Innovating our way through Obstacles
accelerating data science to lead enterprise value creation

A conversation with
Kumar Metlapalli
Founder & CEO
Kuberre Systems
Moderated by **Mike Meriton**
Co-Founder & COO, EDM Council

- Joined EDM Council full-time 2015 to lead Industry Engagement
- EDM Council Co-Founder & First Chairman (2005-2007)
- Former CEO GoldenSource (2002-2015)
- Former Executive for D&B Software and Oracle
- FinTech Innovation Lab – Executive Mentor (2011 – Present)
About Kumar:
• Bachelor’s in Electronics & Telecommunications; Master’s in Computer Science
• Founder & CEO Kuberre Systems, since incorporation in 2001
• US. Patent No.: 7,716,100 - Methods And Systems For Computing Platforms
• Trusted data science leader
• Data processing thought leader (hardware & software)
• Dedicated to creating positive legacy

About Kuberre:
• Data always takes priority over technology
• It is our goal to leave a lasting, positive cultural impact on every client team we interact with.
Challenges: Too much low-value effort

A True Tug of War

Industry Average
Time & $$ Spent

<table>
<thead>
<tr>
<th>Data Engineering</th>
<th>Model Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>~80%</td>
<td>20%</td>
</tr>
<tr>
<td>Collection</td>
<td></td>
</tr>
<tr>
<td>Normalization</td>
<td></td>
</tr>
<tr>
<td>Mapping</td>
<td></td>
</tr>
<tr>
<td>Integrity</td>
<td></td>
</tr>
</tbody>
</table>

Target Balance
Time & $$ Spent

<table>
<thead>
<tr>
<th>Data Engineering</th>
<th>Model Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>~20%</td>
<td>80%</td>
</tr>
</tbody>
</table>
## Data Engineering – Why so much time and $$?

### Existing/Internal Data
- **Data Scope**
  - If not all, production results may vary from sandbox
- **Scalability**
  - If all, will the sandbox need to be the same capacity
- **Data Mart/Lake**
  - Design & Construction
- **Data Timeliness**
  - Keeping in sync with production data

### New/External Data
- **Data Access**
  - File transfers & IT thresholds
- **Scalability**
  - If data is too large, need backend servers, more IT
- **Data Trial/Subs**
  - Vendor origination, legal, etc.
- **Data Interfaces**
  - Loaders/ETL
Avoid These Obvious Landmines

#1- Leave data where it currently lives

#2- Make data available “as if” it were in a common data lake.

#3- Empower quants to use tech they like to run analytics, create derived data, etc.

Frank is a Python expert & creates his own analytics to find alpha

Josh creates his own derived data elements

Nancy maintains downstream uses of data

SQL Connectors, NoSQL, REST,...

Data Virtualization

Uniform API’s

Alternative Data

Files

Portfolio Data

Research Data

Sales Data

Holdings

Account Master

Alphas

Universe Definitions

Alphas

Universe Definitions

#1-

#2-

#3-
Have we really solved the problem with Data Virtualization?

Existing/Internal Data
- Data Scope: If not all, production results may vary from sandbox
- Scalability: If all, will the sandbox need to be the same capacity
- Data Mart/Lake: Design & Construction
- Data Timeliness: Keeping in sync with production data

New/External Data
- Data Access: File transfers & IT thresholds
- Scalability: If data is too large, need backend servers, more IT
- Data Trial/Subs: Vendor origination, legal, etc.
- Data Interfaces: Loaders/ETL

Poll Question #2

Auditability
Enterprise IT Standards
Query Performance
Caching for DV Servers
Data / ID
DB - Coupling

BEWARE: Data Virtualization Presents a “New & Different” set of Landmines
Kuberre’s Unique Learning Journey

Poll Question #3

Consulting Engagements

Productization “QTPy”
Productized Approach… our example

Accelerating research through Data Virtualization and Data Analytics Layers

**EDM**
- Benchmarks Repository
  - Benchmarks
  - Biah
- Portfolio Data
  - Holdings
  - Account Master
- Research Data
  - Alphas
  - Universe Definitions
- Data Lake

**Orchestration**
- Catalog
- Lineage
  - Data Virtualization
- Distributed Computing
- Federated Data Model
- Models
- Data Analytics
  - Data++

**Access**
- Jupyter Hub, Spyder, VS Code, etc.
- REST API
- BI Tools
- Excel
DCAM V2.2: Component 8.0 from EDM Council

Systems Driven

8.3 Aligned with Data Architecture
Data Architecture is Respected

8.4 Aligned with Data Quality
Data Quality is Understood and Improved

8.5 Platform Designed and Operational
Operational Platform Established

8.6 Model Operationalization Established
Models can run in Production

Organization Driven

8.1 The Analytics Function is Established
Analytics function is formalized and funded

8.2 Aligned with Business and Data Strategy
Driven by Business Strategy and supported by Data Strategy

8.7 Culture and Education Needs Managed
Active Management of Culture and Skills
QTPy Core Capabilities – Aligned with DCAM Analytics

Access data from anywhere – now!!
No need to build intermediate data mart

Combine these virtualized data sets with absolute ease

Centralized Source Code – Methods, Models, Calculations are served through a centralized environment to facilitate collaborative development and idea/code share

Scale as per business needs
- Distributed architecture utilizes concurrent servers to support scalability and changing business needs

Ensure Dev-UAT-Prod migration through automatic source code control

Build analytics once, access results from anywhere CONSISTENTLY – Portal, REST API’s, BI Tools, MS Excel
# Results Tell the Story

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance</strong></td>
<td>Model Run-Time: 45 minutes</td>
<td>Model Run-Time: 35 seconds</td>
<td>98%+ Improvement</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td>Add new data set: 3-4 weeks (limited access)</td>
<td>Add new data set: &lt; 2 hours team access</td>
<td>99%+ Improvement</td>
</tr>
<tr>
<td><strong>Alpha creation</strong></td>
<td>Add new models 5 new models/year</td>
<td>Add new models 2 new models/month</td>
<td>~ 5X Improvement</td>
</tr>
<tr>
<td><strong>Collaboration</strong></td>
<td>Whole is Greater... Manual processes for sharing models &amp; factors</td>
<td>Whole is Greater... Version controlled check-in &amp; sharing of models, factors &amp; dynamic schema</td>
<td>Priceless</td>
</tr>
</tbody>
</table>
Questions?

kuberre systems

EDM Webinar
FOR MORE INFORMATION:

Kumar Metlapalli
Founder & CEO
Kuberre Systems
kumar@kuberresystems.com
978-267-4799