

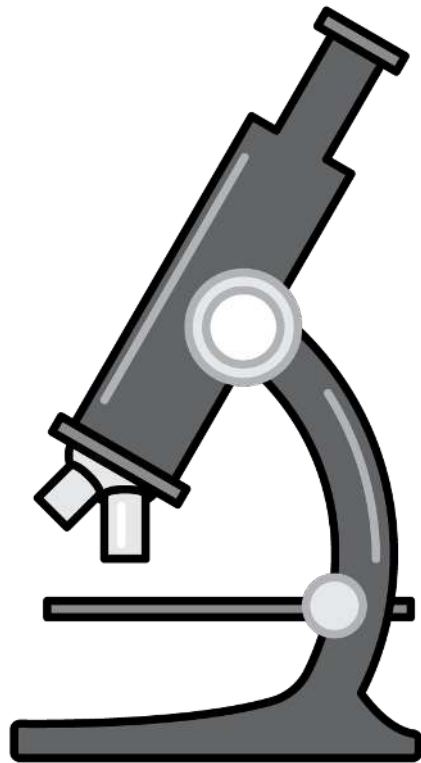
Picking the right AWS backend for your (Java) application

Julien Simon, Principal Technical Evangelist

@jsimon

What to expect

- Writing Java apps on AWS
- Databases
 - Amazon RDS
 - Amazon DynamoDB
- Analytics
 - Hive on Amazon EMR
 - Amazon Athena
 - Amazon Redshift
- Migrating your databases
- Conclusion



Code available at <https://github.com/juliensimon/aws/tree/master/javabackends>

Writing Java apps on AWS

Four deployment options

Amazon EC2

Amazon EC2 Container Service

AWS ElasticBeanstalk

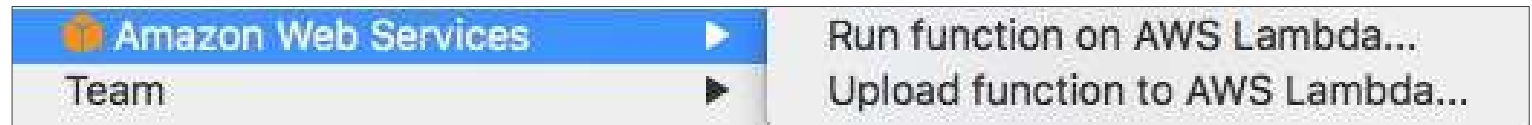
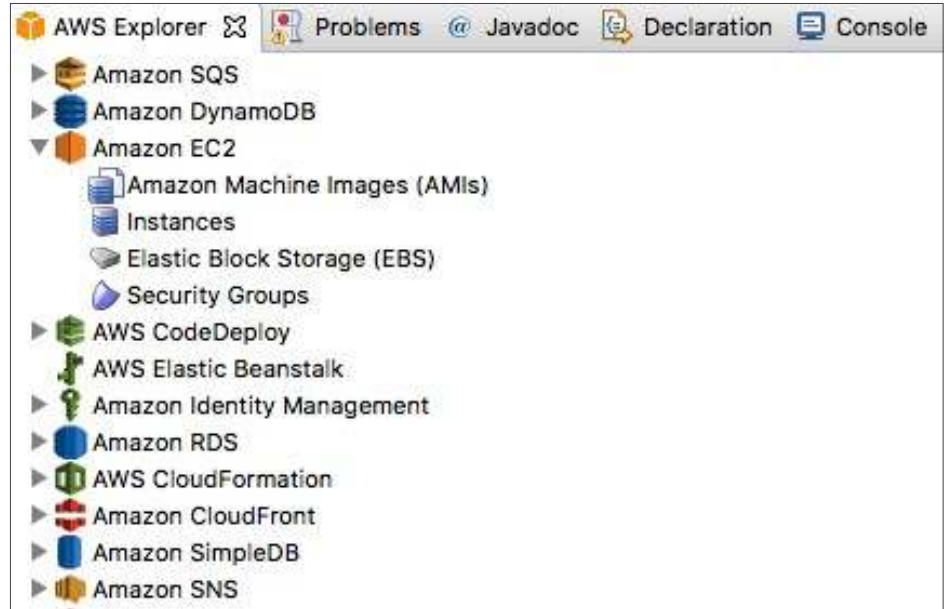
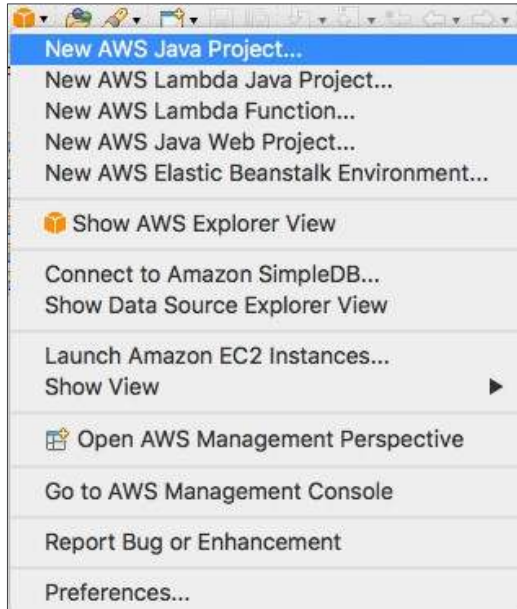
Java 6/7/8 with Tomcat 7/8
Java 7/8 with Glassfish 4

AWS Lambda

Java 8
Open Source frameworks:
Serverless, Gordon, Apex,

Java **SDK** for the AWS API (Java 1.6+)

AWS plugin for Eclipse



3rd party plugins for IntelliJ IDEA

- AWS Elastic Beanstalk Integration

<https://plugins.jetbrains.com/plugin/7274-aws-elastic-beanstalk-integration>

- AWS CloudFormation

<https://plugins.jetbrains.com/plugin/7371-aws-cloudformation>

- AWS Manager –2 years old :-/

<https://plugins.jetbrains.com/plugin/4558-aws-manager>

Managing credentials

- Please **do not hardcode** them in your application
- Please **do not store** them on EC2 instances
- It WILL end in tears

- AWS credentials: use **IAM Roles**
- Backend credentials: use the **EC2 Parameter Store**
 - Automatic encryption with Amazon KMS

<https://docs.aws.amazon.com/sdk-for-java/v1/developer-guide/credentials.html>

<https://docs.aws.amazon.com/AWSJavaSDK/latest/javadoc/index.html?com/amazonaws/auth/AWSCredentialsProvider.html>

<https://docs.aws.amazon.com/systems-manager/latest/userguide/systems-manager-paramstore.html>

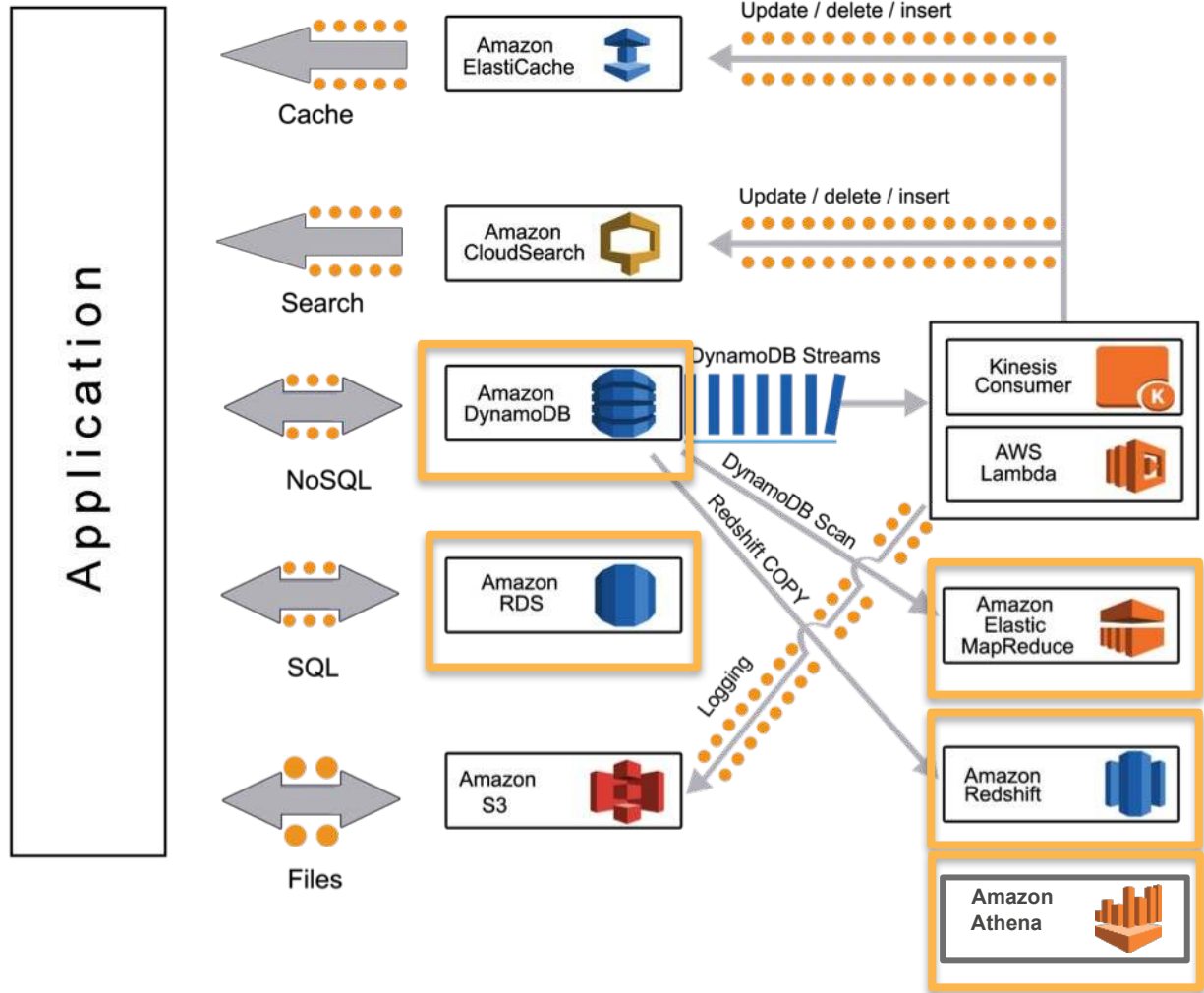




One backend to rule them all?

... and in darkness bind them

Reference architecture



Databases

Amazon Relational Database Service

AWS Free Tier

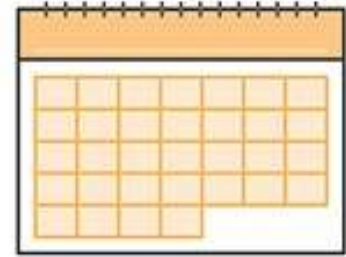
Launch



No infrastructure
management



Application
compatibility

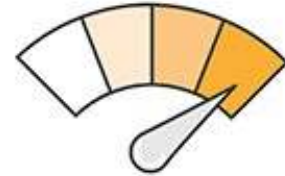


Instant provisioning



Cost-effective

99.95% SLA



Scale up/down

Amazon RDS – 6 Database Engines

- Amazon Aurora
- MySQL 5.5.46 → 5.7.19
- MariaDB 10.0.17 → 10.1.24
- PostgreSQL 9.3.12-R1 → 9.6.5-R1
- Oracle 11.2.0.4.v1 → 12.1.0.2.v9
- SQL Server 2008 → 2016



Selected Amazon RDS customers



vodafone

intuit



SEGA®



Kempinski
HOTELIERS SINCE 1897



OUTBACK
STEAKHOUSE®



Newsweek
& THE DAILY BEAST



P
The
Washington Post
Company

Trinity Mirror plc

coursera



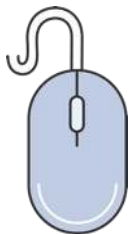
Amazon Aurora demo

Java SDK: <https://docs.aws.amazon.com/AWSJavaSDK/latest/javadoc/com/amazonaws/services/rds/AmazonRDSClient.html>

JDBC drivers: <https://docs.aws.amazon.com/elasticbeanstalk/latest/dg/java-rds.html>

Amazon DynamoDB

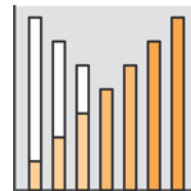
AWS Free Tier



Fully Managed NoSQL



Key-Value, Document, Object



Scales to Any Workload



Fast and Consistent



Access Control



Event Driven Programming

<https://aws.amazon.com/dynamodb/>
http://www.allthingsdistributed.com/2007/10/amazons_dynamo.html
<http://www.allthingsdistributed.com/2012/01/amazon-dynamodb.html>



Case Study – Expedia

“

With DynamoDB we were up and running in a less than a day, and there is no need for a team to maintain it.

Kuldeep Chowhan
Principal Engineer, Expedia



”

Expedia is a leader in the \$1 trillion travel industry, with an extensive portfolio that includes some of the world's most trusted travel brands.

- Expedia's **real-time analytics** application collects data for its “test & learn” experiments on Expedia sites.
- The analytics application processes **~200 million** messages daily.
- Ease of **setup**, **monitoring**, and **scaling** were key factors in choosing DynamoDB.

Amazon DynamoDB demo

Low-level API (key/value): getItem, putItem, updateItem
batchGetItem, batchWriteItem
query, scan

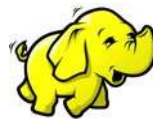
High-level API (ORM-like): DynamoDBMapper

Java SDK: <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/CodeSamples.Java.html>
<https://docs.aws.amazon.com/sdk-for-java/v1/developer-guide/examples-dynamodb.html>
<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/DynamoDBLocal.html>

Analytics

Amazon Elastic Map Reduce (EMR)

- Apache Hadoop, Spark, Hive, etc.
- **Managed** service
- Easy to **start**, **resize** & **terminate** clusters
- Cost-efficient, especially with **Spot Instances**
- Integration with **backends**



Case study – FINRA



FINRA, the primary regulatory agency for stock brokers in the US, uses AWS extensively in their IT operations and has migrated key portions of its technology stack to AWS including Market Surveillance and Member Regulation.

For market surveillance, each night FINRA loads approximately 35 billion rows of data into Amazon S3 and Amazon EMR (up to 10,000 nodes) to monitor trading activity on exchanges and market centers in the US.

Hive demo

Java SDK: <https://docs.aws.amazon.com/emr/latest/ManagementGuide/calling-emr-with-java-sdk.html>

JDBC: <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/HiveJDBCdriver.html>

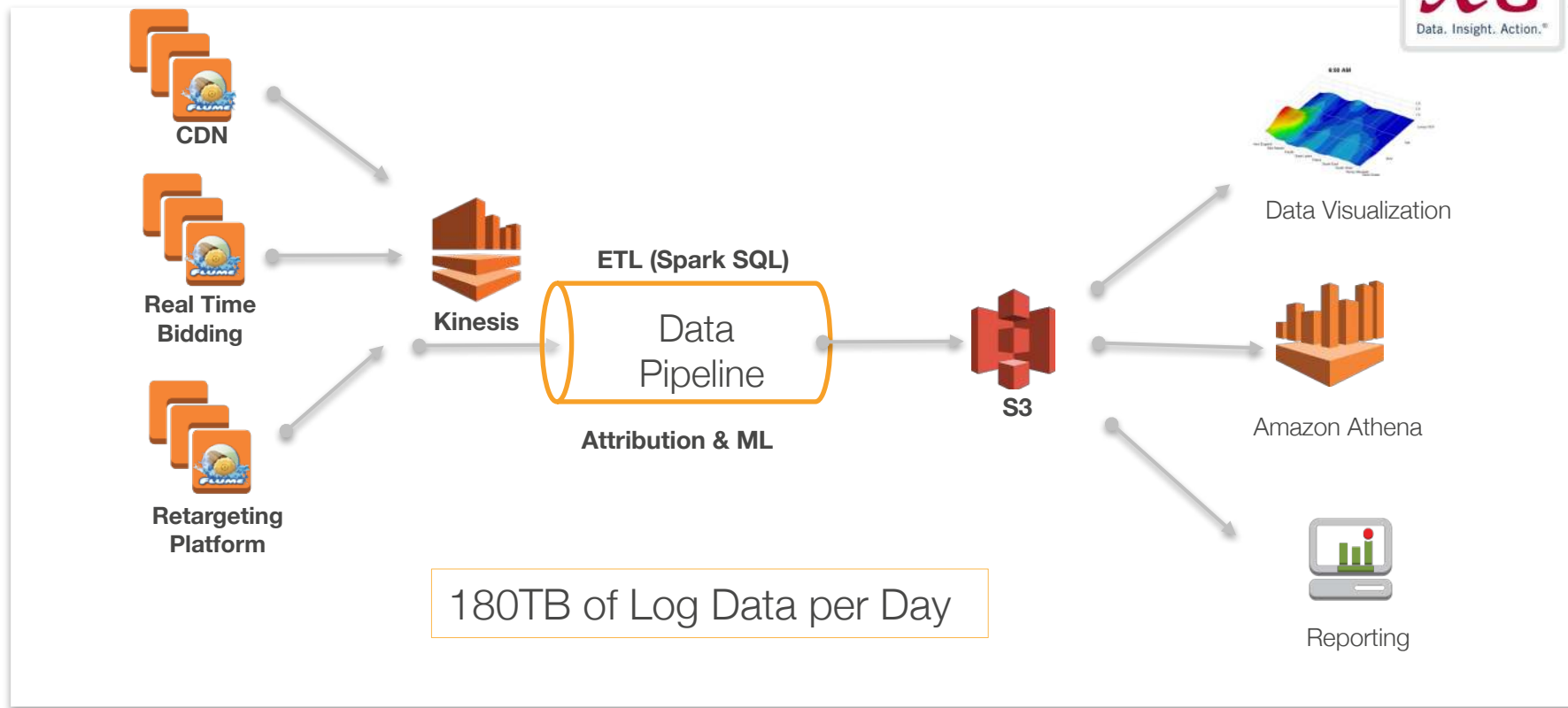
Amazon Athena

- Run read-only SQL queries on S3 data
- No data loading, no indexing, no nothing
- No infrastructure to create, manage or scale
- Service based on Presto
- Table creation: Apache Hive Data Definition Language
- ANSI SQL operators and functions: what Presto supports, with a few exceptions

Data formats supported by Athena

- **Unstructured**
 - Apache logs, with customizable regular expression
- **Semi-structured**
 - delimiter-separated values (CSV, OpenCSV)
 - Tab-separated values (TSV)
 - JSON (including CloudTrail logs)
- **Structured**
 - Apache Parquet <https://parquet.apache.org/>
 - Apache ORC <https://orc.apache.org/>
 - Apache Avro <https://avro.apache.org/>
- **Compression** (LZO, Snappy, Zlib, GZIP) & **partitioning**

Case Study – DataXu



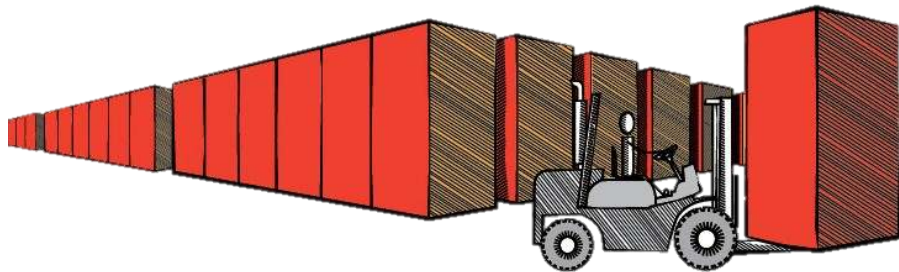
Amazon Athena demo

Java SDK: <https://docs.aws.amazon.com/athena/latest/ug/code-samples.html>

JDBC: <https://docs.aws.amazon.com/athena/latest/ug/connect-with-jdbc.html>

Amazon Redshift AWS Free Tier

- Relational data warehouse
- SQL is all you need to know
- Fully managed
- Massively parallel
- Petabyte scale
- As low as \$1000/TB/year
- Athena-like capabilities with Redshift Spectrum



<https://aws.amazon.com/redshift>

<http://www.allthingsdistributed.com/2012/11/amazon-redshift.html>

Intro to Amazon Redshift Spectrum <https://www.youtube.com/watch?v=gchd2sDhSuY>

What customers says about Amazon Redshift



“Redshift is **twenty times faster than Hive**” (5x – 20x reduction in query times) [link](#)



...[Redshift] performance has blown away everyone here (we generally see **50-100x speedup over Hive**). [link](#)



We regularly process **multibillion row datasets** and we do that in a matter of **hours**. [link](#)



“Queries that used to take **hours** came back in **seconds**. Our analysts are orders of magnitude more productive.” (20x – 40x reduction in query times) [link](#)



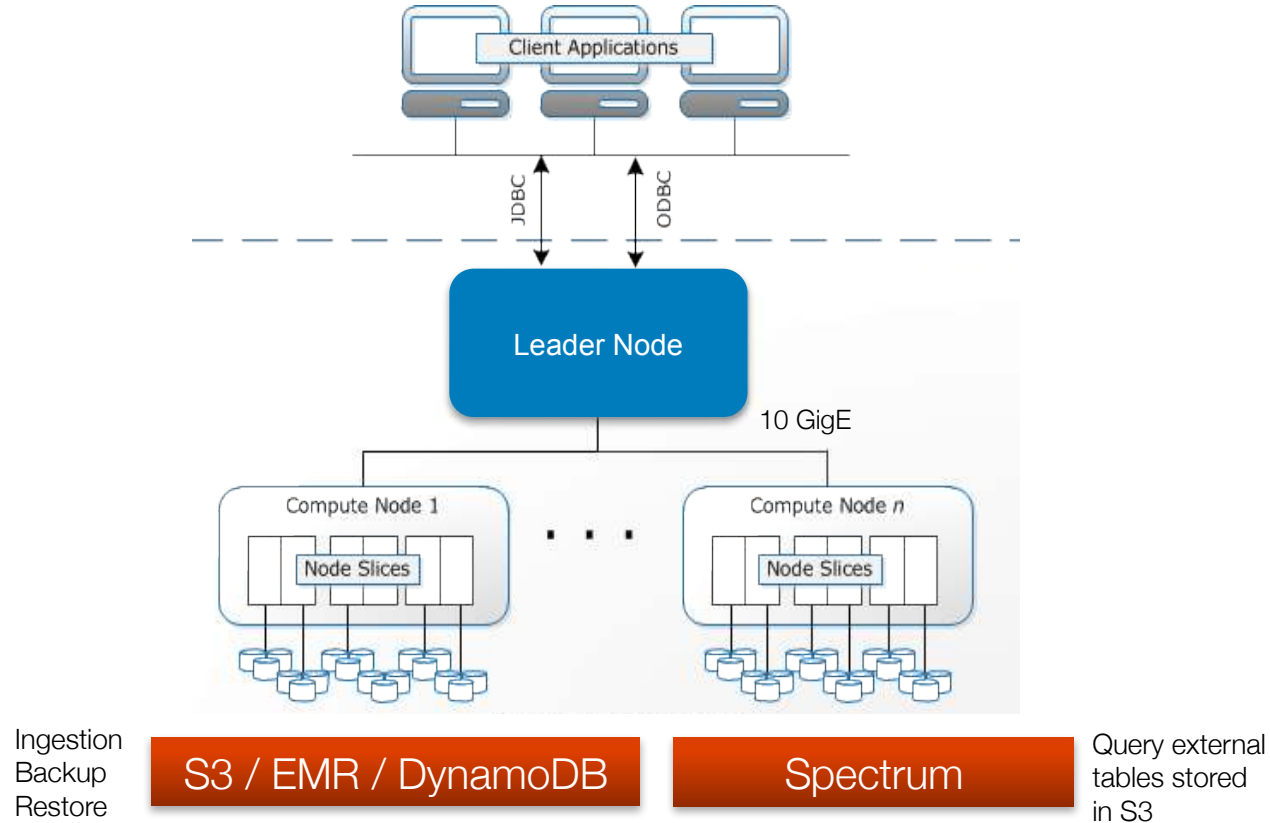
“Did I mention it's **ridiculously fast**? We'll be using it immediately to provide our analysts an **alternative to Hadoop**.”



“Team played with Redshift today and concluded it is ***** awesome.
Un-indexed complex queries returning in < 10s.”



Amazon Redshift Architecture



Amazon Redshift demo

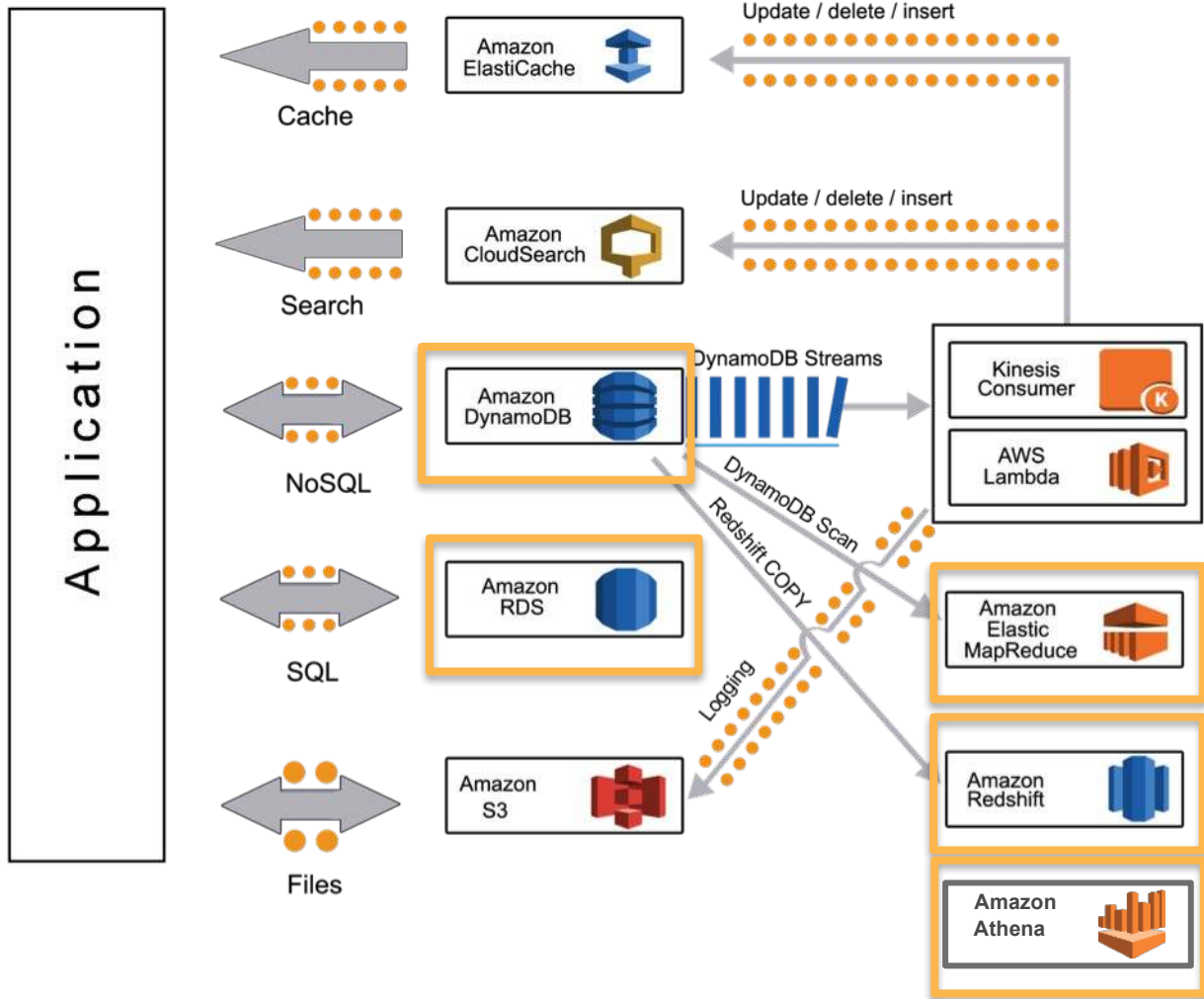
Java SDK: <https://docs.aws.amazon.com/redshift/latest/mgmt/using-aws-sdk-for-java.html>

JDBC driver: <https://docs.aws.amazon.com/redshift/latest/mgmt/configure-jdbc-connection.html>

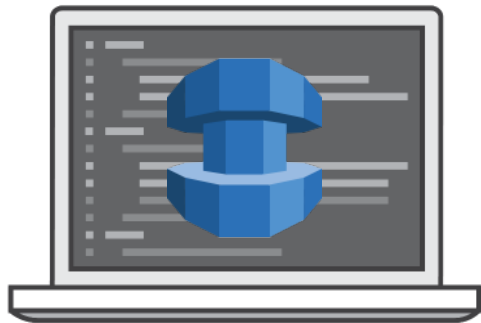
	Amazon Redshift	Amazon Athena	Amazon EMR		
			Presto	Spark	Hive
Use case	Optimized for data warehousing	Ad-hoc Interactive Queries	Interactive Query	General purpose (iterative ML, RT, ..)	Batch
Scale/throughput	~Nodes Automatic (Spectrum)	Automatic / No limits	~ Nodes		
AWS Managed Service	Yes	Yes	Yes		
Storage	Local storage Amazon S3 (Spectrum)	Amazon S3	Amazon S3, HDFS		
Data format	Columnar storage, data compression, and zone maps	CSV, TSV, JSON, Parquet, ORC, Apache Web log	Framework dependent		
Metadata	Amazon Redshift managed	Athena Catalog Manager	Hive Meta-store		
BI tools supports	Yes (JDBC/ODBC)	Yes (JDBC)	Yes (JDBC/ODBC & Custom)		
Access controls	Users, groups, and access controls	AWS IAM	Integration with LDAP		
UDF support	Yes (Scalar)	No	Yes		

Migrating your databases

Reference architecture



AWS Schema Conversion Tool



- Converts your **tables**, **views**, **stored procedures**, and **data manipulation language** to RDS or Amazon Redshift
- **Highlights** where manual edits are needed

Source Database	Target Database on Amazon RDS
Microsoft SQL Server (version 2008 and later)	Amazon Aurora (MySQL or PostgreSQL), Microsoft SQL Server, MySQL, PostgreSQL
MySQL (version 5.5 and later)	Amazon Aurora (PostgreSQL), MySQL, PostgreSQL You can migrate schema and data from MySQL to an Amazon Aurora (MySQL) DB cluster without using AWS SCT. For more information, see Migrating Data to an Amazon Aurora DB Cluster .
Oracle (version 10.2 and later)	Amazon Aurora (MySQL or PostgreSQL), MySQL, Oracle, PostgreSQL
PostgreSQL (version 9.1 and later)	Amazon Aurora (MySQL), MySQL, PostgreSQL

Source Database	Target Database on Amazon Redshift
Greenplum Database (version 4.3 and later)	Amazon Redshift
Microsoft SQL Server (version 2008 and later)	Amazon Redshift
Netezza (version 7.0.3 and later)	Amazon Redshift
Oracle (version 10 and later)	Amazon Redshift
Teradata (version 13 and later)	Amazon Redshift
Vertica (version 7.2.2 and later)	Amazon Redshift

AWS Database Migration Service



- ✓ Move data to the **same** or different **database** engine
- ✓ Move data to RDS, Redshift, DynamoDB or S3
- ✓ Keep your apps **running** during the migration
- ✓ Start your first migration in **10** minutes or less
- ✓ Replicate **within**, **to**, or **from** Amazon EC2 or RDS

<https://aws.amazon.com/dms/>

http://docs.aws.amazon.com/dms/latest/userguide/CHAP_Introduction.Sources.html

http://docs.aws.amazon.com/dms/latest/userguide/CHAP_Introduction.Targets.html

<https://aws.amazon.com/blogs/database/database-migration-what-do-you-need-to-know-before-you-start/>

Conclusion

AWS is a **rich** and **lively** environment for developers

Your choice of backends: relational, NoSQL, Big Data, analytics

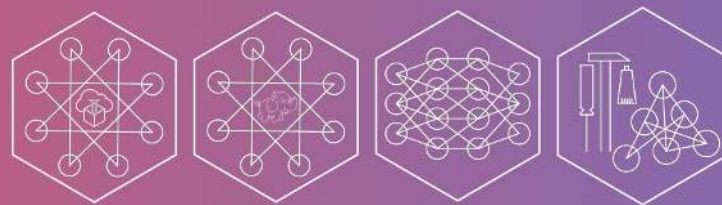
The tools you need, **with less or no infrastructure drama**

Built-in **high availability, scalability, security & compliance**

Focus on **creativity** and **productivity**, not on plumbing

Semaine IA sur AWS

Du 11 au 15 novembre



<https://aws.amazon.com/fr/events/>

Thank you!

<http://aws.amazon.com/evangelists/julien-simon>

@julsimon