

# Data Engineering Workshop

Sanchit Jain  
12th Nov, Saturday  
2:00 PM to 5:00 PM



# Speakers



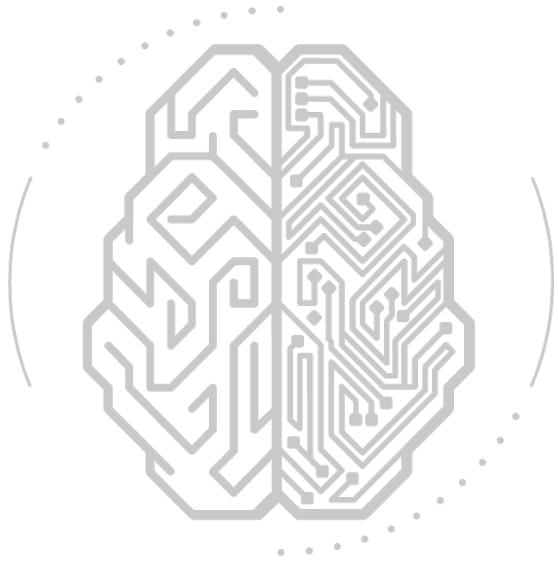
**Sanchit Jain**

Data Analytics & Cloud Practice Lead at Quantiphi  
AWS Hero & AWS Ambassador

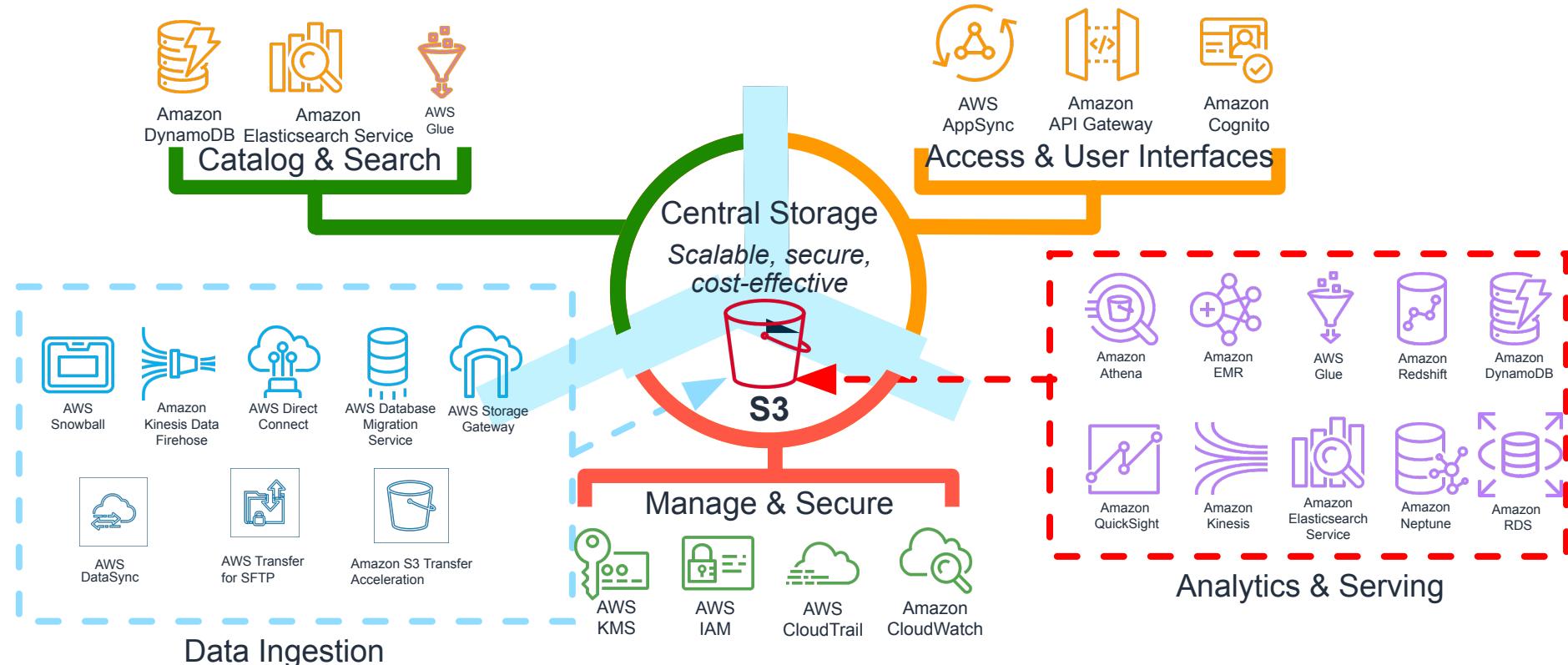
FOLLOW ME



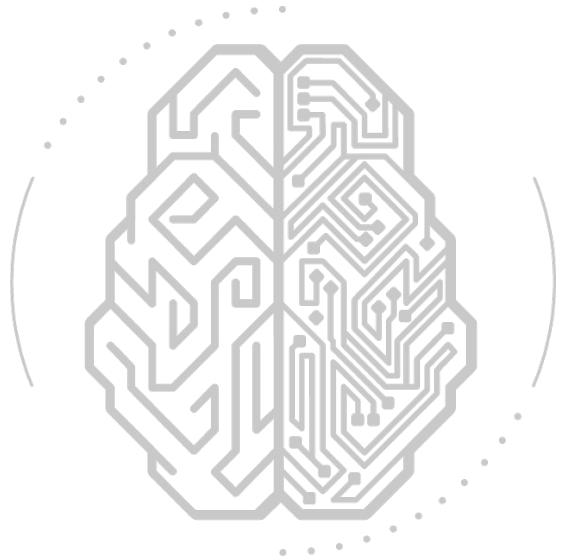
# What we will learn?



# Session's Focus – Query The Data Lake

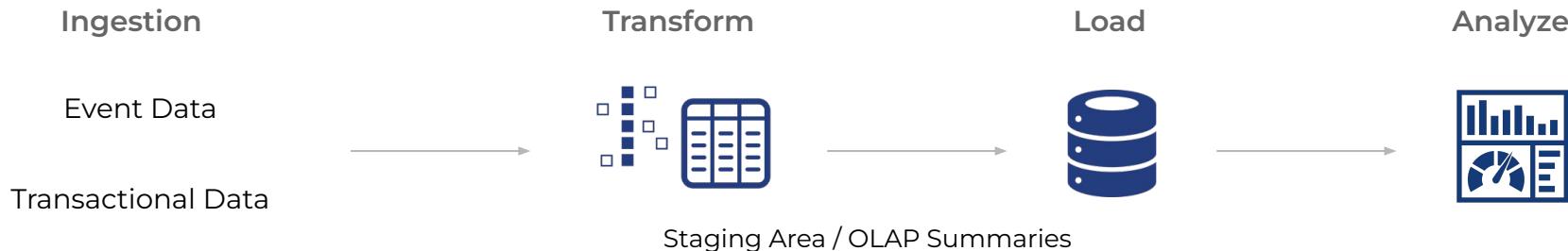


# **How does a traditional ETL process look like?**



# How does a traditional ETL process look like?

## Traditional ETL Migration Process Flow :



## Various Components of ETL Migration Process Pipeline :

### 1 Ingestion

Data is ingested from online transaction processing (OLTP) databases, today more commonly known just as 'transactional databases', and other data sources. OLTP applications have high throughput and they do not lend themselves well to data analysis or business intelligence tasks

### 2 Transform

Data is transformed in a staging area. These transformations cover both data cleansing and optimizing the data for analysis.

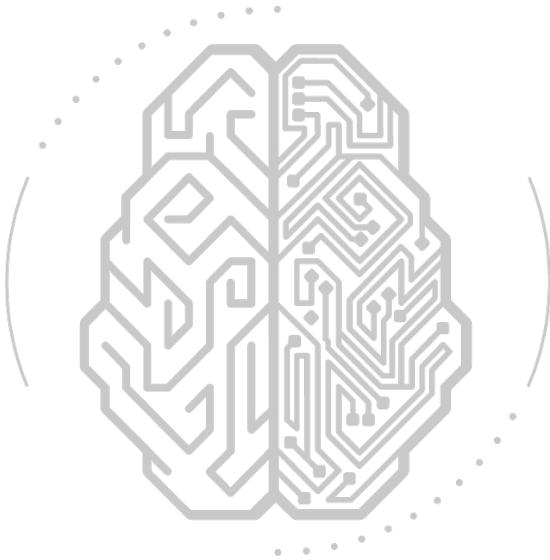
### 3 Load

The transformed data is then loaded into an online analytical processing (OLAP) database, today more commonly known as just an analytics database.

### 4 Analyze

Business intelligence (BI) teams then run queries on that data, which are eventually presented to end users, or to individuals responsible for making business decisions, or used as input for machine learning algorithms or other data science projects

# AWS Glue - Features & Benefits



# AWS Glue Overview



AWS Glue is a serverless data integration service that makes it easy to discover, prepare, and combine data for analytics, machine learning, and application development. AWS Glue provides all the capabilities needed for data integration so that you can start analyzing your data and putting it to use in minutes instead of months.

## FEATURES



Serverless Spark

There is no infrastructure to maintain. Allocate needed compute power and run jobs. Job starts in few seconds and can run at petabyte scale



Cost Effective

All-in-one pricing model includes infrastructure and is 55% cheaper than other cloud data integration options



No Lock-In

Develop data integration pipelines in open source SparkSQL, PySpark and Scala



Data Integration for every user

Development environments catered to different skill sets



Handles complex workloads

Glue connects to 60+ data sources, processes petabytes of data in real-time, batch and event driven modes



More Power

AWS Glue automates much of the effort spent in building, maintaining, and running ETL jobs

# AWS Glue benefits



## Real time analysis

As you process streaming data in a Glue job, you have access to the full capabilities of Spark Structured Streaming to perform real time analysis of data



## Data integration for all users

Development environments catered to different skill sets



## No Lock In

Glue jobs are written open source Spark, Python and Scala



## Cost - Effective



All-in-one pricing model includes infrastructure and is 55% cheaper than other cloud data integration options



## Highly scalable Infrastructure

AWS Glue is highly scalable and being on AWS cloud it scales up as per requirement



## No Specialised skills required

There is visual ETL development for Data Engineers, notebook styled development for Data Scientists and no code development for Data Analysts

### Cost Comparison

**7x**

Glue is 7x cheaper compared to on-premise options

**5x**

Adopting Glue is 5x cheaper than setting up your own Spark cluster

**4x**

Glue reduces maintenance of your self managed Spark clusters by 4x

**55%**

Glue is 55% cheaper compared to other cloud providers

### Impact on Business Processes



**Reduction in human dependency**



**Revenue acceleration**



**Faster data processes**

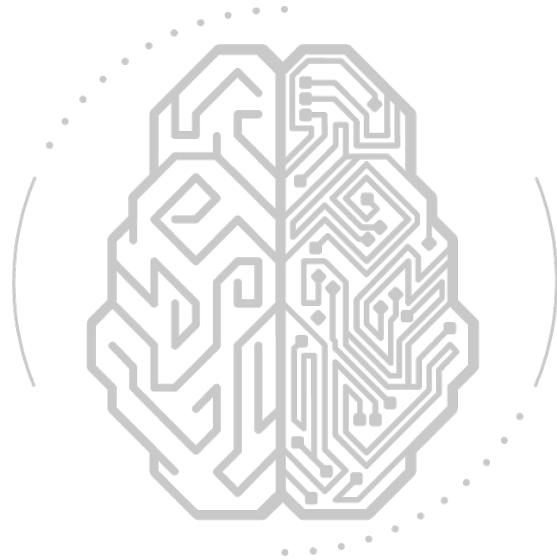


**Better decision making**



**Reduction in time to market**

# AWS Glue - Components

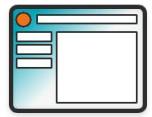


# AWS Glue: Components



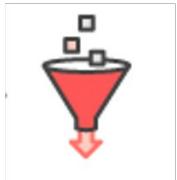
Data Catalog

- Hive metastore compatible with enhanced functionality
  - Hive metastore compatible with enhanced functionality
- Crawlers automatically extract metadata and create tables
  - Crawlers automatically extract metadata and create tables
- Integrated with Athena, Amazon Redshift Spectrum
  - Integrated with Athena, Amazon Redshift Spectrum



Job Authoring

- Auto-generates ETL code
  - Auto-generates ETL code
- Builds on open frameworks—Python and Spark
  - Builds on open frameworks—Python and Spark
- Developer-centric—editing, debugging, sharing
  - Developer-centric—editing, debugging, sharing



Job Execution

- Runs jobs on a serverless Spark platform
  - Runs jobs on a serverless Spark platform
- Provides flexible scheduling
  - Provides flexible scheduling
- Handles dependency resolution, monitoring, and alerting
  - Handles dependency resolution, monitoring, and alerting

# AWS Glue Data Catalog

Manage table metadata through a Hive metastore API or Hive SQL. Supported by tools like Hive, Presto, Spark, etc. AWS added a few extensions:

- **Search** over metadata for data discovery
- **Connection info**—JDBC URLs, credentials
- **Classification** for identifying and parsing files
- **Versioning** of table metadata as schemas evolve and other metadata are updated

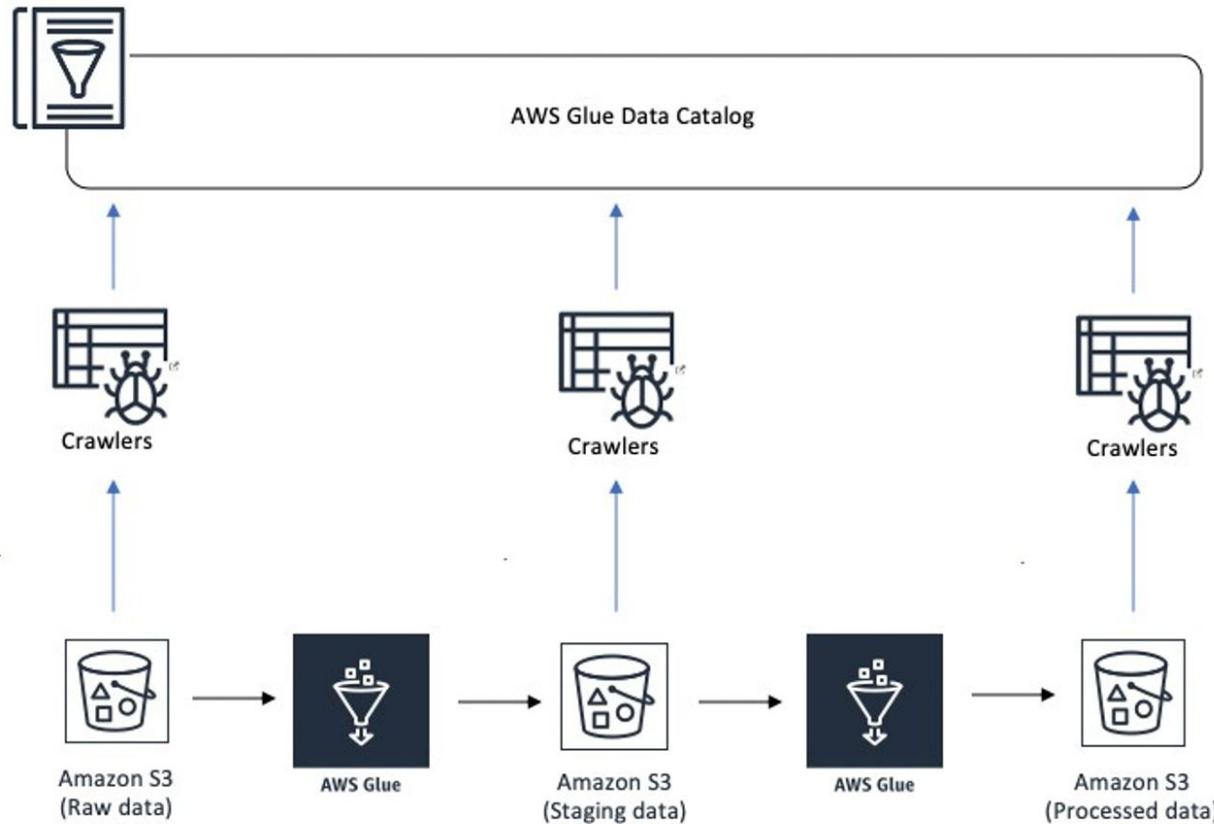
Populate using Hive DDL, bulk import, or automatically through **crawlers**

# AWS Glue Data Catalog: Crawlers

Crawlers automatically build your Data Catalog and keep it in sync

- Automatically discover new data, extract schema definitions
  - Detect schema changes and version tables
  - Detect Hive style partitions on Amazon S3
- Built-in classifiers for popular types; custom classifiers using Grok expressions
- Run ad hoc or on a schedule; serverless—only pay when crawler runs

# AWS Glue In Action



# What is AWS Glue DataBrew ?

AWS Glue DataBrew is a visual data preparation tool that makes it easy for data analysts and data scientists to prepare data with an interactive, point-and-click visual interface without writing code.



## CAPABILITIES OF GLUE DATABREW

Profile

Clean and Normalize

Map Data Lineage

Automate



## NEED FOR DATABREW

*"Up to 80% of data analysis time is spent on preparing data"*

### Time Consuming

- Multi-step process to extract, clean, normalize & load data at scale
- The right tools for the right persona must be integrated

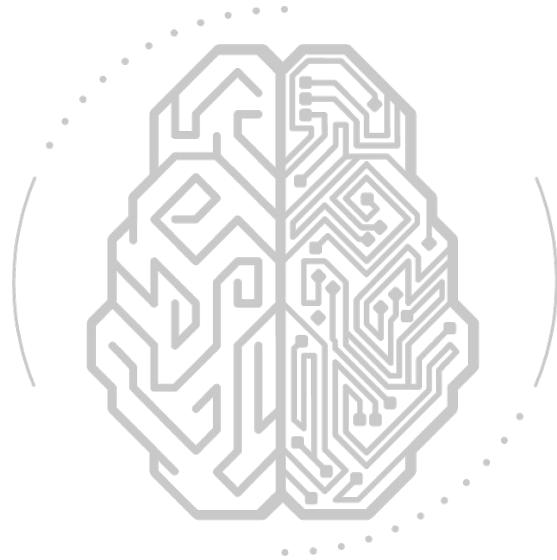
### Expensive

- Costly user licenses & siloed tools that cause rework
- Often requires moving large amount of data into silos

### Manual

- Needs a lot of code-based heavy lifting to work at scale
- Hard to operationalize & build repeatable workflows

# Querying the Data Lake with Amazon Athena





# Amazon Athena

Start querying data instantly. Get results in seconds. Pay only for the queries you run.

An interactive query service that makes it easy to analyze data directly from Amazon S3 using Standard SQL

# Amazon Athena

- Query data in your Amazon S3 based data lake
- Analyze infrastructure, operation, and application logs
- Interactive analytics using popular BI tools
- Self-service data exploration for data scientists
- Embed analytics capabilities into your applications

# What does it look like?

Athena   **Query Editor**   Saved Queries   History   Catalog

Catalog   Sample Queries

DATABASE

default

Table Name

adsads  
cloudfront\_logs  
cloudtritable  
elb\_logs  
ncc  
ncctest  
prestoinstance  
prestostat  
taxinyc\_csv  
taxinyc\_par  
test3  
test4  
wikistats  
language (string)  
page\_title (string)  
hits (bigint)  
retrieved\_size (bigint)  
wikistats\_parq

```
1 select sum(hits) as hits,language from default.wikistats
2 group by language
3 order by 1 desc
4 limit 10
```

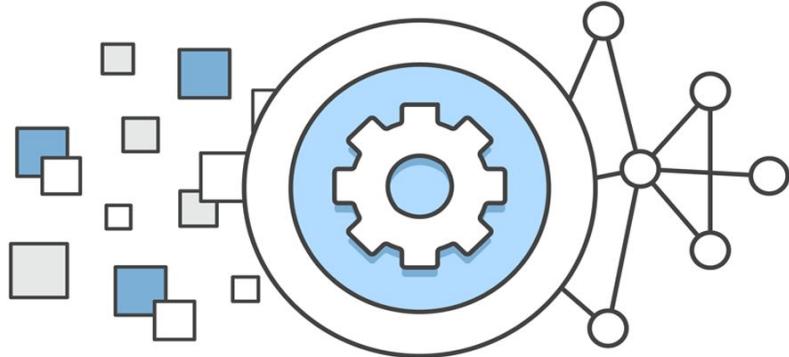
Execute   Save As   or create a   New Query

Recent Queries   Query   Columns   **Results**   Chart

	hits	language
1	213917076	en
2	43673225	ja
3	25593194	es
4	16637343	de
5	10579788	fr
6	7958810	commons.m
7	6664552	pt
8	6108102	ru
9	5857921	pl
10	5783183	it

# Athena is Serverless

- No Infrastructure or administration
- Zero Spin up time
- Transparent upgrades



# Use ANSI SQL

- Support for complex joins, nested queries & window functions
- Support for complex data types (arrays, structs)
- Support for partitioning of data by any key

```
1  WITH q21_tmp1_cached AS
2  (SELECT l_orderkey,
3   count(DISTINCT l_supkey) AS count_supkey,
4   max(l_supkey) AS max_supkey
5   FROM lineitem_parq
6   WHERE l_orderkey IS NOT NULL
7   GROUP BY l_orderkey),
8  q21_tmp2_cached AS
9  (SELECT l_orderkey,
10   count(DISTINCT l_supkey) count_supkey,
11   max(l_supkey) AS max_supkey
12   FROM lineitem_parq
13   WHERE l_receiptdate > l_commitdate
14   AND l_orderkey IS NOT NULL
15   GROUP BY l_orderkey)
16  SELECT s_name,
17   count(*) AS numwait
18  FROM
19  (SELECT s_name
20   FROM
21   (SELECT s_name,
22    t2.l_orderkey,
23    l_supkey,
24    count_supkey,
25    max_supkey
26    FROM q21_tmp2_cached t2
27    RIGHT OUTER JOIN
28    (SELECT s_name,
29     l_orderkey,
30     l_supkey
31     FROM
32     (SELECT s_name,
33      t1.l_orderkey,
34      l_supkey,
35      count_supkey,
36      max_supkey
37      FROM q21_tmp1_cached t1
38      JOIN
39      (SELECT s_name,
40       l_orderkey,
41       l_supkey
42       FROM orders_parq o
43       JOIN
44       (SELECT s_name,
45        l_orderkey,
46        l_supkey
47        FROM nation_parq n
48        JOIN supplier s ON s.s_nationkey = n.n_nationkey
49        AND n.n_name = 'SAUDI ARABIA'
50        JOIN lineitem_parq l ON s.s_supkey = l.l_supkey
51        WHERE l.l_receiptdate > l.l_commitdate
52        AND l.l_orderkey IS NOT NULL) t1 ON o.o_orderkey = t1.l_orderkey
53        AND o.o_orderstatus = 'F') t2 ON t2.l_orderkey = t1.l_orderkey) a
54    WHERE (count_supkey > 1)
55    OR ((count_supkey=1)
56    AND (l_supkey < max_supkey)) b
57    WHERE ((count_supkey IS NULL)
58    OR ((count_supkey=1)
59    AND (l_supkey = max_supkey))) c
60  GROUP BY s_name
61  ORDER BY numwait DESC,
62   s_name LIMIT 100;
```

# Familiar Technologies Under the Covers



## Used for SQL Queries

In-memory distributed query engine  
ANSI-SQL compatible with extensions



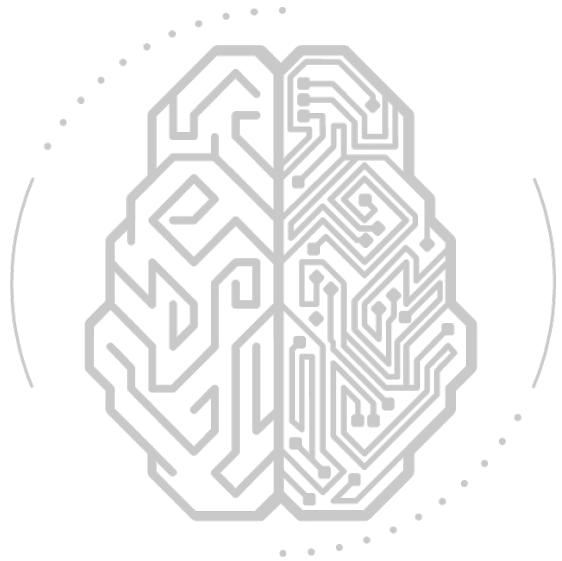
## Used for DDL functionality

Complex data types  
Multitude of formats  
Supports data partitioning

# Amazon Athena is Cost Effective

- Pay per query
- \$5 per TB scanned from S3
- DDL Queries and failed queries are free
- Save by using compression, columnar formats, partitions

# Athena Workgroups



# Athena Workgroups

Athena Workgroups are used to isolate queries between different teams, workloads or applications, and to set limits on amount of data each query or the entire workgroup can process

Workload Isolation

Query Metrics

Cost Controls

# Workgroups – Workload Isolation

Workgroup name\*  Use 1 - 128 characters. (A-Z,a-z,0-9,\_,-,.)

Description  Use up to 1024 characters.

Unique query output location per Workgroup

Query result location  Select Enter a path to an S3 bucket or prefix.

Encrypt results with unique AWS KMS key per Workgroup

Encrypt query results  Encrypt results stored in S3

Encryption type  ⓘ

Encryption key  ⓘ ⓘ Create KMS key

Metrics  Publish query metrics to AWS CloudWatch ⓘ

Override user settings  ⓘ

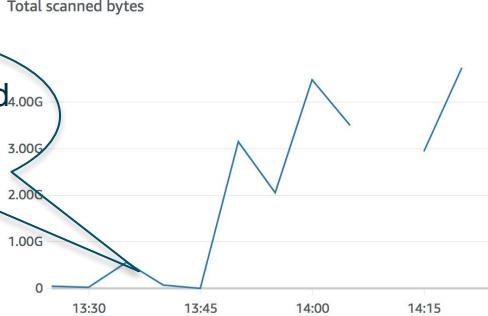
aggregated metrics per Workgroup to AWS CloudWatch

Use Workgroup settings eliminating need to configure individual users

# Workgroups – Metric Reporting

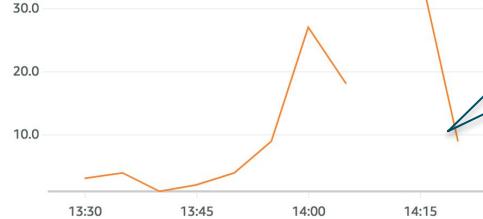
Total scanned bytes

Total bytes scanned per Workgroup



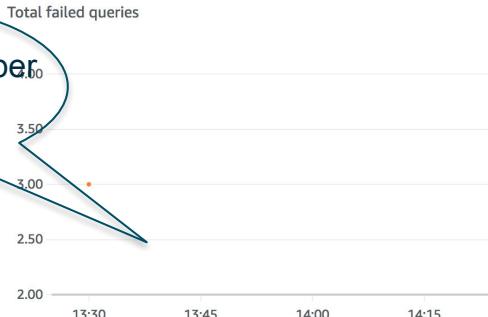
Total succeeded queries

Total successful queries per Workgroup



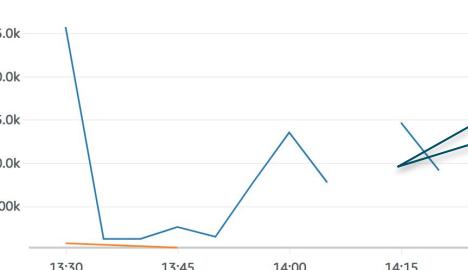
Total failed queries

Total failed queries per Workgroup



Total execution time

Total query execution time per Workgroup



# Workgroups – Cost Controls

- Per query data scanned threshold; exceeding, will cancel query
- Trigger alarms to notify of increasing usage and cost
- Disable Workgroup when all queries exceed a maximum threshold

Any Athena metric

Data limit	Time period	Action	
10 Gigabytes	Not applicable	Query will be cancelled.	
1 Terabytes	24 hours	Send notification to topic : arn:aws:sns:us-east-1:9	9:AthenaAlarm
10 Gigabytes	1 hour	Send notification to topic : arn:aws:sns:us-east-1:9	9:AthenaAlarm

# Workgroups – Usage Notifications

Define a hierarchy of alarms to be alerted as usage

Screenshot of the AWS CloudWatch Metrics & Alarms interface showing the creation and configuration of usage notifications.

**Alarms List:**

- Filter: State is ALARM
- Search Alarms: 1 to 5 of 5 alarms
- Actions: Create Alarm, Add to Dashboard, Actions
- Threshold: Config Status

State	Name	Threshold
<input type="checkbox"/>	2GBAthenaAlarm	ProcessedBytes >= 2 for 1 datapoints within 10 minutes
<input checked="" type="checkbox"/>	1GBAthenaAlarm	ProcessedBytes >= 1 for 2 datapoints within 10 minutes

1 Alarm selected

**Alarm: 1GBAthenaAlarm**

**Details** tab (selected):

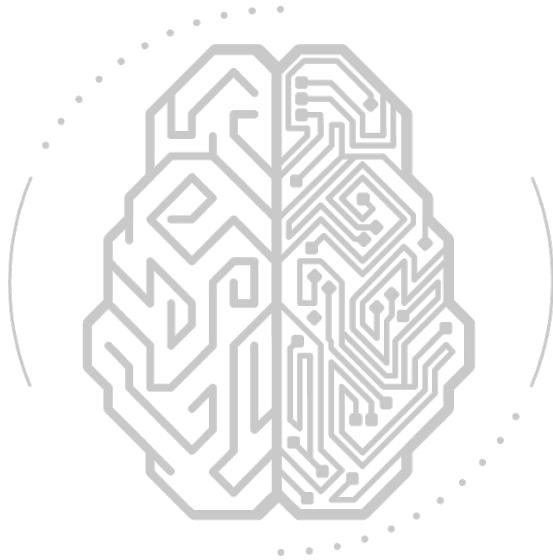
- State Details:** State changed to ALARM at 2018/11/20. Reason: Threshold Crossed: 2 datapoints [5.043427761923077E9 (20/11/18 14:23:00), 3.10345876475E9 (20/11/18 14:18:00)] were greater than or equal to the threshold (1.0).
- Description:**
- Threshold:** ProcessedBytes >= 1 for 2 datapoints within 10 minutes
- Actions:** In ALARM: • Send message to topic "AthenaAlarm" (royon@amazon.com)
- Namespace:** AWS/Athena
- Metric Name:** ProcessedBytes
- Dimensions:** WorkGroup = primary
- Statistic:** Average
- Period:** 5 minutes
- Treat missing data:** missing as:
- Percentiles with low samples:**

**1GBAthenaAlarm** chart:

- Y-axis: Bytes (0, 2.25G, 4.49G)
- X-axis: Time (12:00, 13:00, 14:00)
- Series: ProcessedBytes (blue line)
- Annotation: ProcessedBytes >= 1 for 2 datapoints within 10 min...
- Legend: ProcessedBytes

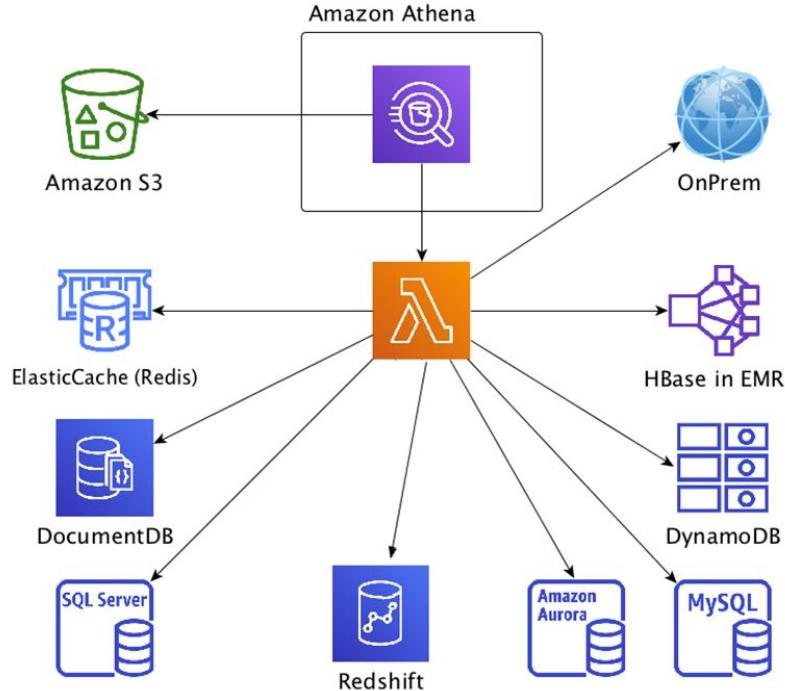
[View in metrics](#)

# Athena federated query

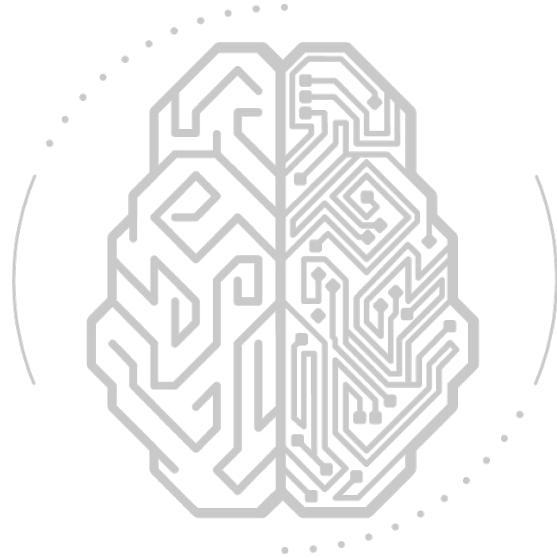


# Athena federated query

- Run query across relational, non-relational, object, or custom data sources
- Run query across On-Premises or cloud data sources
- Can be used for ad-hoc investigations, or complex pipelines, or applications



# Visualizing the Data Lake using Amazon QuickSight



# Why Amazon QuickSight?



## Cloud native = No servers = Auto-Scale

No servers or software to manage, maintain, deploy.  
Start with 10s of users and scale to 10s of 1000s



## Fast, consistent performance

Fast, predictable performance every time.  
Concurrent users or increased interactions do not slow down the system



## Fully integrated with AWS

Build end-to-end analytics in AWS. Secure private VPC access, fine-grained access control, ML integrations



## ML insights

Contextual, relevant insights with ML-powered anomaly detection, forecasting, alerts and customizable narratives



## Secure and global

End-to-end encryption. Native High Availability. 10 Global regions. HIPAA, PCI, ISO, SOC and FedRamp eligibility



## Insights for everyone

Provide access to all users, pay only for usage. No upfront costs, no charges for inactive users



## Easy to develop and maintain

Design with Amazon QuickSight, integrate with APIs. Secure data with row-level security and authenticate seamlessly via single sign-on



## Customize and embed

Embed in applications and enable analytics in hours, not months or years. Use themes to match application/corporate branding

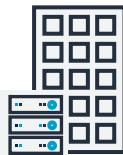


# Connect to your data, wherever it is

QuickSight is natively integrated with AWS data sources, as well as on-premises and hosted databases and third party business applications

## On-premises

Securely connect to on-premise databases and flat files like Excel and CSV



- Excel
- CSV
- Teradata
- MySQL
- SQL Server
- PostgreSQL

## In the cloud

Connect to hosted database, big data formats, and secure VPCs



- Presto
- Spark
- SQL Server
- Postgre SQL
- MariaDB
- Snowflake
- IoT Analytics

## Applications

Connect directly to third party business applications



- Salesforce
- Square
- Adobe Analytics
- Jira
- ServiceNow
- Twitter
- Github

service<sup>now</sup>

# Data Prep

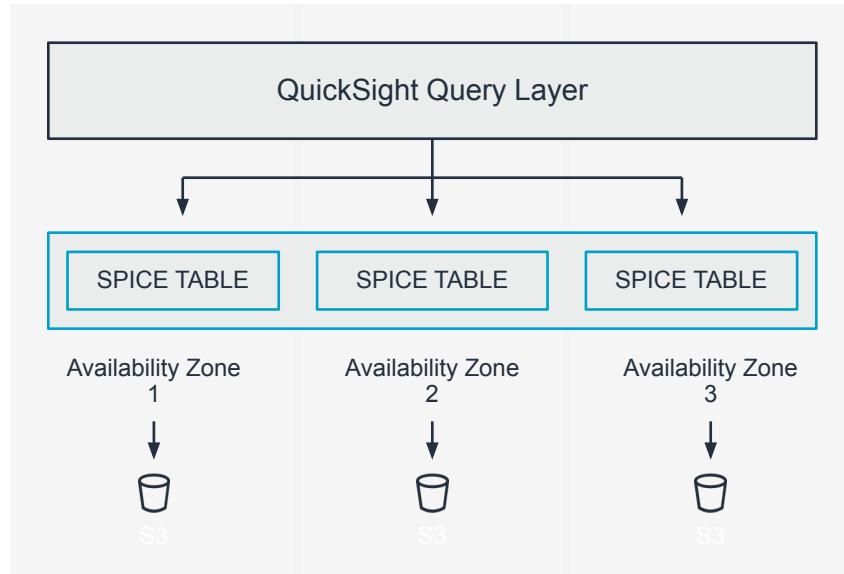
Optional step when creating data sets:

- Preview data
- Rename, remove fields, change data types
- Create new calculated fields
- Filter rows
- Issue direct query or ingest to SPICE
- Push down custom SQL queries
- Join across all data sources supported by QuickSight including file-to-file, file-to-database, and database-to-database joins

The screenshot shows the Amazon QuickSight Data Prep interface. On the left, the 'retail\_sales' dataset is previewed with 1 table selected. The preview table includes columns like cust\_id, cust\_name, cust\_type, city, state, country, zip\_code, region, and age\_group. On the right, a 'Configure join' dialog is open, showing a join between 'all\_flights' and 'airport\_id'. The 'Data sources' section lists 'all\_flights' and 'INNER' join type, with 'airport\_id' listed below it. The 'Join types' section shows options for Inner, Left, Right, and Outer joins. A red oval highlights the 'Configure join' button at the bottom of the dialog.

# SPICE

QuickSight is powered by SPICE, a super-fast calculation engine that delivers performance and scale, regardless of how many users are active.



---

Up to 10X faster (millisecond latency)

---

Fault-tolerant, self-healing

---

Support for high concurrency

---

Backed up in S3 (Write Ahead Log)

---

Instant failover with zero impact

---

# User Types / User Roles

---



## Admin

Manage Users  
Manage SPICE Capacity  
Manage VPC Connections  
Manage Account Settings

---



## Author

Create Data Sets  
Create Analyses  
Create Dashboards

---



## Reader

Consume Dashboards

---

## QS Admin

Sometimes separate from Business Users,  
sometimes the same  
Usually has AWS Console

---

## Analyst

Sometimes in IT, sometimes Business Users  
'Data Analyst'  
'Data Engineer'  
'BI Engineer'

---

## Business User

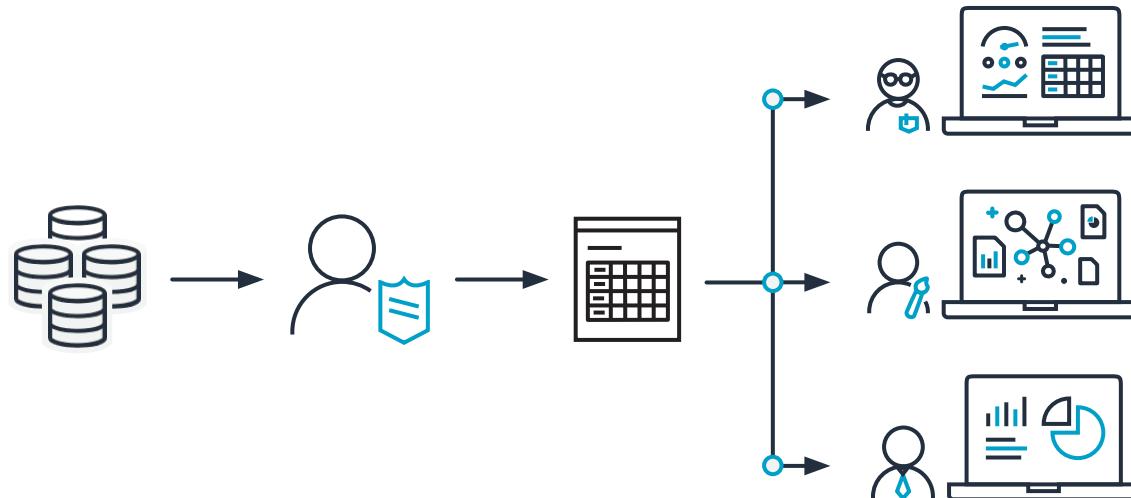
Anyone  
Can be internal or external users  
(customers/partners/3rd

# Data governance

Create managed datasets that give power users and authors the flexibility to perform self-serve analytics on data that you control.

## Create datasets that:

- Can be shared with any user
- Automatically refresh
- Have row level security
- Users cannot modify
- Dynamically update with changes



# Differentiate with natural language and ML



## Auto narratives

Summarize your business metrics in plain language



## Forecasting

Machine learning forecasting with point and click simplicity



## Anomaly detection

Discover unexpected trends and outliers against millions of business metrics



## ML predictions

Visualize and build predictive dashboards with Amazon SageMaker models

# Athena federated query

Contributors <

Daily sales for Region: APAC | Product: Office Supplies increased 20% from \$10K to \$12K on Oct 31, 2018 compared to the day before and was 15% higher than expected.

Last updated on Nov 13th 2018

Top Contributors Configure

Sort by Contribution percentage

Segment	%	DeD
SMB	63%	30% ..
Unicorn	17%	7% ..
Enterprise	13%	5% ..

Channel Online 81% 12% ..

### Anomalies found for Daily Sales on Oct 31st 2018

Region: APAC | Product: Office Supplies

Daily revenue for Segment: SMB increased from \$3K to \$ 4.1K (+25%). It contributed to 53% of the total increase for Region: APAC | Product: Office Supplies, higher than the expected 30% contribution.

Region: North America | Product: Outdoor

The dashboard displays three main sections: a summary of daily sales for APAC Office Supplies, an analysis of SMB segment contribution, and a detailed look at North America Outdoor product sales. Each section includes a line chart showing revenue trends over time, with specific anomalies highlighted on October 31, 2018.

# Forecasting



# Natural Language Narratives

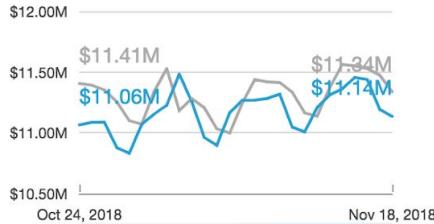
<p><b>Daily Revenue</b></p> <p>Daily revenue <b>decreased -0.51% (-\$57,032.99)</b> on Nov 18, 2018, from \$11.19M to \$11.14M compared to the previous day and is <b>-1.78% (-\$202,111.70) below</b> goal of \$11.34M. We are <b>\$2789.67K (0.334%) above</b> 30-day average revenue of \$8.35M. We're operating at an run rate of <b>\$4.06B</b>.</p>	<p><b>YTD Revenue</b></p> <p>Year-to-date revenue <b>increased by 61.95% (\$1.14B)</b> from \$1.85B to \$2.99B compared to the same period last year and is <b>-1.81% (\$55.03M) below</b> plan of \$3.05B. We are at 98.19% achievement of YTD goal and 84.61% achievement for annual goal.</p>	<p><b>Callouts By Product and Country</b></p> <p>Daily revenue for Baby Products   Russia on Nov 22, 2018 was <b>lower than expected</b> at \$1.170.58.</p>	<p><b>Daily Revenue Forecast</b></p> <p>Daily revenue is predicted to reach <b>\$11.96M</b> by end of the year. We expect to exit the year with an annualized run rate of <b>\$4.37B</b>. Total revenue for 2018 is predicted to reach <b>\$3.47B, \$63.59M (-1.80%) below</b> annual target of \$3.54B.</p>																																								
	<table border="1"> <thead> <tr> <th>Year</th> <th>Revenue</th> </tr> </thead> <tbody> <tr> <td>2018</td> <td>\$3.05B</td> </tr> <tr> <td>2017</td> <td>\$2.31B</td> </tr> </tbody> </table>	Year	Revenue	2018	\$3.05B	2017	\$2.31B	<table border="1"> <thead> <tr> <th>Product</th> <th>Revenue</th> </tr> </thead> <tbody> <tr> <td>Movies</td> <td>\$35M</td> </tr> <tr> <td>Baby Products</td> <td>\$31M</td> </tr> <tr> <td>Office Supplies</td> <td>\$29M</td> </tr> <tr> <td>Industrial</td> <td>\$28M</td> </tr> <tr> <td>Health</td> <td>\$27M</td> </tr> <tr> <td>Books</td> <td>\$20M</td> </tr> <tr> <td>Clothing</td> <td>\$19M</td> </tr> </tbody> </table>	Product	Revenue	Movies	\$35M	Baby Products	\$31M	Office Supplies	\$29M	Industrial	\$28M	Health	\$27M	Books	\$20M	Clothing	\$19M																			
Year	Revenue																																										
2018	\$3.05B																																										
2017	\$2.31B																																										
Product	Revenue																																										
Movies	\$35M																																										
Baby Products	\$31M																																										
Office Supplies	\$29M																																										
Industrial	\$28M																																										
Health	\$27M																																										
Books	\$20M																																										
Clothing	\$19M																																										
<p><b>Top / Bottom Movers by Product</b></p> <p>Top daily revenue <b>increase</b> by products are:</p> <ul style="list-style-type: none"> <li>Electronics increased by <b>\$537.84 (7.61%)</b> from \$7,066.98 to \$7,604.82.</li> <li>Clothing increased by <b>\$484.86 (0.06%)</b> from \$769,561.89 to \$770,046.75.</li> <li>Industrial increased by <b>\$427.49 (0.04%)</b> from \$1,078,710.02 to \$1,079,137.51.</li> <li>Home Services increased by <b>\$114.68 (0.21%)</b> from \$55,338.04 to \$55,452.72.</li> <li>Music increased by <b>\$67.40 (0.67%)</b> from \$12,997.63 to \$13,065.23.</li> </ul>	<p><b>Top / Bottom to Plan Variance by Product</b></p> <p>Top products <b>above</b> plan for today are:</p> <ul style="list-style-type: none"> <li>Movies is <b>\$147,437.42</b> above goal.</li> <li>Financial Services is <b>\$98,111.46</b> above goal.</li> <li>Clothing is <b>\$42,220.37</b> above goal.</li> <li>Computers is <b>\$38,003.06</b> above goal.</li> <li>Outdoors is <b>\$27,404.67</b> above goal.</li> </ul> <p>Top products <b>below</b> plan for today are:</p> <ul style="list-style-type: none"> <li>Digital is <b>-\$236,951.70</b> below goal.</li> <li>Health is <b>-\$147,834.66</b> below goal.</li> </ul>	<p><b>Revenue by Product Category</b></p> <table border="1"> <thead> <tr> <th>Product Ca...</th> <th>Nov 18, 2018</th> <th>Nov 17, 2018</th> <th>N</th> </tr> </thead> <tbody> <tr> <td>Arts</td> <td>\$4,988.40</td> <td>\$4,988.18</td> <td></td> </tr> <tr> <td>Automotive</td> <td>\$52,309.00</td> <td>\$52,493.34</td> <td></td> </tr> <tr> <td>Baby Product</td> <td>\$1,354,243.11</td> <td>\$1,368,901.33</td> <td></td> </tr> <tr> <td>Beauty</td> <td>\$8,114.71</td> <td>\$8,116.06</td> <td></td> </tr> <tr> <td>Books</td> <td>\$1,330,700.80</td> <td>\$1,331,300.71</td> <td></td> </tr> <tr> <td>Business</td> <td>\$38,736.02</td> <td>\$41,916.59</td> <td></td> </tr> <tr> <td>Clothing</td> <td>\$770,046.75</td> <td>\$769,561.89</td> <td></td> </tr> <tr> <td>Collectibles</td> <td>\$709.44</td> <td>\$803.63</td> <td></td> </tr> <tr> <td>Computers</td> <td>\$539,996.53</td> <td>\$540,576.75</td> <td></td> </tr> </tbody> </table>	Product Ca...	Nov 18, 2018	Nov 17, 2018	N	Arts	\$4,988.40	\$4,988.18		Automotive	\$52,309.00	\$52,493.34		Baby Product	\$1,354,243.11	\$1,368,901.33		Beauty	\$8,114.71	\$8,116.06		Books	\$1,330,700.80	\$1,331,300.71		Business	\$38,736.02	\$41,916.59		Clothing	\$770,046.75	\$769,561.89		Collectibles	\$709.44	\$803.63		Computers	\$539,996.53	\$540,576.75		
Product Ca...	Nov 18, 2018	Nov 17, 2018	N																																								
Arts	\$4,988.40	\$4,988.18																																									
Automotive	\$52,309.00	\$52,493.34																																									
Baby Product	\$1,354,243.11	\$1,368,901.33																																									
Beauty	\$8,114.71	\$8,116.06																																									
Books	\$1,330,700.80	\$1,331,300.71																																									
Business	\$38,736.02	\$41,916.59																																									
Clothing	\$770,046.75	\$769,561.89																																									
Collectibles	\$709.44	\$803.63																																									
Computers	\$539,996.53	\$540,576.75																																									

# ML insights

## ABCO Daily Sales Report

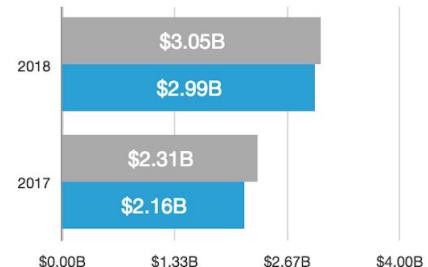
### Daily Revenue

Daily revenue **decreased -0.51%**  
(\$**-\$57,032.99**) on Nov 18, 2018, from \$11.19M to \$11.14M compared to the previous day and is **-1.78% (-\$202,111.70)** below goal of \$11.34M. We are **\$2789.67K (0.334%) above** 30-day average revenue of \$8.35M. We're operating at an **run rate of \$4.06B**.



### YTD Revenue

Year-to-date revenue **increased by 61.95%** (\$**\$1.14B**) from \$1.85B to \$2.99B compared to the same period last year and is **-1.81% (\$55.03M)** below plan of \$3.05B. We are at **98.19%** achievement of YTD goal and **84.61%** achievement for annual goal.



### Callouts By Product and Country

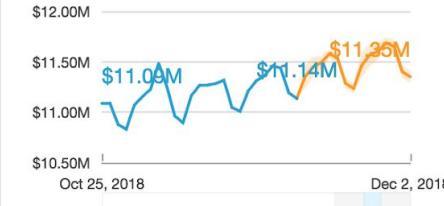
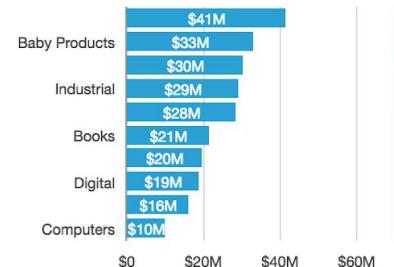
Daily revenue for **Baby Products | Russia** on Nov 22, 2018 was **lower than expected** at \$1.170.58

### Callouts By Customers

Daily revenue for **MULTIDEL INC.** on Nov 22, 2018 was **higher than expected** at \$60.433.95

### Daily Revenue Forecast

Daily revenue is predicted to reach **\$12.02M** by end of the year. We expect to exit the year with an annualized run rate of **\$4.39B**. Total revenue for 2018 is predicted to reach **\$3.47B**, **\$63.06M (-1.78%) below** annual target of \$3.54B.



### Top / Bottom Movers by Product

Top daily revenue **increase** by products are:

### Top / Bottom to Plan Variance by Product

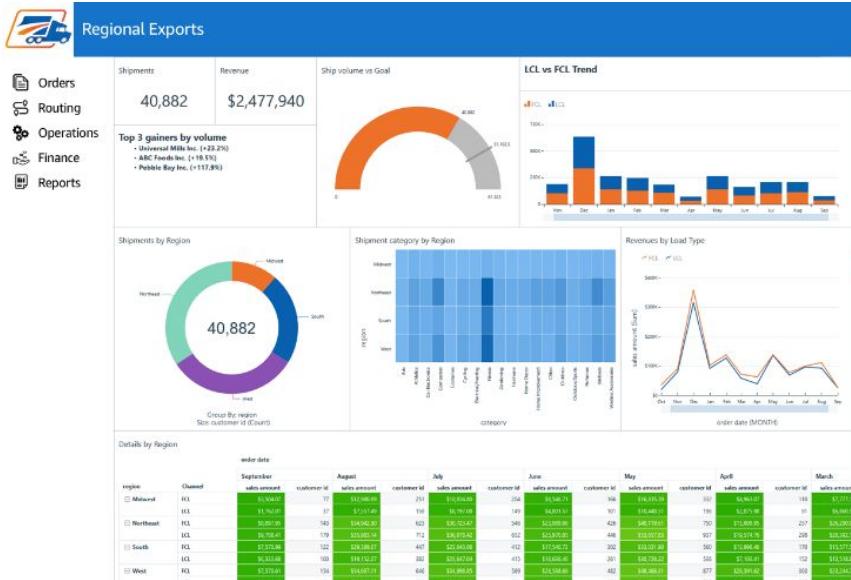
Top products **above** plan for today are:

### Revenue by Product Category

Nov 18, 2018

Nov 17, 2018

# Embed Amazon QuickSight Dashboards



Fully interactive with drill down, filtering, & external links

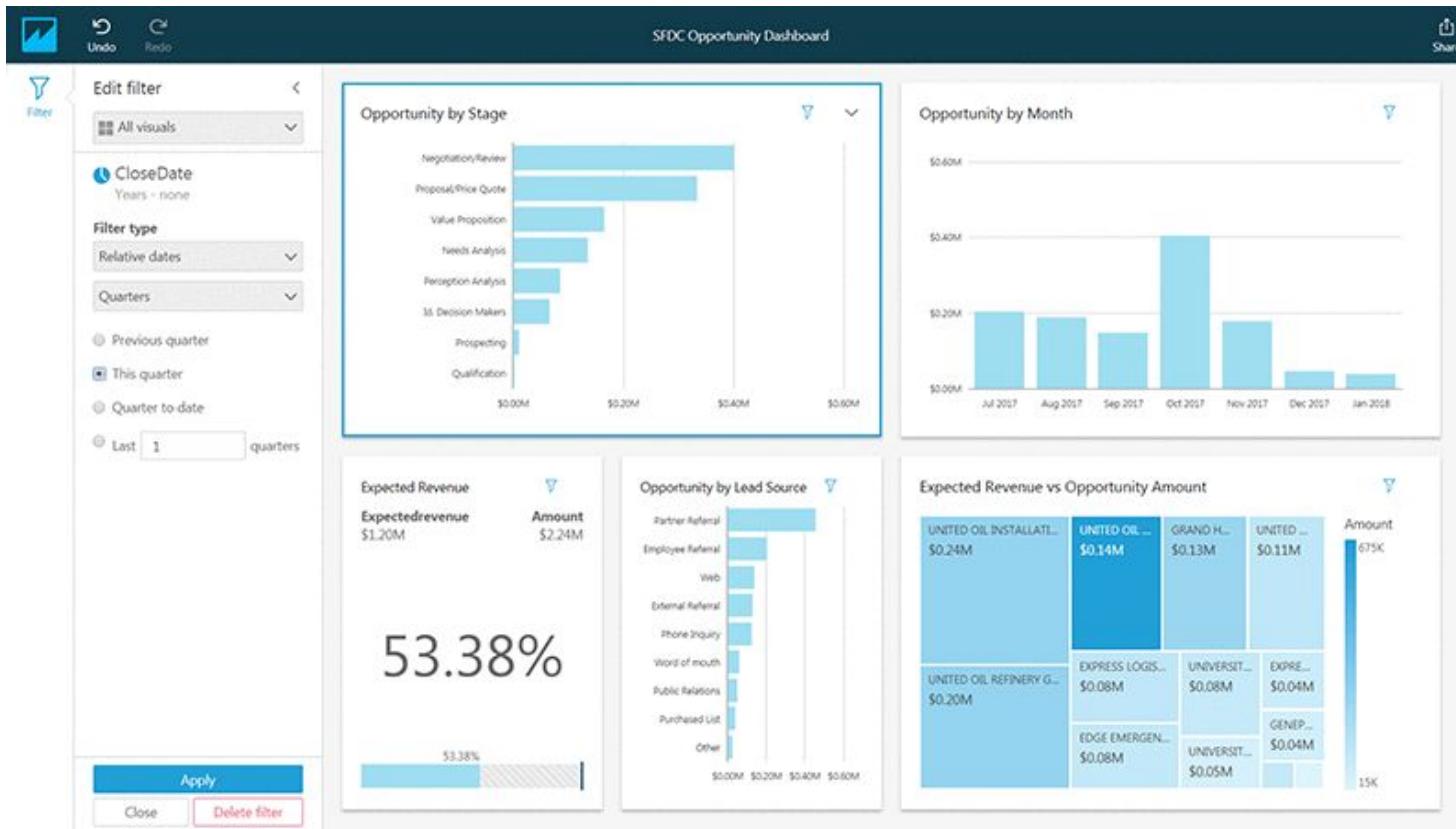
Personalized views with row-level security

No servers to manage, no long-term commitments

Pay for usage with pay-per-session reader pricing

Seamless authentication

# Amazon QuickSight: Examples – Salesforce Analytics



**THANK YOU**