将来における大学・研究機関のデータ管理について

静岡大学創造科学大学院・情報基盤センター
浦野 新

データ管理の重要度

組織や企業を運営するためにヒト・モノ・カネ・情報は不可欠である。コンピュータの高性能化やインターネットが発達した現代においては、データの活用による収益力向上や業務効率化は必須であり、組織が所持するデータの重要度は増している。

一方で、世の中の組織や企業による情報流出の報道は後を絶たない。組織におけるデータの管理は非常に多くの人々が注目し、頭を悩ませる重要なテーマのひとつである。

システム技術分科会会合 2017 概要

2018 年 2 月 1 日に汐留シティセンター6 階にある富士通株式会社プレゼンテーションルームにおいて開催された、サイエンティフィック・システム研究会（以下、SS 研）システム技術分科会会合では、『将来における大学・研究機関のデータ管理について』の題のもと、教育・研究・学務システムそれぞれの分野の講演と意見交換が行われた。

3 名の講演者による組織概要の紹介があり、クラウド利用についての立場や意見を明らかにしながら、データ運用事例の講演が行われた。

クラウド移行派 VS オンプレミス残留派

講演後に行われた参加者全員でのパネルディスカッションでは、クラウド移行派とオンプレミス残留派それぞれの意見が飛び交う。特に多かった意見は、オンプレミスまたはクラウドを選択する際の評価基準についてであった。また、各組織におけるデータ運用の詳細にまで踏み込んだ密度の濃い意見交換が行われた。

持たない、使う

立教大学におけるクラウド活用（メール、ストレージ)事例

立教大学　小川 龍秀氏

立教大学ではメール運用に Gmail と Office 365 メールを利用している。特に Gmail 利用のメリットは障害の少なさである。立教大学では以前、メール障害が多発した経緯があり、障害の頻度軽減が優先事項であった。以上のような背景があり、メールサービスのクラウド化について全学的な理解が得られた。

同大学ではメールのアンチウィルスやアンチスパムも、Gmail や Office 365 の機能を利用し、安全なメール環境を提供している。さらに、Gmail と Office 365 メールを併用することで、メールシステムの冗長化も図られる。新入生の多くは既に Gmail のユーザであり、入学時のオリエンテーションでのメール利用手順の案内などは不要となった。

ストレージの運用に関しては、クラウドとオンプレミス両方を活用している。クラウドサービスは Google 社が提供する G Suite を活用しており、データ共有の利点を生かした議事録の複数人書き込みなどの手法を生み出し、業務効率の改善が図られている。

一方、事務用データはオンプレミスで運用し個人情報は外部へ委託しない運用を行っている。オンプレミス運用の評価基準は費用面と人的ミスによるセキュリティリスクを回避する観点であると述べ、リスクアセスメントの結果からオンプレミスを選択していることを明らかにした。

機密性の観点からメールサービスのクラウド利用について懸念点を問われた小川氏は「原則、重要情報はメールでのやり取りを行わないことを周知している。また、添付ファイルは自動で暗号化するサービスも提供し機密性を保持している。しかし、サービス運用には人の脆弱性が付き物である。そのことを念頭に今後もユーザ教育の徹底や、サービスの利用・運用を行っていく」と答え、データ管理におけるシステム構築と人的運用の重要性を述べていた。

小川 龍秀氏（OGAWA Tatsuhide）
立教大学 メディアセンター

オンプレミス残留派

大規模研究データの保管・配布におけるクラウド利用の課題と期待

宇宙航空研究開発機構　藤田 直行氏

「クラウドは、我々の業務には使えません」と冒頭で述べ、藤田氏の講演は始まった。

宇宙航空研究開発機構ではスーパーコンピュータを用いた解析を行い、解析によって得られた莫大な量の研究データを管理・運用している。保有している研究データ量についての資料には、64GB ほどの研究データファイルが 9 万以上存在し、それらを合計したストレージ容量は 2.5PB に上ると調査されており、参加者からは驚きの声が聞こえた。

藤田氏はクラウド利用について、メリットとデメリットを明確にし、柔軟性や機密性、継続性など複数の観点からリスクアセスメントを行ったうえでオンプレミスを推奨していた。なかでもクラウド利用の大きな障壁となっていたのは費用の問題である。クラウドサービスの、カスタマイズやオプションを選択するにつれてコストが増加する様子を『Black Hole Model』と題し、同機構のデータをクラウドで管理した場合、システム運用にかかる費用が 6 億円を超える調査結果を報告した。

一方で、海外機関ではクラウドを積極的に活用したコミュニティも存在している。EU-Copernicus では衛星観測データを無料で公開し、新しい価値を生み出す取り組みを行っている。EU-Copernicus の公開データを活用したサービスが紹介され、研究データのクラウド活用による利点も述べていた。

クラウドとオンプレミスの選択について意見を求められた藤田氏は「しばらくはハイブリッドクラウドで運用することになるであろう。今後はオープンデータ・オープンサイエンスの流れがやってくる。その観点ではクラウド利用の必要性は出てくる。しかし、研究データを守る観点からはオンプレミスで運用する方が安心できる。今後もクラウドとオンプレミスの共存について考えていかなければならない」と述べていた。

藤田 直行氏（FUJITA Naoyuki）
宇宙航空研究開発機構
スーパーコンピュータ活用課長

災害時安否確認システム『ANPIC』

教育機関の効果的な安否確認システム運用手法 ～静岡大学の事例～

株式会社アバンセシステム/静岡大学　永田 正樹氏

株式会社アバンセシステムでは災害時の安否確認システム『ANPIC』を企業や大学向けに展開している。

ANPIC はクラウドサーバを利用しており、国内外のベンダに基幹システムを構築することで、災害時のシステム冗長性を担保している。安否確認システムは特性上、緊急時にアクセスが増加するが、非災害時のアクセスは非常に少ない。クラウドサービスの多くは従量課金制であるため、非災害時は必要最低限のサーバを待機させておき、災害発生時に必要に応じたサーバ台数の増加を行わなければ、システムの運用コストは肥大してしまう。

そこで同社ではアクセス予測に取り組み、クラウドサーバの増減を正確に予測する研究を重ねている。調査資料によると災害発生後の約 20 分でアクセス数はピークに達し、その後はアクセス数が次第に減少していく。災害時のアクセス数をプロットしたデータが対数正規分布に近いことに注目し、災害発生時のデータを素早く収集しアクセス分析を行うことで、クラウドサーバを最適な台数で稼働する。講演ではアクセスデータと非常に近い予測データを割り出した調査資料も提示された。

複数ベンダでデータの整合性を確保するために、Cassandra によるデータベースを構築する研究開発も行っている。Cassandra は NoSQL の分散データ管理システムであり、書き込み特性に優れている。複数ベンダへの高速書き込みを実現し、災害時のアクセス集中による急激な書き込み処理に対応することでデータの整合性を保つのである。

クラウドによるデータ管理について「私はデータ管理の専門家ではない。情報を守る点からは、データ管理の専門家であるクラウドベンダを活用した方がよいのではないか。」と述べていた。また、データ管理について、企業と大学の違いを問われた永田氏は「大学にとって学生は顧客である」と表現し、企業で用いられるトップダウン式の情報伝達が大学では上手く機能しない事実から、情報の周知と人的運用の大切さを述べていた。

永田 正樹氏（NAGATA Masaki）
株式会社アバンセシステム ASP 事業部長
静岡大学 情報基盤センター 教育研究支援員

パネルディスカッション

講演者 3 名への質疑応答があり、その後は同分科会企画委員まとめ役の佐藤 聡氏(筑波大学)がモデレーターを務め、会場も巻き込んだ参加者全員でのパネルディスカッションが行われた。

パネルディスカッションでは、クラウド移行派とオンプレミス残留派それぞれの声が聞かれた。論点の中心となったのは、クラウドまたはオンプレミスを選択する際の評価基準についてであった。

オンプレミスを推奨する組織は『クラウドはコストが高い』という共通認識があった。次いで『オンプレミスはコストを見積もることができるので安心』との意見も多く聞こえた。また、データ管理をクラウドベンダにアウトソーシングすることで、組織内にデータ管理技術の蓄積ができないとする悩みがあることも明らかになった。

一方でクラウドを推奨する組織は、可用性の観点で非常に優れているとの意見が多かった。また、静岡大学では以前、学内のオンプレミスサーバをクラウドサーバへと移行したことを明らかにし、購入したものの実際に使われていなかったオンプレミスサーバを減らすことで、サーバ管理費用全体の 6 割から 7 割を削減できた事例も紹介された。

ハイブリッドクラウドを採用してデータの重要度に応じて使い分ける組織も多い中、データの性格(重要度や公開可能な範囲)自体が研究フェーズの進行によって変化する場合もあるとの意見もあった。組織を取り巻く環境や経営方針から、データ管理の手段はそれぞれ異なることが明らかになり、データの機密性・完全性・可用性を確保し、かつコストを抑えるデータ管理手法について各組織の努力と苦悩が垣間見えたパネルディスカッションであった。

以上