



WHITE PAPER

# Platform for Cloud BI and Analytics

## A Birst Technical White Paper

## Table of Contents

Introduction . . . . .	4
The Case for Business Analytics . . . . .	4
Product Architecture . . . . .	5
Design Philosophy . . . . .	5
Tiered Architecture . . . . .	5
Virtualization . . . . .	5
Product Components . . . . .	6
Data Connectors . . . . .	6
Birst Connect. . . . .	6
ETL Engine . . . . .	6
Analytical Data Store. . . . .	6
ROLAP Engine . . . . .	6
Multi-Layer Cache . . . . .	6
Predictive Analytics Engine . . . . .	6
Platform Functionality . . . . .	7
Data Extraction and Upload . . . . .	7
Data Transformation . . . . .	7
Data Modeling . . . . .	7
Data Warehouse Automation . . . . .	7
Live Access . . . . .	8
Managed Data Mashups . . . . .	8
Performance and Scalability . . . . .	9
User Experience . . . . .	10
Interactive Dashboards . . . . .	10
Ad Hoc Analysis, Data Visualization and Discovery . . . . .	10
Pixel-Perfect Banded Report Writing. . . . .	11
Report Export, Scheduling and Alerts . . . . .	11

## Table of Contents (Cont.)

Custom Expressions .....	11
Birst for the iPad .....	11
Implementation .....	12
Development Modes .....	12
Deployment .....	13
Birst Cloud .....	13
Appliance .....	13
Amazon Redshift .....	14
Integration .....	14
Single Sign On & Embedding .....	14
Web Services .....	14
Conclusion .....	14

## Introduction

Birst provides a complete Business Analytics platform that is purpose-built for agility, combining the economic benefits of a multi-tenant Cloud solution with the freedom to choose where you host your data, in your private cloud or in the Birst Cloud. Traditional Business Intelligence (BI) solutions require expensive software licenses, integration of multiple different products, long implementation cycles, and costly maintenance. These types of solutions often also yield disappointing results, since high costs and lacking flexibility severely limits the solution's ability to serve ever-evolving business needs effectively.

The Birst platform was designed and developed from the ground up to satisfy the requirements of all types of organizations, small and large. Thanks to patent-pending technology for automating time- and resource-intensive tasks in the BI application lifecycle, Birst provides an unprecedentedly low total cost of ownership (TCO) and fast time to value. Available in a single product, Birst offers complete Business Intelligence and analytics functionality for highly interactive dashboards, pixel-perfect banded report writing, powerful OLAP-style ad hoc analysis and data visualization.

The purpose of this paper, intended for an IT, developer and administrative audience, is to provide a technical overview of the Birst platform, including its design philosophy, architecture, components and capabilities.

## The Case for Business Analytics

Demand for business analytics is strong and growing. For the past several years, Business Analytics has been repeatedly named as the top priority by Chief Information Officers (CIO) in Gartner's annual CIO survey. Putting the right information in the right hands at the right time is an imperative for companies of all sizes across all industries. Yet the promise, of enabling fact-based decision making, remains elusive for many organizations. Too often BI implementations fail to deliver value or adapt quickly to every changing and growing business requirements.

A typical BI implementation using traditional on-premise tools requires the integration of software and hardware from multiple vendors. The process of converting raw data into useful and actionable insights suitable for business decision making involves a significant time and resource investment throughout the implementation process. The acquisition, implementation and maintenance costs quickly add up to levels where BI becomes feasible only for those few companies who have the requisite budget and resources. According to IDC, between 60% and 86% of the total cost of a BI solution over three years is staffing related<sup>1</sup>.

Unlike traditional and Open Source BI platforms, Birst was designed to radically reduce overall costs by providing a fully integrated product that automates much of the manual configuration required by other tools. Birst was developed as an integrated product by one development team with one goal: enabling enterprise business analytics solution with more agility at a lower cost. With a Software-as-a-Service (SaaS) solution like Birst Cloud, hardware and maintenance costs are amortized across many clients and thereby greatly reduced. Through the use of automation, Birst significantly lowers the staffing costs required for a BI project for both the initial build and ongoing development and maintenance of the solution.

<sup>1</sup> IDC. "Demonstrating Business Value: Selling to Your C-Level Executives." Three Year Server TCO. Based on more than 300 interviews conducted across numerous platforms, presented in composite form. April 1, 2007

## Product Architecture

### Design Philosophy

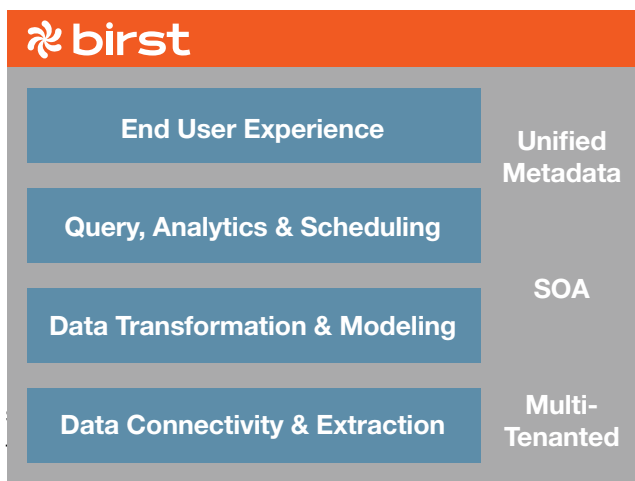
The Birst platform was designed and developed from the ground up to satisfy the varying needs and requirements of organizations of all types and sizes. All of the core components were developed in house and are tightly integrated via a common metadata layer.

Two primary design principles guided and continue to guide the development of the Birst platform:

1. **Automate as many tasks as possible required in the development, deployment and ongoing maintenance of a BI solution.** Through patent-pending technology, Birst automates many of the steps and operations that are manual and repetitive in nature and do not require user direction and control. As a result, Birst allows developers, administrators and end users to focus on executing more valuable tasks.
2. **Provide a superior user experience focused on self-service capabilities without sacrificing functionality and scalability.** Birst provides deep BI and analytics functionality to a broad range of users while reducing users' dependency on IT. Birst shields average users from feature complexity, while providing power users with access to the full breadth and depth of functionality of the product.

### Tiered Architecture

The Birst platform is built on a modern, multi-tenant, service-oriented architecture using the latest web



standards. The architecture comprises four primary tiers that are fully integrated based on a unified metadata model.

- **Data Connectivity and Extraction**  
This tier provides numerous connectors for scheduled extraction of data from both on-premise and Cloud applications and databases. When data extraction is not desired, Live Access is available for connecting to and querying existing data warehouse, marts, cubes and operational data stores in real-time.
- **Data Transformation and Modeling**  
Birst provides extensive data modeling and scripting capabilities for combining, transforming and cleansing raw data sources. Data modeling uses automation to shield developers from repetitive and time consuming tasks.
- **Query and Analytics**  
This tier comprises a ROLAP engine, a predictive analytics engine, a global multi-layer cache and a job/report scheduler.
- **End User Experience**  
Birst supports all major styles of business intelligence, including interactive dashboards, ad hoc analysis, data visualization, pixel-perfect banded report writing and mobile device access.
- **Unified Metadata Model**  
The Birst platform architecture features a unified metadata model. A single, shared layer of metadata ensures consistency across multiple sources of data and eliminates the need to integrate and maintain multiple metadata repositories when using a group of different BI products.
- **Multi-Tenant SOA**  
Both the web application server and the data server tiers are multi-tenanted. Multiple customers reside in a single instance of a web server process. Individual customers reside in a single schema within a shared database instance.

### Virtualization

Birst's multi-tenanted SOA architecture allows administrators and developers to quickly and easily create a new deployment instance, called a Birst space. The ability to easily switch between multiple spaces without having to provision new hardware and install new software provides several benefits, including zero down time while new data is loading. Data loading can be done on a copy of the production environment. Once data loading and processing is complete, the copy can be switched over to become the new production space. In addition, multiple release versions of Birst can run in parallel, allowing seamless upgrades in place.



# Platform for Cloud BI and Analytics

## Product Components

### Data Connectors

Birst provides connectivity and data extract options for on-premise databases, flat, structured and unstructured files as well as popular on-premise and cloud applications. Structured data is extracted and uploaded in a tabular format of columns and rows per data sources. Birst also supports unstructured and semi-structured data.

### Birst Connect

Birst Connect is a Java application that provides the connectivity to various on-premise and cloud databases and applications. Birst Connect is used for both bulk data extraction and for connecting in real-time. See chapter Live Access for details on real-time connectivity.

Birst enables data extraction from all Relational and Analytic databases, including Oracle, DB2, Microsoft SQL Server, Vertica and Amazon Redshift.

Birst capabilities extend to modern and big data sources such as Hadoop Hive, Cloudera Impala and Cassandra.

Birst also enables data extraction from leading applications such as Salesforce.com, Netsuite, Marketo and SAP ERP.

### ETL Engine

The Birst platform includes an engine for transforming and cleansing data. The engine is based on a fully scalable, in-memory middle tier that performs the data transformation processing. It stages intermediate tables and performs transformations on data prior to loading. For optimal performance, the ETL engine evaluates transformation job dependencies and parallelizes the execution of jobs where possible.

### Analytical Data Store

Birst stores data in an analytic database management system. As a multi-tenant system, Birst segregates customer data by creating a separate database schema for each Birst space. Depending on data volumes, anticipated growth rates and other customer requirements, Birst Cloud uses Microsoft SQL Server, ParAccel or InfoBright as the analytical data store. Birst Appliance customers have the option to use these or an Oracle database.

### ROLAP Engine

The Birst ROLAP engine provides full ad-hoc analysis capabilities without the need for physical OLAP cubes,

thereby offloading IT from the resource-intensive and time-consuming task of constantly having to maintain and optimize a growing cube farm. Unlike other OLAP engines, Birst does not restrict dimensional access to the data. Birst constructs a dynamic logical cube of all data that it is mapped to. It provides the full richness, scope and depth of information that can be possibly analyzed.

### Multi-Layer Cache

Birst provides exact and fuzzy cache matching, as well as dynamic cube-like cache structures to help with performance. These indexed data structures provide far better reuse and generate lower database load than traditional caching approaches. The dynamic cache is dynamically partitioned amongst servers to minimize I/O contention and to allow better memory caching, ultimately resulting in a far more scalable solution. The Birst ROLAP engine in combination with its unique caching layer provides a significant performance improvement over traditional OLAP solutions.

### Predictive Analytics Engine

The Birst platform includes a predictive analytics engine. In contrast to conventional data mining environments, data does not have to be moved out of the data warehouse. Data sets for model training and scoring are generated directly from the data warehouse.

Birst's advanced analytics capabilities leverage the ROLAP engine for data preparation. The modeling engine makes full use of aggregates and derived measures. Sophisticated new measures are defined and calculated on the fly as inputs into the modeling process. Share, time-series and dimensional breakout metrics are used to enrich the information contained in the data warehouse. The use of OLAP-style measures for modeling enables the addition of complex and highly predictive behavioral calculations.

For each modeling task, Birst automatically evaluates a comprehensive set of algorithms. Supported algorithms include linear and logistic regression, decision trees, feed-forward neural networks, support vector machines and rules/regression trees. Modeling scores are written directly back to data warehouse tables, ready to be used in ad hoc queries and dashboards or to be fed into additional processing, for example list generation. Both rules-based and model-based recommendations can be combined into complex decisions.

## Platform Functionality

### Data Extraction and Upload

For bulk data extraction from databases, Birst Connect supports the extraction of entire database tables or views or the extraction of subsets of data using custom SQL queries. Birst Connect extraction tasks can be scheduled either using a built-in scheduler or an external OS scheduler. In addition to extraction of data from databases and applications, Birst supports uploading delimited flat files, Microsoft Excel and Access database files. In all cases, data is transferred securely to Birst using strong authentication and compression techniques.

The connector Salesforce.com supports extraction of standard and custom data objects, as well as selective data extraction via SOQL queries. The connector for SAP ERP supports connections via BAPI functions, Query execution or DSO read calls. Birst utilizes the web services APIs of Netsuite and Marketo to extract standard and custom data objects from both these applications.

### Data Transformation

Birst provides a powerful and flexible mechanism to transform and manipulate data in Birst before it is loaded into an analytical data store. Birst ETL capabilities were designed explicitly to provide a high level of flexibility to developers. Data transformation and cleansing jobs are scripted using a SQL-like language.

### Data Modeling

Birst provides a graphical user interface for defining the relationships between uploaded and Live Access-connected data sources.

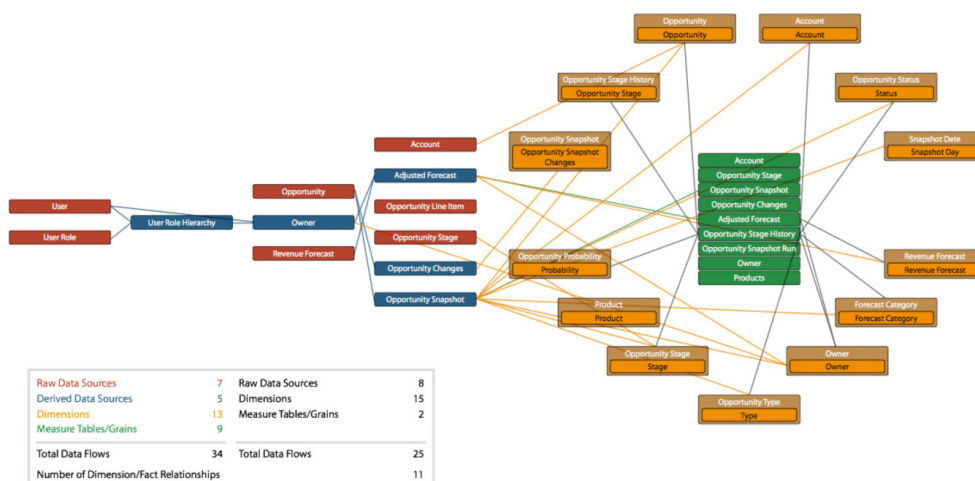
Developers are able to directly map a logical model, including hierarchies and grains, onto the data sources. Birst automatically generates and maintains the corresponding star schema and physical data warehouse.

### Data Warehouse Automation

Birst makes it possible to extract, transform and load data into a versatile data warehouse design in a fraction of the time required when using conventional BI tools. Birst's patent-pending automation features enables IT to quickly serve the analysis and reporting needs of business users without having to give up control over the data.

Using conventional BI tools, designing and creating a data warehouse is a resource-intensive and time-consuming process. The process of creating and maintaining the data warehouse schema, then mapping source data to it has traditionally been manual. Birst on the other hand automatically generates and maintains all the fact tables, dimensions, joins as well as all the required routines for loading data into the analytical data store. The advantage of automation is that it enables a rapid iterative development of a data warehouse.

Once the logical dimensional model has been defined, Birst automatically compiles the model into a fully optimized star schema. Logical measures automatically turn into calculation grains and logical dimensions automatically turn into levels. Fact tables are automatically generated, while dimensions and mini-dimensions are created as needed. Birst generates and



# Platform for Cloud BI and Analytics

manages all key relationships, including surrogate keys, when necessary.

Full and incremental data loading is available automatically. No additional scripting is required for an incremental load. Changes are detected automatically. Birst also provides automatic management of historical data including snapshots.

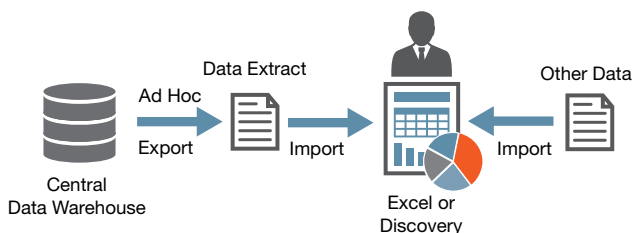
Load optimization allows for calculations to be grouped in the same loading pass. Such load groups enable segmentation and parallelization of the loading process. Each load group can have its own periodicity (e.g. day, week, month) and own specific content (e.g. organization). Time-shifting variants (e.g. month ago, quarter ago, etc.) and rolling-time variants (trailing 3, 6, 12 months, year-to-date, etc.) are automatically supported as well.

## Live Access

Birst Live Access is a major breakthrough for Cloud BI, enabling IT to leverage the economic benefits of the cloud without having to move data to the cloud. IT can connect directly to on-premise data assets using Live Access. Live Access transfers query results securely via SSL.

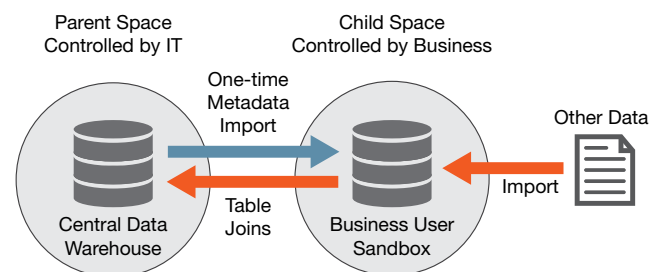
## Managed Data Mashups

In most organizations, centralized IT and Business are at odds over questions of data access and control. IT is incentivized to ensure data quality, enforce data policies and manage risk surrounding the handling of data. Business's need for quick and flexible access to data—in particular the ability to combine centrally managed sources with departmental data—often takes a back seat to IT's data governance charter.



In support of BI applications, IT typically combines all data into a centralized data warehouse. This approach thwarts agility in that it is inherently inflexible, not scalable and causes IT to become a bottleneck in the

organization, unable to meet business users' requests to quickly add new data sources. Lines of business are forced to work around the bottleneck, using Excel or visualization tools to perform one-off analyses on combined data manually extracted from centralized and other data sources. In this far-from-optimal scenario, IT is unable to maintain a single version of the truth across the organization, while Business is hamstrung by stale data and limited analytical capabilities.



Birst Managed Data Mashups promotes agility by providing the only solution in the BI industry that meets the needs of Business for quick and flexible access to data, while also allowing IT to maintain necessary control. Birst Managed Data Mashups delivers self-service capabilities by enabling departments to combine their own data with high-quality data sources that are centrally managed by IT. This activity is safely confined to analytics “sandboxes” created by business users.

To provide business users with personalized data sandboxes, Birst Managed Data Mashups introduces the concept of composite spaces and space inheritance. A composite space integrates data from one or more parent spaces with additional data loaded into the composite space. Data from a parent space and data loaded into a child space are both stored in the same underlying database, allowing joins to occur at the lowest possible level with optimal query performance. A child space only maintains logical references to a parent space. To maintain data integrity, no data from a parent space is copied and duplicated into a child space. Furthermore, none of the actions in a child space can alter the data and metadata in a parent space.



# Platform for Cloud BI and Analytics

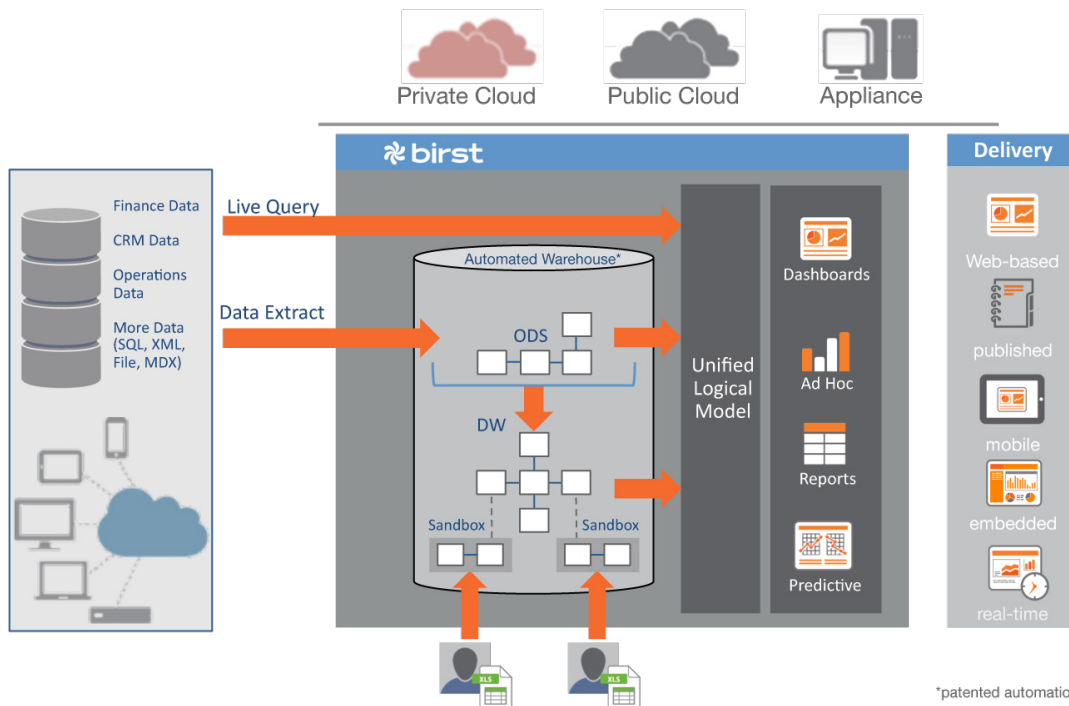
Additionally, Birst Managed Data Mashups allows an administrator of a space to designate staging and warehouse tables, custom formulas and variables as a package for import by other spaces. Once a package is imported into a child space, the imported objects can be combined with other objects and serve as the basis for deriving new objects. Any dependencies, such as hierarchies associated with imported tables, are automatically maintained in the child space. In particular, data security filters that are applied to a source in the parent space are also enforced in the child space.

IT also benefits from the agility of Birst Managed Data Mashups. Due to Birst Managed Data Mashups’s modular design, a monolithic Birst space that combines numerous different data sources at various levels of hygiene can be split into separate, more easily managed spaces. And, to maximize flexibility, Birst Managed Data Mashups allows IT to incorporate data from existing data warehouses that were built, and are still maintained, outside of Birst.

## Performance and Scalability

Birst is a fully multi-tenant solution from both a data processing and data storage perspective. Birst’s web and application server tiers are multi-tenant meaning that users are spread across an infinitely scalable pool of computing resources. Leveraging its shared-nothing architecture, Birst is the only BI solution to provide this level of scalability.

From a user interface standpoint, users can move seamlessly between various styles of BI without having to switch products. Birst provides a single user interface for operational report creation, ad hoc analysis, data visualization and interactive dashboards. The user interface is designed for self-service, allowing users to do more with far fewer dependencies on IT.

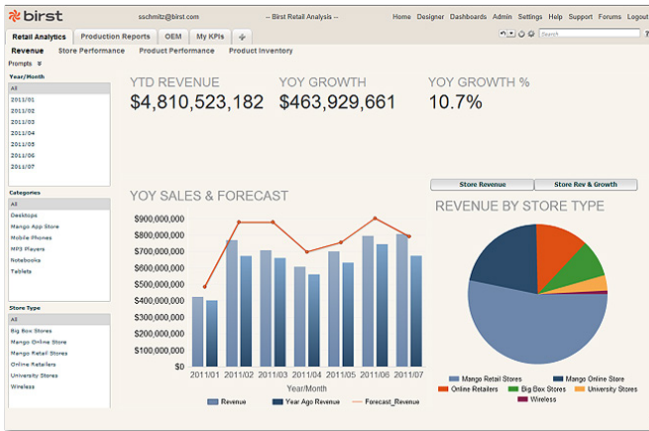
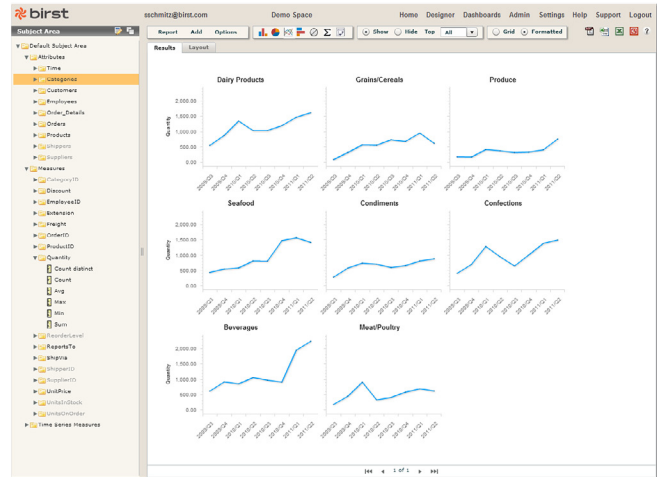


## User Experience

### Interactive Dashboards

Birst's interactive dashboards provide a self-service and easy-to-use interface for business users of all types. With Birst, users can directly interact with the dashboard – or even build new ones – without any formal training or specialized BI expertise.

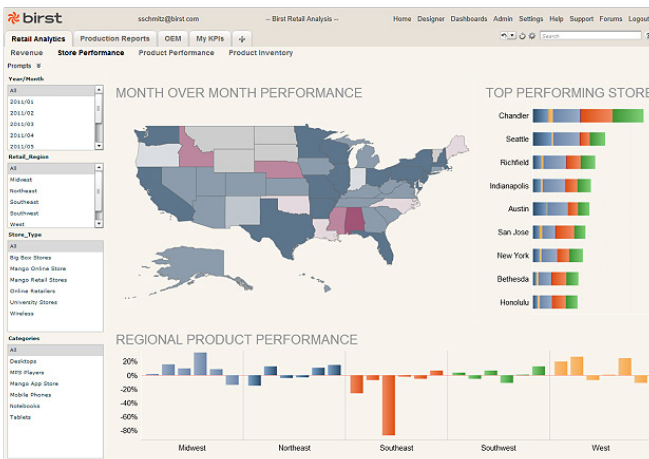
Filtering is an immersive process. Users can filter results via prompts or lasso filters. Birst Dashboards support flexible drill-anywhere capabilities. Both charts and tables can drill across any desired target report, dashboard or external URL.



### Ad Hoc Analysis, Data Visualization and Discovery

Birst Designer enables users to analyze data in an ad hoc fashion, visualizing and discovering trends and patterns. Birst supports a wide variety of visualization techniques, including trellis charts, scatter plots, geo maps, heat grids, tree maps and others.

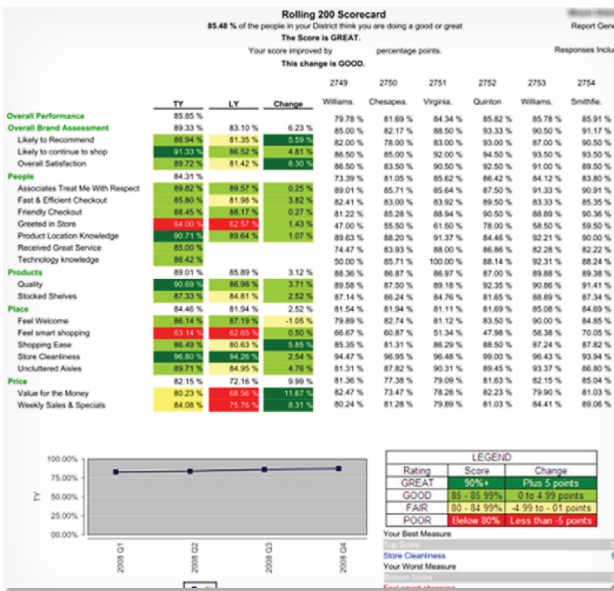
For OLAP-style analysis, Birst supports aggregations, cell-based calculations, slicers and filters. Positional calculations allow users to compare how a data point relates to values elsewhere. All analytic functions including inheritance, business rules, multi-pass calculations, and virtual measures are available via a point and click interface.



## Pixel-Perfect Banded Reporting

Birst also includes a report layout editor for advanced report creation. Formatting includes conditional formatting, conditional display, duplicate suppression, and null value replacement. Embedded images and sub-reports in various bands are supported.

Reports are compiled into Java byte code for fast and direct execution. No interpretation at runtime is required. Performance is enhanced by server-side report caching.



## Report Export, Scheduling and Alerts

Birst reports can be exported to a variety of formats, including PDF, Excel, Powerpoint and CSV. Both business users and administrators can schedule reports for delivery by email, both as an attachment and in-line content. Alerts are supported. At a scheduled time, Birst runs a trigger report to evaluate a specified condition. When the condition is satisfied, (e.g.: a KPI falls below a certain threshold) the alert email will go out.

Birst provides sophisticated report bursting, where a single database pass can be used to serve hundreds to thousands of reports allowing high volume report distribution, without taxing the database. Birst also supports parameterization of bursted reports based on user roles and data visibility rules.

## Custom Expressions

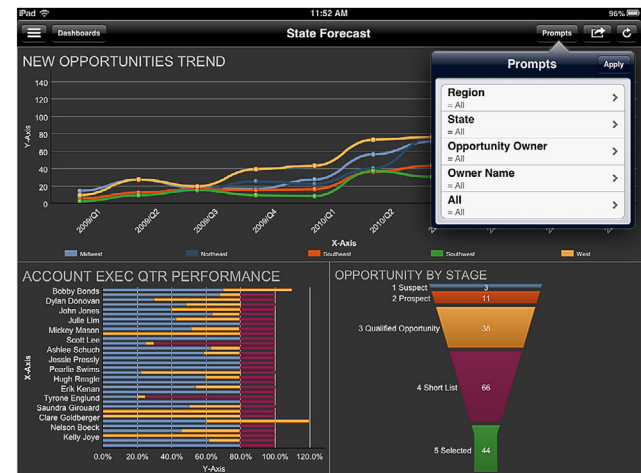
Birst enables users to create powerful custom expressions without the need to get IT involved. Birst's logical query language allows users to define and save both OLAP-style and Excel-style calculations. This includes advanced functions, lookups, transformations and linear regression.

## Birst for Apple iPad

Birst for the Apple iPad provides mobile access to interactive dashboards. Taking advantage of the iPad's touch-enabled features, users can quickly scan through charts and tables, filter to view information of interest, or drill down for more detail.

Users can swipe down to scroll through rows in a report table or pinch to zoom individual reports inside a dashboard. Birst's embedded prompts allow users to easily filter dashboard results.

Birst for the iPad can also be used on an offline mode. Users can pick and choose which dashboards to save for offline viewing.



## Implementation

### Development Modes

Birst offers three types of development modes:

Automatic, Discovery and Advanced. Using the fully automatic mode, Birst scans data uploads to identify key relationships between sources and determine the cardinality of the sources. Based on that information, Birst automatically builds a star schema data warehouse.

Discovery mode is designed for use cases that do not require a star schema data warehouse. In this mode, a developer defines the relationships between multiple data sources where needed.

The Advanced mode provides developers with complete control over the data modeling process. Developers are able to specify the characteristics of each data source and the inter-relationships between them. Based on a logical dimensional model that the developer defines, Birst automatically builds a star schema model and a corresponding data warehouse. Both Automatic and Discovery spaces can be converted to an Advanced space.

# Platform for Cloud BI and Analytics

## Deployment

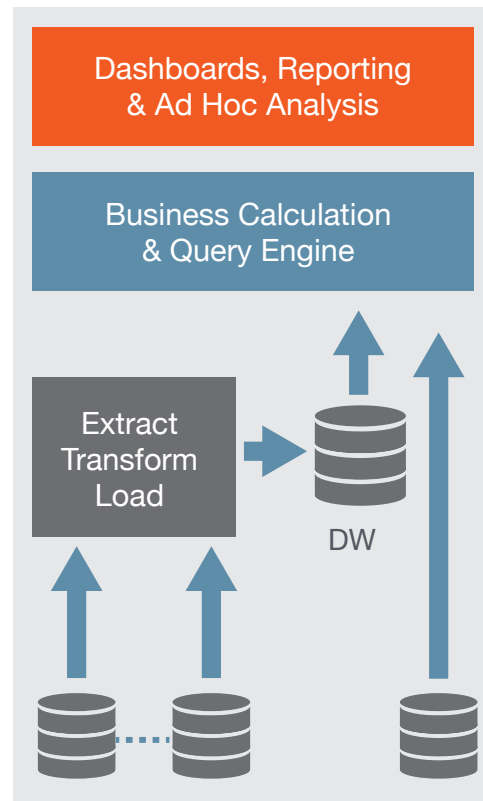
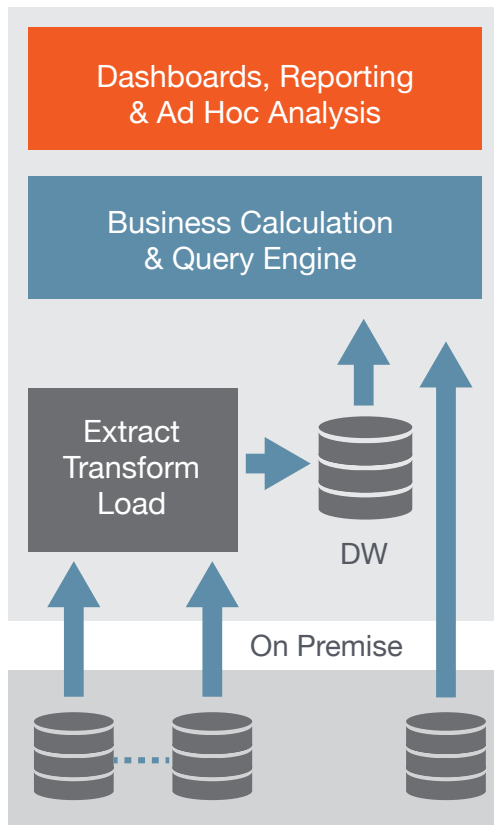
Birst is the industry’s only BI solution that can be deployed on premise or in the Cloud. Users can move from one deployment model to another to meet their strategic and operational goals.

### Birst Cloud

Birst Cloud is a multi-tenant, fully integrated software-as-a-service (SaaS) solution. Users get everything required for advanced business analytics in a subscription-based package delivered in the cloud. With Birst Cloud organizations remain agile while reaping the benefits of SaaS: fast deployment, lower costs and rapid time-to-value. As usage grows, Birst seamlessly expands server capacity to accommodate anywhere from dozens, to hundreds, or even thousands of users. Hosted in its SOC 2 Type II audited data center, Birst Cloud requires no installation of hardware and software. It also includes software upgrades. Users simply access all capabilities through a single web user interface.

## Appliance

The Birst Appliance encapsulates the entire Birst Business Analytics platform as a virtual machine for deployment on premise. It features the same shared-nothing architecture as Birst Cloud, enabling painless horizontal scalability and upgrades in place. The Birst Appliance is quick to setup and requires minimal configuration. Running on any virtual machine or hypervisor that supports the Open Virtualization Format (OVF), Birst Appliance enable easy upgrades and horizontal scalability.



## Amazon Redshift

Whether you need to access and analyze the data in an existing Redshift warehouse or deploy a single, fully managed cloud business analytics and data warehousing service, Birst provides you with that option.

	Use Case	Data Flow
<b>Option 1</b> Redshift Inside: Full Cloud BI around Redshift	You want a complete BI solution, but your data is not yet in Redshift or ready for analysis	Data Flow: Birst manages the entire analysis cycle, uploading your data into Redshift, transforming it into analytic ready form in Redshift, and providing you with rich visualizations
<b>Option 2</b> Live Access: Instant BI against existing Redshift	Your data is in Redshift and it's ready for analysis	Data never leaves Redshift. Simply activate a Birst account, connect to Redshift and start analyzing your data
<b>Option 3</b> Redshift Mash-up: Analyze Redshift data with other data	Your analysis is cross source; some of your data is in Redshift and some in other data sources	Birst extracts all or some of your Redshift data into Birst, merges it with other sources in a logical layer and provides you with rich visualizations

## Integration

### Single Sign On & Embedding

Birst provides the ability to embed reports and dashboards in other applications. The SSO framework supports session parameters to dynamically control access privileges and data visibility to those logged into the application. For authentication, Birst also supports OpenID and SAML. To ensure a seamless integration with other application, Birst offers customizable colors, fonts and labels.

### Web Services

Birst web services enable programmatic administration of a Birst solution and tight integration into another

application or portal. The Birst web services API is SOAP-based and can work with any programming language that supports web services. Web services range from calls to manage users and metadata, to services for running queries.

### Conclusion

With a complete business analysis platform that is purpose-built for agility, Birst offers a simpler and more flexible solution for applications of all types.