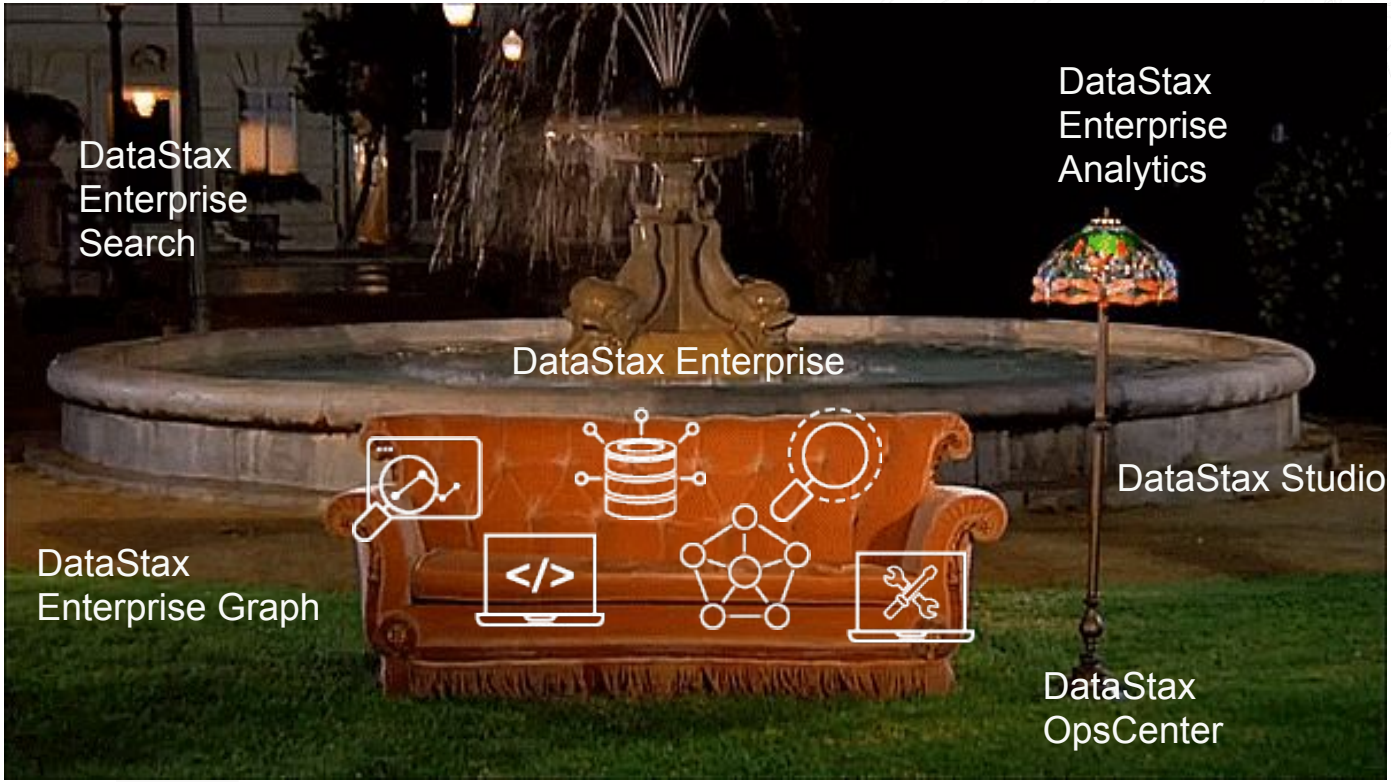
Meet my Friend DataStax Enterprise!

DataStax Meetup
Jan. 15, 2019
Amanda Moran
Dev Advocate

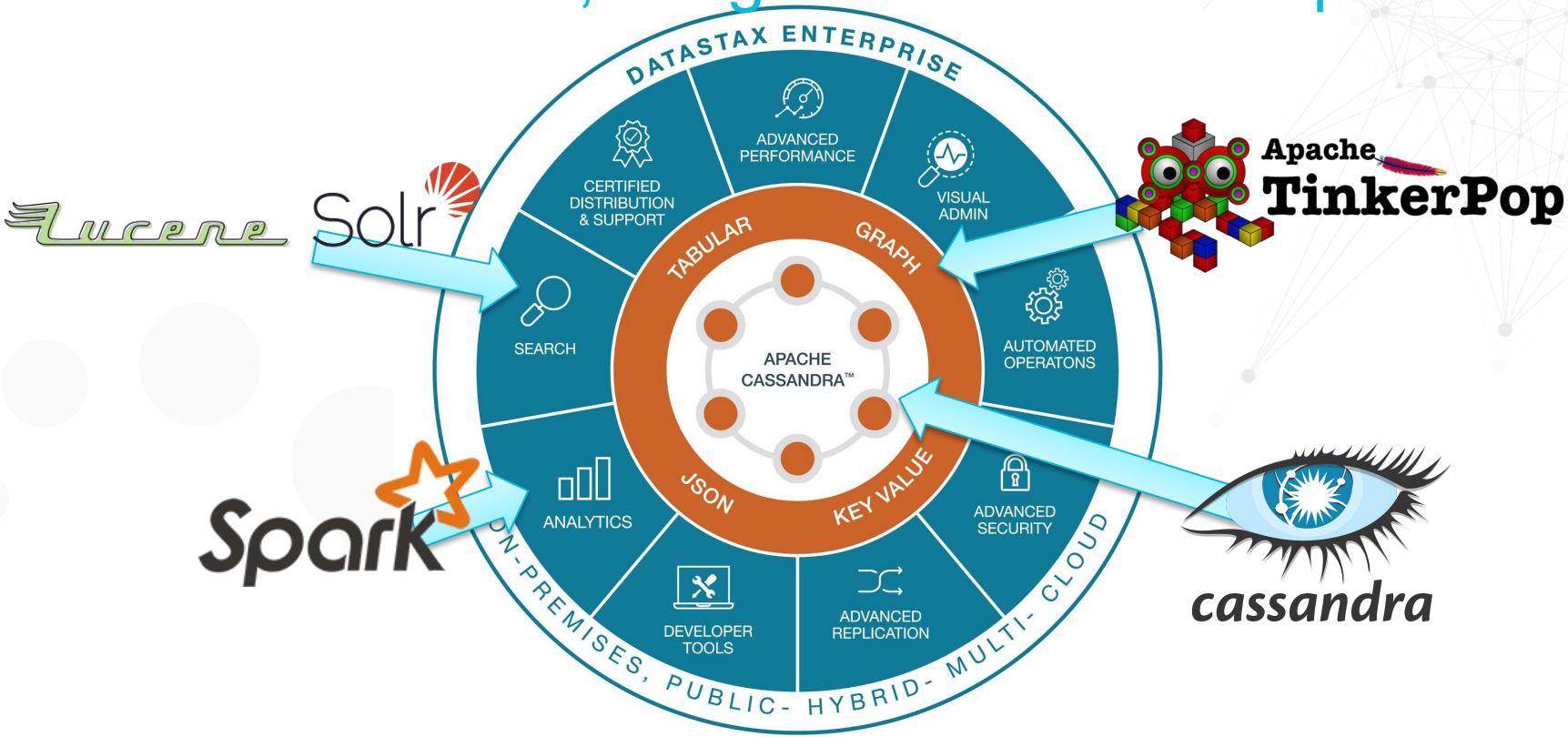
But first… A Little About Amanda

- Graduated with MS in Computer Science and Engineering from Santa Clara University in 2012
- Worked as a Software Engineer for 6 years at Lockheed Martin, HP, Teradata, Esgyn and now DataStax as a Developer Advocate
- Apache Committer, PMC Member, and initial contributor to all installation and deployment work for Apache Trafodion
- Keywords: Disney, Cloud, Dogs, Veggies, Linux, Databases, Big Data, Analytics, Testing, and Running

What Are We Here to Learn About?



OSS Foundations, Integrated for the Enterprise



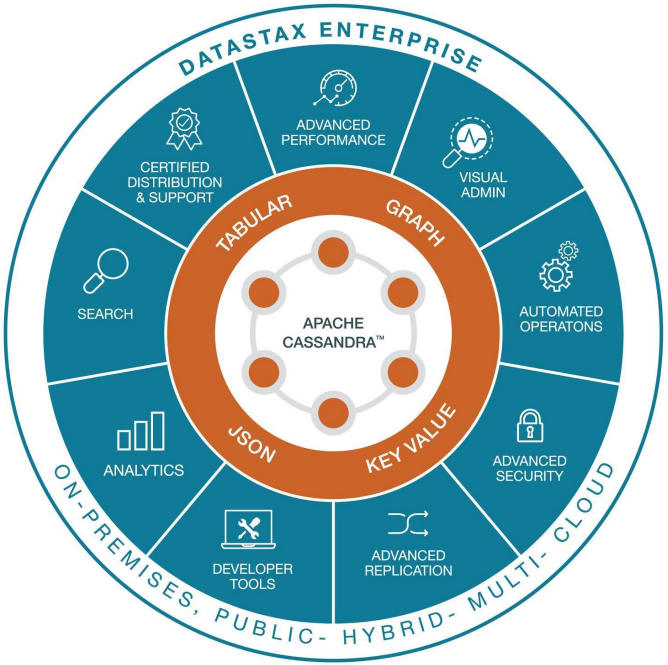
Apache Cassandra, Spark, Lucene, Solr, TinkerPop ® Apache Software Foundation

DataStax Enterprise and Rachel

- Everything is based around Rachel just like everything in DataStax Enterprise is based on our core: Apache Cassandra

Integrated Data Platform for Cloud Applications

- DataStax Enterprise
  - Core
  - Search
  - Analytics
  - Graph
- DataStax Studio
- DataStax Drivers
- DataStax OpsCenter



What is DataStax Enterprise?

- The best distribution of Apache Cassandra™
- Production certified Cassandra -- Enterprise Ready!
- Advanced Performance -- 2x Reads and >2x Writes
- Advanced Security
- Advanced Replication --NodeSync
- Advanced Testing

DataStax Enterprise Search and Ross

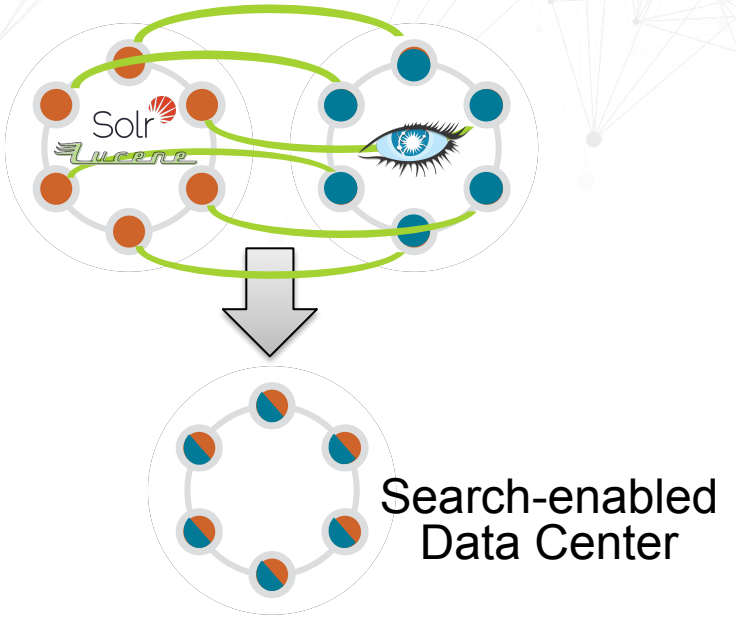
- Ross is a people pleaser, he will do for you what others won't
- DataStax Enterprise Search gives you the ability to do queries you couldn't before -- it is the ultimate people pleaser (just not annoying)

Lucene Solr

DSE Search: Apache Solr / Lucene + Cassandra

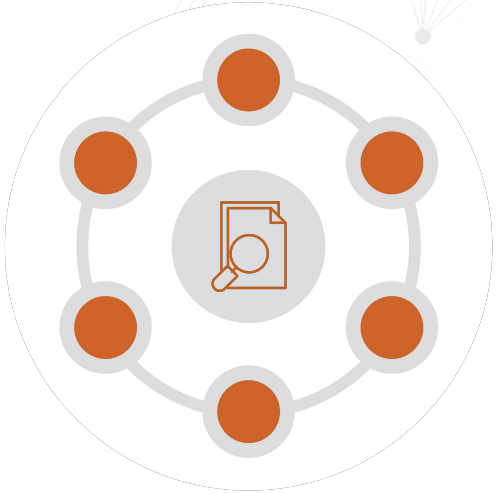
- Supports ad-hoc queries not supported by Cassandra
- Full text search, Faceting, Stemming
- Geospatial Search
- Live indexing engine
- No separate search cluster
- No ETL or sync to build and maintain

- Search indexes co-located with Cassandra



DSE Search Enhancements -- in 6.0++

- CQL-enabled search
- Like queries
- Eliminate Solr syntax in CQL
- Support for satisfying standard CQL index queries with DSE Search
  - Predicates
  - Sorting
  - Limits



DataStax ENTERPRISE 6

DSE Search Queries

- SOLR Based Queries (releases through DSE 5.1)

```
SELECT title, release_year FROM killrvideo.videos WHERE solr_query =
'title:inception';
```

- CQL Search Queries (new in DSE 6)

```
SELECT title, release_year FROM killrvideo.videos WHERE title="inception";
```

- New additional keywords IN, LIKE supported in CQL

```
SELECT * FROM killrvideo.videos WHERE title LIKE "incept%";
```

DataStax Enterprise Search: Demo

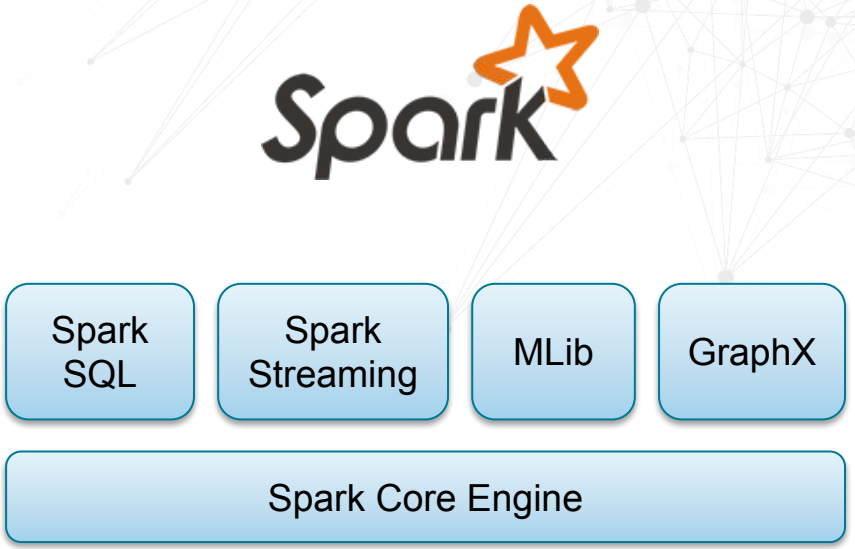
- Let's find Central Perk with DSE Search

DataStax Enterprise Analytics and Chandler

- Chandler was a Data Scientist before we even knew what that meant
- DataStax Enterprise Analytics gives you the ability with Apache Spark to do full SQL queries and ML

Apache Spark at a Glance

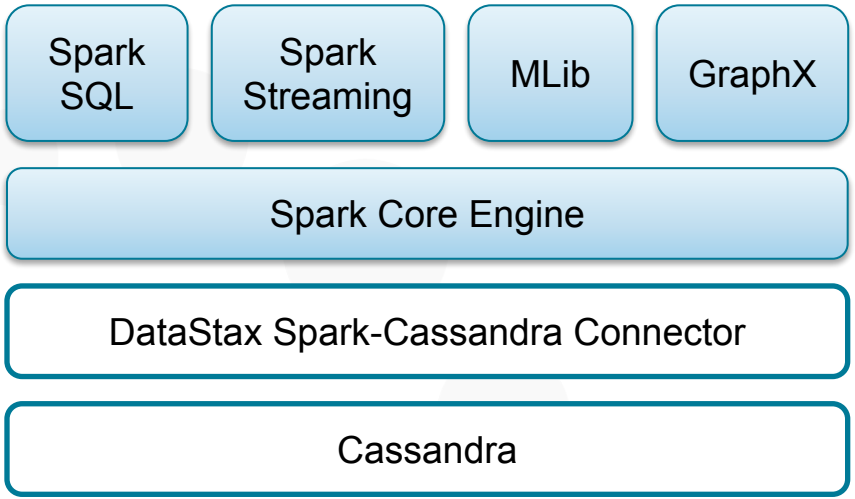
- Distributed computing framework
- Similar cluster structure to Cassandra
- Generalized DAG execution
- Easy dataset abstraction
- Integrated SQL Queries
- Libraries for Streaming, Machine Learning



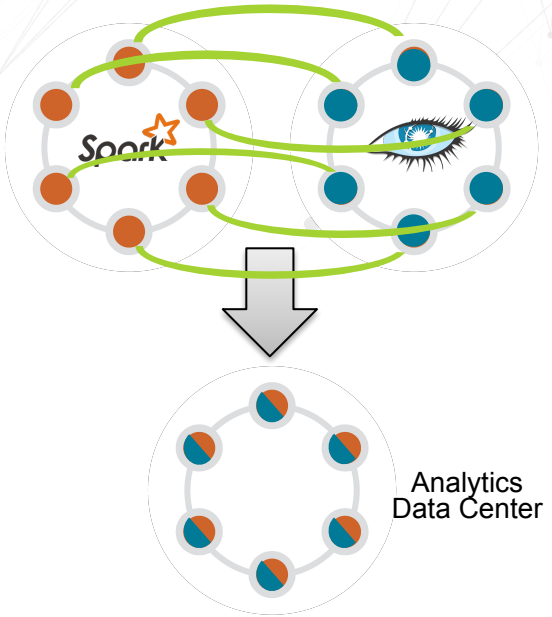
Apache Spark ® Apache Software Foundation

DSE Analytics: Spark + Cassandra

Read/write Cassandra data from Spark via DataStax Connector



Co-located Spark with Cassandra



DSE Analytics: SparkSQL

- SQL query engine on top of Spark
  - Hive compatible (JDBC, UDFs, types, metadata, etc.)
  - Support for in-memory processing
  - Pushdown of predicates to Cassandra when possible
- Provide a single interface for working with structured data including Apache Hive, Parquet, JSON files and any DataSource
- Friendly abstraction layer for Spark Batch

```
SELECT merchant,
  sum(amount) AS total_amount,
  count(amount) AS total_trans
  FROM dsbank.transactions
  GROUP BY merchant
  ORDER BY total_amount desc
  limit 10;
```

DataStax Enterprise Analytics: Demo

- Let’s do some SQL style queries on our Friends tables

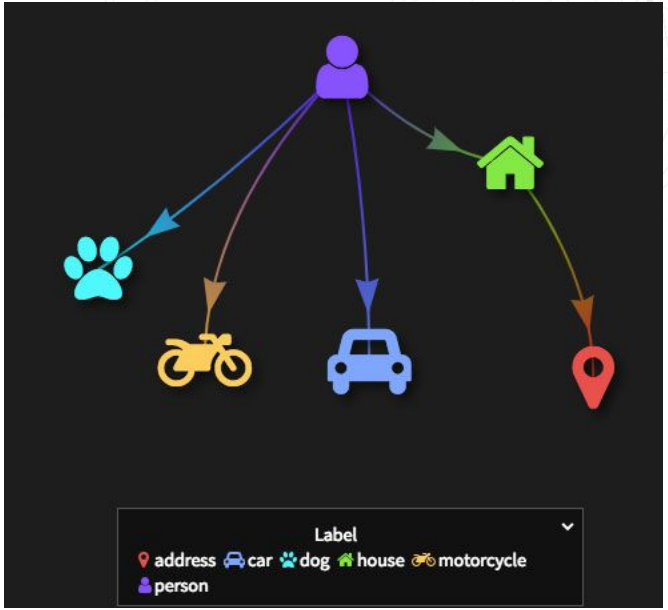
DataStax Enterprise Graph and Phoebe

- Phoebe looks at the world a bit differently
- Modeling your data with a graph database like DataStax Enterprise Graph will allow you to use and see your data in a whole different way.

Apache TinkerPop

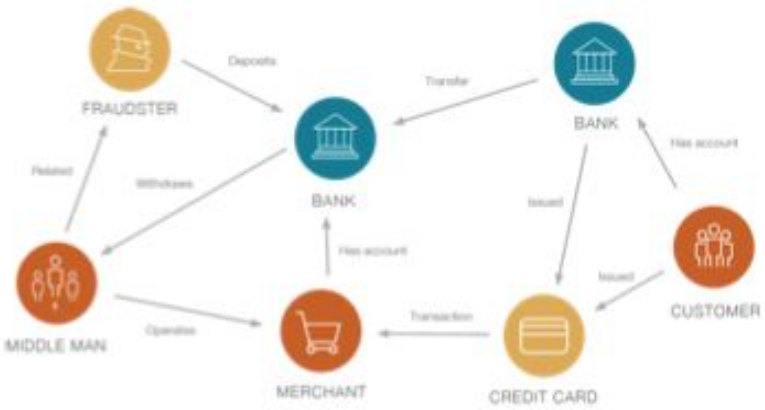
DSE Graph

- Scalable, distributed graph DB
- Optimized for storing, traversing, and querying
- DSE Analytics and DSE Search integrated
- Value data between relationships
- Use cases: Customer 360, Recommendations, Fraud Detection



Gremlin Query Language (GQL)

```
g.V().has("name","gremlin").
      repeat(in("manages")).until(has("title","ceo")).
      path().by("name")
>> The management chain from Gremlin to the CEO
```



Comparing SQL and Gremlin

```
SELECT TOP (5) [t14].[ProductName]
  FROM (SELECT COUNT(*) AS [value],
        [t13].[ProductName]
    FROM [customers] AS [t0]
  CROSS APPLY (SELECT [t9].[ProductName]
        FROM [orders] AS [t1]
      CROSS JOIN [order details] AS [t2]
      INNER JOIN [products] AS [t3]
          ON [t3].[ProductID] = [t2].[ProductID]
      CROSS JOIN [order details] AS [t4]
      INNER JOIN [orders] AS [t5]
          ON [t5].[OrderID] = [t4].[OrderID]
       LEFT JOIN [customers] AS [t6]
          ON [t6].[CustomerID] = [t5].[CustomerID]
      CROSS JOIN ([orders] AS [t7]
            CROSS JOIN [order details] AS [t8]
            INNER JOIN [products] AS [t9]
                ON [t9].[ProductID] = [t8].[ProductID])
        WHERE NOT EXISTS(SELECT NULL AS [EMPTY]
                FROM [orders] AS [t10]
              CROSS JOIN [order details] AS [t11]
              INNER JOIN [products] AS [t12]
                  ON [t12].[ProductID] = [t11].[ProductID]
                WHERE [t9].[ProductID] = [t12].[ProductID]
                  AND [t10].[CustomerID] = [t0].[CustomerID]
                  AND [t11].[OrderID] = [t10].[OrderID])
          AND [t6].[CustomerID] <> [t0].[CustomerID]
          AND [t1].[CustomerID] = [t0].[CustomerID]
          AND [t2].[OrderID] = [t1].[OrderID]
          AND [t4].[ProductID] = [t3].[ProductID]
          AND [t7].[CustomerID] = [t6].[CustomerID]
          AND [t8].[OrderID] = [t7].[OrderID]) AS [t13]
      WHERE [t0].[CustomerID] = N'ALFKI'
    GROUP BY [t13].[ProductName]) AS [t14]
ORDER BY [t14].[value] DESC
```

vs

```
g.V().has("customerId", x).as("customer").
      out("ordered").out("contains").
      out("is").aggregate("products").
      in("is").in("contains").in("ordered").
      where(neq("customer")).out("ordered").
      out("contains").out("is").
      where(without("products")).
      groupCount().order(local).by(valueDecr)
```

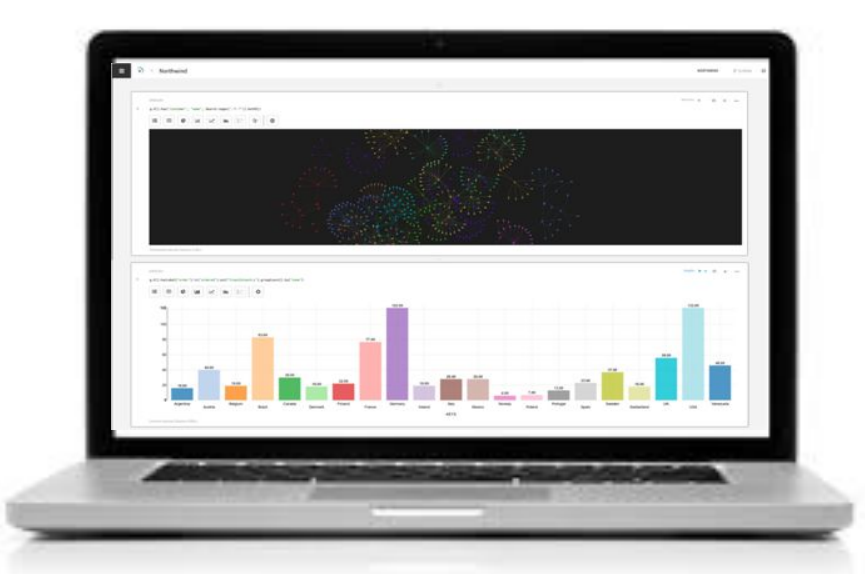
DataStax Enterprise Graph: Demo

- Let's use DSE Graph to analyze the relationships between the Friends

DataStax Enterprise Studio and Joey

- Joey is an actor --Okay basically I got nothing here
- Studio is a notebook style interface to help you present and troubleshoot queries.

DataStax Studio



- Notebook style interface
- CQL, Gremlin, Spark SQL
- Data visualization
- Auto-completion

#DataStaxMeetup

DataStax Enterprise OpsCenter and Monica

- Organization++ describes Monica. Clean and organized to the max
- DataStax OpsCenter provides a UI to monitor your cluster

DSE OpsCenter

Visual management tool for DSE

- Create / upgrade clusters visually
- Built-in automatic failover
- Monitoring and alerts
- Backup and restore --new destinations in 6.7
- Capacity planning

Information and Links

- Get this demo: https://github.com/amandamoran
- Learn more about DataStax: https://academy.datastax.com/
- Follow me on Twitter: @AmandaDataStax
- Connect with me on Linkedin

DATASTAX ACADEMY

DATASTAX
The power behind the moment.

Amanda
@AmandaDataStax

@DataStaxAcademy

#DataStaxMeetup

Thank you!

Questions?

@DataStaxAcademy

#DataStaxMeetup