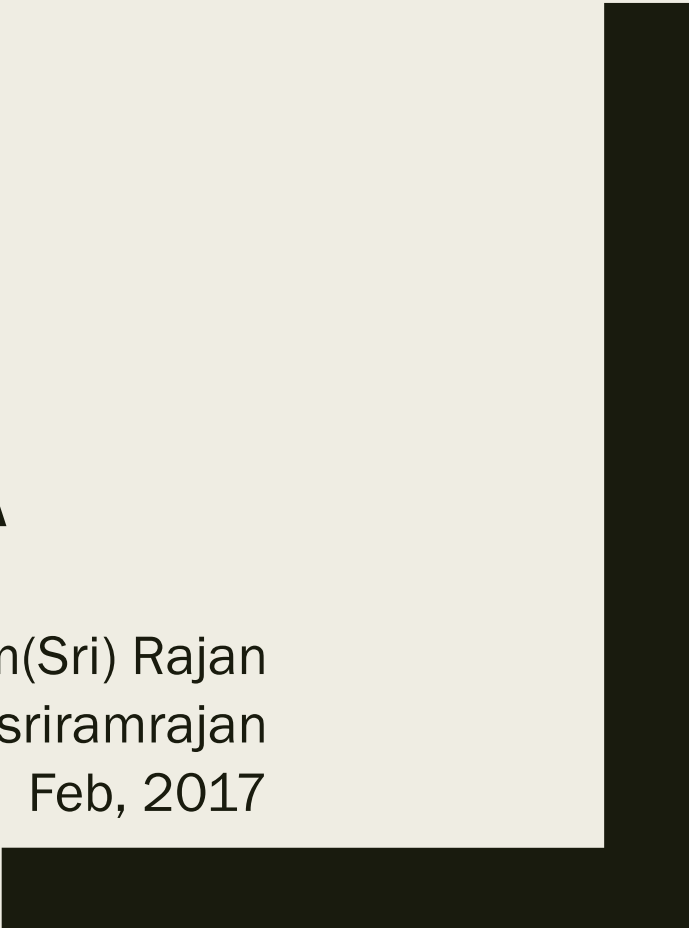


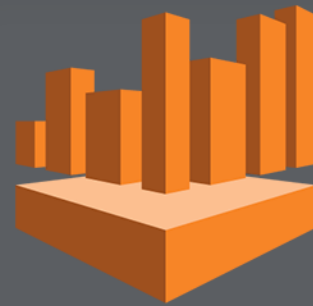


AWS ATHENA

Sriram(Sri) Rajan
@sriramrajan
Feb, 2017



```
[~]$ whoami
```



Amazon Athena





Amazon RDS



DynamoDB



Amazon Redshift



Amazon EMR



Amazon Athena

Athena - Basics

- *Serverless* interactive query service
- PrestoDB (Facebook) implementation
- Works on data stored in S3
 - But is still quite fast

Athena - Basics

- Use Apache Hive Data definition language (DDL)
- Supports ANSI SQL
- Pay as you go model (like Lambda)
 - Charges based on data scanned by the query. 5\$ per TB of data scanned
 - No charge for creating tables
 - S3 costs are separate

Athena – Concepts

- Tables - Metadata that describes your data similar to traditional database tables.
- Tables are like views.
 - For e.g. You can delete table definitions without impacting the underlying S3 data

Athena – Concepts

- Databases - Logical grouping of tables. (catalog or a namespace)
- ****SerDe**** - Serializer/Deserializer - libraries that tell Hive how to interpret data formats.
 - Apache Web Logs, CSV, JSON, ORC (Optimized Row Columnar) etc
 - Data can also be compressed in GZIP to save costs

Athena Limitations

- No support for transactions. (This includes any transactions found in Hive or Presto)
- When you create, update, or delete tables, those operations are guaranteed ACID-compliant.

Athena Limitations

- Limited to 2 AWS regions
 - You can use Athena to query underlying Amazon S3 bucket data that's in a different region
- API/SDK support (Limited to JDBC driver)

Athena – Use cases

- Querying of any data in S3. If latency is not critical and queries can be run in the background
 - For e.g analysing log data in S3

Athena – Use cases

- Go serverless across the stack. API gateway can be used to accept requests which are handled by Lambda which in turn can leverage Athena for queries. The only persistent service used will be S3.

Athena – Use cases

- Mix and match AWS services. Use an on-demand EMR cluster to process data and dump results to S3. Then use Athena to create adhoc tables and run reports.

Athena – Use cases

- “Facebook uses Presto for interactive queries against several internal data stores, including their 300PB data warehouse. Over 1,000 Facebook employees use Presto daily to run more than 30,000 queries that in total scan over a petabyte each per day”



Athena – Takeaways

- <https://github.com/srirajan/athena>



- AWS Free Tier



References & Links

- <http://docs.aws.amazon.com/athena/latest/ug/what-is.html>
- <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>