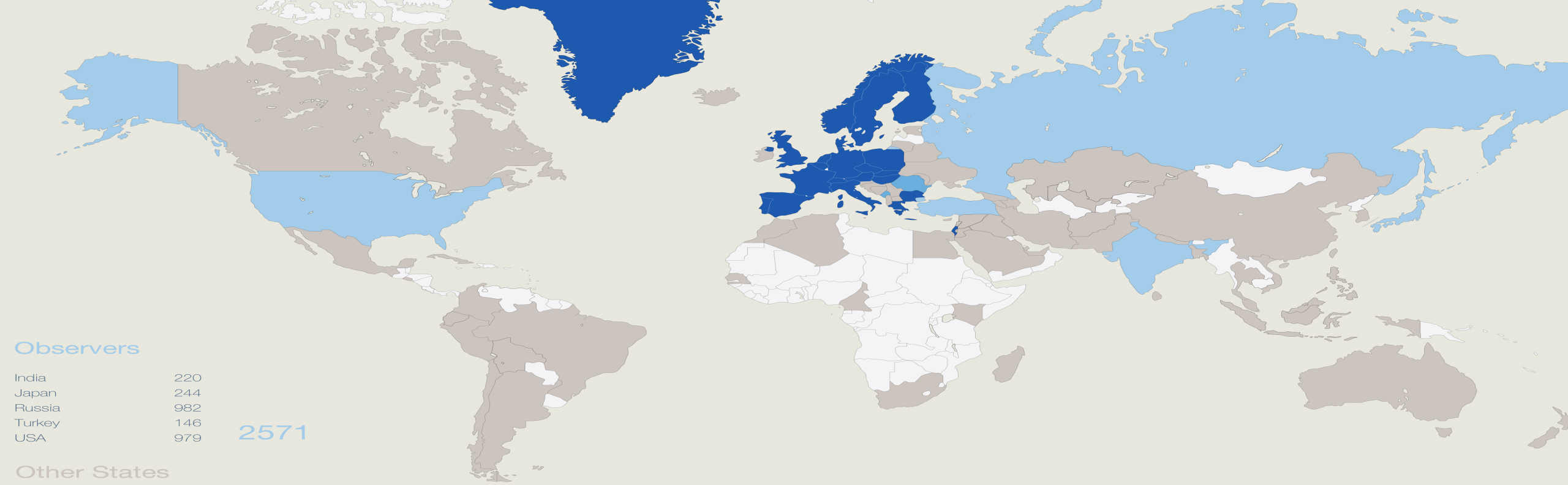
Information Discovery

Exploring Data-Driven Decision-Making for Improving the Control of CERN's Accelerator Complex

Antonio Romero Marín

What is CERN

- CERN - European Laboratory for Particle Physics
- Founded in 1954 by 12 countries for fundamental physics research in a post-war Europe
  - “Science for Peace”



Observers

India	220
Japan	244
Russia	982
Turkey	146
USA	979

2571

Other States

Afghanistan	1	El Salvador	1	Pakistan	41
Albania	2	Estonia	16	Palestine (O.T.).	4
Algeria	8	Georgia	36	Peru	8
Argentina	11	Gibraltar	1	Philippines	1
Armenia	25	Hong Kong	1	Saudi Arabia	3
Australia	25	Iceland	4	Senegal	1
Azerbaijan	8	Indonesia	1	Singapore	2
Bangladesh	4	Iran	28	Sint Maarten	2
Belarus	47	Ireland	22	Slovenia	27
Bolivia	3	Jordan	2	South Africa	16
Bosnia & Herzegovina	1	Kenya	1	Sri Lanka	5
Brazil	108	Korea, D.P.R.	1	Syria	2
Cameroon	1	Korea Rep.	117	Thailand	12
Canada	134	Kuwait	1	T.F.Y.R.O.M.	1
Cape Verde	1	Lebanon	12	Tunisia	6
Chile	12	Lithuania	19	Ukraine	55
China	280	Luxembourg	4	Uzbekistan	4
China (Tapei)	45	Madagascar	4	Venezuela	9
Colombia	30	Malaysia	15	Viet Nam	9
Croatia	35	Mauritius	1	Zimbabwe	2
Cuba	7	Mexico	64		
Cyprus	16	Montenegro	3		
Ecuador	3	Morocco	12		
Egypt	19	Nepal	5		
		New Zealand	7		

1415

Member States

Austria	99	Greece	152	Slovakia	88
Belgium	106	Hungary	68	Spain	337
Bulgaria	75	Israel	51	Sweden	75
Czech Republic	202	Italy	1686	Switzerland	180
Denmark	53	Netherlands	153	United Kingdom	640
Finland	87	Norway	61		
France	751	Poland	229		
Germany	1150	Portugal	109		

6352

Candidate for Accession

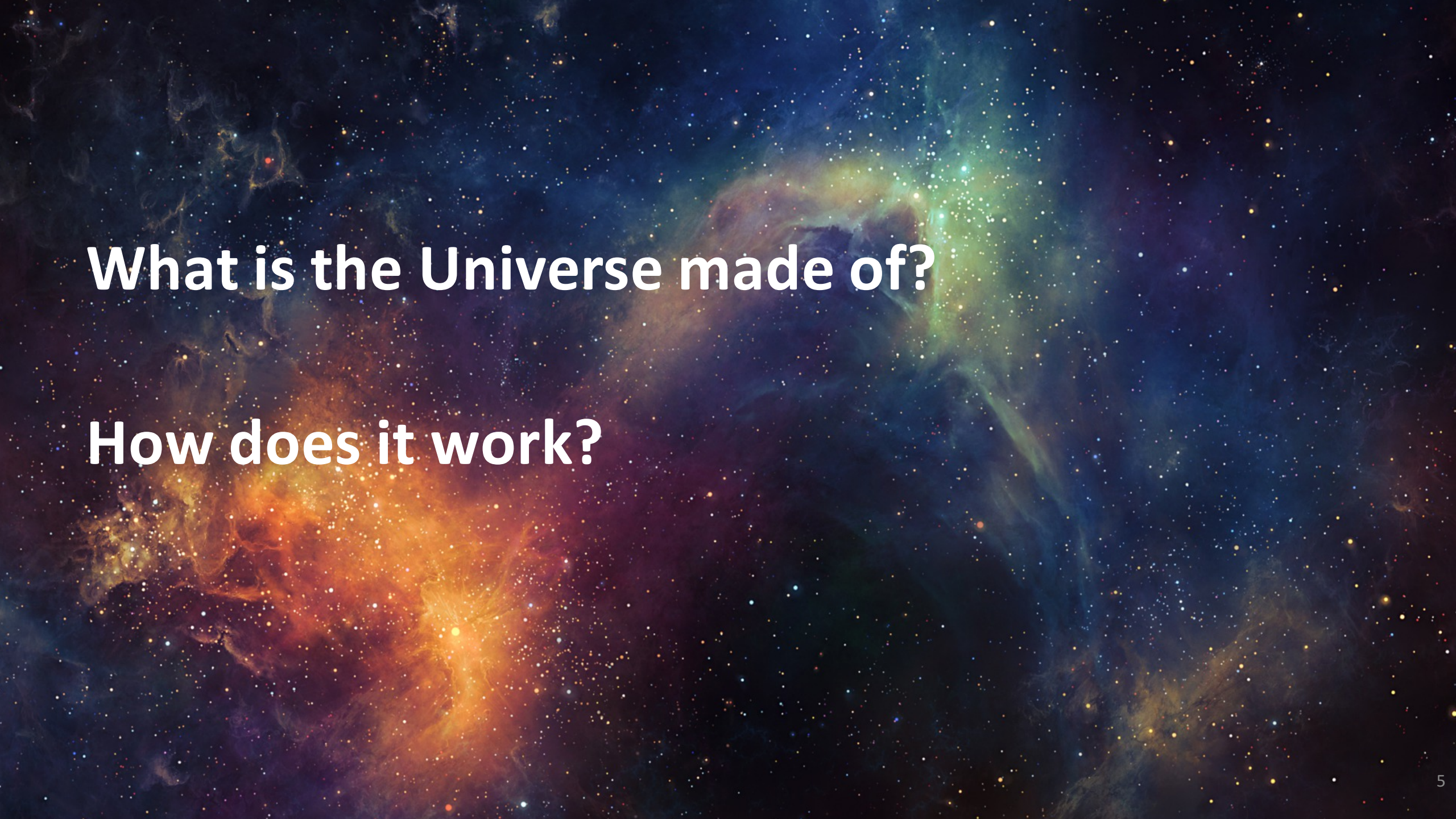
Romania	118

Associate Members in the Pre-stage to Membership

Serbia	41

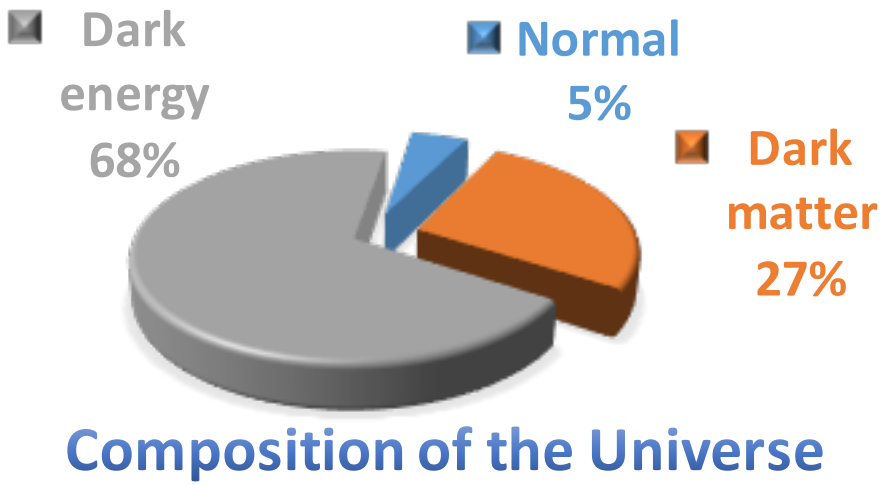
Distribution of All CERN Users by Nationality on 14 January 2014

CERN Mission


What is the Universe made of?
How does it work?

Fundamental Research

- Why do particles have mass?
- What is dark matter and dark energy?

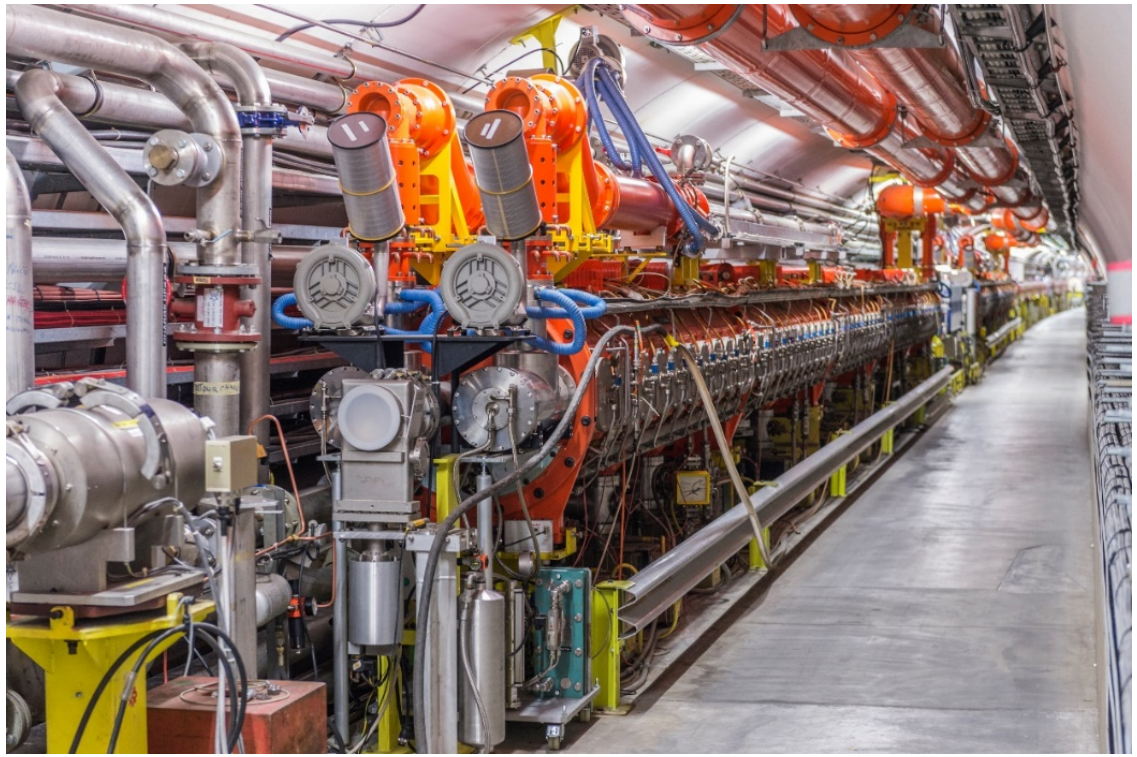


- Why is there far more matter than antimatter in the universe?
  - Big Bang should have created equal amounts of matter and antimatter

CERN Instruments

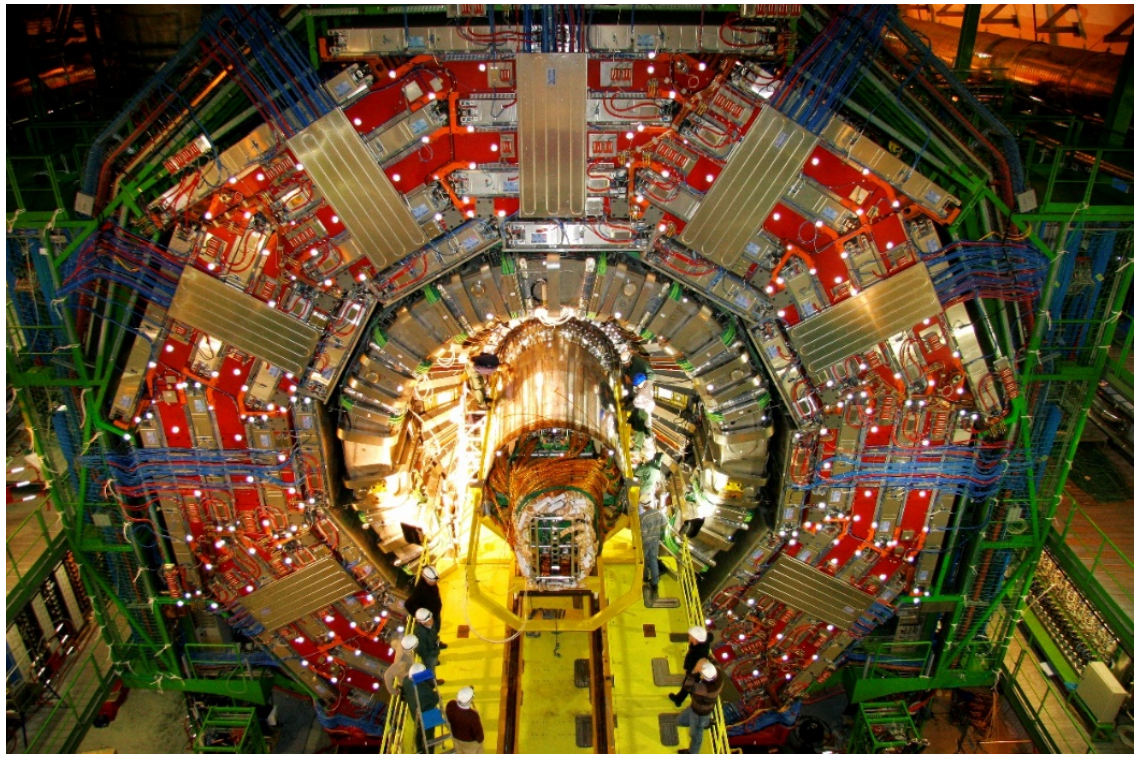
Accelerators

Boost particles to high energies and speed to collide



Detectors

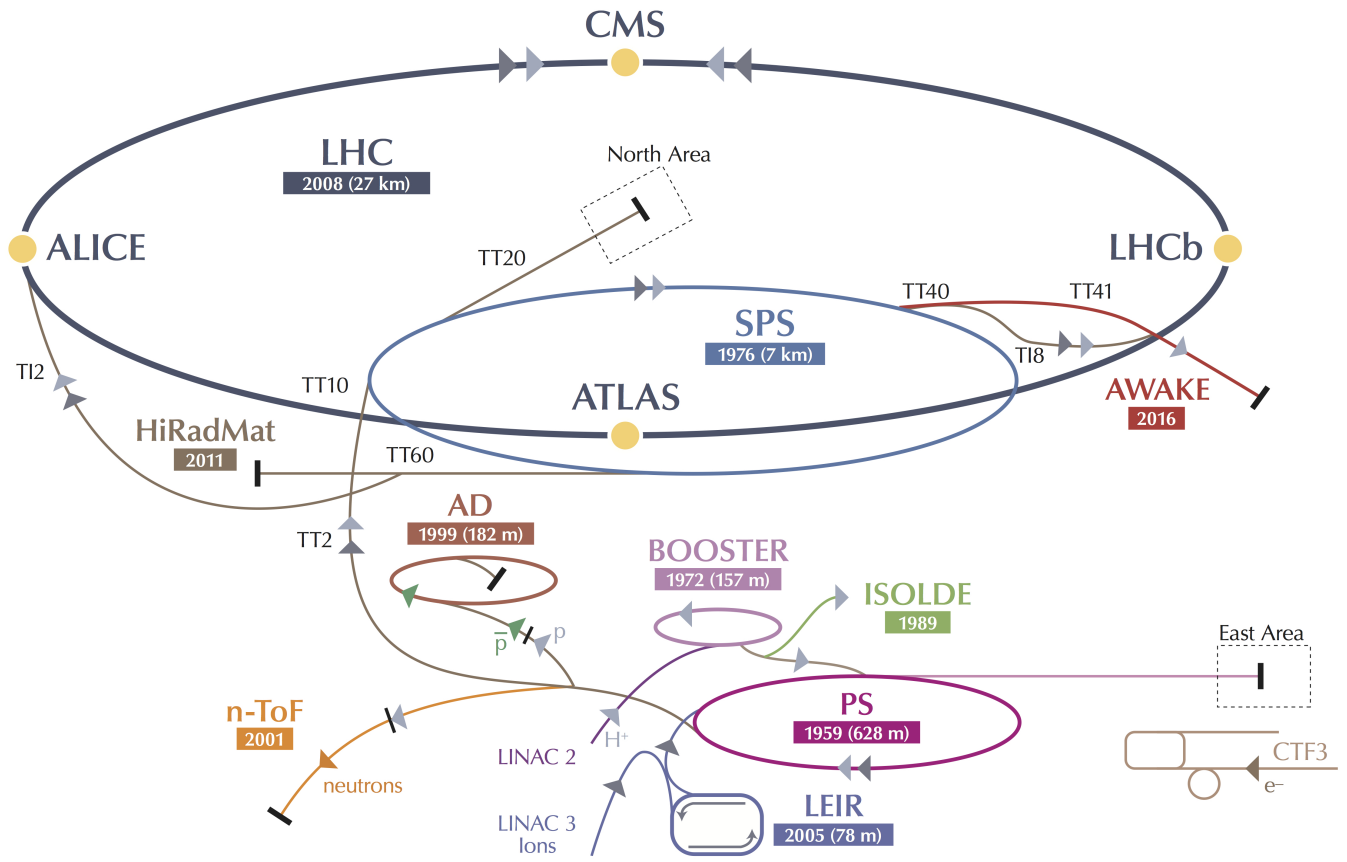
Observe and record the results of these collisions



The Large Hadron Collider (LHC)

- Largest machine in the world
  - 27km, 6000+ superconducting magnets
- 600 million collisions per second
  - Generating approximately one petabyte of data per second
- One of the coldest places on Earth
  - Main magnets operate at a temperature of 1.9 K (-271.3°C)
- Hottest spot in the galaxy
  - During Lead ion collisions create temperatures 100000x hotter than the heart of the sun

The CERN Accelerator Complex



LHC Large Hadron Collider **SPS** Super Proton Synchrotron **PS** Proton Synchrotron

AD Antiproton Decelerator **CTF3** Clic Test Facility **AWAKE** Advanced WAKefield Experiment **ISOLDE** Isotope Separator OnLine DEvice

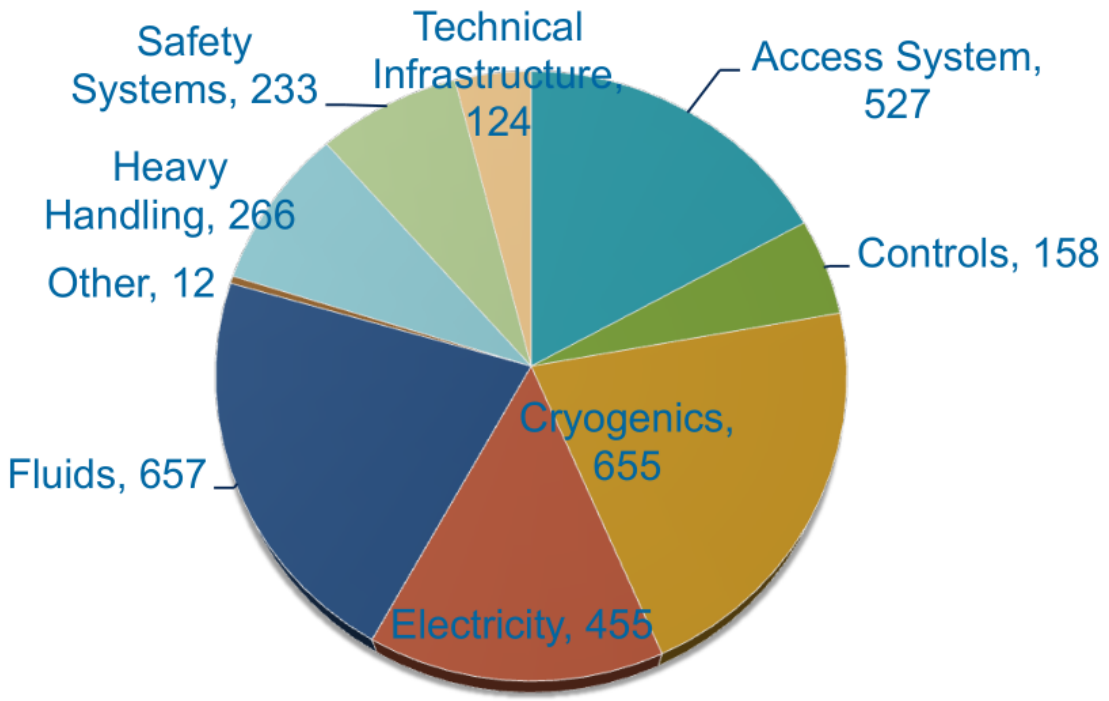
LEIR Low Energy Ion Ring **LINAC** LINear ACcelerator **n-ToF** Neutrons Time Of Flight **HiRadMat** High-Radiation to Materials

CERN is an extreme data environment

- Control and operations
  - Millions of sensors, signals
  - Large number of control devices
  - Equipment
- Monitoring and logging
- Supporting IT infrastructure
  - Databases
  - Network
  - Services
- CERN has great monitoring and logging systems
  - Large amount of data has been stored over years

Data Analytics Challenges

- Some faults cannot be avoid
- Decrease the availability for running physics
- Corrective interventions needed

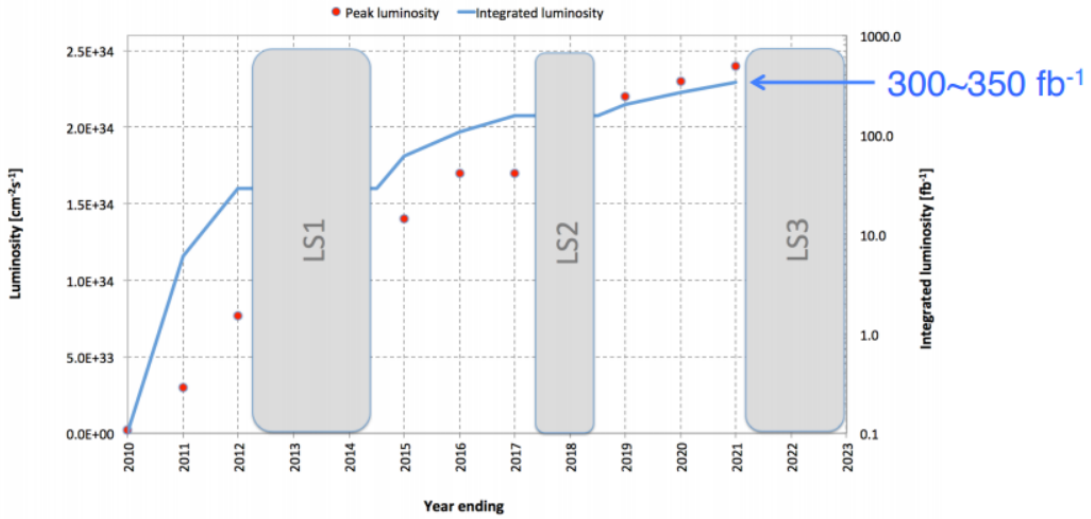


A look into the Future

- LHC upgrades will further increase luminosity
  - Computing resources needs will be higher
  - Data generated will increase drastically

Table 17: HLT Rate Evolution

Hz	ALICE	ATLAS	CMS	LHCb
2012	400 Hz 330 MB/s (p-p) 540 MB/s (p-Pb)	550 Hz 440 MB/s	460+360 Hz 328 MB/S	5000 Hz 300 MB/s
2015	500 Hz 525 MB/s (p-p) 810 MB/s (p-Pb) 3750 MB/s (Pb-Pb)	1000 Hz 800-1000 MB/s	1000 Hz 600 MB/S	10000 Hz 750 MB/s



- Next accelerators
  - Future Circular Collider (80-100 km)

openlab Data Analytics Project

- Optimize our systems
  - Reducing and predicting faults and corrective interventions
  - Increase the availability and operations efficiency
- Profit from CERN data investment by using data analytics
  - Extract knowledge
  - Discover useful information
  - Suggest conclusions
  - Support decision making
- Control and Monitoring Systems
  - Proactive
  - Predictive
  - Intelligent

CERN openlab

- Public-private partnership between CERN and leading ICT companies
- Accelerate cutting-edge solutions to be used by the worldwide LHC community
- Designed to create and disseminate knowledge
  - Publication of reports and articles
  - Workshops or seminars
  - CERN openlab Student Programme

CERN openlab

Partners

HUAWEI

intel

ORACLE

SIEMENS

Contributors

rackspace
the open cloud company

Associates

Yandex

Areas of investigation

- Data extraction, transformation and loading (ETL)
- Predictive maintenance & Anomaly Detection
- System Optimization
- Data Visualization and Discovery

Data Discovery

- Interactive and visual analytics
  - Get **new insights** from data
- Intended to be used by the end users
  - Enabling them to use their intuition and knowledge of the data
- Powerful customization of dashboards and visualizations
  - Without intervention of IT
- Integrate multiple data sources
  - Analyze information of any type and any source

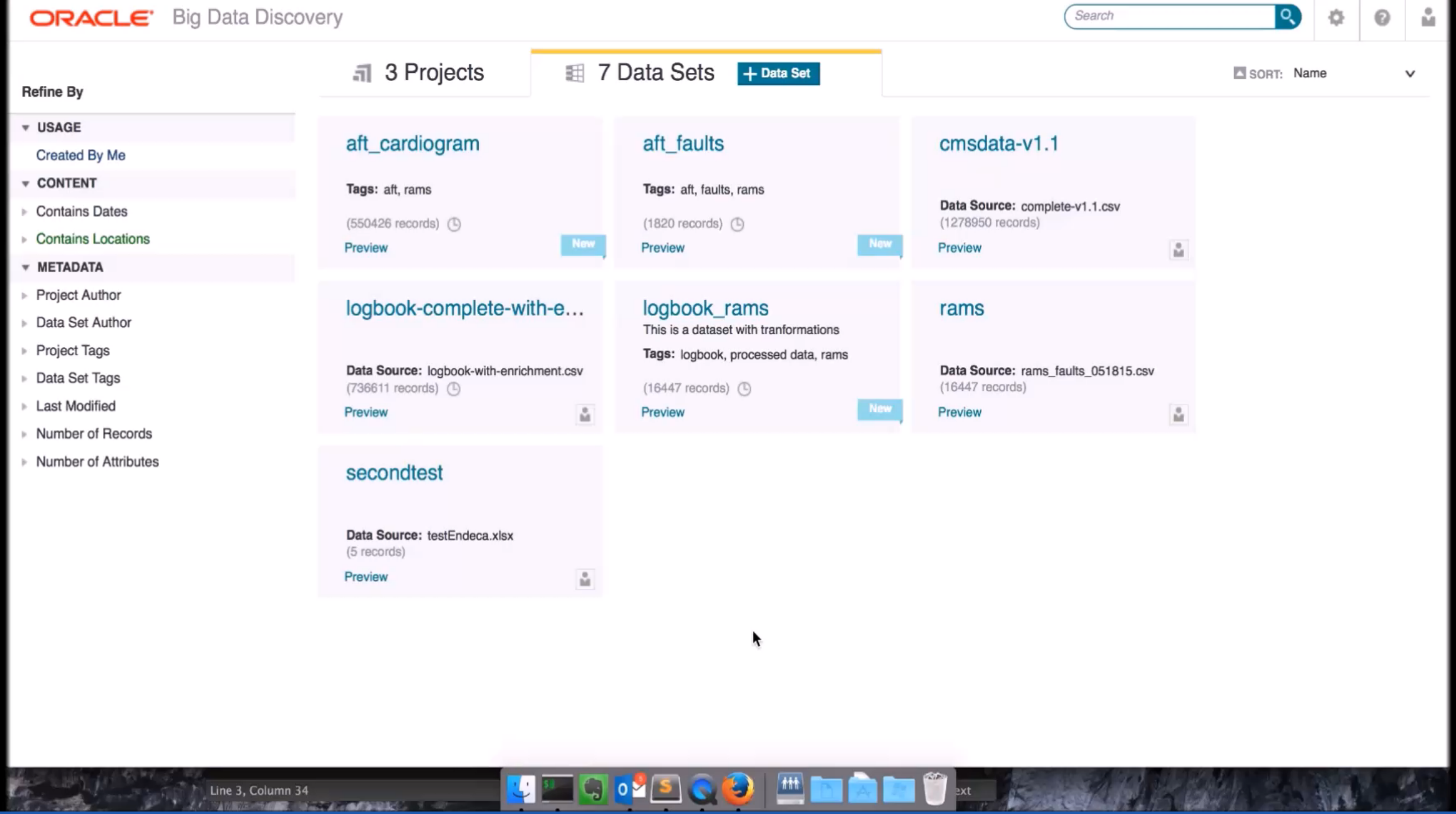
Oracle Big Data Discovery

- Scale-out architecture
  - Integration with Hadoop ecosystem
- Flexible and user-friendly interface
  - Powerful and dynamic visualizations
- Advanced Analytics
  - Text analytics
  - Correlation
- User-driven ETL
- Collaborative environment

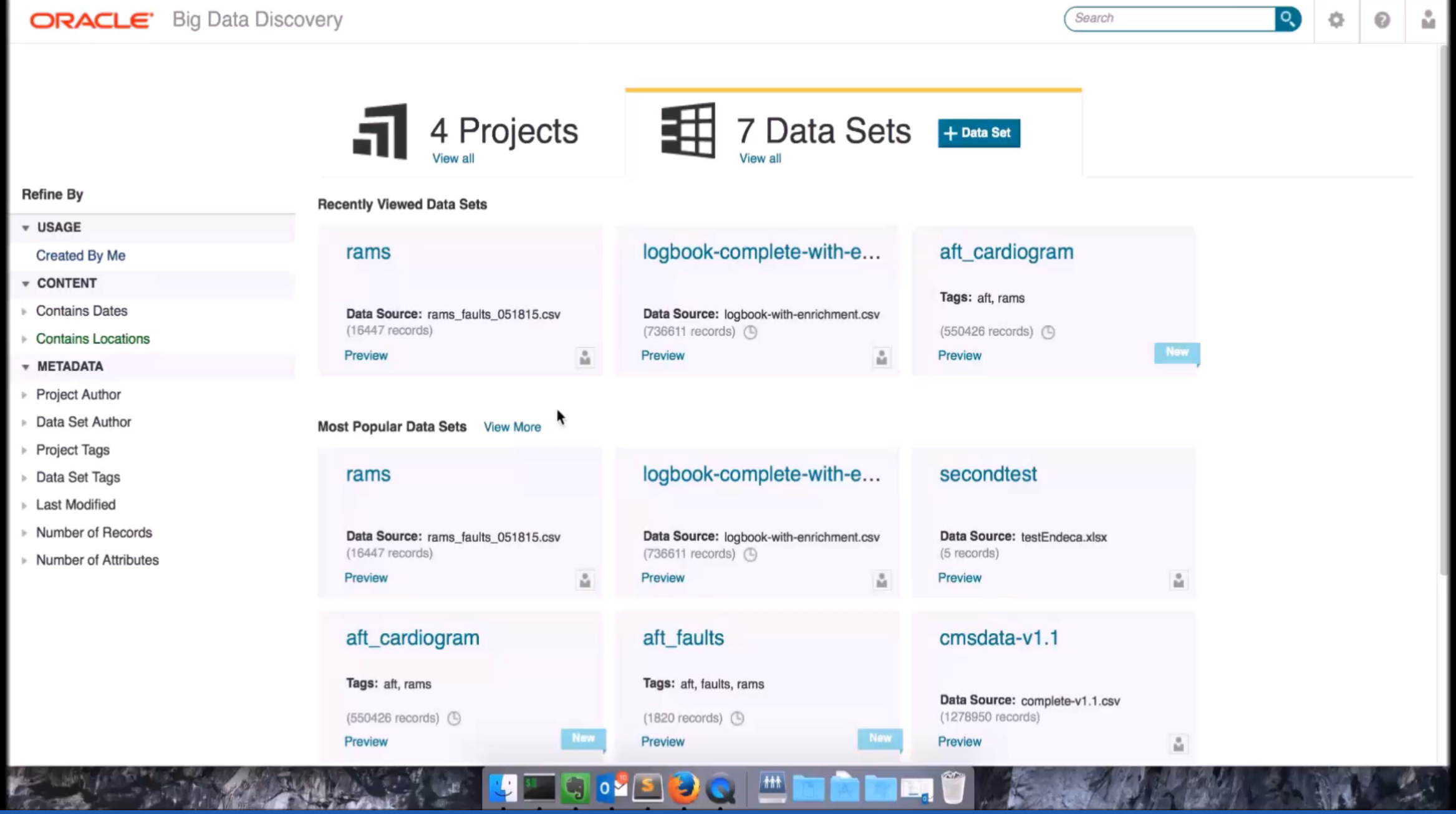
PoC – FCC RAMS studies

- Reliability, Availability, Maintainability and Safety (RAMS) studies for the Future Circular Collider (FCC)
- Increase the reliability and availability of the LHC
- Use RAMS findings to assess the feasibility and needs of FCC
- Data distributed across multiple sources
  - Operations e-logbook
  - Accelerator Fault Tracking project
  - Accelerator logging
  - Power Converters, cryogenics, more…

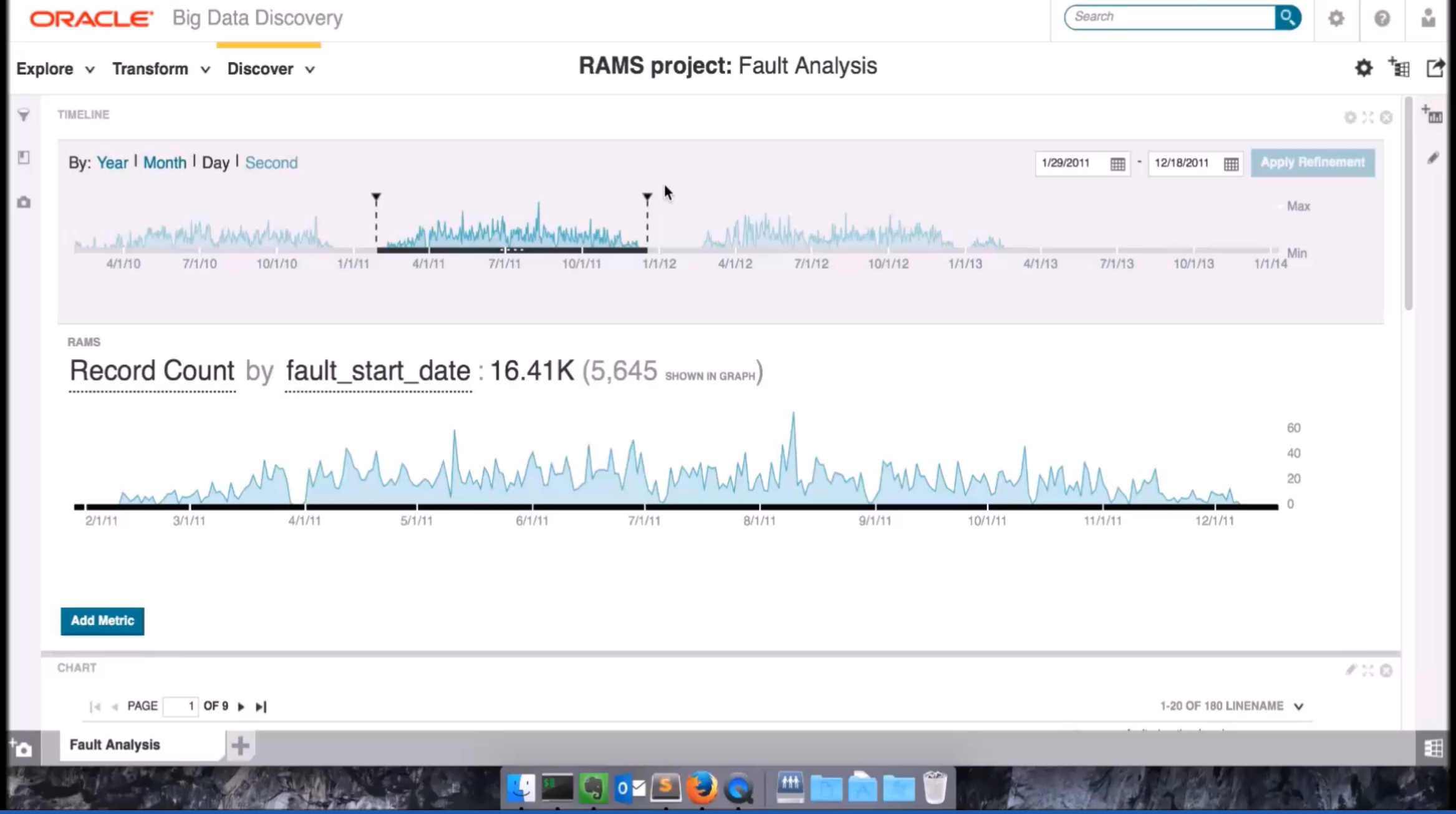
Data Exploration


ORACLE Big Data Discovery
Search
3 Projects
7 Data Sets
+ Data Set
SORT: Name
Refine By
USAGE
Created By Me
CONTENT
Contains Dates
Contains Locations
METADATA
Project Author
Data Set Author
Project Tags
Data Set Tags
Last Modified
Number of Records
Number of Attributes
aft_cardiogram
Tags: aft, rams
(550426 records)
Preview
New
aft_faults
Tags: aft, faults, rams
(1820 records)
Preview
New
cmsdata-v1.1
Data Source: complete-v1.1.csv
(1278950 records)
Preview
logbook-complete-with-e...
Data Source: logbook-with-enrichment.csv
(736611 records)
Preview
logbook_rams
This is a dataset with tranformations
Tags: logbook, processed data, rams
(16447 records)
Preview
New
rams
Data Source: rams_faults_051815.csv
(16447 records)
Preview
secondtest
Data Source: testEndeca.xlsx
(5 records)
Preview
Line 3, Column 34

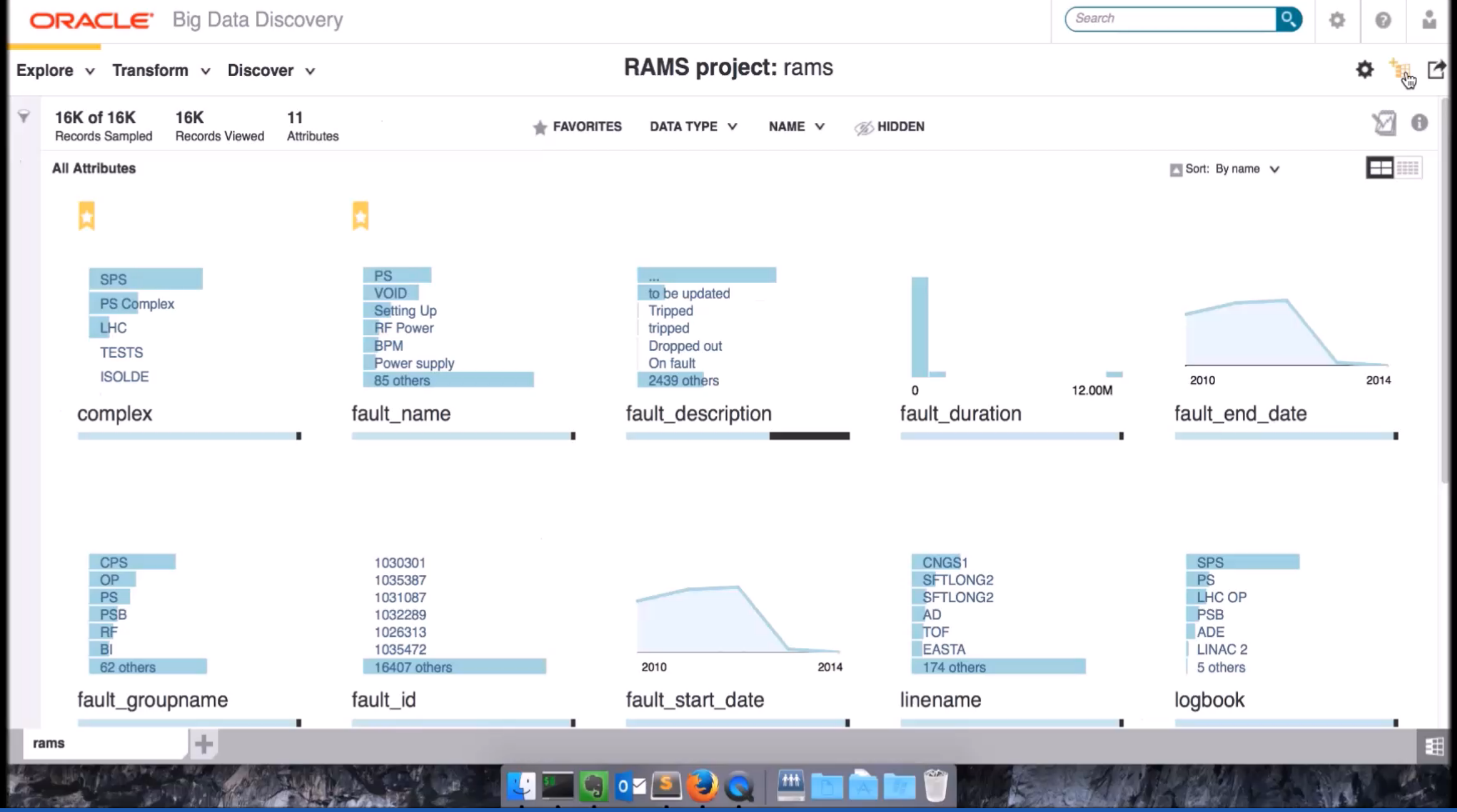
Data Transformation


ORACLE Big Data Discovery
Search
4 Projects
View all
7 Data Sets
View all
+ Data Set
Refine By
USAGE
Created By Me
CONTENT
Contains Dates
Contains Locations
METADATA
Project Author
Data Set Author
Project Tags
Data Set Tags
Last Modified
Number of Records
Number of Attributes
Recently Viewed Data Sets
rams
Data Source: rams_faults_051815.csv
(16447 records)
Preview
logbook-complete-with-e...
Data Source: logbook-with-enrichment.csv
(736611 records)
Preview
aft_cardiogram
Tags: aft, rams
(550426 records)
Preview
New
Most Popular Data Sets
View More
rams
Data Source: rams_faults_051815.csv
(16447 records)
Preview
logbook-complete-with-e...
Data Source: logbook-with-enrichment.csv
(736611 records)
Preview
secondtest
Data Source: testEndeca.xlsx
(5 records)
Preview
aft_cardiogram
Tags: aft, rams
(550426 records)
Preview
New
aft_faults
Tags: aft, faults, rams
(1820 records)
Preview
New
cmsdata-v1.1
Data Source: complete-v1.1.csv
(1278950 records)
Preview

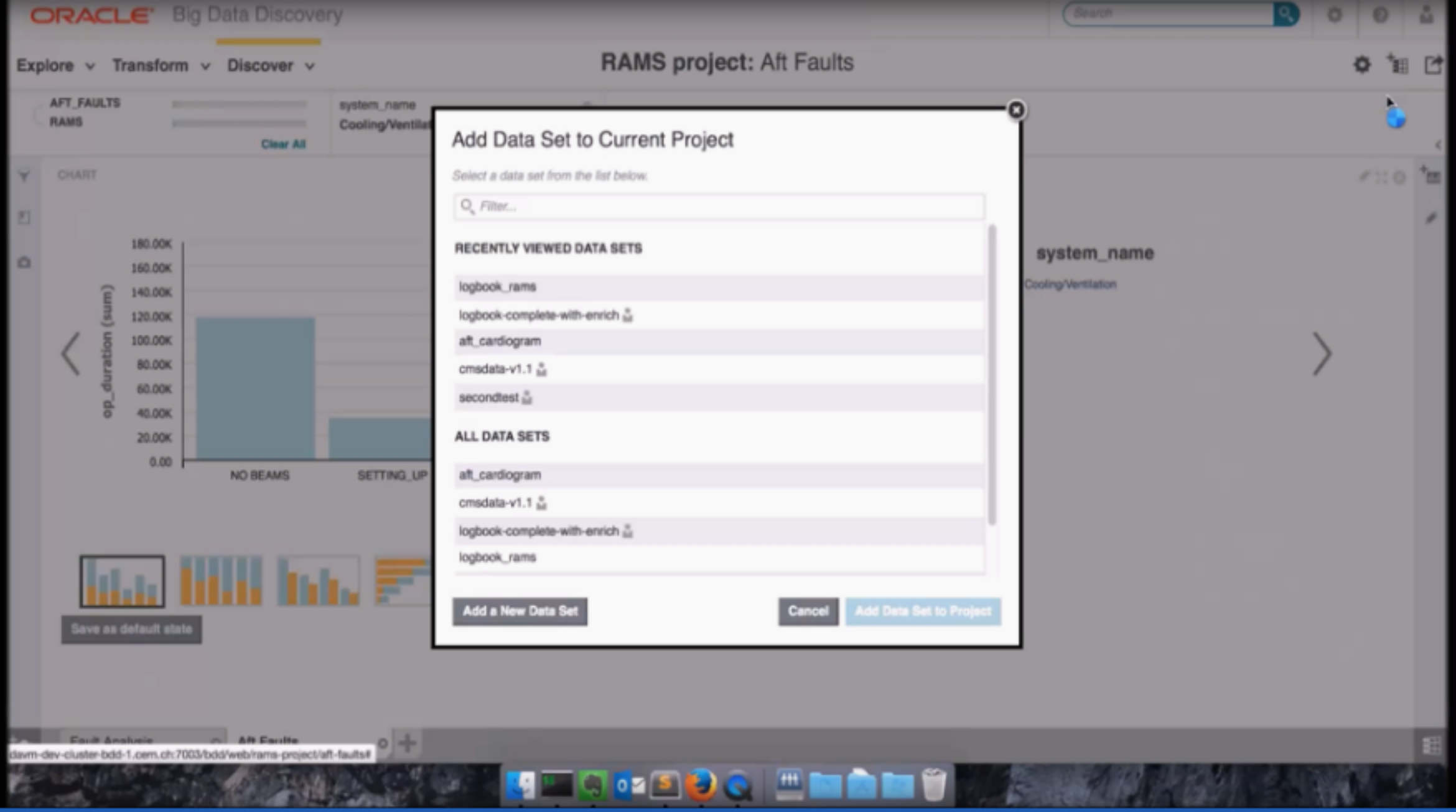
Data Discovery: Getting insight


ORACLE Big Data Discovery
Search
Explore
Transform
Discover
RAMS project: Fault Analysis
TIMELINE
By: Year | Month | Day | Second
1/29/2011
12/18/2011
Apply Refinement
Max
Min
4/1/10
7/1/10
10/1/10
1/1/11
4/1/11
7/1/11
10/1/11
1/1/12
4/1/12
7/1/12
10/1/12
1/1/13
4/1/13
7/1/13
10/1/13
1/1/14
RAMS
Record Count by fault_start_date : 16.41K (5,645 SHOWN IN GRAPH)
60
40
20
0
2/1/11
3/1/11
4/1/11
5/1/11
6/1/11
7/1/11
8/1/11
9/1/11
10/1/11
11/1/11
12/1/11
Add Metric
CHART
PAGE 1 OF 9
1-20 OF 180 LINENAME
Fault Analysis

Data Discovery: Analyzing multiple data sets


ORACLE Big Data Discovery
Search
Explore
Transform
Discover
RAMS project: rams
16K of 16K
Records Sampled
16K
Records Viewed
11
Attributes
FAVORITES
DATA TYPE
NAME
HIDDEN
All Attributes
Sort: By name
SPS
PS Complex
LHC
TESTS
ISOLDE
complex
PS
VOID
Setting Up
RF Power
BPM
Power supply
85 others
fault_name
...
to be updated
Tripped
tripped
Dropped out
On fault
2439 others
fault_description
0
12.00M
fault_duration
2010
2014
fault_end_date
CPS
OP
PS
PSB
RF
BI
62 others
fault_groupname
1030301
1035387
1031087
1032289
1026313
1035472
16407 others
fault_id
2010
2014
fault_start_date
CNGS1
SFTLONG2
SFTLONG2
AD
TOF
EASTA
174 others
linename
SPS
PS
LHC OP
PSB
ADE
LINAC 2
5 others
logbook
rams

Data Discovery: Linking with your own data


ORACLE Big Data Discovery
Explore
Transform
Discover
RAMS project: Aft Faults
AFT_FAULTS
RAMS
Clear All
system_name
Cooling/Ventilat
CHART
op_duration (sum)
180.00K
160.00K
140.00K
120.00K
100.00K
80.00K
60.00K
40.00K
20.00K
0.00
NO BEAMS
SETTING_UP
system_name
Cooling/Ventilation
Save as default state
Add Data Set to Current Project
Select a data set from the list below.
Filter...
RECENTLY VIEWED DATA SETS
logbook_rams
logbook-complete-with-enrich
aft_cardiogram
cmsdata-v1.1
secondtest
ALL DATA SETS
aft_cardiogram
cmsdata-v1.1
logbook-complete-with-enrich
logbook_rams
Add a New Data Set
Cancel
Add Data Set to Project

Conclusions

- Data visualization and discovery is an important area in data analytics
  - Facilitates users to visualize and explore their data
  - Find correlations, extract insight and useful information
- Important points
  - Flexible and user-friendly platform
  - Advanced data visualization and exploration
  - Collaborative
- Application to different domains
  - Controls and Operations
    - Accelerator Fault Tracking
    - Diagnostics and Monitoring
  - IT Infrastructure Monitoring
    - Server logs analysis
    - Database latency
  - Human Resources

www.cern.ch