Salesforce Technology White Paper

Wave: The New Architecture of Business Intelligence

Understand the technology behind the analytics platform that makes every organization smarter.

salesforce analytics cloud

Executive Summary

Organizations have an unprecedented opportunity to learn more about their businesses, markets, and customers from the explosion of data being generated from a wealth of sources – from sensors to apps, software to websites. The need to explore, analyze, and gain insights from this data has never been more pressing. With legacy business intelligence and analytics tools, the underlying technology is based on structured, relational databases. Relational databases lack the agility, speed, and true insights necessary to transform data into value.

Salesforce has revolutionized business intelligence technology by taking an innovative approach to analytics, one that combines a non-relational approach to heterogeneous data forms and types with a search-based query engine, a sticky and engaging interface and mobile-friendly experience. Salesforce's open and agile Wave Platform is built to enable business users to explore data in a fast, self-service, agile way – without dependency on data scientists, cumbersome data warehouse schemas, and slow, resource-intensive IT infrastructure.

Today's data requires a new era of business intelligence.

> " The data revolution has created tremendous demands on business intelligence. In 2013, it was estimated that 90% of the world's data had been created in the past 12 months. "

Business intelligence (BI) systems have played a role in business decision-making for five decades. In that time BI and analytics tools have grown more complex, powerful, and visual. They have also grown in importance to organizations – and now require large, costly infrastructures to fuel BI needs.

The data revolution has created tremendous demands on business intelligence. In 2013, it was estimated that 90% of the world's data had been created in the past 12 months.[1] However, less than 5% of the world's useful metadata has been analyzed, according to IDC.[2] These are astounding quantifications of the opportunity that exists with big data. Organizations are generating and accessing vast amounts of data, more than ever before, coming from a multitude of sources: log data, location data, behavioral data, sensor data. This flood of data is not only voluminous but comes in many forms, from unstructured to structured and every variation in between. Harnessing this explosion of data is key to a company's competitive advantage. Yet few companies have to date been able to truly exploit this data as a strategic asset.

A report by Accenture found that only 20% of enterprises they studied were using analytics across the organization – but only when the entire enterprise is relying on analytics for information and insight about the past, present and future can data be considered a strategic asset for that organization.[3]

Meanwhile, the way business users are wired to explore and investigate business problems and questions has completely changed in the past two decades. A Columbia University study found a phenomenon that researchers dub "The Google Effect" has altered people's way of accessing information, making us

mentally dependent on instant access to computerized information.[4] Business users have become adept at searching for answers in a free-form way. That often means that they start with a question, but quickly discover that it is the wrong question or find context that allows them to narrow the scope of the question or investigate it from a different angle.

Legacy BI restricts speed, agility, and its use is limited to IT and analysts.

Interestingly, though the complexity of BI tools has evolved over the years, the fundamental architectural approach to BI and analytics largely remains static. When an enterprise sets out to explore a problem or question, the BI team addresses the query by building a relational database or data warehouse. True to Codd's original rules of relational databases published back in the early 1970s, data warehouses contain relational databases that add and store data in tables of rows and columns, with each piece of information captured as a value in the table. Relationships between tables develop into snowflake or star-shaped schemas. Each new addition of data adds new rows and new dimensions to the schema. Once the structure has been created, it is rigid and prohibits new data from being added to it; adding new data requires building a new schema from the ground up.

The relational database model continues to work well for many types of applications, namely transactional operations with highly structured data. However, sweeping changes in technology, data volume and variety, and dynamic markets over the past decade have created a chasm between legacy business intelligence and analytics capabilities – based on traditional relational database design – and the needs of businesses today.

Analyst Cindi Howson, author of *Successful Business Intelligence,* found that nearly 76% business users were somewhat or largely dissatisfied with how BI was working

The deficiencies of legacy BI in meeting businesses' needs stem largely from the architecture of BI solutions as well as the internal processes required to analyze a business problem with the current BI model. Typically, business users or managers approach the BI team with a problem they want to solve or question they want to investigate. The BI team then decides what type of data to include; locates, collects, scrubs, and structures that data into a snowflake or star schema; and builds out the interface to allow business users to query the data. For more complex queries and datasets, the data architects will build an Online Analytical Processing (OLAP) model that creates multiple dimensions or a cube of data, which allows users to drill down and pivot

The relational database model presents a number of challenges in today's business environment:

User challenges

- **The model reduces agility.** The waterfall nature of traditional BI development acts as an impediment to uncovering new ways of doing business, handicaps the ability for team members to constructively challenge current process, and keeps the personnel who have the most access to customers and the market from asking their own questions and exploring and modeling their own innovative ideas for improving the business.

- **It does not represent the interactive way users explore information.** Traditional BI projects do not allow the agility to refine the question or add new data for additional context. Users ask a question and then wait weeks or months for an answer; if they discover the original question was the wrong one, the schema build-out must start all over again. Legacy BI additionally pre-aggregates the data, which limits insights.

- **It forces compromise.** A typical BI deployment strikes a balance between anticipated queries and performance. The compromise leads to dissatisfaction. For example, data is typically "rolled up" to a higher grain to provide acceptable query performance, but this prevents users from answering second- or third-order questions. They then must go back to IT or use different tooling to answer their questions.

Business challenges

- **The model does not operate at the speed of business.** Building out a BI schema can often take weeks or months, depending on its size and complexity – and that does not account for time internal customers must wait in the queue for BI or IT resources to become available. At best, this delay represents slow time to value for BI investments; at worst, it puts severe limitations on the business, which is often depending on insights from BI to move forward with an initiative or decision and may be threatened competitively by failure to act rapidly.

- **It is resource-intensive.** The current way of developing BI tools requires an army of experts – IT architects, business analysts, and data scientists, not to mention project managers – to manage the BI needs of an enterprise. These teams are often highly compensated and in great demand because organizations are so dependent on business intelligence.

- **It does not exploit data as an asset for the whole company.** The limited, resource-intensive and cumbersome nature of BI data warehouses the way they are built today makes it prohibitive for widespread adoption and usage across the enterprise, and end users cannot nimbly engage in data-driven decision-making.

for their needs.[5] Companies that use legacy BI collect a proliferation of transactional reports that give limited views of data at a given point in time. Many of these thousands of reports are virtually useless. IT leaders are revisiting the value of these reports and recognized that self-service data discovery is more efficient and valuable to users.

Yet in spite of these drawbacks, enterprises have made considerable financial and resource investments in procuring and implementing expensive, legacy business intelligence analytics solutions – because in the past, they were the predominant solutions available. Often those responsible for BI strategy in an organization are reluctant to consider alternative solutions because of the emotional and accounting burden of capital costs. This reluctance may be reinforced by the realization that solutions have not always delivered what vendors promised they would, and that BI investments have failed to gain widespread internal adoption.

Turning business intelligence on its head for fast, agile, end-user exploration.

A number of emerging solutions in recent years have attempted to address the challenges outlined above. Many of them, however, have continued to rely at least partially on the same architecture and technology approach that have caused the challenges in the first place. For example, one innovation that has emerged is the use of columnar or in-memory databases, adopted by BI vendors over the past decade. While they moved the needle forward, they were still hampered by the relational model and its associated limitations.

But Salesforce has developed and unveiled an analytics platform that turns business intelligence on its head. The

Wave Platform dismisses most of the pre-conceived principles of data warehousing and database design, instead taking a "Google-inspired" approach to business analytics. It combines a proprietary, non-relational data store, search-based query engine, advanced compression algorithms, columnar in-memory computing, and high-speed visualization engine.

The resulting analytics platform embraces the complexity of heterogeneous data, the fluidity of questions and problems business users are trying to solve, and the end user's proclivity for exploring data with agility – all without limitations on time and information. Wave was architected from the ground up to allow enterprises to quickly find value in data. The platform was built first for a native mobile app, allowing users to rapidly find answers and take action using their smartphones.

Technology principles of the Wave Platform.

A number of key principles differentiate the Wave Platform from other business intelligence solutions on the market today. To understand how Wave works, it is important to understand the nine underlying principles of the platform:

1. Agility
Wave doesn't discriminate among data types. It on-boards data by accommodating any data structure, type, or source, and making it available immediately – without a lengthy ETL process.

2. Search-based exploration
Data is searched using an inverted index – similar to the Google search engine – allowing for query results within seconds.

3. Columnar, in-memory aggregation
Quantitative data is spun up and queried in a columnar store in RAM across Salesforce's cloud instead of in the row structure of a relational database on disk.

4. Speed
Heavy compression, optimization algorithms, parallel processing and other strategies allow sub-second and highly efficient queries on extremely large datasets.

5. Actionability
Once a user has discovered an insight or made an important decision, they can instantly take the next best action right from within Wave.

6. Interactivity
Fast, intuitive, visualization promotes user adoption and contextual understanding – bringing true self-service analytics to every business user.

7. Mobile-first design
Wave was designed with smartphones in mind, enabling salespeople and other business users to access information easily from anywhere, in meetings, with customers and on the go – further promoting user adoption. The platform actually enables data creation right from the mobile device: for example, the ability to ingest an Excel/CSV file using a smartphone and immediately explore the data, and even build an analytical dashboard on the fly.

8. Open, scalable cloud platform
Wave is an open, scalable, and extensible platform. With easy-to-use APIs, Wave's architecture enables deep relationships with third-party tools and complements existing BI solutions. It is also deeply integrated with Salesforce so you can see your Sales Cloud and Service Cloud data like never before, collaborate, and take action from within Salesforce.

9. Security
The Wave Platform inherits Salesforce's proven, multi-layered approach to data availability, privacy and security, with the additional benefit that data on the Salesforce platform need not move outside of Salesforce servers to be available for analytics.

Salesforce also offers Wave Apps, a suite of sophisticated analytic applications built on the Wave Platform. The first two offerings – Sales Wave and Service Wave – are end-to-end apps that bring the power of Wave Analytics to Sales Cloud and Service Cloud. They deliver a new level of insight directly to any device by bringing all crucial sales and service KPIs together in one place. These apps help managers quickly gain organizational visibility, track team performance, and uncover new opportunities to sell and service smarter.

"Wave uses a search-based query engine that is similar in its design to modern, commercial search engines such as Google and Bing."

The sections below provide a detailed explanation of each of these eight principles, along with a summary of how each principle uniquely benefits enterprises from a business and technology perspective.

Agility: Ingest, index, and begin analyzing data immediately.

The traditional way of designing a data warehouse is a waterfall approach to gather requirements, figure out relationships, pre-determine the data structure, scrub the data set, add a semantic layer to the data – and finally to ingest the data. Depending on the size and complexity of the dataset, the process can take many months to complete.

The Wave Analytics Platform reverses this process. It treats data ingestion not as an exercise in "extract, transform, and load" (or ETL, the traditional way of ingesting data into a database), but as ELT – data is extracted, loaded, indexed, and made available immediately for analysis or additional transformations.

Wave accommodates heterogeneous data of any form, type, or source. The platform enables immediate search and exploration of the raw data, allowing the analytics tool to detect patterns and relationships instead of requiring a lengthy data normalization process. Data is loaded into a proprietary, non-relational store, with a dynamic, horizontally scalable key-value pair approach. The workflow engine applies small, inline transformations upon ingestion – pruning, filtering, partitioning, and augmenting – but largely stores the data in its native form. The benefit is that you gain rapid access to your data, and can immediately determine in what ways the data is relevant to your needs – without weeks or months of investment in "cleaning up" data before exploring it. Once you determine the applications of the data, you can specify more transformations to make it easier and richer for end-users to consume.

This makes self-service data exploration rapid and iterative, putting the ability to understand relationships between

data in the hands of the end users and allowing enterprises to dramatically shorten the path to innovation. Users can access a rich dataset with meaningful attributes and context, which might have otherwise been limited in the process of normalizing and fitting data into a pre-ordained structure. Then, users can easily connect that data with other types and forms of data – combining information from their CRM with data from their ERP platform, or joining values from spreadsheets with machine-generated location data – for new opportunities for revenue, investigation and exploration.

Search-based exploration: Process queries of large, heterogeneous datasets in seconds.

Wave uses a search-based query engine that is similar in its design to modern, commercial search engines such as Google and Bing. Data is ingested and stored as key-value pairs in a non-relational inverted index, permitting variable numbers of dimensions and attributes for data and the accommodation of text strings and unstructured data, as well as data sets with variable levels of completeness or characterization. Unlike traditional relational databases, key-value pairs only store non-empty data values, which, in the case of really sparse data, adds to data storage efficiency and speed. Wave's query engine is highly optimized, using proprietary techniques such as differential encoding, vector encoding, and incremental encoding to compress data and make queries on compressed data as fast and efficiently as possible.

Columnar, in-memory aggregation and calculation: Gain incredible speed by dramatically optimizing the query.

The Wave Platform queries quantitative data in an in-memory columnar store, rather than against rows and tables on disk, optimizing the size of the dataset and the query process itself, as the engine does not need to process rows of data and can avoid reading columns not related to a query.

Speed: Get instant answers from free-form navigation and exploration.

The benefit of search-based exploration is, quite simply, speed. Performance of a query depends on a combination of data structure and query strategy, and Wave brings both together. With relational databases, a query on a large dataset requires the analytics engine to process each value in each row of a very large set of data. Business analytics users often share the experience of starting a query and going to fetch a cup of coffee while waiting for the process to finish,

which can sometimes take 30 minutes to an hour or more. With the inverted index, Wave permits datasets equivalent to up to a billion rows to be queried in seconds.

In addition to the inverted index, Wave combines other strategies to achieve unparalleled speed. For one thing, it heavily compresses data. Using a variety of proprietary compression techniques, the workflow engine that ingests and transforms data is able to store it at a significantly smaller size than the original source data – at compression ratios that can exceed 90%. Wave's specialized compression and the algorithms that operate on a compressed data platform perform significantly better on modern CPUs, resulting in more efficient computation.

Additionally, the Wave query engine relies on the massively parallel processing architecture of Salesforce's cloud – with more than 40,000 available processors in the cloud, further powering computations at near-instantaneous speed. However, because of the heavy compression, Wave uses as few resources as possible to reach answers to queries.

Actionability: take the next logical step right within Wave.

The normal workflow within legacy business intelligence tools is to find an answer, and then switch to a different tool to share the insight with your team or direct their next steps. But Wave Analytics comes with built-in tools called Wave Actions that allow you to quickly go from question to answer to action without logging into a separate solution. Now when you find an answer, you can immediately create a task, update a record, log a call, and more – and all without waiting for legacy technology or data analysts.

"According to a much-cited report by Gartner, 70% to 80% of enterprise business intelligence projects fail to meet their business objectives "

Interactivity: encourage adoption and exploration with a powerful visualization layer.

The secret to user adoption and stickiness is the ability to visualize data and customize views in ways that are intuitive and interactive and that encourage faceting and exploration. The Wave visualization layer was designed with gamification principles in mind, using simple, beautiful graphics that immediately engage users and fire up their curiosity to play and discover.

Wave draws SVG graphics within the browser, and uses an animation engine when the graphics shift as users filter and change views, resulting in fast-rendering graphics

that connect the relationships between views of data – and demand few resources from the user's device or the server. Images animate contextually, demonstrating how relationships change as users interact with the data. Users can easily combine datasets and create custom dashboards to explore questions from different angles and using different combinations of information. And Wave is designed so users can immediately turn answers into action and take the next logical workflow steps.

Mobile-first: put information in users' hands, literally.

More than 328 million people are expected to use smartphones at work globally by 2017.[6] In particular, analysts have strongly advocated smartphone enablement for sales teams to drive adoption of sales force automation software investments.

Legacy analytics tools have been largely designed to spit out dense spreadsheets or 8.5x11 reports. The Wave Analytics Platform is natively built for smartphones. Users can visualize and interact with data, run queries, and develop dashboards as rapidly and intuitively as they can on a PC, with equally rapid response times. Salesforce does not leave data locally on mobile devices or laptops, so enterprises do not need to be concerned about data theft when a device or laptop is stolen or lost.

Open, scalable cloud platform: gain value faster and get more from BI investments.

Salesforce has built Wave as an open, scalable platform that allows flexibility and agility by interfacing easily with third-party tools. With the Wave Platform, enterprises can complement and supplement existing BI solutions and build new ones as custom needs arise, so they can derive more value from their existing BI and IT investments. A complete, rich set of APIs invites third parties and enterprises to build their own analytics tools and applications atop Wave. Further, Salesforce has worked closely with third-party vendors – including ETL vendors such as Informatica – to build interfaces with market-leading BI and data solutions right out of the gate for seamless data portability.

The Wave Platform is an integral part of Salesforce's Customer Success Platform. The inherent benefit of being part of Salesforce's multi-tenant cloud is immense scalability – the ability for enterprises to scale up and down in their data analytics usage without investing in hardware and on-site IT resources – as well as high performance and

speed. Native integration with Salesforce Sales Cloud, Service Cloud, App Cloud, and Community Cloud enables enterprises to rapidly achieve time to value by layering on additional data to begin understanding customers, sales performance, and markets on a deeper level with data from any other source.

Security: trust is inherent with Salesforce's secure cloud.

With more than 100,000 customers now using Salesforce products, the company has made trust, the privacy and security of customer data, its top priority. The company's world-class privacy program and security infrastructure extends to the Wave Platform, including a multi-layered approach to protecting, monitoring, and staying ahead of security challenges.

Salesforce security and privacy measures include a robust information security governance model, security coding and testing through each layer of development, investment in network defense, and comprehensive physical and operational security at all Salesforce facilities.

A unique approach to BI delivers faster time to value.

Salesforce has combined a variety of differentiated approaches – a non-relational inverted index data store, a fast and powerful query engine, intuitive and engaging visualization, mobile-first technology, and the secure, scalable, high-performance power

of the cloud – to deliver an open platform and agile, self-service solution for enterprise business intelligence.

Recognizing that many organizations have invested heavily in BI technology, Salesforce has created Wave to supplement existing solutions and interface seamlessly with third-party data tools, and enable enterprises to customize their own analytics applications with ease. This new BI analytics platform supports organizations' quest to achieve faster time to value with BI solutions. Wave additionally supports a unified data governance strategy, fosters enterprise-wide adoption, and frees IT teams up from resource-intensive and low-value data retrieval and preparation tasks to focus on more strategic initiatives.

As the world enters the third phase of computing – from today's systems of engagement to tomorrow's systems of intelligence – the open Wave Platform positions Salesforce and its partners to continuously innovate and add layers of intelligence to help business users gain insights even faster, through automated analytics. Wave is the foundation for a future of true business intelligence: fast, agile, insightful, and able to not only capture where the business and customers have been, but smart enough to see where they're going next.

Schedule a demo of Salesforce Analytics Cloud.

To learn more about Salesforce Wave Analytics and schedule a demo, contact your Salesforce Account Executive.

References

[1] Dragland, Åse. "Big Data – for better or for worse." SINTEF, May 22, 2013.
http://www.sintef.no/home/corporate-news/Big-Data--for-better-or-worse

[2] The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things by IDC, April 2014.

[3] Harris, Jeanne G. How to turn data into a strategic asset. Accenture, June 2010.
http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_Outlook_Turn_data_into_strategic_asset_analytics.pdf

[4] Sparrow, Betsy, Jenny Liu, Daniel M. Wegner. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips." Science, August 2011. http://www.sciencemag.org/content/333/6043/776.abstract

[5] Howson, Cindi. Successful Business Intelligence: Unlock the Value of BI & Big Data. Mcgraw-Hill Osborne Media, 2013.

[6] Fried, Ina. "As More Workers Take Their Smartphones to Work, Business Use of Mobile Apps Rises Sharply." Re/Code, February 12, 2014.
http://recode.net/2014/02/12/business-use-of-mobile-apps-continues-to-rise-while-iphone-gains-ground-in-q4/