

Data Science Technologies for Financial Services

Presented by

Dr. Suen, Wai Mo

Founder & CEO, ClusterTech Ltd.
Professor, Washington University
Honorary Professor, CUHK

孫緯武

聯科集團創辦人兼首席執行官
華盛頓大學理論物理教授
中文大學榮譽教授

14 Dec 2021

Technology Forum – Accelerate Adoption of Innovation to Enhance Public Services





Introduction - ClusterTech

CT was established in year 2000 to provide Advanced Computing Services, in High Performance Computing (HPC), AI & Analytic, Big Data, and Cloud Computing. We have served over 500 institutional clients, in Hong Kong/China/Asia-Pacific, including:



Cloud + High Performance Computing + AI & Analytics + Big Data
 Business Intelligence · Financial Engineering · Environmental Science · Smart City

IT Solutions and Data Analytics Services in 4 Areas

 <p>High Performance Computing</p>	<p>Enabler---Accelerate application processing, and enhance accuracy of analysis and planning by :</p> <p>designing, implementing and maintaining large-scale HPC clusters (on-cloud / on-premises)</p>
 <p>Big Data Infrastructure</p>	<p>Volume---Unlock value in big data by :</p> <p>constructing customised big data infrastructures with built-in analytics</p>
 <p>AI & Analytic</p>	<p>Insight---Automate workflows, and uncover insights to improve operational efficiency by :</p> <p>offering data strategy consulting and AI/ML solution development services (on-cloud / on-premises)</p>
 <p>Cloud Computing</p>	<p>Cost Effectiveness---Provide scalable, lightning fast, cloud-based services by :</p> <p>offering private, public, and hybrid cloud solutions, with 24 x 7 local support</p>

Pain Points in Financial Services

1. Sufficient, relevant and reliable data for assessment



香港特区政府认可的
资讯科技专业服务活跃
供应商

2. Handling large, diverse and complex sets of data



2012年度深港澳
资讯科技奖
十年领军企业奖



3. Manage large volume of documents and files



2016
香港工商業獎
HONG KONG AWARDS
FOR INDUSTRIES

科技成就優異證書
TECHNOLOGICAL
ACHIEVEMENT
CERTIFICATE OF MERIT

4. High computational demand for big-data analytics and simulation



2012年度
中国云计算数据中心

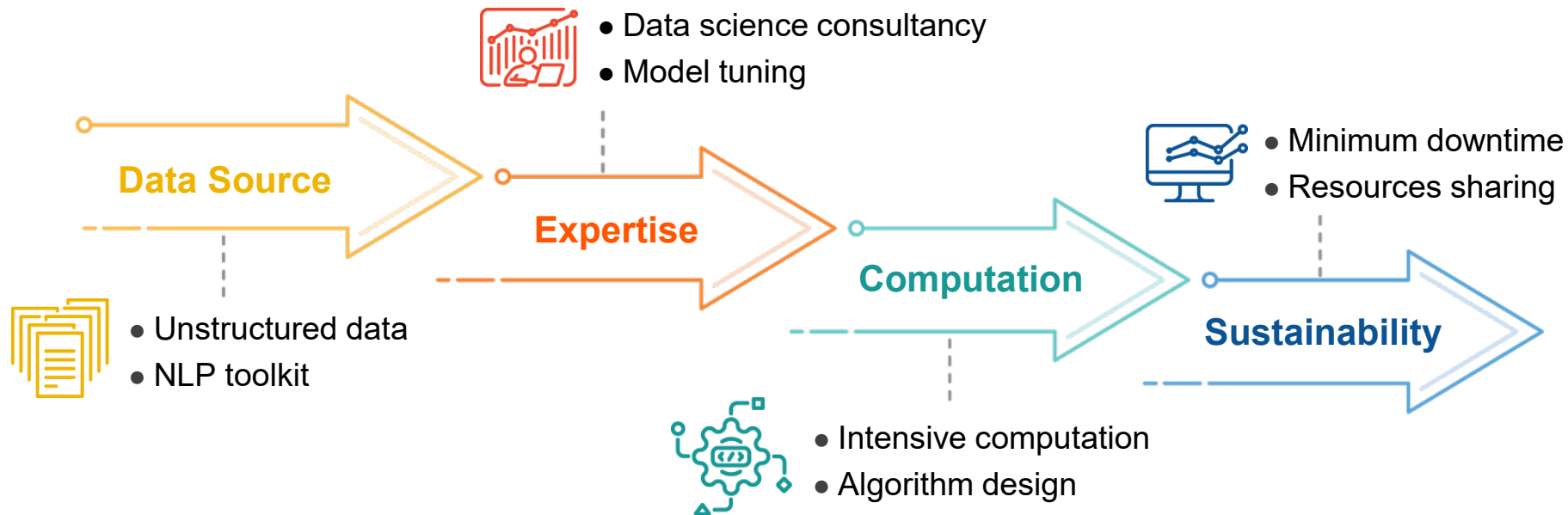
5. Infrastructure supporting high service level



IS 679580

6. Fulfill security and privacy in auditing and policy requirement

Key Components for Financial Services Solution



Public Sector Financial Big Data Case Study:

The Need:

Financial records growing at a rate of more than 1 TB per day. Need to examine last 3 years of data. Data queries beginning to take too long to meet the service level requirements.

Our Solution:

A Hadoop platform to facilitate distributed storage and computation across clusters of computers, allowing scalability by splitting files into blocks and distributing them to the nodes in the cluster.

Data Analyst

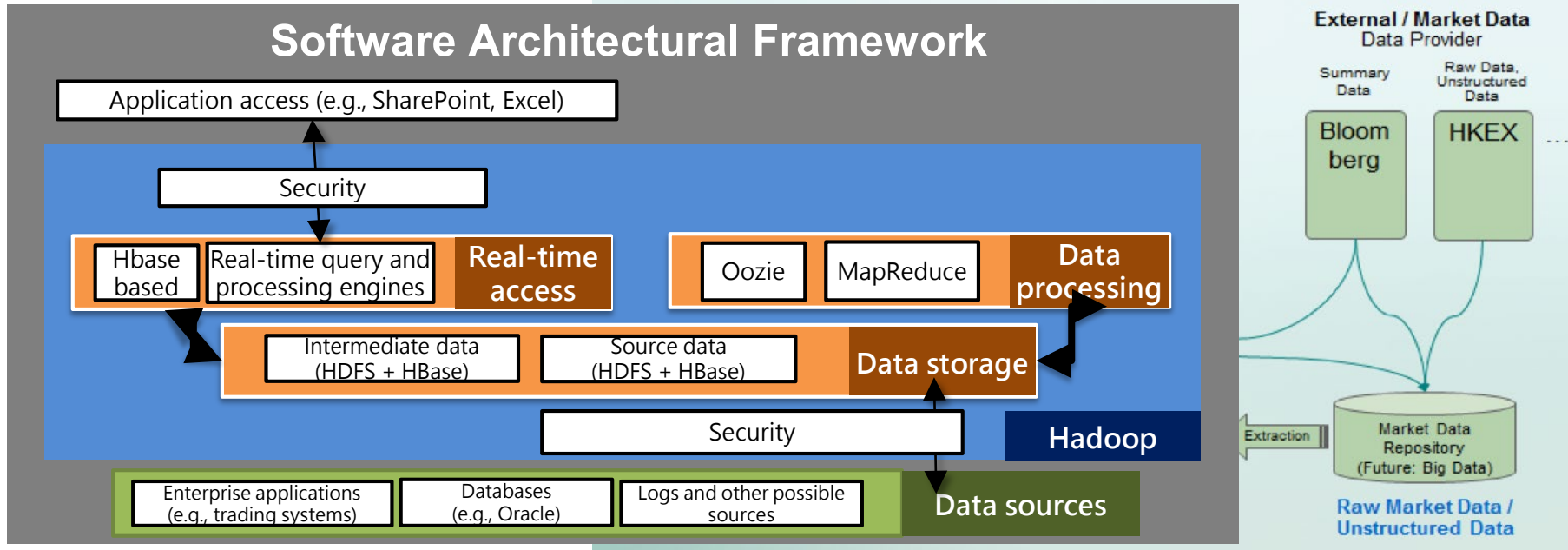


growing at a rate of more than 1 TB per day !



Case Study: Financial Big Data Solution

Structure Overview



Case Study:

Financial Big Data Solution

Key Challenge: Integration with HPC capabilities

1. Integrate: our Solution + Existing Legacy Systems

- Minimize change on the frontend
- Maximize improvement on the backend

2. Integrate: Domain Knowledge + HPC System Design Expertise

To design a scalable system, need to master how the system is to be used

3. Integrate: System Expertise + HPC Software Development Expertise

For a stable, secure and easy to maintain system

4. Integrate: Data Science Expertise + HPC Software Expertise

To validate results and develop data analytics

Case Study:

Financial Big Data Solution

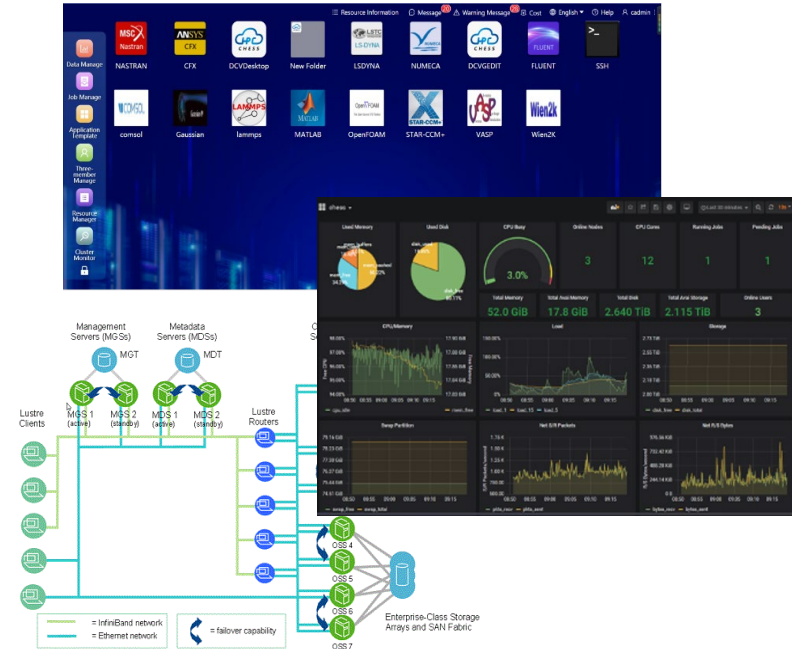
Results:

- ✓ Enabled the client to meet the service level requirements in data handling:
 - Response time
 - Length of period to examine
 - Types of data included
- ✓ Enabled the client to adopt more advanced machine learning models to:
 - Reduce risks
 - Analyze fraud patterns
 - Identify problem behavior
 - Perform better customer segmentation

Tools for Financial Services (Highlight)

HPC Tools for Financial Services (highlight)

- **CHES--to solve big computation needs:**
Cluster Management & Job Scheduler
 - Batch jobs execution
 - Schedule jobs to vacant servers
 - Billing, reporting, and cost sharing
 - Avoid overload during peak hours
- **Parallel File System--to solve big data needs:**
 - Aggregate disk systems and computers
 - Automatic failover & recovery
- **Reliable Tools for:**
daily clearing, monthly statements, analytics, etc.

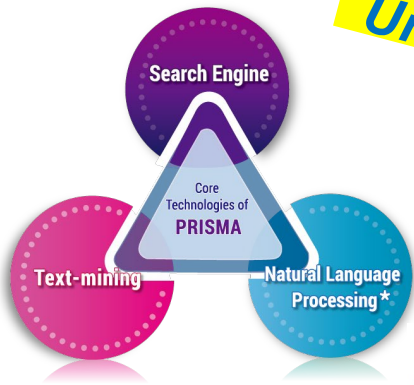


Mining Tool for Financial Services (highlight): PRISMA



- a Text Analytics Solution

3 Core Technologies:



- Capable of processing trilingual texts simultaneously including
 - ✓ *Traditional Chinese*
 - ✓ *Simple Chinese*
 - ✓ *English*
- **Tailor-made** for the multi-cultural Hong Kong market (兩文三語)(你個system, 成日down機, 咁豆沙架!)
- **Silver medal winner** in the category of Best Smart Hong Kong Award (Big Data Application) of Hong Kong ICT Awards 2016

PRISMA - Trilingual Text Processing Tool

- Utilise unstructured data in addition to structured data (e.g. financial indices, financial news...)
- 3 core technologies:
 - Text Mining
 - Search Engine
 - Trilingual Natural Language Processing (NLP)
- Problem solving:
 - Topic Classification (Document, News...)
 - Entity Extraction (Organization, Stakeholder...)
 - Relation Extraction (Related parties in graphic form)



A.I. Knowledge -based System

- Solve the right problem with data science expertise

- Examples:

- Sentiment analysis for financial news, emails and reports
- Machine Learning Extraction (not just rule-based, but also training-based extraction) of person and company entities with positions highlighted



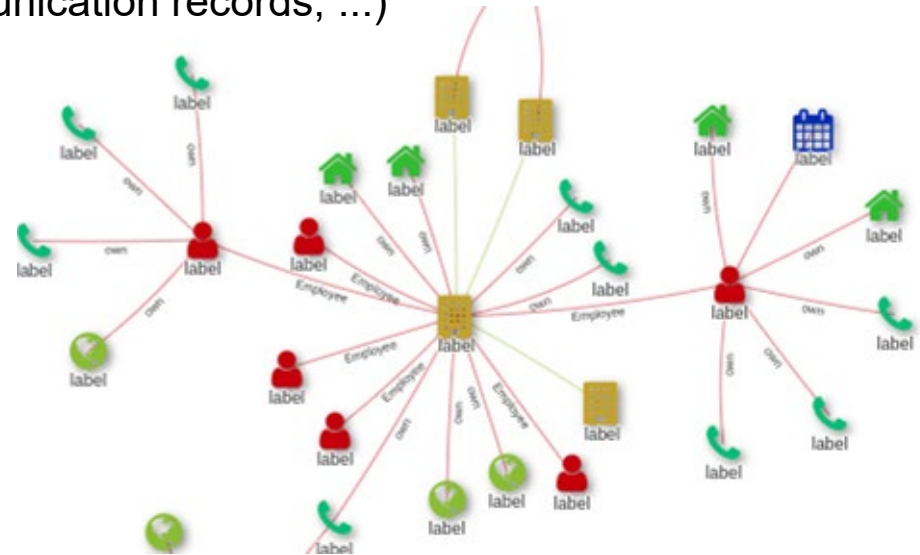
Person S-b Organization S-h Date S-t

蘇漢章先生，60歲，於二零一一年四月二十八日獲委任為獨立非執行董事。蘇先生自二零零七年九月起為加拿大上市公司 CINS Holding Corp. (前稱為 Genius World Investments Limited) 之董事。於本公告日期，董事會包括六名董事：兩名執行董事為 蘭夙女士及 謝聲宇先生；一名非執行董事為 羅頌霖先生；及三名獨立非執行董事為 譚澤之先生、劉永勝先生及 陳嘉洪先生。

A.I. Knowledge -based System - Relation Analytics

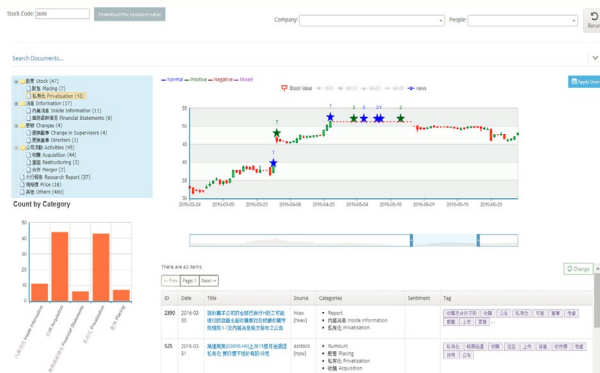
Infer Large-scale Relation Analytics from a large collection of documents (financial news, statements, communication records, ...)

- ✓ System-generated
- ✓ Big-Data Scale
- ✓ Customized entity type and relation type based on Deep Neural Network models
- ✓ Allow user to explore interactively (e.g., discover path of connection...)
- ✓ Editable visuals
- ✓ Export and share



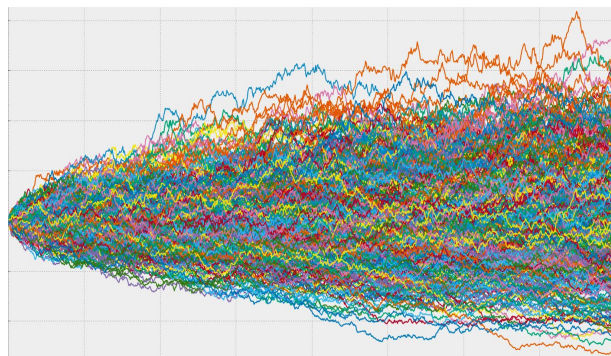
1. Listed Company Risk Assessment

- Correlate with the change of financial indices with sentiment in company-related news
- Link company and stakeholder in network
- Alerts for company in financial crisis



2. Financial Computing in Cluster Environment

- Monte-Carlo simulations
- Daily clearance operations
- Export ad-hoc analytics
- Spread jobs across many hosts
- Avoid overload during peak hours



Take Home Messages:

1. Modern Data Science, based on High Performance Computing (HPC), has opened up new opportunities in financial services:

---HPC: Let you handle **large volume of data** and **huge amount of text** documents like never before....Big Data Infrastructure to store, curate, transfer, search, query, analyze, process and visualize the data

---HPC: Let you carry out **innovative applications** (not just statistics): in-depth investigation, relation discovery, Deep Learning, algorithmic transactions....

---HPC: Robustness and High Availability (HA), for mission critical public services

2. **Security and Privacy**: a must for public financial services



.....Important considerations in choosing service provider, while everyone claims to be AI expert

To support HKGov's plan to adopt advanced technology in daily operations, we are now offering free consultancy services to HKGov departments.



FREE Consultation

Thank you!



+852 2655 6100



enquiry@clustertech.com



www.clustertech.com