported.

The Agile Advanced Analytic Platform (A3P) serves as this platform. A3P is a fast, robust analytic environment that supports full-spectrum analytics from data exploration & discovery through advanced and predictive analytics. It represents the combination of software and hardware that has been professionally engineered, optimized, developed and deployed to support the (Big Data) analytic challenges faced today.

A3P empowers analysts to explore, test and evaluate with ease and speed. Oracle's platform and process re-balances the time and energy of analysts to where it is most valuable. Oracle's approach and solutions are optimized for Big Data Analytics and support the performance and scale required for high volumes and high velocities of highly varied data. A key differentiated objective is to empower analysts to explore, test, and evaluate with speed in a self-service fashion without data movement thus reducing the need for costly programmers and data scientists needed for every analyst that needs to ask a new question of the data.

The Analytical Challenge

Advanced Analytics is the corner stone to producing good Intelligence. However, analytics means many things to many people. This is natural because the Intelligence production process requires the use of many and varied tools. To illustrate the wide use of analytics, consider the bullets below that represent the topics that one might naturally entertain in a discussion where the subject was simply the word "analytics."

- **Search & Discovery**
- **Reporting**
- **Ad-hoc Query**
- **Predictive Analysis**
- **Statistical Analysis**
- **Pattern Matching**
- **Data Mining**
- **Big Data Analytics**
- **Fast Data Analytics**
- **Spatial & Geotagging**
- **Time-series**
- **Natural Language Processing**
- **Entity Extraction**
- **Latent Semantic Indexing**
- **Sentiment Analysis**
- **Graph Analytics**
  - **Property Graphs**
  - **Resource Description Framework**
  - **Network Data Models**

Chart 1 – The word analytics can mean something different to everyone. The above represent some of the popular analytics.

Unfortunately, the overloaded use of the word "analytics" creates not only ambiguity in conversation but also incomplete and inefficient solution architectures. The reason has to do with the fact that the analytical tools used to perform the above processes vary as widely as the definition of analytics itself and typically involve separate analytical servers depending on the use case's requirements. The reality is that the different tools and different ways we store and manage data for analytics creates impedance in doing higher level, result-oriented advanced analytics and adds significant delays, inefficiencies and security exposures due to the movement of data between data management and data analysis servers and systems.

Data Fragmentation

Analysis often involves bringing together data from different data sources. The data is formatted differently, processed and analyzed differently. For example, geospatial data usually contains the geospatial specifics e.g., latitude and longitude, along with some meta-data of what is significant about this location or area. The processing and analysis of geospatial data is a combination of the elements of interest (e.g., a building) in the framework of a geospatial question.

For example, a list of all buildings, their addresses and what purpose the building serves may be stored in a special purpose geospatial database. This will allow someone to ask questions such as the following:

- Show me all buildings in this area that contain emergency responders
- Show buildings which contain hazardous substances and are within 10km of schools or daycare facilities
- Show the shortest route between two locations (A, B) that include a gas station, a grocery store and a post office.

In the geospatial example above, special geospatial algorithms work together with the semantics associated with the geospatial metadata to provide the answers needed to for the analyst's questions. Likewise, consider a database that stores documents. This library of information may be useful as a Wiki or other online document repository. The processing and analysis required here may consist of rapid, relevant word search (with synonyms) as well as a way to group, categorize and recommend similar and relevant documents. This database must be good at performing those and other related content management functions. Note that while two people can be doing "analytics" – one on geospatial data and one for document search – the databases and their capabilities vary widely.

Today, to handle this myriad of different data types and their respective formats and processing, we typically find a different data store for each of the different types; one data store used for geospatial, a different one used for documents, perhaps another for processing transactional, log or scalar data. Because the data is stored separately, the data is effectively siloed. From an enterprise level, we would say the data is fragmented based on the data types.

The siloed data creates a serious challenge for analytics that ask questions across different data. The reason for this is that the individual data stores can only manage and process one (or very few) specific type of data and nothing more. A challenge arises when we need to access multiple disparate databases with a single analytical question because neither database can support the question individually. Consider the example databases from above; an analyst might want to know which documents refer to buildings that are within 2km of each other. This question would be hard to answer since the answer is a fusion of data from multiple databases. However, advanced analytics often requires us to ask questions and fuse information that comes from various sources.

It is the correlations and insights amongst this data and the inter-relationships that exist from different data sources that give us the intelligence we are seeking.

Today, to achieve any type of cross data type analysis, an analyst has to write (or have a programmer write) a program to translate their question into queries for each of the data stores. The results from each data store would then be manually or programmatically combined or discarded, summarized, correlated and so forth. Often there is a series of interim steps where the results from one query are used as inputs to a subsequent query. To facilitate anything but the simplest of questions, a program will have to perform a series of operations and then store and dynamically query based on these interim results. This is difficult to do, difficult to manage and verify, and often extremely sub-optimized and inefficient in execution.

The figure below graphically depicts this data fragmentation challenge.

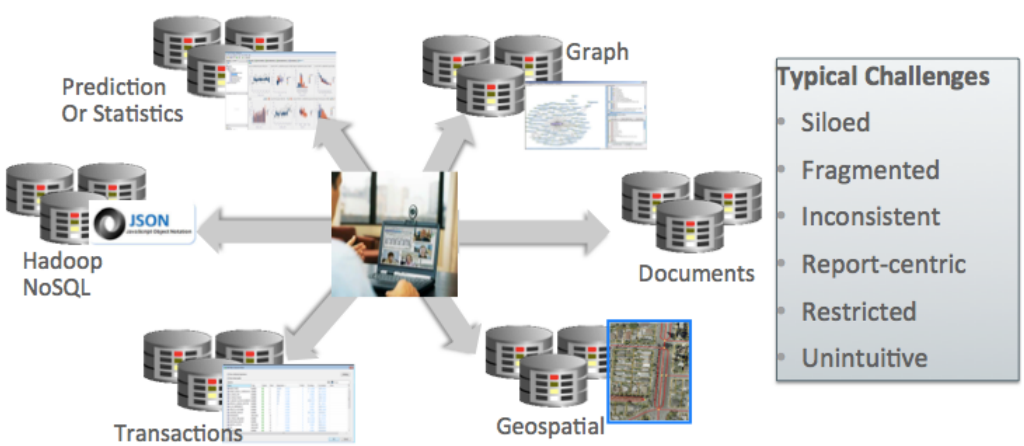

Figure 1: A different data store for different data types hinders the analytic process.

To overcome these challenges, one requires a platform that can intelligently store, process and analyze the various data. By moving the data fusion, correlation, integration and management to a platform that can intelligently handle and process this data, analyst can then work more rapidly, with more agility and in the most efficient mode possible.

The Pareto Principle of Analytics

Rarely is there a scenario in which new data sets arrive and the data is instantaneously "analyzed." Data almost always has to go through a preparatory stage or series of stages. In today's world of big data in which the data may have unknown value, there can be as many as seven stages as indicated in the figure below. We can refer to the stages as the analytic lifecycle.

Figure 2: Seven step summary to find, understand, prepare and analyze data to derive intelligence.

There's another aspect to the analytic process that often impedes our ability to efficiently derive the insights and actionable intelligence from our data. Some conservative estimates place 80% of today's efforts for "big data analysis" on simply preparing data for the analysis itself. Data wrangling[1] – a term often used to describe this process – consists of preparing the data through formatting, normalization, standardization, transformations etc.

[1] See http://en.wikipedia.org/wiki/Data_wrangling

[2] http://en.wikipedia.org/wiki/Agile_software_development

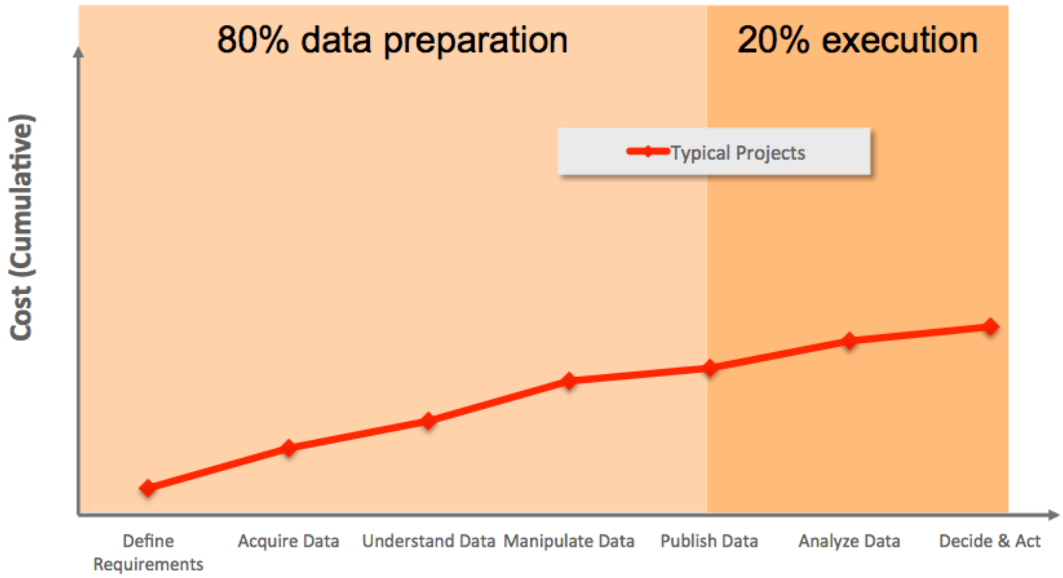

Figure 3: Typical projects today incur 80% of cost just getting data to the point that it can be utilized and analyzed. The goal is to shift these costs from 80% time in data wrangling to 80% actual analysis execution.

Note that if you are dealing with high volumes and high velocities of Big Data, this 80/20 ratio may be more like 90/10 as Big Data exacerbates this problem. Once data has been properly prepared, then and only then can it be used for the actual analysis activity. Our objective is to create a more efficient platform to handle the analytic stages. Combined with the challenges around data fragmentation and conducting analysis across different data types stored in different data stores, we see the requirements for an analytical platform. We call this platform the Agile Advanced Analytic Platform or A3P and describe its characteristics below.

The Agile Advanced Analytic Platform – A3P

Our goal with the A3P was to create a platform that addresses the myriad of most all of the possible analytic processes, in an efficient manner, that ultimately drives better actionable intelligence. To put it briefly, Oracle has committed to a strategy of "bringing the algorithms to the data" versus the traditional approaches of moving the data to separate analytical servers. As the volume of data grows, at some point, the traditional approaches will fail. Algorithms are small. Data can be large. Move the algorithms and not the data!

To establish the platform, we had to assess the key factors A3P needed to make it successful. Subsequently, in assessing the analytic challenges faced, we formed some conclusions about desirable traits that would do a lot to improve the efficiency, cost, time and overall success in performing analytics.

Agile

Agile Development is a current coding and development practice that has yielded much success by more efficiently creating programs that satisfy their intended objectives. The Agile Software Development[2] process provides a few best practice principles from which we saw relevance and analogs to in considering an A3P. A few principles with direct applicability include the ability to **support a rapid, iterative and adaptive approach to the analytical lifecycle**. Let's discuss these principles and how to approach them in an A3P context.

The desire to do a job quickly and efficiently is an often easy-to-state objective that is hard to actualize. For analytics, there are several ways to improve the speed to analytics.

A self-service approach to analytics will allow the analyst to first gain access to the "right data" and then quickly experiment with new ideas and ask new questions without the reliance on programmers or other data specialists

To facilitate a self-service environment, one needs a data management and data analysis platform that has tools designed to empower the user in a self-service construct. The resulting solution platform would have quick ways to view data, model data, construct data flows as well as execute analytical functions.

In achieving a more rapid ability to execute analysis, we can then think about our data management and analytical platform as supporting two desirable concepts: Sprint analytics and disposable analytics. With **sprint analytics** – adopted from the Agile Development process – the analytics need to be done quickly and then evaluated. That is, you are trying out ideas to see if there is value in the data or value in the analytic tool or process. This "fail fast/succeed fast" approach ensures that you obtain the valued analytical operations quickly.

Disposable analytics simply means there is a low cost capability in evaluating an analytic. Costs here can be defined not with absolute dollars per se, but measured by amount of effort, time, and other resources required. The premise is that if the costs are extremely low in my sprint analytics, it is acceptable to fail.

With the ability to support rapid analysis through self-service, the platform supports the iterative nature of analysis and has a cumulative and positive impact to the analytic lifecycle.

Another definition of agile is adaptive. **Adaptive** in this case means that the platform can change to meet your specific analytical needs. It is not necessary to have one-size fits all approach to establishing a platform. More importantly is that the platform can conform and adjust to your specific requirements. Therefore the platform supports agility in allowing the data, processes, techniques and tools to be adjusted as needed.

Advanced

Next, we desire to have a platform that supports all the common definitions and types of simple, basic, and advanced analytics in use. We need a fast, fully complete analytic environment. It needs to support the entire analytic lifecycle from data exploration and discovery through advanced statistical and predictive analytics.

[2] http://en.wikipedia.org/wiki/Agile_software_development

In addition to analytic capabilities, the platform also must satisfy the common "enterprise grade" requirements of consistent security, manageability, scalability and interoperability. This can be a challenge today because each of the various data stores probably has its own approach to doing these things. Sometimes the desired capability – e.g., enforcing rigorous security – consists of incomplete and inconsistent implementations.

To accomplish these goals, we need a platform that is intentionally designed to support large-scale analytics in a National Security context where security, scalability, and interoperability are non-optional. For security and manageability, this does not happen by accident. One has to "instrument" the platform to allow for this. By instrument, we meant that the components and systems have been designed and implemented not as stand alone, single function components, but rather as parts of a larger system with interoperable subsystems.

Platform

In the past, papers such as this have focused exclusively on the software layers. In today's world of Big Data Analytics, this is no longer a viable option. The hardware matters because the software is extremely reliant on the hardware to achieve the necessary performance, scale and reliability demands placed upon it. The platform instrumentation discussed above has to occur in the software and the hardware. By addressing both the software and hardware, A3P provides the ability to manage and secure the entire environment.

To build an A3P, a variety of individual systems will be used in which the engineering and design of these systems will support a larger purpose-built system for doing analytics. Integration must occur at both the physical interface level and the logical software levels. As you will see, Oracle's Engineered Systems provide the components to build such a platform.

Additionally, the platform needs to be professionally supported. This means that diagnosis and remedy of problems – whether they are performance, integration, security, etc.– can occur. Oracle is one of the few if not the only company that can provide such a capability because we engineer, build and support all the individual systems that make up the platform.

In the end, the overall platform cannot be a set of independent parts that, when working together in unison to achieve an analytical objective, fail to support the enterprise-grade requirements. What is needed is a 'system of systems' in which the data management and analytic processes can be satisfied in an agile manner with the support of security, scalability, manageability and interoperability.

A3P was designed to be such a system of systems. It represents a comprehensive analytic environment. It supports full-spectrum analytics on a combined software and hardware infrastructure that addresses big data performance and scale, security, manageability. Perhaps most importantly, the A3P consists of systems that have been engineered, developed with the necessary hooks for integration, scale, security and manageability that is professionally certified and supported by Oracle Corporation.

In summary, the following concisely represents the desired characteristics of A3P:

- Agile: Self-service analytics that is easy to adjust
    - Sprint analytics: Fail fast/Succeed fast
    - Disposable analytics: Low effort, time & costs for rapid iterations
- Advanced: Complete yet simple solution that fully maintains consistent security, manageability, scalability and interoperability

Platform: A 'system-of-systems' that integrates, is complementary and is professionally supported

Software Functional Layers

A3P supports a familiar architecture. Figure 4 below represents a logical layering of the software for A3P.

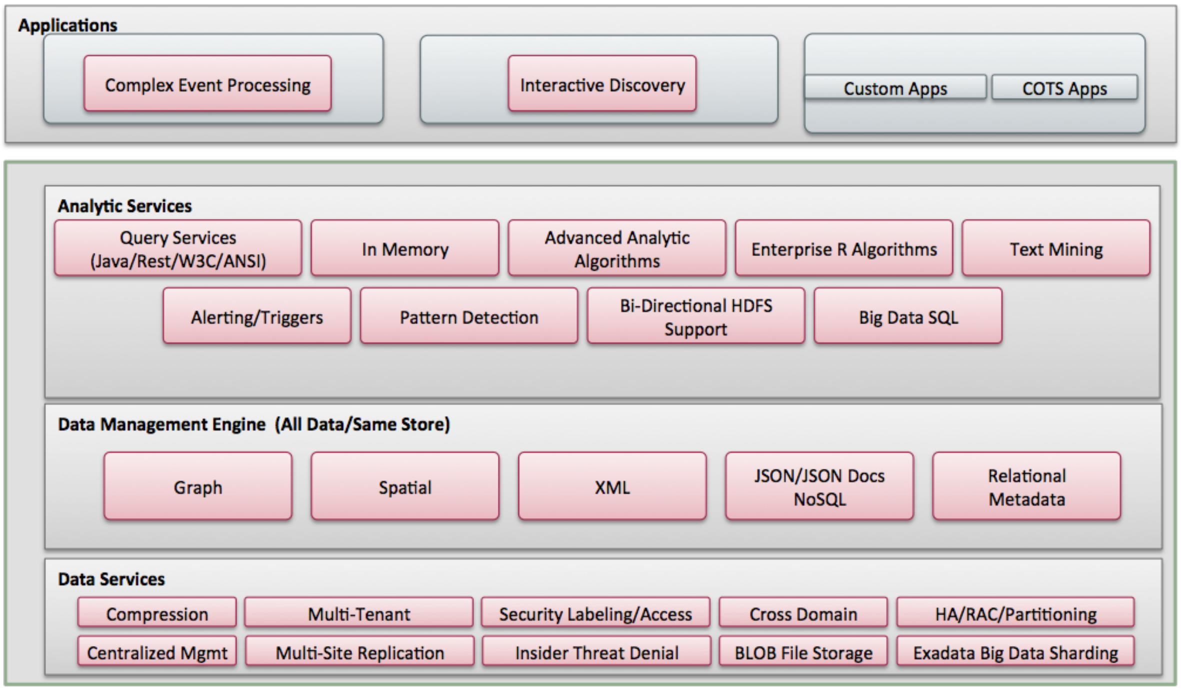

Figure 4: Software Functional Layers of the Agile Advanced Analytic Platform

Applications

The top tier of this functional diagram illustrates the Applications tier. As the name implies, both custom and commercial applications interact with the A3P from here.

Since integration, interoperability and avoidance of technology lock-in is always a prime objective, the Applications tier utilizes standards whenever possible. As you see in Figure 5 below, different data types have standards e.g., XML, SPARQL and OGC. There also exist protocol standards that the Applications tier should adhere to e.g., JDBC, REST, and SOA Web Services.

Recognizing the A3P may grow in capability to include support for new data, data types and analytics, it is important to ensure that the Applications and the rest of the A3P remain true to the use of standards for integration and interoperability. Where standards don't exist, standard APIs with high quality documentation can be used.

Data Services

At the lowest level in the diagram are the data services. The data services depicted describe the general capabilities needed for many if not most data. The service titles are self-descriptive. Note however there is a critical reliance on the proven ability of the hardware to help efficiently support these capabilities. For example, the offloading of data encryption and compression to the hardware is critical if the platform is to support scalability.

Data Management Engine

The Data Management Engine is the centerpiece in resolving our data fragmentation and cross analysis conundrum described earlier. The engine not only supports the different types of data, but also natively stores, manages, secures and processes the data.

Figure 5 below shows the Data Management Engine data types and inherent analytic processing support within A3P.

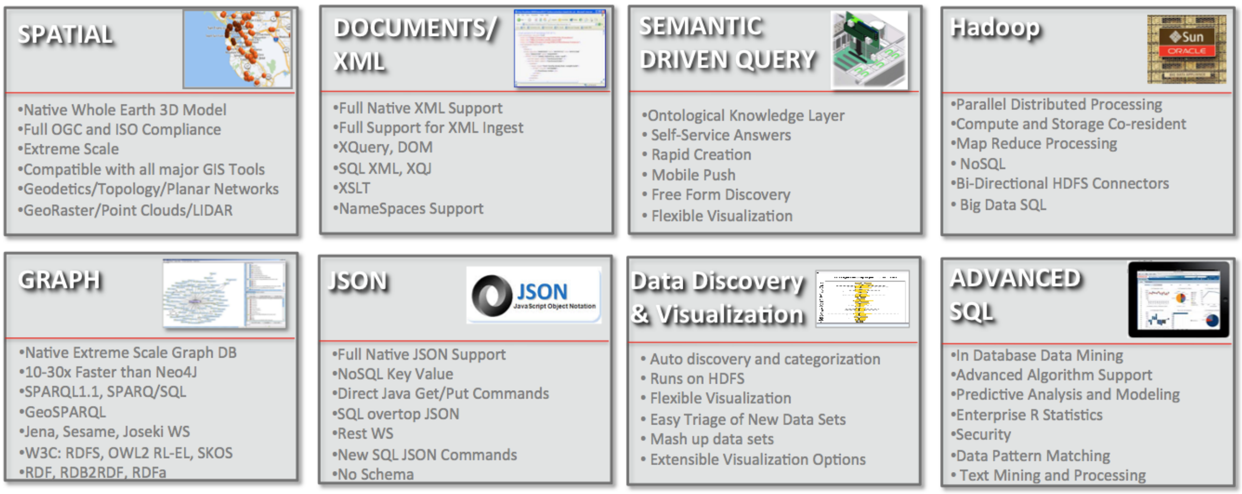

Figure 5: Co-resident data management and processing capabilities solves the problems seen in the siloed approaches used today

Analytical Services

The Analytical Services describe a combination of the types of processing that often occurs over the data from the Data Management Engine. Once again, the figure is intended to be self-describing.

Within A3P itself, there have been very conscious decisions on what capabilities to provide and where and how that should happen. There are very deliberate interactions between the Analytic Services, the Data Management Engine, and the Data Services. In A3P, all three functional layers are working together in such a way that the intended functions and capabilities are realized in an efficient and practical manner.

For example, by putting the security elements in the Data Services, the Data Management Engine and Analytic Services do not have to worry about (re)implementing core and basic security functions. As another example, in A3P, the ability to enact the Analytic Services within the Data Management Engine physical environment means that one is not moving vast amounts of data to a separate processing tier.

Understanding the Scope of A3P

For clarity, A3P is not a product or engine that just answers questions with no preparation or development. A3P is a system of systems environment for customers to conduct analytical activities. It allows customers to integrate their data, perform analytical functions, utilize advanced algorithms and visualize their data in various and meaningful ways. A3P is a platform for customers to build and integrate the advanced analytics necessary to meet mission objectives in an efficient, secure, agile and performant manner. While not realistic to state that A3P can answer any or all analytical questions, A3P does attack the common patterns, processes and architectures that limit many of today’s analytical environments.

A3P Infrastructure Overview

As mentioned earlier, an effective and efficient analytics platform has to be a system of systems. It needs to address the variety of data types and analytic processing required. A3P is a system that is built from a series of applied, interoperable, complementary, smaller systems. It is important to note that A3P is a co-design of specific software applied to specific hardware. A3P engineers considered not only the logical software capabilities but also the necessary and supporting hardware that will make the software scale to support the Big Data analytics in use today.

Figure 6 below illustrates one instance of an A3P.

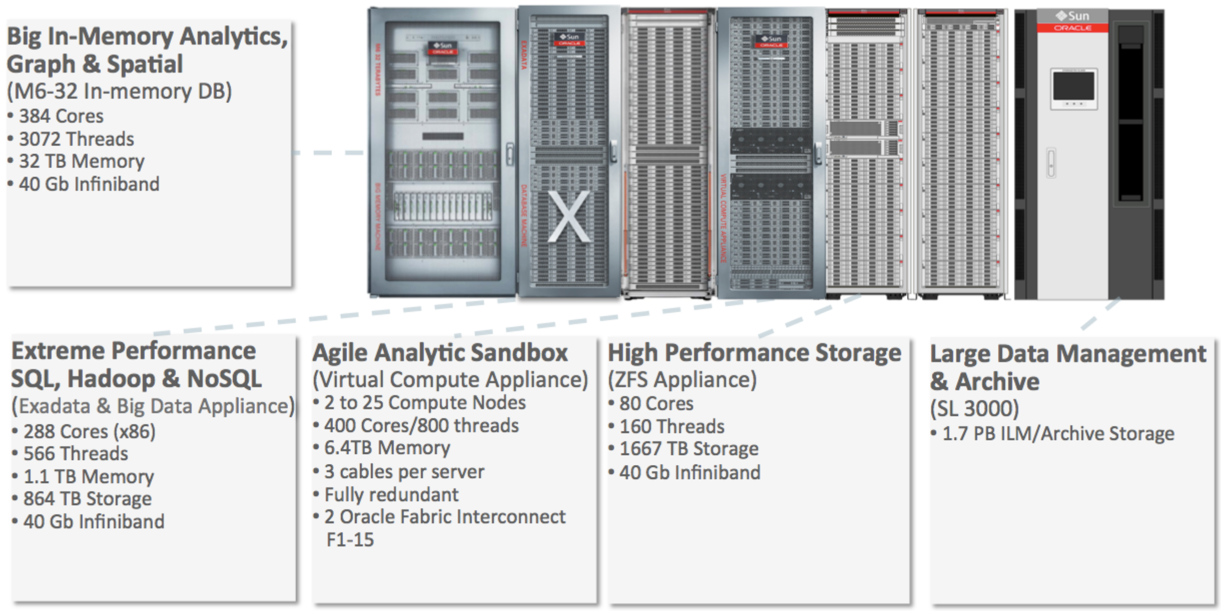

Figure 6: A3P Infrastructure work together to form a 'system of systems' to handle today's varied analytical requirements.

A3P Platform Details

Oracle's goal is to give A3P customers flexibility to meet mission analytic demands. Every analytical challenge is different, which will require differing compute, storage and tools. This requirement led us to our system of systems approach. A3P can be expanded at any particular element to meet a more specific need. For example, if there is a larger requirement for a more robust Big Data system, then more Hadoop nodes can be added by extending the Big Data Appliance without concern or disruption to any of the other systems.

The software tools are also flexible. If a particular analytical capability is not needed, it can be removed from the A3P blueprint and pricing is adjusted accordingly.

A System of Systems

System of systems design is critical to reduce the complexity and efficiency in order to focus resources on mission analytics and not on infrastructure, integration, and "glue-ware". A3P's system elements are as follows:

- Interoperable – Each element in A3P has standard interfaces to the others. For example, data that requires both Hadoop and structured Relational can exist separately, be merged physically or merged logically. The copious use of standards also allows you to un-plug and plug other complementary technologies that are not part of the A3P baseline.

- Complementary – Each system element in A3P is designed for joint agile analytic workloads, each supporting various methods, tools and processes for the right approach to each analytic challenge.
- Manageable – A3P is managed as one system. Using Oracle Enterprise Manager, the A3P hardware and software layers are all monitored and managed from the same console.
- Integrated – A3P system elements are integrated via a private Infiniband network. This gives A3P the bandwidth needed for advanced analytic workloads. Integration is not just at the network. Oracle integrates at the architectural layers and tests integration between layers (see the figure below.)

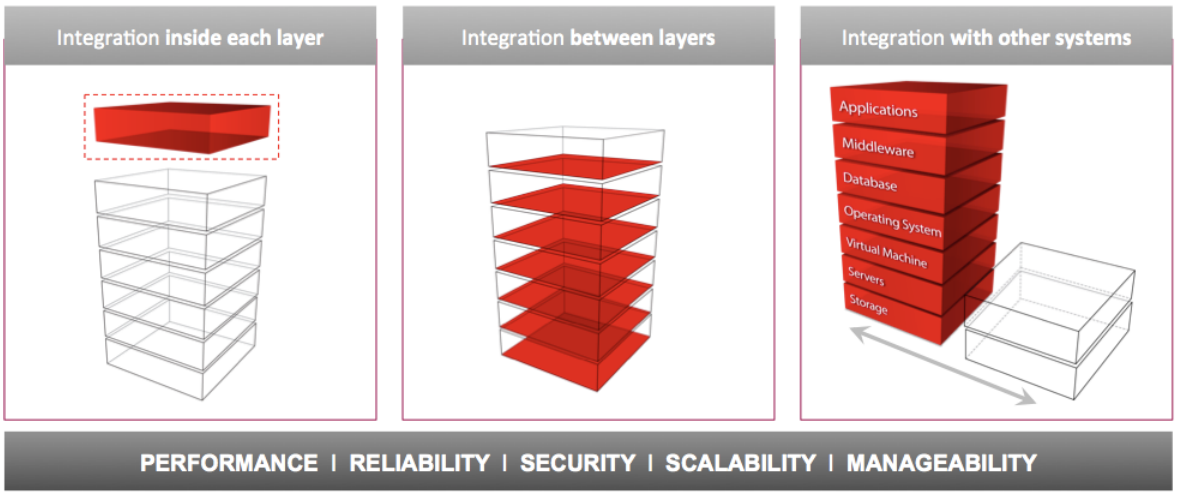

Figure 7: Vertical and horizontal integration means lower costs, faster time to mission, better reliability, faster performance and simplified management

A3P Ingredients & Self-Service Analytics

In this section, we will discuss the ingredients that make up A3P. In doing so, we will refer back to the original requirements and approach discussed earlier in the paper with the intent of mapping the actual technologies to the solution areas.

Sprint Analytics and the interactive speed of thought nature of analysis require an analytical environment that is self-service. All too often analysts are tied to a score of developers and map-reduce programmers providing custom coding services. This is a slow tedious process of analytical growth and doesn't give the analyst the type of "trial and error" needed to get to a mission outcome.

Data-driven visualizations transform the way an analyst absorbs and understands the information at hand thereby accelerating their time to valuable insights, and enabling them to make better decisions. Oracle Big Data Discovery for unstructured data and Oracle Answers for structured data are the technologies used in A3P to provide self-service, rapid interactive analysis.

To realize true analytic value from Hadoop, we need to take a fundamentally new approach. First we need a single user interface that is highly intuitive, and visually compelling. Next, we need a set of integrated, end-to-end capabilities that allows anyone, not just the data scientist, to quickly find, explore, transform and analyze data leveraging the power of Hadoop on the data in place.

And finally, the results of data transformations and new business analytics should be shared back into the ecosystem so they can be leveraged by advanced statistical tools, queried by BI technologies, native Hadoop tooling or even integrated into an operational data analytics store. The goal is to remove the complexity of Hadoop and increase overall analytic agility, as well as maximizing the chances of wider adoption throughout the organization.

Figure 8: Big Data Analytics requires a new approach that allows analysts to interact with data and collaborate on their findings

Once a potential data set is found, quickly understanding its value is important. A3P helps you understand the shape of any data set by visualizing its attributes. A3P provides the capability to easily sort these attributes by information potential so the most interesting ones appear first. Analysts can also get an overall perspective on the data quality to determine if it is sufficient in rendering a proper analysis. The analyst can quickly and easily experiment with attribute combinations and understand important correlations to ultimately decide whether any particular data set is worthy of further investment.

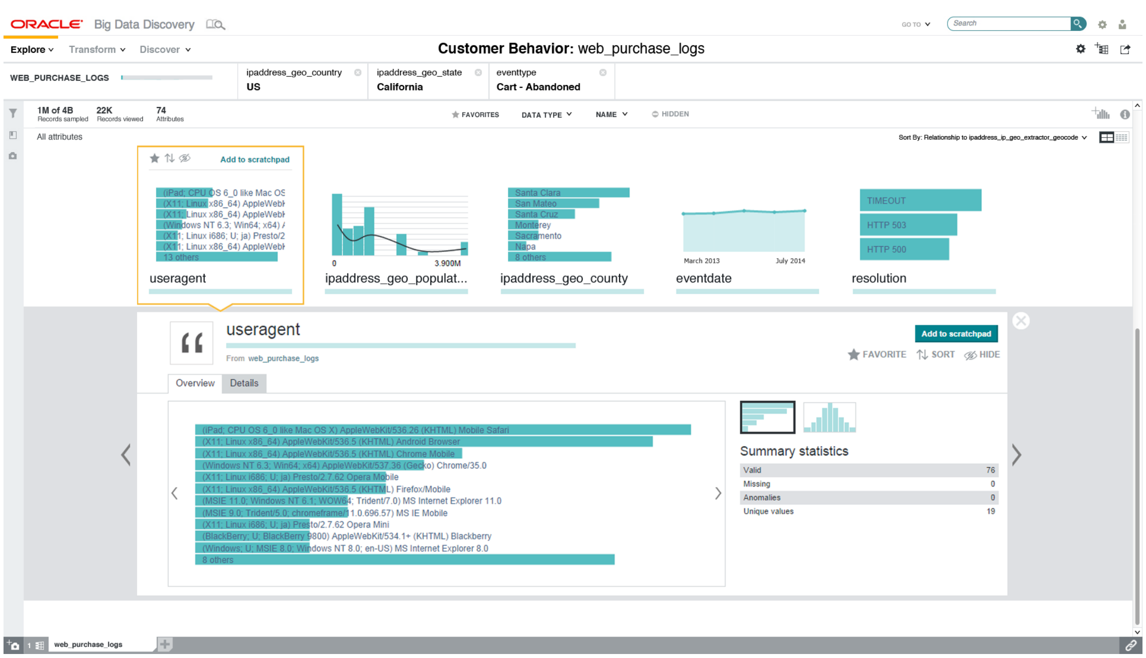

Figure 9: Self-service analytics allows analysts to quickly discover and visualize important attributes and correlations in Big Data

From self-service monitoring of performance and processes, to easy-to-use ad-hoc queries and reports, to forward looking predictive analysis, A3P delivers on the promise of self-service. Analyst access to information can be through multiple channels such as web-based user interfaces, industry standard portals, mobile devices, and the Microsoft Office Suite of applications.

Figure 10: Oracle Answers enable analysts to visually interact with vast data sets through simple and intuitive interfaces

Geospatial Analytics

Many data sets have a location or a derived location as a key element used in analytical processes. Many feel geospatial data is the analytic glue that ties data sets together. Geospatial analytics is integral to the A3P architecture and provided by the industry's most widely used geospatial data management and analytical engine: Oracle Spatial.

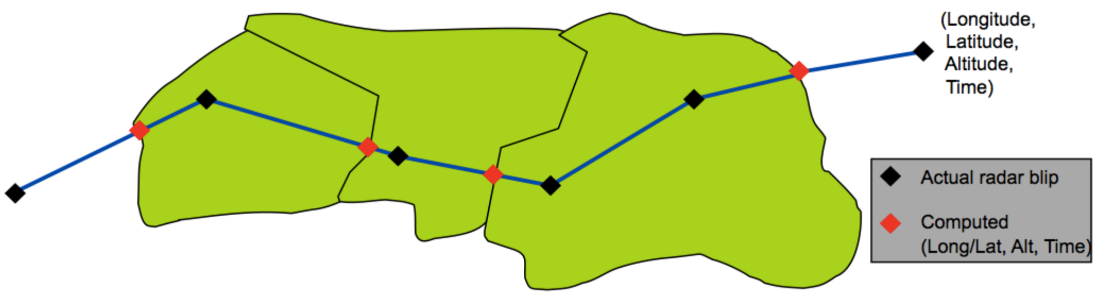

Figure 11: Spatial functions calculate intersection of flight paths and airspaces using Linear Referencing to interpolate time and altitude for entry/exit. While mathematically complex, A3P gives the analyst an easy way to leverage geospatial analytics.

Oracle Spatial supports complex geographic information systems (GIS) applications for visualization, enterprise applications and location services. Oracle Spatial provides spatial query and analysis features with advanced spatial analysis and processing in the Oracle Database. It supports all major spatial data types and models, addressing challenging analytical requirements including the following:

- Vector Performance Acceleration – this turbo-charged spatial performance is up to 50-100 times faster execution of operations and functions such as spatial joins, touch, contains, overlaps;
- Geocoding and routing engines, spatial analysis and mining functions;
- GeoRaster features to store and process geo-referenced raster data, such as satellite imagery and gridded data; now supports raster algebra and virtual mosaics;
- 3D, point cloud and LIDAR data management and analysis;
- Topology data model and linear referencing for land management applications

Oracle Spatial provides over 400 functions to perform calculations on geometries, such as area of a polygon, length or perimeter. These functions are used, for example, to determine the total area of all conflicts around a given county, the length of a roadway, or the length of a provincial border. Functions can also generate new geometries such as buffers, unions, intersections, and much more. For example, they can be used to define regions by creating a 5-mile buffer around all targets, identify the new geometry representing the union of two regions, or find the intersection between two regions. Other functions include interior point, concave hull, and generation of triangulated irregular networks through Delaunay triangulation.

Oracle Spatial is a whole earth geometry model and takes into account the curvature of the Earth's surface when performing calculations on geodetic data. Thus, Oracle Spatial functions return accurate lengths and areas for both projected and geodetic data. It supports over 30 of the most commonly used distance and area units, including foot/square foot, meter/square meter, kilometer/square kilometer.

Oracle Spatial supports the Open Geospatial Consortium (OGC) and ISO TC211 standards including: OGC OpenLS 1.1, Web Feature Service – Transactional (WFS-T) 1.0, Web Feature Service 1.0, and Catalogue Service 2.0, on a variety of client technologies and platforms.

Oracle Spatial and Graph includes full support for database transactions on WFS-T feature tables through SQL without restriction. It also supports Workspace Manager versioning and WFS feature tables. Java and PL/SQL client APIs are provided.

Most importantly, as A3P encompasses such a broad range of data management and data analysis capabilities, each of the components can feed each other. For example, we can ask a geospatial question to determine the number of times an event occurred within 5 KM of a specific location and automatically feed that result into a predictive model that determines the likelihood of a specific behavior correlated with that event type.

Graph Analytics

The Graph data model is used in numerous ways within the Intelligence Community. Uses include linked data, relationship networks, social relationships, and inference data. Each graph model contains a set of subject – predicate – object expressions organized as a graph of directed, labeled edges. The edge is the link (or relationship) that connects a subject node to an object node and is labeled by a predicate (property).

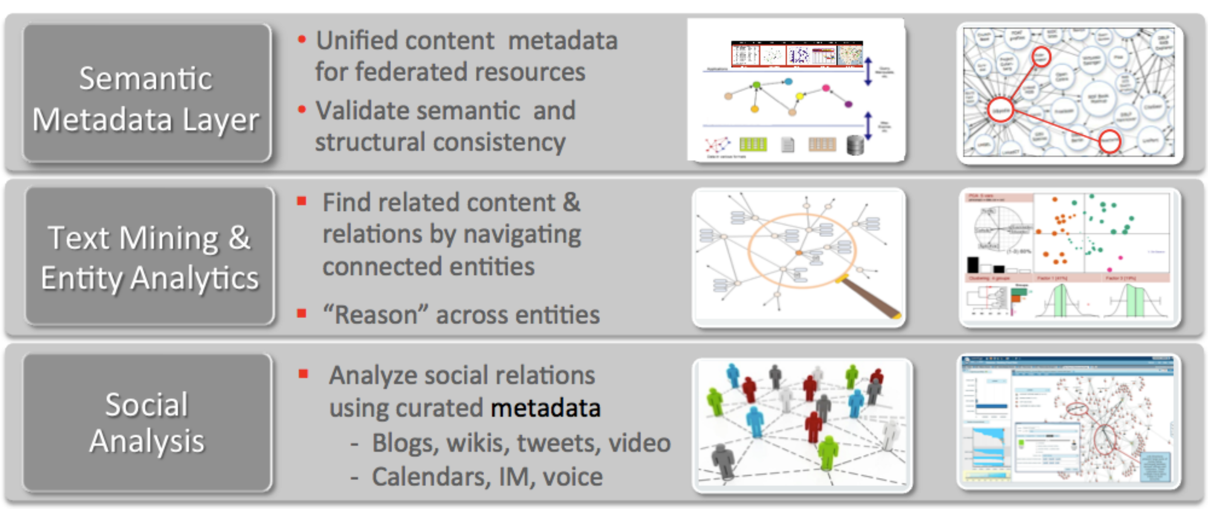

Figure 12: Typical Uses for Graph Analytics

The RDF Semantic Graph feature of A3P provides complete semantic data management. A3P has native support for World Wide Web Consortium (W3C) standards RDF and OWL for representing and defining complex, semantically related data and SPARQL. (SPARQL is a query language designed specifically for graph analysis.) Oracle Graph also stores and loads RDF/OWL data and ontologies, inferencing using OWL 2 and user-defined rules, and SPARQL 1.1 query and update.

Oracle Graph has proven scalability to over 1 trillion triples (LUBM benchmark) and can scale up to 8 petabytes of triples. It has native, forward-chaining, persistent inferencing using any combination of RDF, RDFS, and OWL 2 RL and EL profiles, as well as user-defined rules for specialized inference capabilities. It also provides a plug-in framework to support third party specialty reasoners.

A3P's space-efficient storage saves up to 60% disk space for scalable and performant loading, querying, and inferencing. Graph's often get quite large and grow very rapidly causing issues with ingest, query and analytic performance. Oracle combats these challenges with robust in-memory and sharding approaches to query using A3P hardware accelerations.

Numerous vendor products (example: Cytoscape, Tom Sawyer) using standard interfaces can accomplish graph visualization from data in A3P Graph. A3P Graph also support label security on Graph Nodes and Edges as well as Spatial Location through the use of GeoSPARQL and Oracle Spatial.

Advanced Analytics

A3P empowers analysts to extract knowledge, discover new insights and make predictions—working directly with large data volumes in the Oracle Database. A3P includes Oracle Advanced Analytics that offers a combination of powerful pre-built in-database algorithms and open source R algorithms. This self-service capability is easily accessed via a range of GUI and IDE options as well as standards-based SQL tools and the R statistical graphics and computing language.

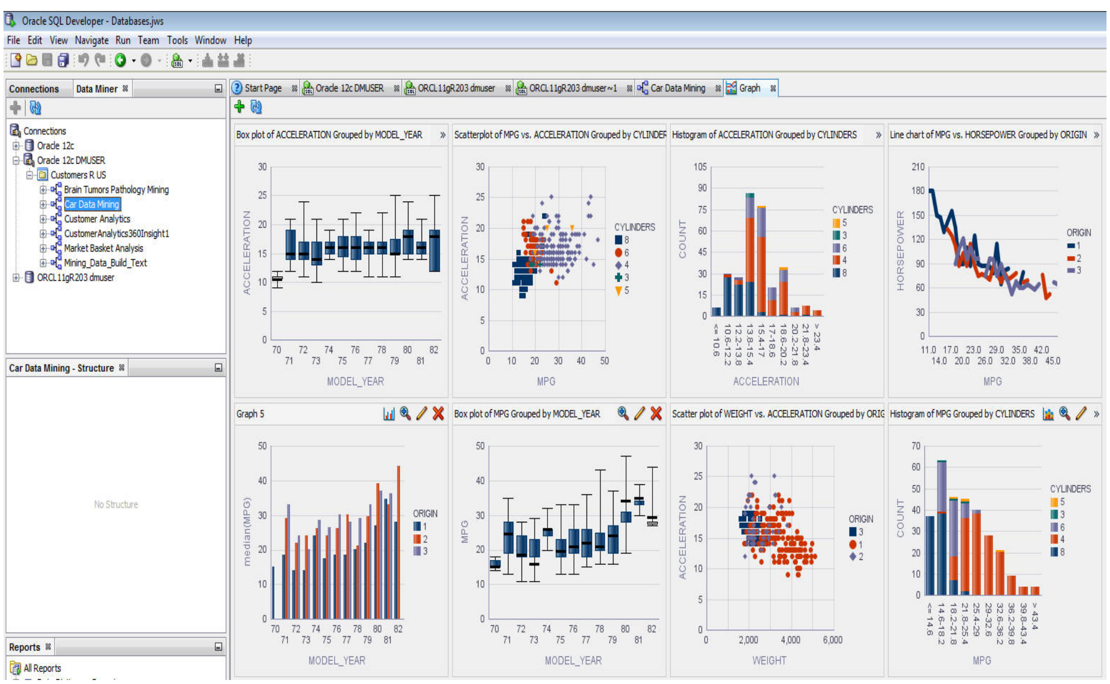

Figure 13: SQL Developer's Oracle Data Miner "workflow" GUI allows analysts to quickly experiment with data through point-and-click visuals and generates SQL scripts to support immediate deployment.

A3P in-database implementation of high performance, data mining and statistical algorithms allows you to discover patterns hidden in massive data volumes, discover new insights, make predictions and immediately transform raw data to actionable insights. By doing the mining in the database it avoids the hassles and slow processes and loss of security of moving data to separate analytical servers such as SAS, SPSS, and Matlab.

Data Mining

A3P data mining functions analyze data in its natural form to include star schema data and tables, views, transactional data, unstructured data (CLOBs), geospatial data and correlations including nested relationships that represent a 360-degree view of an object of interest. A3P can analyze more than a thousand input attributes and trillions of records. A3P provides in-database implementation of the following data mining algorithms:

Classification

- Naïve Bayes
- GLM (Generalized Linear Models)—Logistic Regression
- Decision Trees
- Support Vector Machines (SVM)

Regression

- GLM (Generalized Linear Models)—Multiple Regression
- Support Vector Machines (SVM)

Clustering

- Hierarchical K-Means Clustering
- Expectation Maximization Clustering
- Orthogonal Partitioning Clustering

Anomaly Detection

- 1-Class Support Vector Machines

Associations

- A Prior algorithm
- Attribute Importance
- Minimum Description Length
- SQL statistical functions for correlations e.g. Pearson's, Chi-square, etc.

Feature Creation and Feature Selection

- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)
- Non-Negative Matrix Factorization

Oracle Data Miner, provides data scientists and analysts with an easy to use workflow environment to explore data, build, evaluate and apply predictive models and save, share and deploy. Because analysts are working directly with the data, they can more rapidly explore various analytical strategies, build and evaluate simple and combinatorial analytical methodologies and if satisfied with the results, immediately deploy them as SQL scripts. The Oracle Data Miner workflow UI generates SQL scripts for analytical methodologies so analysts can rapidly move from analytical concept to enterprise-wide deployment—saving time and money. Analysts can also interact with visualization options to discover the best way to understand results.

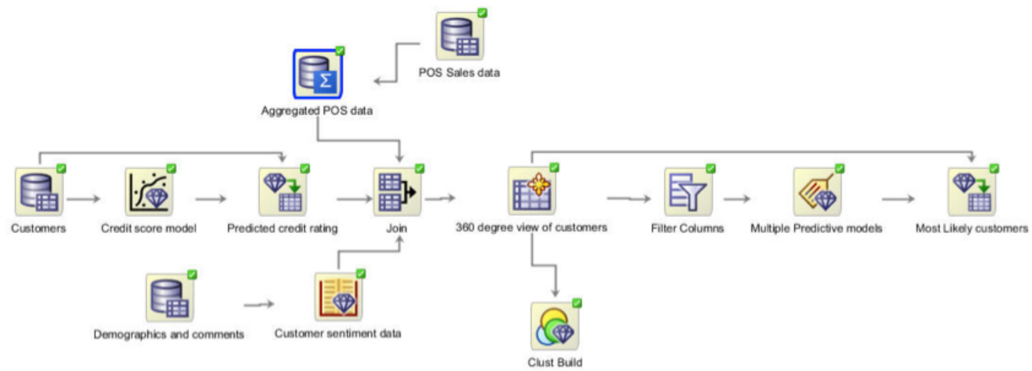

Figure 14: Data Miner UI allows users to create, inspect and deploy information flows to support their analysis

Oracle R Enterprise

Oracle R Enterprise, a component of the Oracle Advanced Analytics Option, makes the open source R statistical programming language and environment ready for the enterprise and Big Data. Designed for problems involving large amounts of data, R users can run R commands and scripts for statistical and graphical analyses on data stored in the Oracle Database. R users can develop, refine, and deploy R scripts that leverage the parallelism and scalability of the database to automate data analysis. Data analysts and data scientists can run R packages and develop and operationalize R scripts for analytical applications in one step—without having to learn SQL.

Oracle R Enterprise performs function pushdown for in-database execution of base R and popular R packages. Because it runs as an embedded component of Oracle Database, Oracle R Enterprise can run any R package either by function pushdown or via embedded R while the database manages the data served to the R engines.

Basic in-Database Statistical Functions

Included in A3P is a compelling array of statistical functions accessible from SQL. These include descriptive statistics, hypothesis testing, correlations analysis, test for distribution fit, cross tabs with Chi-square statistics, and analysis of variance (ANOVA). Oracle continues to expand its support for advanced analytics functionality in the database. The Oracle Advanced Analytics Option, a combination of Oracle Data Mining and Oracle R Enterprise, delivers predictive analytics, data mining, text mining, statistical analysis, advanced numerical computations and interactive graphics inside the database. It brings powerful computations to the database resulting in dramatic improvements in information discovery, scalability, security, and savings.

The statistical functions in the database can be used in a variety of ways, for example, users can call Oracle's DBMS_STAT_FUNCS packages to obtain basic count, mean, max, min and standard deviation information for their dataset; or users can determine the strength of relationships using hypothesis testing statistics such as a t-test, f-test or ANOVA. Users are able to not only complete a wide range of statistics, but also include these results in more advanced SQL queries and analytical pipelines. Features include:

- Descriptive statistics
- Hypothesis testing
- Correlations analysis (parametric and nonparametric)
- Ranking functions
- Cross Tabulations with Chi-square statistics
- Linear regression
- ANOVA
- Test Distribution fit
- Window Aggregate functions
- Statistical Aggregates
- LAG/LEAD functions
- Reporting aggregate functions
- SQL Pattern Matching

Pattern Matching with SQL

Recognizing patterns in a sequence of rows has been a capability that is widely desired, but not possible with SQL until now. There were many workarounds, but these were difficult to write, hard to understand, and inefficient to execute. A3P's Oracle Database 12c supports data pattern matching through standard SQL. As a simple case of pattern matching, consider the stock price chart illustrated below.

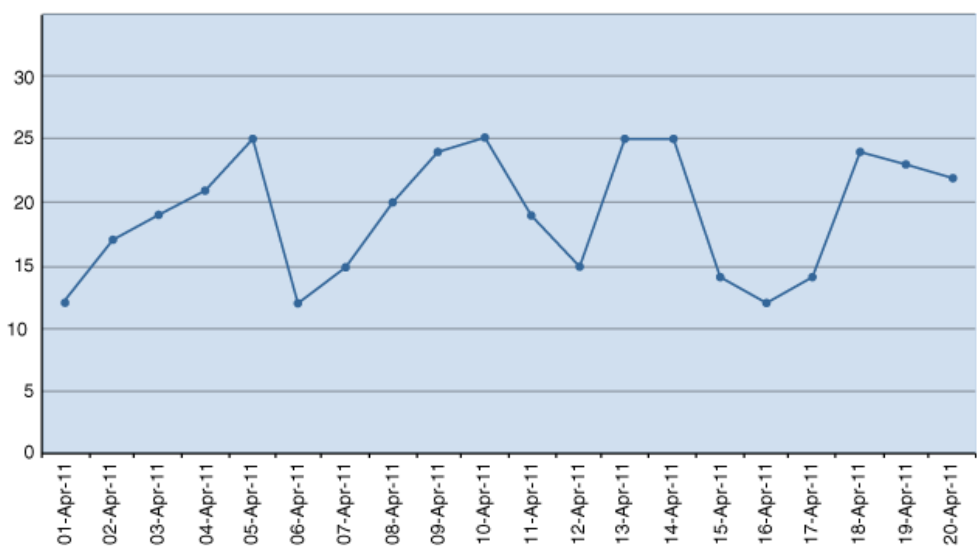

Figure 15: This chart of a closing stock's value over time readily shows patterns. Writing code to do the same has, until now, been very complicated and error prone.

Pattern matching can let you identify price patterns, such as V-shapes and W-shapes illustrated above, along with performing many types of calculations. For example, your calculations might include the count of observations or the average value on a downward or upward slope.

When combined with the other Agile Analytics in-database capabilities, SQL Patterns can serve to both detect patterns and to serve as additional input attributes to predictive modeling.

Big In-Memory Analytics

Fast data speeds, big data size and rapid analytic answers have been pushing the envelope of analytical tools and infrastructure. Proper infrastructure is essential to achieving successful analytics. Due to the extreme data volumes, algorithmic memory, compute core requirements and I/O requirements, a low cost infrastructure approach often does not support advanced analytics. For this reason A3P is not just a software solution but is a joint software and hardware approach to achieving success doing advanced analytics. For example, in-memory analytic workloads such as fast data, SMP algorithms, and graph analytics requires big memory and in-memory technologies.

Fast Data Analytics

A3P supports an emerging industry term called Fast Data Analytics. Fast Data Analytics occurs when data analysis has to be done very quickly (i.e. with extremely low latency) after the data arrives or becomes available. This is generally achieved through the use of in-memory analytics to include rules engines, event detections, alerts, workflows and data references. High volumes and high velocity data feeds obviously exacerbate the challenges to doing Fast Data Analytics.

A3P uses Oracle Event Processing (OEP) as a Fast Data platform for building applications to filter, correlate and process events in real-time so that downstream applications, service oriented architectures and event-driven architectures are driven by true, real-time intelligence.

A3P's OEP provides a complete top-down solution for designing, defining, developing and implementing Fast Data Analytics applications that not only meet business requirements, but perform to the highest levels of performance. Built on the latest industry-standards, including ANSI SQL, Java, Spring DM and OSGi, OEP provides an open architecture for sourcing, processing, and publishing complex events. To include both a visual development environment as well as standard Java-based tooling, OEP ensures that applications are developed and deployed without the hurdle of specialized training or unique skill-set investments.

M6-32 Big Memory Machine

The M6-32 Big Memory Machine is designed to meet big memory, multi-threaded analytic use cases. Used for large in-memory graph analytics, graph query and inferencing, or large in-memory vertical data access this system delivers the scale needed. It is a highly configurable server enclosure that is optimized for scale, reliability, availability, and serviceability.

The M6-32 can expand up to 32 processors (384 Cores), 32 TB of system memory, 32 internal disk storage devices, and 64 PCIe 3.0 I/O expansion slots. There is an extra large shared 48 MB level 3 cache. Unmatched system virtualization capabilities include dynamic domain hardware partitions, and logical domains. This allows for data scientists and analysts to work on multiple problems in different hardware partitions such as an HPC algorithm or graph algorithm.

The M6-32 provides the highest level of enterprise-class RAS features including redundancy and a hot-pluggable capability of most major components. Other key high-availability features include end-to-end ECC memory protection, and Oracle Fault Management. These capabilities enable server self-healing, and no single system failure can prohibit system recovery, which is a requirement for mission-critical computing.

Oracle Database In-Memory

Oracle Database In-Memory transparently accelerates analytics by orders of magnitude while simultaneously speeding up mixed-workload OLTP. With Oracle Database In-Memory, users get immediate answers to questions. Oracle Database In-Memory is fully integrated with the Oracle Database's renowned scale-up, scale-out, storage tiering, availability, and security technologies.

The technology optimizes analytics and mixed workload OLTP, delivering outstanding performance for transactions while simultaneously supporting real-time analytics and reports. This breakthrough capability is enabled by the "dual-format" architecture of storing data in memory in row and column format.

There are also many In-Memory optimizations, including In-Memory storage indexes, In-Memory compression and vector processing. When tied with large DRAM memory in the Big Memory Machine in A3P, the data scientist and analyst have 32 TB of memory to load vertical data sets, spatial data sets or graph sets into memory for rapid scans.

Oracle In-Memory Database performed a benchmark in 2013 where it achieved a record throughput scanning 341 Billion records per second on a single M6-32 machine. Hardware matters to Analytics!

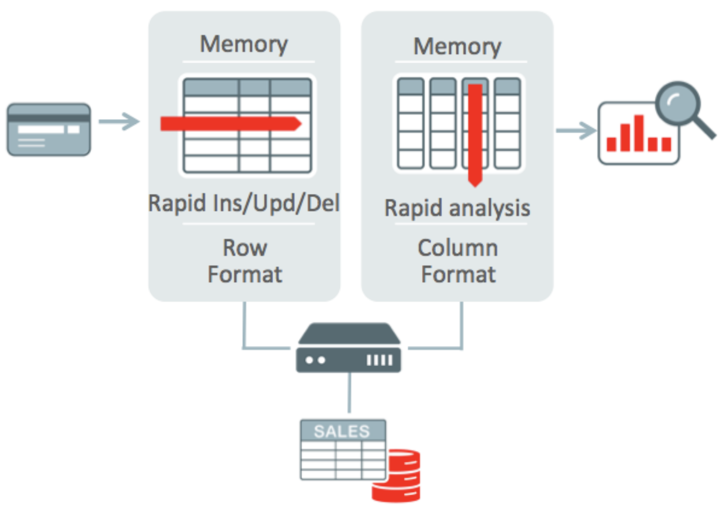

Figure 16: Oracle In-Memory Database transparently supports high-speed record access and high-speed columnar analytics.

Parallel Graph Analytics: PGX

PGX is a fast, parallel, in-memory graph analytic framework and takes advantage of the large memory in the A3P architecture. Analysts can load up their graphs into main-memory, run popular graph algorithms, explore their results, and export them back into the file system.

Loading graphs into memory – PGX is an *in-memory* graph analytic framework that needs to load the graph instance into main-memory before running analytic algorithms on the graph. PGX supports a few popular graph file formats for convenient data loading.

Running built-in graph algorithms – PGX provides built-in implementations of many popular graph algorithms. The user can easily apply these algorithms on their graph data sets by simply invoking the appropriate methods.

Running custom graph algorithms – PGX is also able to execute custom (i.e. user-provided) graph algorithms. Users can write graph algorithms in the Green-Marl domain specific language (DSL). Green-Marl is a DSL created specifically for graph data analysis that allows users to encode their algorithms in intuitive ways while leaving the parallel execution performance to the compiler.

Mutating Graphs – Complicated graph analyses often consist of multiple steps, where some of the steps require graph-mutating operations. For example, one may want to create an undirected version of the graph, to renumber the nodes in the graph, or remove repeated edges between nodes. PGX provides fast, parallel built-in implementations of such operations.

Browsing and exporting results – Once the analysis is finished, the users can browse the results of their analysis and export them into the file system.

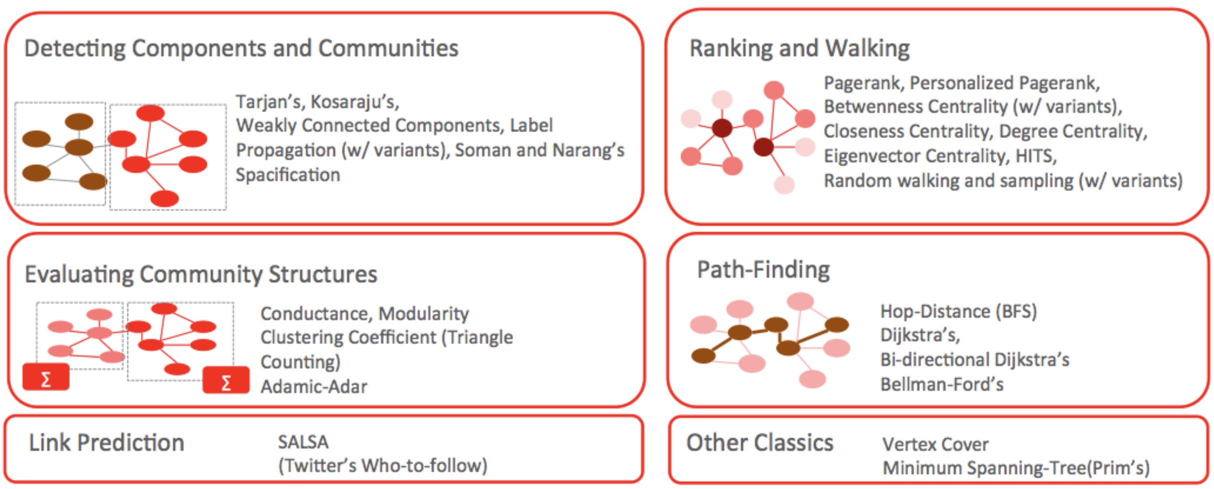

Figure 17: Built-in Algorithms supported by A3P's PGX Engine

PGX is a new product in early trial distribution. The current early adopter trial version is included in the A3P architecture.

Extreme Performance Hadoop and SQL

Too often technologists focus on processing at the expense of supporting an advanced analytic. Oracle is pioneering the fusion of processing, query and analysis across both the HDFS and SQL domains. This allows the analyst or data scientist to do analysis regardless of data type or data residence. This capability is provided in A3P in the Big Data Appliance and Exadata. It exists as embedded software functions known as Oracle Big Data SQL (see below).

Big Data Appliance (BDA)

Oracle Big Data Appliance is an open, multi-purpose engineered system for Hadoop and NoSQL processing. It runs a diverse set of workloads – from Hadoop-only workloads (MapReduce 2, Spark, Hive etc.) to interactive, all-encompassing interactive SQL queries using Oracle Big Data SQL. It is a highly optimized systems using Infiniband to reduce I/O bottlenecks during sort/shuffle processes. The BDA supports numerous analytical workloads such as graph analytics, natural language processing and fraud detection and is fully installed with Cloudera's CDH Hadoop distribution.

In addition to providing Oracle Big Data SQL and the full Cloudera Distribution for Hadoop (CDH) software platform, Big Data Appliance utilizes Oracle Big Data Connectors to simplify data integration and analytics. Big Data Connectors provide high-speed access to data in Hadoop from Oracle Exadata and Oracle Database – with data transfer rates on the order of 15 TB/hour.

Big Data Connectors also enable integrated, highly scalable analytics – providing native access to Hadoop data and parallel processing using Oracle R Distribution. Finally, Oracle XQuery for Hadoop facilitates standard XQuery operations to process and transform documents in various formats (JSON, XML, Avro and others), executing in parallel across the Hadoop cluster.

Exadata – The Big Data Database Machine

The Oracle Exadata Database Machine is engineered to be the highest performing and most available platform for running the Oracle Database. Exadata is a modern architecture featuring scale-out industry-standard database servers, scale-out intelligent storage servers, and an extremely high speed InfiniBand internal fabric that connects all servers and storage. Unique software algorithms in Exadata implement database intelligence in storage, PCI based flash, and InfiniBand networking to deliver higher performance and capacity at lower costs than other platforms. Exadata runs all types of database workloads including Online Transaction Processing (OLTP), Data Warehousing (DW) and consolidation of mixed workloads.

A single rack Exadata can have up to: 192 cores of compute, 168 cores of storage compute, 44.8TB of flash cache and 300 TB of uncompressed storage, and often over a 1PB compressed storage. Performance characteristics are also very impressive including up to: 2.6M transactions per second, 100 Gb/Sec IO bandwidth and 10-100x query performance improvements. Exadata is capable of handling relational, graph, spatial, xml, nosql/JSON, text, imagery and video data types and schemas.

Big Data SQL

Oracle Big Data SQL is an innovation that uses the Exadata and Oracle Big Data Appliance. It is a new architecture for SQL on Hadoop, seamlessly integrating data in Hadoop and NoSQL with data in Oracle Database. Using Oracle Big Data SQL, organizations can do the following:

- Combine data from Oracle Database, Hadoop and NoSQL in a single SQL query
- Query and analyze data in Hadoop and NoSQL
- Integrate big data analysis into existing applications and architectures
- Extend security and access policies from Oracle Database to data in Hadoop and NoSQL
- Maximize query performance on all data using Smart Scan
- ”Score” or apply existing Oracle Advanced Analytics SQL data mining models on the Big Data Appliance.

Note: this leverages the same “smart scan” technology available in Exadata Engineered Systems.

Oracle Big Data SQL radically simplifies integrating and operating in the big data domain through two powerful features: newly expanded External Tables and Smart Scan functionality on Hadoop.

Using new external table types, data in Hadooop and NoSQL is exposed to Oracle Database users. These tables, once defined, automatically discover Hive metadata including data location and data parsing requirements (i.e. SerDes and StorageHandlers). This enables SQL queries to access the data in its existing format leveraging native parsing constructs.

Oracle's unique Smart Scan capability brings the proven storage processing innovations of Oracle Exadata to Oracle Big Data Appliance. The biggest performance penalties in data processing are typically the result of excess data movement. Instead of sending all scanned data to the compute resources, Smart Scan on Hadoop radically minimizes data movement to the compute nodes by applying the following techniques at the storage level:

Data-local scans
- Hadoop data is read using native operators at the source

Column projection
- Only relevant columns are returned from the source

Predicate evaluation
- Only relevant rows are returned from the source

Complex function evaluation
- SQL operators on JSON and XML types applied at the source
- Model scoring and analytical operators evaluated at the source

Agile Analytic Sandbox

A3P embraces a concept called Sprint Analytics. Sprint Analytics supports the need for data scientists and analysts to try new tools, algorithms, and experimentation on data to improve prediction and mission results. In order to accomplish either success or retrial quickly, A3P includes an agile sandbox for creating development, test or an operational application hosting environment for the A3P users. This agility is delivered through the Oracle Virtual Compute Appliance (OVCA).

Oracle Virtual Compute Appliance is an integrated, "wire once", software-defined converged infrastructure system designed for rapid deployment of both x86 infrastructure hardware and application software. Whether running any Linux, Oracle Solaris, or Microsoft Windows, Oracle Virtual Compute Appliance supports a large range of OS versions hosted in a converged server, network, and storage environment to enable general purpose, and mission-critical application workloads. The OVCA has an Infiniband designed network, and comes complete with Linux and virtualization software.

The OVCA gives the A3P user the ability to rapidly spin up a template environment and share templates with others.   Templates might include test or development versions of items like: graph analytic algorithm, R statistical visualization, or an advanced analytical workflow

High Performance Storage

Analytic workload variety necessitates the need for performance-oriented storage. Data sets such as streaming video, imagery, spatial data, social media data sets and large documents all need a location to reside for pre-processing, staging, transformation and archiving.

The Oracle ZFS Storage Appliance is based on a integrated hardware and software architecture, including a highly intelligent multithreading storage OS that makes the most of modern enterprise hardware, enabling you to run multiple workloads and advanced data services without performance detriment.

Oracle' s ZS3-2

CPU: 4 x 8Core Intel Xeon
L1 Cache: 1024 GB DRAM
L2 SSD Based Cache: 12.8 TB
Max Capacity (uncompressed, RAW): 1.5 PB
Connectivity: 16Gbit FC, 10Gbit Ethernet, 40Gbit InfiniBand (same for both systems)
Supported Protocols: FC, iSCSI, iSER, SNS, Samba (Cifs), NFS, FTP , HTTP, WebDav

Oracle' s ZS3-4

CPU: 8 x 10 core Intel Xeon
L1 Cache: 2048 GB
L2 SSD Based Cache: 12.8 TB
Max Capacity (uncompressed, RAW): 3.4 PB
Connectivity: 16Gbit FC, 10Gbit Ethernet, 40Gbit InfinBand (same for both systems)
Supported Protocols: FC, iSCSI, iSER, SNS, Samba (Cifs), NFS, FTP , HTTP, WebDav

Unsurpassed Storage Efficiency	• Hybrid Columnar Compression, 4 levels of compression, de-duplication and thin provisioning offered as standard, reducing required storage
High Performance	• ZS3 Series Storage employs the latest multi-core CPUs and more than 10X more cache than competing systems • World record performance proven by public benchmarks
Low TCO	• Combination of symmetric multiprocessing, up to 1TB of Cache on each controller and the latest software release provides the best price/performance in the market
Streamlined Management	• DTrace Analytics provides a highly-visual interface for provisioning, management, real-time analysis and monitoring

Chart 2: Snapshot of the Oracle's ZS3 which provides standards-based, high performance storage for A3P

Its unique Hybrid Storage Pool architecture automatically caches data on dynamic random access memory (DRAM) or flash to provide optimal performance and exceptional efficiency, while ensuring that data remains safely stored on reliable and high capacity hard-disk drive (HDD) storage. This enables heavily accessed data to be served mostly from cache for extremely high performance without spindle speed limitations, while securing and storing data on cost-effective HDDs.

High-availability features such as active-active controller clustering for failover, a self-healing file system architecture that ensures end-to-end data integrity, and a rich set of enterprise-class data services make the Oracle ZFS Storage Appliance an ideal choice for enterprise storage. Furthermore, the aggressive price point makes it extremely attractive to employ in environments where extreme performance and superior efficiency are required. The ZFS-ZS3 storage is also connected into A3P via Infiniband for better I/O bandwidth between A3P elements.

Large Data Management

Data growth is a constant in today's analytical environments. Evolving data storage requirements can expose your analytical mission to performance risk if your solution doesn't give you the flexibility to keep up, no matter how fast data is growing. In some cases data sets in an analytical environment age, and are less useful over time. These data sets use precious disk storage resources but in many cases the data set can't be thrown away. In other cases, the data set might need to be retrieved later to compare and contrast with newer data looking for longer term trends.

The cost-effective, eco-efficient StorageTek SL3000 modular library system offers flexibility, scalability, and high availability in a midrange archive storage solution that solves these issues. Oracle's StorageTek SL3000 modular library system offers an innovative, eco-efficient approach to midrange storage with scalability up to 50 PB uncompressed to accommodate growth. The StorageTek SL3000 modular library system enables you to scale from 200 to just under 6,000 cartridge slots and from 1 to 56 tape drives in a footprint that grows linearly in a rack environment. The "real time" growth capability enables you to install physical capacity in advance and tap into it incrementally, with Capacity on Demand. You can grow at your own pace and pay only for the capacity you need. This approach delivers seamless, non-disruptive scalability with no physical parts to install.

A3P Configurations

A3P was designed to be agile and adaptable to specific customer analytical requirements. This agility allows for capabilities to be removed or modified to fit very specific needs. Most modifications are made by Oracle at our manufacturing facilities before they even arrive on site for installation.

Example of Configuration changes include:

- Additional memory in key hardware components
- Removal of a A3P element and scale out of other elements
- Additions of other A3P like capability such as Exalytics
- Additional software tools

Working with an Oracle account team it is possible to discuss analytical use cases and mission profiles that might lead to recommendations to add or remove elements of A3P. For example, a particular customer might not need In-Memory Graph capability but need expanded spatial analytical capability. This type of requirement would lead to a modification before the A3P installation.

A3P Expansion

As analysts start to realize the positive effect of the impact to mission from A3P, expansion might become necessary. Expansion requirements will also be driven by data growth, increasing user demands and future algorithm development.

The agility in A3P also allows for scale out. Any A3P ingredient can be expanded individually or as a set. For example, the Big Data Appliance can expand to 8 racks before another switch has to be added. Likewise, Exadata can expand to 8 racks with the same effect. Storage, Archive and Compute Appliances are also expandable without the need to add another switch.

Figure 18: A3P’s agility allows you to customize your analytics. Pay for what you need for and grow in-place when needed.

Summary

The word "analytics" can mean many things. It can include everything from basic word search to complex and multi-stage predictive analytics. Valued Intelligence results from the correlations and insights amongst data of different types that come from different sources. Frequently, different data types are managed in un-integrated and separated data stores because the analytics performed on the data vary. Information fragmentation occurs because data is stored separately without ways to efficiently combine and correlate the data. This hinders the desired analysis or potentially renders it useless or unachievable.

What is needed is an efficient and robust platform that can manage and process many data types in a secure, agile and interoperable way. The Agile Advanced Analytic Platform (A3P) serves as this platform. A3P is a fast, fully complete data management and data analysis environment that supports full-spectrum analytics from data exploration & discovery through advanced and predictive analytics and deployment. It represents the combination of software and hardware that has been professionally engineered, optimized, developed and deployed to support the (Big Data) analytic challenges faced today.

The characteristics of A3P include the following:

- Agile: Self-service analytics that is easy to adjust
    - Sprint analytics: Fail fast/Succeed fast
    - Disposable analytics: Low effort, time & costs for rapid iterations
- Advanced: Complete yet simple solution that fully maintains consistent security, manageability, scalability and interoperability

A3P represents the compilation of many systems working together to perform the analytic functions. The entire 'system of systems' is professionally engineered, certified and supported. The result is that A3P empowers analysts to explore, test and evaluate with ease and speed. Oracle's platform and process re-balances the time and energy of your analysts to where it is most valuable.

A3P combines enterprise-grade capabilities with the latest advances in analytics. For example A3P supports R statistical programming language for advanced statistical analysis. A3P also supports the latest in developments for graph analysis in Parallel Graph AnalytiX (PGX). In the Appendix below, you will find a detailed list of algorithms and capabilities supported.

As A3P is "agile" as it supports dynamic expansion and adaptation to meet your customized and personalized analytical needs. It does all this with a consistent alignment to standards to ensure integration, interoperability and mitigates technology lock-in.

Above all, A3P is about driving efficiencies. A3P reduces the time-to-value, reduces the costs required to design, build, deploy and maintain the necessary infrastructure required to do today's advanced analytics.

Appendix A: Detailed List of A3P Analytical Algorithms and Capabilities

Graph Data Management and Analytics
Linked Data, RDF, Property Graphs
Multi-Trillion+ Node Graphs
SPARQL, SPARQL/SQL, GeoSPARQL
W3C OWL/RDF Standards
Spatial Analytics on Nodes
Link and Node Level Label Security Enforcement
In Memory PGX Analytic Engine (GreenMarl Compiled Algorithms)
- Spacification, Tarjan's, Weakly Connected Components
- Propagation with Variants, Soman/Narang
- Conductance, Modularity, Clustering, Coefficient, Triangle Counting
- Link Prediction, Pagerank, Betweeness, Centrality, Closeness
- Degree Centrality, Eigenvector Centrality, HITS
- Random Walking and Sampline, Hop Distance, Dijkstra
- Bellman's-Ford, Vertex Cover, Spanning-Tree
- Pattern Matching, Isomorphism

Spatial Temporal Analytics
Full OGC Compliance and Support
Points, Lines, Polygons, Point Clouds
Imagery, LIDAR
Network Modeling
3D Whole Earth Modeling
Coordinate Transforms
Geocoding
Within Distance
Complex Spatial Search
Shortest Path, Least Risk Weight Path
Link Prediction, Pagerank, Betweeness, Centrality, Closeness
Edge Detection
Virtual Mosiac
Topology Modeling and Linear Referencing
Ontology Referenced Query

In-Database Predictive Advanced Analytics
Classification
Clustering
Anomaly Detection
Regression
Prediction
Attribute Importance
Feature Extraction
Associations
Decision Trees
K-Means
Support Vector Machine
Neural Networks
SQL Patterns
Enterprise R Algorithms/Statistics

Big Data Hadoop Processing and Analytics
Full CDH Install including Support
- Hbase, Cassandra, Accumulo
Oracle NoSQL
Oracle Big Data SQL (SQL Engine on Big Data)
Oracle Loaders for Hadoop
Enterprise R
Oracle NoSQL GRAPH
Map Reduce Processing

Analyst Driven Answers and Semantic Query
Semantic Layer Discovery
Map to any source
Interactive Textural and Point and Click Analysis
Numerous Graph and Visualization Options
Pixel Perfect Reporting Options
Mobile Version

Extreme 12c Analysis and Processing
Big Data/Sharded Oracle 12c
Cell Offload SQL Smart Scan
Scans in Flash
In Memory (Columnar) Extreme Performance
Full JSON NoSQL Schema-less support
Accelerated Graph and Spatial
Complex Joining and SQL
SQL Pattern detection
Text Mining
XML and Document Management
Imagery and Large Object Management
Massive Data Compression
Accelerated Performance during joint ingest and query

Big Data Discovery and Visualization
Endeca Big Data Discovery
Direct Auto Detection and Classification
Discover structures, patterns and relations
Extensive Visualization and Graphing
Analyze Data without moving it
Data Set summaries and quick reference gauges
Complex discovery without spending time in ETL
Support Data Triage and Understanding from new data sets

Agile Advanced Analytic Platform
October 2014
Author: Mark Tatum, David Knox

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com